Cowles Foundation Paper 720

Contemporary Mathematics
Volume 80, 1988

MULTIPLE REGRESSION WITH INTEGRATED TIME SERIES

Peter C. B. Phillips[1]

ABSTRACT. Recent work on the theory of regression with integrated processes is reviewed. This work is particularly relevant in economics where many financial series and macroeconomic time series exhibit nonstationary characteristics and are often well modeled individually as simple ARIMA processes. The theory makes extensive use of weak convergence methods and allows for integrated processes that are driven by quite general weakly dependent and possibly heterogeneously distributed innovations. The theory also includes near integrated time series, which have roots near unity, and cointegrated series, which move together over time but are individually nonstationary. A general framework for asymptotic analysis is given which involves limiting Gaussian functionals and extends the LAN and LAMN families of conventional asymptotic theory. An application to the Gaussian AR(1) is reported.

1. INTRODUCTION

The subject matter of this paper is regression theory for nonstationary time series. This subject is of general interest and importance in statistics, but it is particularly relevant in economics where time series are widely believed to be intrinsically nonstationary. Under this heading come various financial and commodity market price series, which behave as if they have no fixed mean, and many macroeconomic aggregates like real output and consumption expenditure, which display secular growth characteristics. Upon investigation all of these series are found to be individually well explained by integrated processes in the ARIMA class, usually with a single unit root. Indeed, a large and growing literature in econometrics now uses integrated processes to model such series in preference to trend stationary processes (processes which are stationary about deterministic trends).

1980 AMS Mathematics Subject Classification: Primary 62M10,

[1]This paper was presented at the AMA, IMS SIAM Conference on Statistical Inference from Stochastic Processes held at Cornell University in August 1987. The research reported here was supported by the NSF under Grant No. SES 8519595. My thanks go to Glena Ames for her skill and effort in keyboarding the manuscript of the paper.

Regression theory for integrated processes turns out to be very different from the traditional theory of regression for stationary time series. These differences affect in a fundamental way the interpretation of regression coefficients, significance tests and residual diagnostics. They need to be understood if the regressions are to be meaningfully used in applied work.

The present paper attempts to review some work that I have been doing recently on this topic in econometrics. The natural starting point in the analysis is asymptotic theory but I shall also include some results on higher order properties. The central idea in the development of the asymptotics is simple and uses the theory of weak convergence on function spaces. We transport the time series we observe and sample moments of them into B-valued random elements (random variables that live in Banach spaces). This enables us to work with rather general integrated processes and to capture in a simple way one of the distinguishing features of non ergodic time series: that sample moments of the time series converge weakly to random variables, which are often simply written as the sample moments of a stochastic process; and that objects like the hessian of the likelihood converge weakly to random matrices rather than constants, reflecting the presence of random information in the limit.

Some notational economies are used to simplify the presentation of the results. Stochastic processes such as the Brownian motion $B(r)$ on $[0,1]$ are simply written as B whenever it is convenient to do so. Similarly, we write integrals with respect to Lebesgue measure such as $\int_0^1 B(s)ds$ more simply as $\int_0^1 B$. Vector Brownian motion with covariance matrix Ω is written "$BM(\Omega)$" We routinely use the following symbols: "$\Rightarrow$" to signify weak convergence, "$\equiv$" to signify equality in distribution, "$>$" to signify positive definite when applied to matrices, "L" to signify the back shift operator, "Δ" $(= 1-L)$ to signify the first difference operator, and "$\|A\|$" to signify the Euclidean matrix norm $(tr(A'A))^{1/2}$.

2. INTEGRATION, NEAR-INTEGRATION AND COINTEGRATION

2.1. *Integrated Time Series*. We call an n-vector time series $\{y_t\}$ an integrated process of order k and write $y_t = I(k)$ if the time series $\{\Delta^k y_t\}$ is weakly stationary (written $\Delta^k y_t = I(0)$). Most of our attention will focus on $I(1)$ processes, which are generated by accumulating innovations from an initialization that is taken to occur at $t = 0$. Thus,

$$y_t = y_{t-1} + u_t = \Sigma_1^t u_j + y_0 = S_t + y_0. \tag{1}$$

Here, $\{S_t\}$ is a partial sum process and $\{u_t\}$ is a weakly dependent innovation sequence. When u_t is iid (respectively, martingale difference sequence) y_t is a random walk (martingale). More generally, u_t will be assumed to satisfy certain moment and weak dependence conditions to ensure the validity of a functional central limit theory for standardized partial sums.

Examples of I(1) processes satisfying (1) are:

(a) ARIMA models: $(1-L)A(L)y_t = B(L)e_t$, e_t iid$(0,\Sigma)$ with u_t following the stationary and invertible ARMA model

$$A(L)u_t = B(L)e_t$$

(b) ARIMAX models: $(1-L)A(L)y_t = B(L)x_t + C(L)e_t$, e_t iid$(0,\Sigma)$ with u_t following the stationary ARMAX model

$$A(L)u_t = B(L)x_t + C(L)e_t, \quad x_t = I(0)$$

(c) Linear processes: $(1-L)y_t = \Sigma_{j=-\infty}^{\infty} B_j \epsilon_{t-j}$, ϵ_t iid$(0,\Sigma)$ with absolutely summable coefficients

$$\Sigma_{-\infty}^{\infty} \|B_j\| < \infty$$

In these examples $A(L)$, $B(L)$ and $C(L)$ are all matrices of finite degree polynomials in the lag operator L .

In recent years empirical evidence in support of such representations with a single unit root have been found by various authors. Box and Jenkins [1] give examples from several subject areas. In economics there are now many different studies that lend support to I(1) model specifications. These include theoretical studies of efficient markets (Shiller [34]) and models of optimizing behavior by representative agents (Hall [11]) as well as a wide range of empirical studies (Granger and Newbold [9], Nelson and Plosser [17], Campbell and Mankiw [2], Stock and Watson [35], Perron and Phillips [21] to mention a few).

2.2. *Near-Integrated Arrays*. Many of the studies just mentioned give strong evidence of the presence of root in the neighborhood of unity. However, it is very difficult to discriminate between a root at unity and a root near unity because the power of unit root tests is very low in the vicinity of the null hypothesis of a unit root. One way of accommodating the possibility of roots that are near unity is through a near integrated array. We write in place of (1)

$$y_{t,T} = A_T y_{t-1,T} + u_t , \quad t = 1, \ldots, T \tag{2}$$

where the coefficient matrix $A_T = \exp(T^{-1}C) \sim I + (1/T)C$ gives alternatives that are local to unit roots as the sample size $T \to \infty$. $\{\{y_{t,T}\}_1^T\}_1^\infty$ is a time series array with roots near unity. In particular, each row of the array for fixed T is an autoregressive series generated by the weakly dependent sequence u_t and with a coefficient matrix A_T whose latent roots are close to unity for T large. C is a matrix of noncentrality parameters. For example, when $C = \mathrm{diag}(c_1, c_2, \ldots, c_n)$ then the i'th component of the series is near explosive for $c_i > 0$, near stationary when $c_i < 0$ and I(1) when $c_i = 0$.

Near integrated arrays generated by (2) have been studied extensively in Phillips [25, 26] and in the scalar case by Chan and Wei [4] also.

2.3. *Cointegrated Time Series.* While individual time series may display nonstationary characteristics, it is often the case that several related time series tend to move together over time as if there were a common stochastic trend involved in each of the series. Multiple time series with this characteristic are said to be cointegrated if the individual series are I(1) processes and yet some linear combination(s) of the series is (are) I(0). Cointegrated systems of this type have recently attracted a good deal of attention in econometrics. Some of the ideas go back to Frisch [6], but the concept was formally introduced by Granger [7] and has been systematically explored in subsequent work by various authors. My own attention has concentrated on the statistical issues of testing for cointegration (Phillips and Ouliaris [32, 33]) and on the development of an asymptotic distribution theory for estimators and tests in the presence of cointegration (Phillips and Durlauf [30]), Phillips [26] and Park and Phillips [19, 20]).

To fix ideas, let $\{y_t\}$ be an n-vector multiple time series and suppose $y_t = I(1)$. If h is a constant n-vector for which $h'y_t = I(0)$ then we say y_t is cointegrated with cointegrating vector h. Note that h annihilates the stochastic trend in y_t, indicating that the components of y_t have a common stochastic trend of dimension less than n.

2.4. *Economic Applications.* Cointegration may be regarded as a statistical embodiment of ideas from economic theory. The hypothesis of cointegration is particularly important in terms of well established notions of long run equilibrium in economics. We give several examples:

(a) steady state growth theory.

According to this theory many economic aggregates like output, consumption and investment grow together over time along paths that are determined principally by common factors such as technical progress and population change. Thus, in the long run aggregate expenditure may be expected to display only stationary fluctuations about some fraction of aggregate income. Another example stems from the quantity theory of money, according to which real income, money and prices may be expected to be in long run balance with the velocity of circulation displaying only stationary fluctuations.

(b) present value models.

In these models one variable (Y_t) represents the discounted present value of expectations (E_t) of future realizations of another variable (X_t) using today's information set. Symbolically,

$$Y_t = \theta(1-\delta)\Sigma_{i=0}^{\infty}\delta^i E_t X_{t+1}$$

where Y_t, $X_t = I(1)$. This gives

$$Y_t - \theta X_t = \theta\Sigma_{i=1}^{\infty}\delta^i E_t \Delta X_{t+i} = I(0) .$$

Note that if X_t is a martingale then $E_t(X_{t+i}) = X_t$ and $Y_t = \theta X_t$ in long run balance. Models of this type are used in theories of the term structure of interest rates, stock prices and dividends and in the permanent income theory of consumption. Campbell and Shiller [3] provide a detailed discussion.

(c) purchasing power parity.

According to this theory in international financial economics the dollar value of goods produced abroad and the dollar value of goods produced domestically should be in long run balance. Thus, in the long run we should expect only stationary fluctuations about the equation $P_t = S_t P_t^*$ which relates the level of domestic prices (P_t) to foreign prices (P_t^*) and the spot exchange rate (S_t) .

3. WEAK CONVERGENCE OF SAMPLE MOMENTS

3.1. *Functional Limit Theory and Convergence to Stochastic Integrals.* We transport partial sums of the innovations in (1) into B-valued random elements using

$$X_T(r) = T^{-1/2}S_{[Tr]} \in D[0,1]^n , \quad 0 \le r \le 1$$

where $D[0,1]^n$ is the product space of n copies of $D[0,1]$, the space of right continuous functions with finite left limits endowed with the Skorohod topology. Under very general conditions the functional central limit theorem

$$X_T(r) \Rightarrow B(r) = BM(\Omega) \tag{3}$$

holds [5, 12, 30] and the covariance matrix of the Brownian motion is

$$\Omega = \lim_{T\to\infty} T^{-1}E(S_T S_T') .$$

which reduces to

$$\Omega = 2\pi f_{uu}(0) = E(u_0u_0') + \Sigma_{k=1}^{\infty}\{E(u_0u_k') + E(u_ku_0')\}$$

$$= \Sigma + \Lambda + \Lambda'$$

when $\{u_t\}$ is stationary with spectral density matrix $f_{uu}(\lambda)$. Note that Ω may be interpreted as the (scaled) long run variance of the process y_t. Note also that the estimation of Ω is an important and interesting topic in itself, which has a substantial bearing on empirical work in this field. ([18] and [30] provide some discussion of the problem and available methods).

In addition to (3) we frequently need a theory of weak convergence to stochastic integrals. Under quite general conditions we do indeed have the following result:

$$T^{-1}\Sigma_1^{[Tr]}S_{t-1}u_t' \Rightarrow \int_0^r BdB' + r\Lambda . \tag{4}$$

The limit process here is a matrix stochastic integral with bias given by $r\Lambda$. (4) is proved in [27, 28] for the finite dimensional case $r = 1$. When u_t is a martingale difference sequence a proof of (4) is given in [37]. The proof in [28] relies on martingale approximation methods and applies for fairly general stationary sequences.

3.2. *Limit Processes for Sample Moments.* We deal first with integrated time series generated by (1). In this case sample moments have the following asymptotic behavior:

$$T^{-3/2}\Sigma_1^{[Tr]}y_t \Rightarrow \int_0^r B , \quad T^{-2}\Sigma_1^{[Tr]}y_ty_t' \Rightarrow \int_0^r BB' \tag{5}$$

$$T^{-1}\Sigma_1^{[Tr]}y_tu_t' \Rightarrow \int_0^r BdB' + r\Delta , \quad \Delta = \Sigma + \Lambda . \tag{6}$$

Note that the limit processes in these formulae are simple linear and quadratic functionals of Brownian motion. In effect, the limits of the sample moments of the time series are stochastic processes which are themselves just the sample moments of a vector Brownian motion. Proofs of these results follow simply from (3) and (4), in most cases by direct appeal to the continuous mapping theorem. See [23] and [30] for details.

For near-integrated arrays generated by (2) we have in place of (3) the limit:

$$T^{-1/2}y_{[Tr],T} \Rightarrow K_C(r) = \int_0^r \exp((r-s)C)dB(s) . \tag{7}$$

Here, K_C is a vector diffusion process which satisfies the stochastic differential equation system

$$dK_C(r) = CK_C(r)dr + dB(r) , \quad K_C(0) = 0 .$$

Again, sample moments of the array converge to corresponding sample moments of K_C . We find

$$T^{-3/2}\Sigma_1^{[Tr]}y_{t,T} \Rightarrow \int_0^r K_C \ , \quad T^{-2}\Sigma_1^{[Tr]}y_{t,T}y_{t,T}' \Rightarrow \int_0^r K_C K_C'$$

$$T^{-1}\Sigma_1^{[Tr]}y_{t,T}u_t' \Rightarrow \int_0^r K_C dB' + r\Delta$$

(see [25, 26] for proofs and applications).

3.3. *Filtered Processes and Projections.* In many cases time series such as (1) and (2) are filtered prior to their use in regression. In other cases the time series are effectively filtered by the inclusion of additional variables such as time trends in a regression. The effects of such filtering can often be simply determined by looking at the filtered series as regression residuals. For example, in the case of a series such as (1) that is detrended by a polynomial trend we construct the regression residual process $\underline{y}_t$ from the least squares regression:

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 t + \ldots + \hat{\beta}_p t^p + \underline{y}_t \ .$$

Then sample moments of $\underline{y}_t$ have the following limits

$$T^{-3/2}\Sigma_1^T \underline{y}_t \Rightarrow \int_0^1 \underline{B} \ , \quad T^{-2}\Sigma_1^T \underline{y}_t \underline{y}_t' \Rightarrow \int_0^1 \underline{B}\ \underline{B}'$$

$$T^{-1}\Sigma_1^T \underline{y}_t u_t' \Rightarrow \int_0^1 \underline{B} dB' + \Delta$$

where

$$\underline{B} = QB$$

= projection of B in $L_2[0,1]^n$ on the orthogonal complement of the space spanned by $\{0(r),\ 1(r),\ \ldots,\ p(r);\ j(r) = r^j\}$.

Thus, $\underline{B}(r)$ is simply detrended Brownian motion or the residuals from the continuous time regression

$$B(r) = \hat{\alpha}_0 + \hat{\alpha}_1 r + \ldots + \hat{\alpha}_p r^p + \underline{B}(r)$$

where the $\hat{\alpha}_i$ minimize the least squares criterion in L_2 norm

$$\int_0^1 \left\| B(r) - \alpha_0 - \alpha_1 r - \ldots - \alpha_p r^p \right\|^2 dr \ .$$

Explicit formulae are easy to obtain. We give the following examples:

$$p = 0 \qquad \underline{B} = B(r) - \int_0^1 B = \text{demeaned BM}$$

$p = 1 \quad \underline{B} = B(r) - \hat{\alpha}_0 - \hat{\alpha}_1 r = \text{detrended BM}$

$$\text{with} \quad \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{bmatrix} = \begin{bmatrix} 1 & \int_0^1 s \\ \int_0^1 s & \int_0^1 s^2 \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 B \\ \int_0^1 sB \end{bmatrix} .$$

Similar results apply in the case of near-integrated time series. Using the same notation, we get:

$$T^{-3/2}\Sigma_1^T \underline{y}_{t,T} \Rightarrow \int_0^1 \underline{K}_C \ , \quad T^{-2}\Sigma_1^T \underline{y}_{t,T'} \, \underline{y}_{t,T} \Rightarrow \int_0^1 \underline{K}_C \underline{K}_C'$$

$$T^{-1}\Sigma_1^T \underline{y}_{t,T} u' \Rightarrow \int_0^1 \underline{K}_C dB' + \Delta \ .$$

3.4. *Applications to Vector Autoregressions (VAR's)*. The above results have many applications. In particular, a regression theory for time series generated by (1) or (2) is straightforward. Thus if y_t is a near-integrated time series (corresponding to the t'th row of the array in (2), but we drop the second subscript for convenience) then a least squares regression yields:

$$y_t = \hat{A} y_{t-1} + \hat{u}_t \ , \quad \hat{\Sigma} = T^{-1}\Sigma_1^T \hat{u}_t \hat{u}_t'$$

where $\hat{A} = (\Sigma_1^T y_t y_{t-1}')(\Sigma_1^T y_{t-1} y_{t-1}')^{-1}$. We find

$$T(\hat{A} - A_T) \Rightarrow \left(\int_0^1 dBK_C' + \Lambda'\right)\left(\int_0^1 K_C K_C'\right)^{-1} \tag{8}$$

a matrix quotient of quadratic functionals in the diffusion K_C . Of course, $\hat{A} - A_T \to_p 0$ and since $A_T \to I$ we obtain $\hat{A} \to_p I$ as $T \to \infty$.

Since $\hat{A}$ is consistent the residuals $\hat{u}_t$ are consistent estimates of the errors u_t in (2) and the asymptotic distribution theory for the variance estimator $\hat{\Sigma}$ is given by:

$$T^{1/2}\text{vec}(\hat{\Sigma}-\Sigma) \Rightarrow N(0,V)$$

with

$$V = P_D \Sigma_{k=0}^{\infty} [\Psi_k - (\text{vec } \Sigma)(\text{vec } \Sigma)'] P_D$$

where

$$\Psi_k = \begin{cases} E(u_0 u_0' \otimes u_0 u_0') \ , & k = 0 \\ E\{(u_0 u_k' \otimes u_0 u_k') + (u_k u_0' \otimes u_k u_0')\} \ , & k = 1, 2, \ldots \end{cases}$$

and $P_D = D(D'D)^{-1}D'$ is the projection onto the natural support in R^{n^2} of the covariance matrix $\hat{\Sigma}$ i.e. the range space of D , the matrix which duplicates the nonredundant elements σ of Σ through the mapping

$\text{vec}\ \Sigma = D\sigma$.

When the innovations $\{u_t\}$ are iid $N(0,\Sigma)$ the covariance matrix reduces to the simpler

$$V = 2P_D(\Sigma \otimes \Sigma)P_D = 2P_D(\Sigma \otimes \Sigma) .$$

The reader is referred to [26, 30] for more details of these results.

3.5. *Power Functions for Unit Root Tests*. Another application of the theory is to obtain power functions for unit root tests. Here we use the near-integrated array (2) and the matrix C provides the noncentralities in the asymptotic distribution theory under local alternatives. We demonstrate by taking the scalar case. Set $n = 1$, $A = a$, $\Omega = \omega^2$, $\Sigma = \sigma^2$, $\Lambda = \lambda$, $C = c$ and

$$K_C(r) = \int_0^r e^{(r-s)C} dB(s) = \omega\int_0^r e^{(r-s)c} dW(s)$$

$$= \omega J_c(r) \ , \quad \text{say}$$

where $B = \omega W$ and $W = BM(1)$, standard Brownian motion.

Suppose we wish to test for the presence of a unit root in the array (2). We test $H_0 : a = 1$ against the sequence of local alternatives $H_1 : a_T = e^{c/T} \sim 1 + c/T$. From (8) the asymptotic theory for the simple regression coefficient $\hat{a}$ is given by:

$$T(\hat{a}-1) \Rightarrow c + \left(\int_0^1 J_c^2\right)^{-1} \left(\int_0^1 J_c dW + \lambda/\omega^2\right) . \tag{9}$$

To test H_0 we may use the statistic

$$Z_a = T(\hat{a}-1) - \left(T^{-2}\Sigma_1^T y_{t-1}^2\right)^{-1}\hat{\lambda}$$

suggested in [23], where $\hat{\lambda}$ is a consistent estimator of λ . From (9) we now have

$$Z_a \Rightarrow c + \left(\int_0^1 J_c^2\right)^{-1} \int_0^1 J_c dW \quad \text{under } H_1 \tag{10}$$

$$= \left(\int_0^1 W^2\right)^{-1} \int_0^1 W dW \qquad \text{under } H_0 .$$

Power functions given by (10) may now be calculated using numerical inversion of the joint characteristic function of $(\int_0^1 J_c dW, \int_0^1 J_c^2)$, which is given in [25]. These ideas are easily extended to models with fitted drift and trend [34] and to multivariate models and tests [19, 20, 26].

3.6. *Limits of the Near-Integrated Theory*. Again we take the scalar case of the near integrated array generated by (2) with noncentrality parameter c .

When $c = 0$ we have the conventional unit root theory. When $c \neq 0$ but is fixed we have asymptotics which are local to the unit root theory. We may also consider the limits of the near integrated theory as $c \to \pm\infty$, the natural limits of its domain of definition. Heuristically, we can associate $c \to +\infty$ with explosive alternatives since when $a_T = e^{c/T} > 1$ is fixed we have $c = T \ln a_T \to \infty$ on a diagonal sequence for the pair (c,T). Similarly, when $a_T = e^{c/T} < 1$ we have $c = T \ln a_T \to -\infty$ on a diagonal sequence for (c,T).

It is particularly interesting to study the asymptotic behavior as $c \to \pm\infty$ of the regression coefficient $\hat{a}$ and the associated t-ratio $t_a = (\hat{a}-a)/S_a$ where $S_a = \{(T^{-1}\Sigma_1^T\hat{u}_t^2)(\Sigma_1^T y_{t-1}^2)^{-1}\}^{1/2}$ is the traditional regression standard error of $\hat{a}$. Taking the t-ratio first, we have:

$$t_a \Rightarrow \frac{\omega}{\sigma}\,\frac{\int_0^1 J_c dW + \lambda/\omega^2}{(\int_0^1 J_c^2)^{1/2}} \quad \text{as} \quad T \to \infty$$

$$\Rightarrow \left\{\begin{matrix} N(0,1) & \text{if} \ \lambda = 0 \\ \text{diverges} & \text{if} \ \lambda \neq 0 \end{matrix}\right\} \quad \text{as} \quad c \to \pm\infty \tag{11}$$

and for the regression coefficient itself (with $\lambda = 0$, $\omega^2 = 1$):

$$g(c)^{1/2}T(\hat{a}-a) \Rightarrow g(c)^{1/2}\int_0^1 J_c dW/\int_0^1 J_c^2$$

$$\Rightarrow \begin{cases} N(0,1) & \text{as} \quad c \to -\infty \\ \text{Cauchy as} & c \to +\infty \end{cases} . \tag{12}$$

Here

$$g(c) = E(\int_0^1 J_c^2) = -(1/2c)\{1 + (1/2c)(1 - e^{2c})\}$$

is the limiting information under local alternatives. In conventional asymptotics for fixed $a_T = a$ in (2) we have instead:

$$g_T = E(T^{-2}\Sigma_1^T y_{t-1}^2) \sim \begin{cases} \dfrac{1}{T(1 - a^2)} & |a| < 1 \\[2ex] \dfrac{1}{T^2(a^2 - 1)} & |a| > 0 \end{cases}$$

Using g_T to restandardize $T(\hat{a}-a)$ in place of $g(c)$ we can interpret (12) in terms of results for the stationary and explosive cases, giving:

$$T^{1/2}(\hat{a}-a)/(1 - a^2)^{1/2} \Rightarrow N(0,1) \quad \text{stationary AR(1)}$$

$T(\hat{a}-a)/(a^2 - 1) \Rightarrow \text{Cauchy}$ explosive Gaussian AR(1) .

These results were obtained originally by White [39]. The idea of exploring the limits of the near-integrated theory was developed in Phillips [25] and, for the t-ratio case, independently also by Chan and Wei [4].

4. SPURIOUS AND COINTEGRATING REGRESSION ASYMPTOTICS

4.1. *Spurious Regressions.* Studies in spurious regressions go back to early work by Yule [40, 41] in the 1920's. The topic has attracted attention in econometrics because many economic time series have strong trend components and the potential for spurious regressions is thought to be high. A general asymptotic theory for regressions of this type has recently been given by the author in [22].

To fix ideas, let

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \begin{matrix} 1 \\ m \end{matrix} = I(1) \ , \quad n = m+1 \tag{13}$$

be an n-vector I(1) process whose partial sums in (1) satisfy the invariance principle (3). We partition the limit Brownian motion conformably as:

$$B(r) = \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} = BM(\Omega) \tag{14}$$

with

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix} > 0 \ .$$

Next consider the least squares regression

$$y_{1t} = \hat{\beta}' y_{2t} + \hat{v}_t \tag{15}$$

which is spurious because the variables y_{1t} and y_{2t} have stochastic trends but may not otherwise be related if $\omega_{21} = 0$. Even if $\omega_{21} \neq 0$, however, the relationship between the series is not in general strong enough to permit consistent estimation of a regression coefficient by $\hat{\beta}$. Only when Ω is singular will this occur.

In general, we have (from [22])

$$\hat{\beta} \Rightarrow \left(\int_0^1 B_2 B_2'\right)^{-1} \left(\int_0^1 B_2 B_1\right) \tag{16}$$

a matrix quotient of quadratic functionals of B . Note that we can write

$$B_1 = \omega_{21}'\Omega_{22}^{-1}B_2 + \ell_{11}W$$

where

$$\ell_{11}^2 = \omega_{11\cdot 2} = \omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21}$$

and $W = BM(1)$ is independent of B_2. As shown recently in [29] the distribution of (15) in the limit is a simple mixture of normals. Specifically,

$$\left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1 B_2B_1\right) = \int_{V>0} N(\Omega_{22}^{-1}\omega_{21},\ \omega_{11\cdot 2}V(B_2))dP(V) \tag{17}$$

with

$$V(B_2) = \left(\int_0^1 B_2B_2'\right)^{-1}\left(\int_0^1\int_0^1 B_2(r\wedge s)B_2'\right)\left(\int_0^1 B_2B_2'\right)^{-1} \tag{18}$$

(17) can be further reduced to the scalar mixture:

$$\int_{v>0} N(\Omega_{22}^{-1}\omega_{21},\ \omega_{11\cdot 2}\Omega_{22}^{-1}v)dP(v)\ .$$

Interestingly, as shown in [29], this distribution does not lie in the conventional LAMN family.

The asymptotic behavior of tests and regression diagnostics from (14) is also easy to obtain. Write the sample second moment matrix of the Brownian motion B in partitioned form as

$$G = \begin{bmatrix} g_{11} & g_{21}' \\ g_{21} & G_{22} \end{bmatrix} = \begin{bmatrix} \int_0^1 B_1^2 & \int_0^1 B_1B_2 \\ \int_0^1 B_2B_1 & \int_0^1 B_2B_2' \end{bmatrix}$$

Then, since $\hat{v}_t$ is a (random) linear combination of I(1) variates the standard error of regression diverges. We have

$$s^2 = T^{-1}\Sigma_1^T\hat{v}_t^2 \uparrow \text{(diverges)}$$

and

$$T^{-1}s^2 \Rightarrow g_{11} - g_{21}'G_{22}^{-1}g_{21} = g_{11\cdot 2}$$

the conditional sample variance of the sample path of B_1 given B_2. Similarly, we find that the standard error of estimated regression coefficient $\hat{\beta}_i$ converges to zero, giving a spurious impression of precision in the estimate $\hat{\beta}_i$. That is

$$s_{\hat{\beta}_i}^2 = s^2\left[\left(\Sigma_1^T y_{2t}y_{2t}'\right)^{-1}\right]_{ii} \xrightarrow{p} 0$$

Finally, using (15) and (17) we deduce that

$$t_{\beta_i} = \hat{\beta}_i / s_{\hat{\beta}_i} \quad \text{diverges}$$

as $T \to \infty$, corroborating the experimental evidence in [8] that we observe high rejection rates in significance tests in spurious regressions like (14). The reader is referred to [22, 26] for a detailed discussion of this phenomenon and further analytic results along these lines.

4.2. *Cointegrating Regressions.* These regressions may be regarded as the natural limit of a spurious regression like (15) in which the residuals form a near integrated triangular array whose limit along a certain diagonal sequence is I(0). Let y_t be an n-vector time series partitioned as in (13) and generated by the system

$$y_{1t} = \beta' y_{2t} + u_{1t}, \quad y_{2t} = y_{2t-1} + u_{2t} \tag{19}$$

where

$$u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \begin{matrix} 1 \\ m \end{matrix} = I(0) . \tag{20}$$

We assume that the partial sums of the innovations u_t satisfy the invariance principle (3) and we partition the limit Brownian motion $B(r) = BM(\Omega)$ conformably with (20), as in (14) above.

Note that y_{1t} and y_{2t} are cointegrated in (19) since the linear combination $y_{1t} - \beta' y_{2t} = I(0)$. Moreover, the signal from the regressor $y_{2t} = I(1)$ in (19) is stronger by an order of magnitude in the sample size than the contemporaneous (and serial) correlations between y_{2t} and u_{1t}. It follows that linear least squares regression in (19) yields a consistent estimator $\hat{\beta}$, in contrast to the spurious regression (15). In fact, the regression asymptotics (from [29, 30]) are

$$\begin{aligned} T(\hat{\beta}-\beta) &= \left(T^{-2} Y_2' Y_2\right)^{-1} (T^{-1} Y_2' u_1) \\ &\Rightarrow \left(\int_0^1 B_2 B_2'\right)^{-1} (\int_0^1 B_2 dB_1 + \delta_{21}) \\ &= \int N(\Psi \Omega_{22}^{-1} \omega_{21} + \Theta^{-1} \delta_{21}, \; \omega_{11\cdot 2} \Theta^{-1}) dP(\Psi, \Theta) \end{aligned} \tag{21}$$

where

$$\Psi = \left(\int_0^1 B_2 B_2'\right)^{-1} (\int_0^1 B_2 dB_2')$$

$$\Theta = \int_0^1 B_2 B_2'$$

(21) shows that the limit distribution is a mean and covariance matrix mixture of normals.

4.3. *Testing for Cointegration.* To distinguish spurious regressions empirically from cointegrating regressions it seems natural to test the residuals $\hat{v}_t$ in (15) for nonstationarity. In effect, we may test the null hypothesis of a spurious regression (or absence of cointegration) by testing whether $\hat{v}_t$ has a unit root against the alternative hypothesis that $\hat{v}_t = I(0)$. To do this, we can use any unit root test. A simple test recommended recently in [33] is based on the Z_a statistic considered earlier in Section 3.5. Here we define

$$Z_a = T(\hat{a}-1) - \left(T^{-2}\Sigma_1^T\hat{v}_{t-1}^2\right)^{-1}\hat{\lambda}$$

where $\hat{a}$ is the regression coefficient in

$$\hat{v}_t = \hat{a}\hat{v}_{t-1} + \hat{k}_t$$

and

$$\hat{\lambda} = (1/2)(2\pi\hat{f}_k(0) - T^{-1}\Sigma_1^T\hat{k}_t^2)$$

where $\hat{f}_k(0)$ is any consistent spectral estimate for $f_k(\ )$ at the origin $\omega = 0$.

The asymptotic theory for this statistic is derived in [33]. We have

$$Z_a \Rightarrow \int_0^1 RdR \tag{22}$$

where

$$R(r) = Q(r)/\left(\int_0^1 Q^2\right)^{1/2} ,$$

$$Q(r) = W_1(r) - (\int_0^1 W_1W_2')\left(\int_0^1 W_2W_2'\right)^{-1} W_2 ,$$

$$W = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \begin{matrix} 1 \\ m \end{matrix} = BM(I_n) .$$

In the above, $Q(r)$ is the projection in $L_2[0,1]$ of W_1 on the orthogonal complement of the subspace spanned by the elements of W_2 . Further, R lies on the unit sphere in $L_2[0,1]$ and is therefore a random element in a Hilbert manifold. Also, we note that Q , and hence R , depends only on the dimension of the system. It is otherwise free of nuisance parameters. Tabulations of the initial values of the limit distribution (22) are given in [33].

Finally, we observe that when the dimension of the system $n = 1$ we have $m = 0$, $Q = W_1$ and the limit distribution reduces to

$$\int_0^1 RdR = \int_0^1 W_1 dW_1 / \int_0^1 W_1^2$$

that is, the conventional limit distribution of the autoregressive coefficient in a random walk [23, 39]. Thus (22) includes this traditional unit root theory as a special case.

5. REGRESSION WITH COINTEGRATED REGRESSORS

5.1. *Examples.* Two commonly occurring examples in which the regressors are themselves cointegrated are:

(a) VAR's with unit roots and several lags, such as:

$$\begin{aligned} y_t &= B_1 y_{t-1} + B_2 y_{t-2} + u_t , \quad B_1 + B_2 = I \\ &= (B_1 + B_2) y_{t-1} + B_2 (y_{t-2} - y_{t-1}) + u_t . \end{aligned}$$

Here y_{t-1} and $(y_{t-2} - y_{t-1})$ are trivially cointegrated since $y_{t-1} - y_{t-2} = I(0)$.

(b) VAR's with common stochastic trends: Suppose $y_t = y_{t-1} + u_t = I(1)$ but that for some $n \times n_1$ matrix J_1 with orthonormal columns $J_1' y_t = w_{1t} = I(0)$. Define the orthogonal matrix $J = [J_1, J_2]$ and write

$$\begin{aligned} y_t &= JJ' y_{t-1} + JJ' u_t \\ &= J_2 x_{2t} + J_1 w_{1t} . \end{aligned}$$

Here

$$x_{2t} = J_2' y_{t-1} + J_2' u_t = J_2' y_t$$

represents the common stochastic trend in y_t . In effect, y_t is an I(1) process with fewer than n stochastic trends.

5.2. *General Formulation and Asymptotic Results.* Suppose the n-vector time series y_t and the m-vector time series x_t are cointegrated and satisfy the equation

$$y_t = Ax_t + u_{1t} \tag{23}$$

where $u_{1t} = I(0)$. Next let $H = [H_1, H_2]$ be an orthogonal $m \times m$ matrix which rotates the coordinates of the regressor space so that $x_{1t} = H_1' x_t = I(0)$ and $x_{2t} = H_2' x_t = I(1)$. We write

$$\begin{aligned} y_t &= AHH' x_t + u_{1t} \\ &= A_1 x_{1t} + A_2 x_{2t} + u_{1t} \end{aligned} \tag{24}$$

and set

$$x_{1t} = u_{2t} \ , \quad \Delta x_{2t} = u_{3t}$$

$$u_t' = (u_{1t}', u_{2t}', u_{3t}') \ .$$

We assume that partial sums of the innovation sequence u_t satisfy the invariance principle (3) with $\Omega > 0$ and we partition the limit Brownian motion and associated matrices Ω , Σ , Λ and Δ conformably with u_t . Thus,

$$B(r) = \begin{bmatrix} B_1(r) \\ B_2(r) \\ B_3(r) \end{bmatrix} = BM(\Omega) \ ,$$

$$(\Omega_{ij}) = (\Sigma_{ij}) + (\Lambda_{ij}) + (\Lambda_{ji}') \ ,$$

$$(\Delta_{ij}) = (\Sigma_{ij}) + (\Lambda_{ij}) \ .$$

Partitioned least squares regression on (24) yields in conventional regression notation the following estimates of the submatrices A_1 and A_2 :

$$\hat{A}_1 = Y'Q_2X_1(X_1'Q_2X_1)^{-1} \ , \quad \hat{A}_2 = Y'Q_1X_2(X_2'Q_1X_2)^{-1}$$

As shown in [20] these estimates have quite different asymptotic behavior which we may characterize as:

$$\sqrt{T}(\hat{A}_1 - \bar{A}_1) \Rightarrow \zeta_{12}\Sigma_{22}^{-1} - \Sigma_{12}\Sigma_{22}^{-1}\zeta_{22}\Sigma_{22}^{-1} \tag{25}$$

$$= \zeta \equiv N(0,M)$$

and

$$T(\hat{A}_2 - A_2) \Rightarrow (\int_0^1 dB_0B_3' + \Delta_{30}')\left(\int_0^1 B_3B_3'\right)^{-1} \tag{26}$$

with

$$B_0 = B_1 - \Sigma_{12}\Sigma_{22}^{-1}B_2 \ , \quad \Delta_{30}' = \Delta_{31}' - \Sigma_{12}\Sigma_{22}^{-1}\Delta_{32}' .$$

In (25)

$$\bar{A}_1 = A_1 + \Sigma_{12}\Sigma_{22}^{-1}$$

and the estimator $\hat{A}_1$ is consistent to A_1 iff $\Sigma_{12} = 0$, that is iff $E(u_{1t}u_{2t}') = 0$ when u_t is covariance stationary. The limiting distribution of $\sqrt{T}(\hat{A}_1 - \bar{A}_1)$ is matrix normal and both the numerator sample moments and the denominator sample moments in $\hat{A}_1$ contribute to the limit distribution (through ζ_{12} and ζ_{22} , respectively) when $\Sigma_{12} \neq 0$. Thus, the estimated coefficients of the stationary components x_{1t} in (24) are asymptotically normal.

On the other hand, the estimated coefficients of the I(1) components x_{2t} are always consistent and have a limit distribution of the general form (21). However, as seen in (26), the Brownian motion B_0 involves contributions from the error u_{1t} (giving B_1) and the component $u_1'X_1(X_1'X_1)^{-1}X_1'X_2$ (giving $\Sigma_{12}\Sigma_{22}^{-1}B_2$). In other words, the presence of the I(0) regressor x_{1t} in (24) does influence the asymptotic distribution of $\hat{A}_2$.

These results for $\hat{A}_1$ and $\hat{A}_2$ can be used to obtain the asymptotic behavior of $\hat{A} = Y'X(X'X)^{-1}$, the matrix of least squares regression coefficients in (23). We have

$$\sqrt{T}(\hat{A}-\bar{A}) \sim \sqrt{T}(\hat{A}_1 - \bar{A}_1)H_1' \Rightarrow \zeta H_1' = N(0, (I \otimes H_1)M(I \otimes H_1'))$$

where

$$\bar{A} = [\bar{A}_1, A_2]H' = [A_1, A_2]H' + \Sigma_{12}\Sigma_{22}^{-1}H_1'$$

$$= A + \Sigma_{12}\Sigma_{22}^{-1}H_1' .$$

Thus, $\hat{A}$ is consistent to $\bar{A}$ and is asymptotically normal but has a singular covariance matrix, corresponding to the fact that $\sqrt{T}(\hat{A}-\bar{A})H_2 \underset{p}{\rightarrow} 0$.

We also observe singularities in the limits of the sample moment matrix of the regressors and its inverse. Interestingly both exist and are singular. We have:

$$T^{-2}X'X \Rightarrow H_2(\int_0^1 B_3B_3')H_2'$$

since not all components of x_t are I(1); and

$$T(X'X)^{-1} \underset{p}{\rightarrow} H_1\Sigma_{22}^{-1}H_1'$$

since not all components of x_t are I(0).

5.3. *Regressors with Deterministic and Stochastic Trends.* One might expect similar results for regressors that involve deterministic as well as stochastic trends. However, some important differences do arise. These are well illustrated by using the general formulation (23) as before but allowing the regressors to include a non-zero drift. Thus, we write

$$x_t = \pi + x_{t-1} + u_{2t} = \pi t + x_t^0 , \quad \text{say}$$

where $x_t^0 = I(1)$ and $\Delta x_t = u_{2t}$. Then

$$T^{-3}\Sigma_1^T x_t x_t' \underset{p}{\rightarrow} \tfrac{1}{3}\pi\pi'$$

which is singular for $m > 1$. It is the singularity of this sample moment matrix in the limit that makes the application of traditional theory difficult and that causes a degeneracy in the limit distribution. We handle this degeneracy by simply rotating coordinates in the regressor space to isolate components of x_t with signals of different orders of magnitude in T. Start by defining $h_1 = \pi/(\pi'\pi)^{1/2}$ and construct the orthogonal matrix $H = [h_1, H_2]$. Next, rotate the regressor space using H to give:

$$\begin{aligned} y_t &= AHH'x_t + u_{1t} \\ &= a_1 x_{1t} + A_2 x_{2t} + u_{2t} \end{aligned} \tag{27}$$

where

$$x_{1t} = h_1'x_t = (\pi'\pi)^{1/2}t + h_1'x_t^0$$

$$x_{2t} = H_2'x_t = H_2'x_t^0 .$$

Define the new (effective) error vector

$$\underline{u}_t = \begin{bmatrix} u_{1t} \\ \underline{u}_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ H_2'u_{2t} \end{bmatrix}$$

and assume that partial sums of $\underline{u}_t$ satisfy the invariance principle (3) with limit Brownian motion

$$\underline{B}(r) = \begin{bmatrix} B_1(r) \\ \underline{B}_2(r) \end{bmatrix} = BM(\underline{\Omega})$$

where

$$\underline{\Omega} = \begin{bmatrix} \Omega_{11} & \underline{\Omega}_{12} \\ \underline{\Omega}_{21} & \underline{\Omega}_{22} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12}H_2 \\ H_2'\Omega_{21} & H_2'\Omega_{22}H_2 \end{bmatrix} .$$

Least squares regression on (27) yields $\hat{a}_1$ and $\hat{A}_2$. These fitted coefficients have the following asymptotics behavior, both involving ratios of functionals of Brownian motion:

$$T^{3/2}(\hat{a}_1 - a_1) \Rightarrow (\int_0^1 dB_1\zeta + \delta)\left(\int_0^1\zeta^2\right)^{-1} \tag{28}$$

$$T(\hat{A}_2 - A_2) \Rightarrow (\int_0^1 dB_1\eta' + \underline{\Delta}_{21}')\left(\int_0^1\eta\eta'\right)^{-1} \tag{29}$$

where

$$\zeta(r) = (\pi'\pi)^{1/2}\left\{r - (\int_0^1 s\underline{B}_2')\left(\int_0^1\underline{B}_2\underline{B}_2\right)^{-1}\underline{B}_2\right\}$$

$$\eta(r) = \underline{B}_2(r) - \left(\int_0^1 \underline{B}_2 s\right)\left(\int_0^1 s^2\right)^{-1} r$$

$$\delta = -\underline{\Delta}_{21}'\left(\int_0^1 \underline{B}_2 \underline{B}_2'\right)^{-1}\left(\int_0^1 B_2 s\right)$$

$$\underline{\Delta}_{21} = H_2'\Delta_{21} .$$

Both $\hat{a}_1$ and $\hat{A}_2$ are consistent. The asymptotic distribution of $\hat{a}_1$ involves the process ζ which is the projection in $L_2[0,1]$ of the scaled time trend $(\pi'\pi)^{1/2}r$ on the orthogonal complement of the space spanned by the elements of $\underline{B}_2$. In effect, ζ is the L_2 function space equivalent of the Euclidean space projection. Q_2x_1 that occurs in the partitioned regression formula for $\hat{a}_1$, viz. $\hat{a}_1 = (Y'Q_2x_1)(x_1'Q_2x_1)^{-1}$. Similarly, the process η is the projection of $\underline{B}_2$ on the orthogonal complement in $L_2[0,1]^{m-1}$ of the space spanned by the simple time trend r.

Results (28) and (29) combine to give us

$$T(\hat{A}-A) = T[\hat{a}_1 - a_1, \hat{A}_2 - A_2]H' - T(\hat{A}_2 - A_2)H_2'$$

$$\Rightarrow \left(\int_0^1 dB_1\eta + \underline{\Delta}_{21}'\right)\left(\int_0^1 \eta\eta'\right)^{-1} H_2'$$

a distribution whose support is a subspace of dimension $nm-n$. When $m = 1$ this distribution is degenerate and we are left with the vector of coefficients of x_{1t} giving in this case ($m = 1$, $\Omega_{21} = 0$, $\delta = 0$, $\zeta(r) = \pi r$) :

$$T^{3/2}(\hat{a}-a) \Rightarrow \left(\int_0^1 dB_1\zeta\right)\left(\int_0^1 \zeta^2\right)^{-1} = N\left(0,\ \omega_{11}\left(\pi^2\int_0^1 r^2\right)^{-1}\right)$$

$$= N(0,\ 3\omega_{11}/\pi^2) .$$

The reader is referred to [19, 20] for further analysis of regressions with deterministic and stochastic trends.

5.4. *Special Results with Exogenous Regressors.* Stronger results can be obtained for regressions with strictly exogenous regressors. To illustrate, we take the multiple regression model

$$y_t = \beta' x_t + u_{1t} \ ; \quad x_t = x_{t-1} + u_{2t} \tag{30}$$

where the time series $\{u_{1t}\}$ and $\{u_{2t}\}$ and independent. If $u_t' = (u_{1t}, u_{2t}')$ has partial sums which obey (3) with limit Brownian motion

$$B(r) = \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} \begin{matrix} 1 \\ m \end{matrix} = BM(\Omega)$$

then

$$\Omega = \begin{bmatrix} \omega_{11} & 0 \\ 0 & \Omega_{22} \end{bmatrix}$$

and B_1 and B_2 are independent.

If $\hat{\beta}$ is the least squares regression coefficient in (30) then

$$T(\hat{\beta}-\beta) \Rightarrow \left(\int_0^1 B_2 B_2'\right)^{-1} \left(\int_0^1 B_2 dB_1\right)$$
$$= \int_{V>0} N(0, \omega_{11} V(B_2)) dP(V)$$

with $V = (\int_0^1 B_2 B_2')^{-1}$. This limit distribution is a simple covariance matrix mixture of normals.

In addition to (31) it is easily shown that random normalization of $\hat{\beta}$ gives

$$(X'X)^{1/2}(\hat{\beta}-\beta) \Rightarrow \left(\int_0^1 B_2 B_2'\right)^{-1/2} \int_0^1 B_2 dB_1$$
$$= N(0, \omega_{11} I) \ .$$

As shown in [31], this theory allows us to examine and modify traditional regression significance tests. For example, the usual Wald statistic for testing the linear hypothesis

$$H_0 : R\beta = r \quad R(q\times m) \text{ of rank } q$$

is

$$W = (R\hat{\beta} - r)\left[R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta} - r)/s^2 \ , \quad s^2 = T^{-1}\Sigma_1^T \hat{u}_{1t}^2$$
$$= \gamma_T' \gamma_T / s^2$$

where

$$\gamma_T = \left[R(X'X)^{-1}R'\right]^{-1/2}(R\hat{\beta} - r) \Rightarrow N(0, \omega_{11} I_q)$$

under H_0 . Hence

$$W \Rightarrow (\omega_{11}/\sigma_{11})\chi_q^2$$

leading to tests with the wrong asymptotic size when $\omega_{11} \neq \sigma_{11}$ (i.e. when u_{1t} displays serial correlation). To accommodate the general case we simply need to replace s^2 in W by $\hat{\omega}_{11}$, any consistent estimator of ω_{11} , giving

$$\hat{W} = (R\hat{\beta} - r)\left[R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta} - r)/\hat{\omega}_{11}$$
$$\Rightarrow \chi^2_q$$

under H_0 .

Questions of asymptotic efficiency can also be examined in the context of models like (30). Writing (30) in the linear model format (with vectors and matrices embodying T observations) we have:

$$y = X\beta + u_1 \ , \quad var(u_1) = V \quad (T \times T) \ , \quad say.$$

Write $\hat{\beta} = (X'X)^{-1}X'y$ for ordinary least squares (OLS) and $\tilde{\beta} = (X'V^{-1}X)^{-1}(X'V^{-1}y)$ for generalized least squares (GLS). The conditional covariance matrices of these estimators are just:

$$var(\hat{\beta}|X) = (X'X)^{-1}X'VX(X'X)^{-1}$$
$$var(\tilde{\beta}|X) = \left(X'V^{-1}X\right)^{-1}$$

and appropriately scaled these have limits

$$\text{OLS:} \quad T^2(X'X)^{-1}X'VX(X'X)^{-1} \Rightarrow \omega_{11}\left(\int_0^1 B_2B_2'\right)^{-1}$$
$$\text{GLS:} \quad T^2\left(X'V^{-1}X\right)^{-1} \Rightarrow \omega_{11}\left(\int_0^1 B_2B_2'\right)^{-1}$$

which are the same. The unconditional asymptotic covariance matrix is

$$\omega_{11}E\left(\int_0^1 B_2B_2'\right)^{-1} \ .$$

This shows that OLS and GLS are asymptotically equivalent when the regressors in (30) are I(1) and strictly exogenous and the errors in (30) are stationary with continuous spectrum at the origin giving $\omega_{11} = 2\pi f_{u_1}(0)$. This result has recently been shown by Phillips and Park [31] for the case of autoregressive errors u_{1t} . It extends the theory of Grenander and Rosenblatt [10] on the efficiency of least squares in time series regressions with deterministic regressors.

6. ASYMPTOTIC EXPANSIONS

The above theory deals with first order asymptotics. Higher order asymptotics may also be developed. Since this involves an extension of functional limit theory such as (3) to accommodate correction terms the mathematical theory is difficult and different from the conventional theory of Edgeworth expansions in Euclidean space asymptotics. In order to find the first correction term, however, the theory is not difficult in many cases and has been discussed and applied recently in [24].

To fix ideas we take a scalar model with a unit root

$$y_t = ay_{t-1} + u_t \; ; \quad a = 1 \, , \quad y_0 = 0 \tag{32}$$

and $\{u_t\}$ stationary with zero mean, variance σ^2, zero third cumulants and spectra:

$$f^{(1)}(\lambda) = \text{spectral density of } \{u_t\}$$

$$f^{(2)}(\lambda) = \text{spectral density of } u_t^2 - E(u_t^2)$$

$$= 2\int_{-\pi}^{\pi} f^{(1)}(\lambda-\alpha) f^{(1)}(\alpha) d\alpha \quad \text{if } \{u_t\} \text{ is Gaussian.}$$

In this case it is shown in [24] that the expansion of the distribution of $T(\hat{a}-1)$, where $\hat{a}$ is the least squares regression coefficient in (32), can be written in the form:

$$T(\hat{a}-1) = \frac{\int_0^1 BdB + \lambda}{\int_0^1 B^2} - \frac{1}{2\sqrt{T}} \frac{\xi}{\int_0^1 B^2} + O_p(T^{-1}) \tag{33}$$

where

$$\xi = N(0, \, 2\pi f^{(2)}(0)) \quad \text{and independent of } B$$

$$B = BM(\omega^2) \, , \quad \omega^2 = 2\pi f^{(1)}(0)$$

$$\lambda = (2\pi f^{(1)}(0) - \sigma)/2 \, , \quad \sigma^2 = E(u_t^2) \; .$$

To give an example, we consider the first order moving average

$$u_t = \epsilon_t + \theta\epsilon_{t-1}$$

where $\{\epsilon_t\}$ is iid $N(0,1)$. Then

$$\lambda = \theta \, , \quad 2\pi f^{(1)}(0) = (1+\theta)^2$$

$$2\pi f^{(2)}(0) = 2(1 + 4\theta^2 + \theta^4) \; .$$

We write $B = \omega W$, $W = BM(1)$ and we have:

$$T(\hat{a}-1) = \frac{\int_0^1 WdW + \theta/(1+\theta)^2}{\int_0^1 W^2} - \frac{1}{\sqrt{2T}} \frac{\left(1 + 4\theta^2 + \theta^4\right)^{1/2}}{(1+\theta)^2} \frac{Z}{\int_0^1 W^2} + O_p(T^{-1}) \tag{34}$$

where $Z = N(0,1)$ and is independent of W.

The correction terms in (33) and (34) may be used to examine the adequacy of the first order asymptotics for different regions of the parameter space. Note in (34) that as $\theta \to -1$ the correction terms grow large very quickly. At $\theta = -1$ the asymptotics fail since there is a common factor in (32) and the data generating mechanism reduces to

$$y_t = \varepsilon_t = I(0) \ .$$

The reader is referred to [24] for a more detailed analysis and extensions to vector autoregressions.

7. LIMITING GAUSSIAN FUNCTIONAL FAMILIES OF DISTRIBUTIONS

All of the limit distributions considered in earlier sections of this paper may be written in a simple form involving a matrix ratio of quadratic functionals of certain stochastic processes. This formulation suggests that the criterion function that underlies the estimator may itself admit a related asymptotic approximation that involves the same stochastic processes. This approach and some of its connections with the LAN and LAMN families of LeCam [16] and Jeganathan [13, 14] have been studied recently in Phillips [29].

Let $\Lambda_T(h)$ denote a sample objective criterion suitably centered and scaled so that its argument h measures scaled deviations from some fixed parameter value, say θ_0 . Optimization of Λ_T then leads to an optimization estimator, say $\hat{\theta}$, and the associated deviation $h = \delta_T^{-1}(\hat{\theta} - \theta_0)$ for some sequence of scale factors δ_T . In a case of consistent estimation $\delta_T \to 0$ but for estimators that converge with probability zero we can set $\delta_T = 1$ for all T .

In [29] we defined a limiting Gaussian functional (LGF) family as follows. We say that $\Lambda_T(h)$ satisfies the LGF condition if:

$$\Lambda_T(h) - (h'W_T - (1/2)h'S_T h) \underset{p}{\to} 0 \tag{35}$$

for some n-vector W_T and $n \times n$ matrix S_T ; and

$$(W_T, S_T) \Rightarrow (\textstyle\int_0^1 MdN + \lambda, \int_0^1 MM') \tag{36}$$

where the elements of M are square integrable and lie in $D[0,1]^n$. N(r) is a Gaussian random function with sample paths in C[0,1] , the space of continuous functions on [0,1]. λ is a constant vector.

A simple example is the Gaussian AR(1):

$$y_t = \theta_0 y_{t-1} + u_t \ ,$$

with u_t iid N(0,1) and $y_0 = 0$. Writing $\theta = \theta_0 + \delta_T h$ we have

$$\Lambda_T(h) = \ell n\{pdf(y;\theta)/pdf(y;\ \theta_0)\}$$

$$= -(1/2)\Sigma_1^T(y_t - \theta y_{t-1})^2 + (1/2)\Sigma_1^T(y_t - \theta_0 y_{t-1})^2$$

$$= h(\delta_T\Sigma_1^T y_{t-1}u_t) - (1/2)h^2(\delta_T^2\Sigma_1^T y_{t-1}^2)\ .$$

Here

$$\delta_T = \begin{cases} T^{-1/2} & , \ |\theta_0| < 1 \\ (\theta_0^2 - 1)/\theta_0^T & , \ |\theta_0| > 1 \\ T^{-1} & , \ \theta_0 = 1 \end{cases}$$

and

$$\Lambda_T(h) \Rightarrow \Lambda(h)$$

with

(i) $\Lambda(h) = hY(\theta_0)Z - (1/2)h^2Y(\theta_0)^2\ , \ |\theta_0| < 1$
with $Z \equiv N(0,1)\ , \ Y(\theta_0) = (1 - \theta_0^2)^{-1/2}\ ,$

(ii) $\Lambda(h) = hYZ - (1/2)h^2Y^2\ , \ |\theta_0| > 1$
with $Z \equiv N(0,1)$ and independent of $Y \equiv N(0,1)$

(iii) $\Lambda(h) = h\int_0^1 WdW - (1/2)h^2\int_0^1 W^2\ , \ \theta_0 = 1$
with $W(r) \equiv BM(1)$.

To include these three cases in (36) we write:

(i) $|\theta_0| < 1$: $M(r) = 1(r)Y(\theta_0)$ with $1(r) = 1\ , \ 0 \le r \le 1$
$N(r) = W(r) \equiv BM(1)$

(ii) $|\theta_0| > 1$: $M(r) = 1(r)Y\ ;\ N(r)$
$N(r) = W(r) \equiv BM(1)$

(iii) $\theta_0 = 1$: $M(r) = N(r) = W(r) \equiv BM(1)$

and

$$\Lambda(h) = h\int_0^1 MdN - (1/2)h^2\int_0^1 M^2\ .$$

Note that when $\theta_0 = 1$ and when $\theta = 1 + h/T$ the quadratic term in (35) satisfies

$$S_T(\theta) = T^{-2}\Sigma_1^T y_{t-1}^2 \Rightarrow \int_0^1 J_h^2 = S(\theta_0,\ h)$$

where

$$J_h = \int_0^r e^{(r-s)h}dW(s)$$

is the diffusion process introduced in Section 3.2 above. Indeed, when $\theta = 1 + h/T$ the process y_t (strictly, $y_{t,T}$ since θ depends on T)

forms a triangular array of near integrated time series. Since

$$S(\theta_0, h) \neq S(\theta_0) - \int_0^1 W^2$$

for all $h \neq 0$, we see that the quadratic approximation to $\Lambda_T(h)$ varies over arrays with different h. This leads in the limit to what we call in [29] variable random information in the limit. In effect, the curvature of $\Lambda(h)$ in this case depends on $\int_0^1 J_h^2$ and this random information varies for different h.

By contrast in case (i) the Fisher information is constant and in case (ii) it is random but independent of h. In both these cases, the LAMN theory applies. However, the LAMN theory does not apply in case (iii) and the limit distribution is not mixed normal. This case has recently been discussed by Jeganathan [15] and Phillips [29].

BIBLIOGRAPHY

1. Box, G. E. P. and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco, 1976.

2. Campbell, J. Y. and N. G. Mankiw, "Permanent and transitory components in macroeconomic fluctuations," *AEA Papers and Proceedings*, 77 (1987), 111-117.

3. Campbell, J. Y. and R. J. Shiller, "Cointegration and tests of present value models," mimeographed. Cowles Foundation for Research in Economics, Yale University, 1987.

4. Chan, N. H. and C. Z. Wei, "Asymptotic inference for nearly nonstationary AR(1) processes," *Annals of Statistics*, 15 (1987), 1050-1063.

5. Eberlain, E., "On strong invariance principles under dependence assumptions," *Annals of Probability, 14 (1986), 260-270.*

6. Frisch, R., *Statistical Confluence Analysis by Means of Complete Regression Systems*, University Economics Institute, Oslo, 1934.

7. Granger, C. W. J., "Some properties of time series data and their use in econometric model specification," *J. of Econometrics*, 16 (1981), 121-130.

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

9. Granger, C. W. J. and P. Newbold, *Forecasting Economic Time Series*, Academic Press, New York (2nd Edition), 1986.

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

11. Hall, R. E., "Stochastic implications of the life-cycle permanent income hypothesis: Theory and evidence," *J. of Political Economy*, 86 (1978), 971-987.

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

13. Jeganathan, P., "An extension of a result of L. LeCam concerning asymptotic normality," *Sankhya Series A*, 42 (1980), 146-160.

14. Jeganathan, P., "On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal," *Sankhya Series A*, 44 (1982), 173-212.

15. Jeganathan, P., "On the asymptotic behavior of least squares estimators in AR time series with roots near the unit circle: Parts I and II," University of Michigan Technical Report #150, 1987.

16. LeCam, L., *Asymptotic Methods in Statistical Decision Theory*, Springer, New York, 1986.

17. Nelson, Charles R. and Charles I. Plosser, "Trends and random walks in macroeconomic time series: Some evidence and implications," *J. of Monetary Economics*, 10 (1982), 139-162.

18. Newey, W. K. and K. D. West, "A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55 (1987), 703-708.

19. Park, J. Y. and P. C. B. Phillips, "Statistical inference in regressions with integrated processes: Part 1," Cowles Foundation Discussion Paper No. 811, 1986.

20. Park, J. Y. and P. C. B. Phillips, "Statistical inference in regressions with integrated processes: Part 2," Cowles Foundation Discussion Paper No. 819, 1987.

21. Perron, P. and P. C. B. Phillips, "Does GNP have a unit root? A reevaluation," *Economics Letters* (1987).

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

23. Phillips, P. C. B., "Time series regression with a unit root," *Econometrica*, 55 (1987), 277-301.

24. Phillips, P. C. B., "Asymptotic expansions in nonstationary vector autoregressions," *Econometric Theory*, 3 (1987), 45-68.

25. Phillips, P. C. B., "Towards a unified asymptotic theory of autoregression," *Biometrika*, 74 (1987), 535-547.

26. Phillips, P. C. B., "Regression theory for near integrated time series," *Econometrica* (1988), forthcoming.

27. Phillips, P. C. B., "Weak convergence to the matrix stochastic integral $\int_0^1 BdB'$," *J. of Multivariate Analysis* (1988), forthcoming.

28. Phillips, P. C. B., "Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations," Cowles Foundation Discussion Paper No. 846, July 1987.

29. Phillips, P. C. B., "Partially identified econometric models," Cowles Foundation Discussion Paper No. 845, August 1987.

30. Phillips, P. C. B. and S. N. Durlauf, "Multiple time series regression with integrated processes," *Review of Economic Studies*, 53 (1986), 473-496.

31. Phillips, P. C. B. and J. Y. Park, "Asymptotic equivalence of OLS and GLS in regressions with integrated regressors," *J. of the American Statistical Society* (1988), forthcoming.

32. Phillips, P. C. B. and S. Ouliaris, "Testing for cointegration using principle components methods," Cowles Foundation Discussion Paper No. 809R, August 1987.

33. Phillips, P. C. B. and S. Ouliaris, "Asymptotic properties of residual based tests for cointegration." Cowles Foundation Discussion Paper No. 847, August 1987.

34. Phillips, P. C. B. and P. Perron, "Testing for a unit root in time series regression," *Biometrika* (1988), forthcoming.

35. Shiller, R. J., "The volatility of stock market prices," *Science*, 235 (1987), 33-37.

36. Stock, J. H. and M. W. Watson, "Does GNP have a unit root," *Economics Letters* (1987).

37. Stock, J. H. and M. W. Watson, "Testing for Common trends," Harvard Institute of Economic Research Discussion Paper #1222, 1987.

38. Strasser, H., "Martingale difference arrays and stochastic integrals," *Probability Theory and Related Fields*, 72 (1986), 83-98.

39. White, J. S., "The limiting distribution of the serial correlation coefficient in the explosive case," *Annals of Mathematical Statistics*, 29 (1958), 1188-1197.

40. Yule, G. U., "On the time correlation problem with especial reference to the variate difference correlation method," *J. of the Royal Statistical Society*, 84 (1921), 497-526.

41. Yule, G. U., "Why do we sometimes get nonsense-correlations between time series"--A study in sampling and the nature of time series," *J. of the Royal Statistical Society*, 89 (1926), 1-64.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06520

Current Address:
COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
BOX 2125 YALE STATION
NEW HAVEN, CONNECTICUT 06520-2125