COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station

New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 802

ASYMPTOTIC EQUIVALENCE OF OLS AND GLS IN REGRESSIONS WITH INTEGRATED REGRESSORS

Peter C.B. Phillips & Joon Y. Park

September 1986

ASYMPTOTIC EQUIVALENCE OF OLS AND GLS IN REGRESSIONS

WITH INTEGRATED REGRESSORS*

by

P. C. B. Phillips and Joon Y. Park

Cowles Foundation for Research in Economics
Yale University

0. ABSTRACT

In the multiple regression model $y_t = x_t'\beta + u_t$ where $\{u_t\}$ is stationary and x_t is an integrated m-vector process it is shown that the asymptotic distributions of the ordinary least squares (OLS) and generalized least squares (GLS) estimators of β are identical. This generalizes a recent result obtained by Krämer (1986) for simple two variate regression. Our approach makes use of a multivariate invariance principle and yields explicit representations of the asymptotic distributions in terms of functionals of vector Brownian motion. Some useful asymptotic results for hypothesis tests in the model are also provided.

Key Words: Asymptotic efficiency; Integrated regressors; Invariance principle; Multiple regression; Vector Brownian motion.

August 1986

---

*Our thanks go to Glena Ames for her skill and effort in typing the manuscript of this paper and to the NSF for research support under Grant No. SES 8519595. This paper was completed while Joon Park was an Alfred P. Sloan Doctoral Dissertation Fellow.

1. INTRODUCTION

Conditions under which least squares regression is efficient have been of interest to statisticians for many years. Necessary and sufficient conditions for the equivalence of OLS and GLS in finite samples were given by Kruskal (1968), Zyskind (1967) and Rao (1967). These conditions are of great theoretical importance, but are less important in practice since they are so seldom satisfied, particularly in time series regressions. In infinite samples, on the other hand, the situation is rather different. Here the central result is due to Grenander and Rosenblatt (1957). These authors showed that for a regression with fixed regressors and stationary errors least squares is asymptotically efficient if and only if the spectrum of the error process is constant on the elements of the regression spectrum. This condition is known to be satisfied in many cases of importance in time series, including regressions on polynomial and trigonometric functions of time. Thus, if a time series is stationary about a deterministic trend an investigator may detrend the series by a least squares regression on a polynomial of time and then analyze the resulting series without any loss of (asymptotic) efficiency.

Frequently, we are interested in regressions that involve stochastic regressors of independent interest instead of deterministic functions. In economics for example, long run regularities between various macroeconomic variables often suggest formulations of regressions in terms of the levels or log levels of the relevant time series. Since such time series are typically nonstationary and nonergodic the results of Grenander and Rosenblatt (1957) on the efficiency of least squares do not strictly apply. However, if the errors in a regression relating such time series are stationary and if the regressors are integrated processes of the ARIMA type then we might expect least squares still to be asymptotically efficient. Intuitively, this is because the regression spectrum has a singularity at the origin in this case, so that power is effectively concentrated at a single point--the zero frequency. As we shall show in the present paper, this intuition is correct and the Grenander-Rosenblatt result does indeed extend to this type of regression.

A simple example of this phenomenon was recently discovered by Krämer (1986). Krämer studied a two variable regression model driven by a stationary AR(m) error process and with a regressor generated by an ARIMA(p,1,q) model. He demonstrated the asymptotic equivalence of OLS and GLS in this regression. But he did not find the limiting distribution of these estimators and his method of derivation does not easily generalize to multiple regressions.

In the present paper we shall deal directly with the multiple regression case. Our method of proof relies on the theory of weak convergence and yields generalizations of Kramer's results in a very straightforward manner. Proofs are given in the Appendix to the paper.

2. EFFICIENCY OF OLS

We consider the regression model

(1) $$y_t = x_t'\beta + u_t \; ; \quad t = 1, 2, \ldots$$

where $\{u_t\}_{-\infty}^{\infty}$ follows a zero mean stationary AR(p) process and $\{x_t\}_0^{\infty}$ is an m-dimensional multiple time series that is generated recursively by

(2) $x_t = x_{t-1} + v_t \; ; \quad t = 1, 2, \ldots \; .$

We assume that the innovation sequences $\{u_t\}_{-\infty}^{\infty}$ and $\{v_t\}_1^{\infty}$ in (1) and (2) are statistically independent, so that the regressors in (1) are strictly exogenous. Our results do not depend on the initialization of (2). We allow x_0 to be any random variable (with a fixed probability distribution) including, of course, a constant.

We define $w_t' = (u_t, v_t')$ and we require only that the partial sum process $S_t = \Sigma_1^t w_k$ satisfies a multivariate invariance principle. More specifically, if

$$X_T(r) = T^{-1/2} S_{j-1} \; , \quad (j-1)/T \le r < j/T$$

then

(3) $$X_T(r) \Rightarrow B(r) \quad \text{as} \quad T \uparrow \infty \; .$$

Here, T denotes the sample size, the symbol " $\Rightarrow$ " signifies weak convergence of the associated probability measures and $B(r)$ is n-vector Brownian motion $(n = m+1)$ with nonsingular covariance matrix

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

(4) $$= \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \Sigma_2 \end{bmatrix} .$$

Since $\{u_t\}$ and $\{v_t\}$ are independent we have $B(r)' = (B_1(r), B_2(r)')$ where $B_1(r)$ and $B_2(r)$ are independent Brownian motions of dimension one and m , respectively, with variance matrices σ_1^2 and Σ_2 .

Multivariate invariance principles of this type have recently been proved by Eberlain (1986) and Phillips and Durlauf (1986). They apply for a very wide class of innovation sequences $\{w_t\}$ that are weakly dependent and possibly heterogeneously distributed. Following Hall and Heyde (1980, p. 146), they may also be shown to apply to a large class of linear processes, including those generated by all stationary and invertible ARMA models.

When $\{w_t\}$ is stationary with spectral density matrix $f_{ww}(\lambda)$ then (4) may be written in the form:

$$\Sigma = 2\pi f_{ww}(0) = 2\pi \begin{bmatrix} f_u(0) & 0 \\ 0 & f_{vv}(0) \end{bmatrix} .$$

If $\{u_t\}$ is generated by

$$(5) \qquad \Sigma_{j=0}^{p} \rho_j u_{t-j} = \varepsilon_t , \quad \rho_0 = 1$$

where $\{\varepsilon_t\}$ is $\mathrm{iid}(0, \sigma^2)$ and the roots of $\Sigma_{j=0}^{p} \rho_j z^j = 0$ lie outside the unit circle then

$$\sigma_1^2 = 2\pi f_u(0) = \left(\Sigma_{j=0}^{p} \rho_j\right)^{-1} \sigma^2 .$$

For the purposes of this section of the paper we assume that (5) is the model generating $\{u_t\}$.

We now write (1) in conventional matrix form for a sample of T observations as $y = X\beta + u$. In this set up, the asymptotic distribution of the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ is easily obtained. In fact, this is just a special case of a more general result in Phillips and Durlauf (1986, Theorem 4.1). In particular we have:

LEMMA 2.1. *As* $T \uparrow \infty$

$$T(\hat{\beta}-\beta) \Rightarrow \left[\int_0^1 B_2(r)B_2(r)'dr\right]^{-1} \left[\int_0^1 B_2(r)dB_1(r)\right] \tag{6}$$

where $B(r)' = (B_1(r), B_2'(r))$ *is n-vector Brownian motion with covariance matrix* (4).

In this Lemma we represent the asymptotic distribution of the OLS estimator as a simple functional of vector Brownian motion. The integral $\int_0^1 B_2 dB_1$ in (6) is interpreted as a vector of stochastic integrals with respect to the univariate Brownian motion $B_1(r)$. The matrix $\int_0^1 B_2 B_2' dr$ is a quadratic functional of the vector Brownian motion $B_2(r)$ and is nonsingular with probability one.

The representation (6) is very useful in what follows. Not only does it enable us to obtain a very elegant demonstration of the asymptotic efficiency of OLS in the model (1). It also leads us easily to some interesting consequences concerning the distribution of statistical tests. These are explored in the following section. Finally, we note from (6) that $\hat{\beta} = \beta + O_p(T^{-1})$ and $\hat{\beta}$ is, of course, a consistent estimator of β .

The GLS estimator of β in (1) is given by $\tilde{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y)$ where $E(uu') = \sigma^2\Omega$. As is well known, $\tilde{\beta}$ can be regarded as the OLS estimator of the coefficient vector in the transformed model $y^* = X^*\beta + u^*$ where y^* , X^* , u^* are obtained from y , X , u by premultiplying a nonsingular matrix R such that $R'R = \Omega^{-1}$. Our first main result follows from a direct application of Lemma 2.1 to this transformed model.

THEOREM 2.2. $T(\hat{\beta}-\beta)$ *and* $T(\tilde{\beta}-\beta)$ *have the same limiting distribution as* $T \uparrow \infty$.

Theorem 2.2 can be viewed as an extension to the multiple regression case of Kramer's (1986) theorem 3. However, our result is stronger and more clearly demonstrates the asymptotic efficiency of OLS than Krämer's theorem 3. Krämer shows the asymptotic equivalence of $(X^{*\prime}X^{*})^{1/2}(\hat{\beta}-\beta)$ and $(X^{*\prime}X^{*})^{1/2}(\tilde{\beta}-\beta)$ for the case of a scalar coefficient β. He does not establish the asymptotic equivalence of $T(\hat{\beta}-\beta)$ and $T(\tilde{\beta}-\beta)$; and he does not obtain the limiting distribution of the estimators.

THEOREM 2.3. *As* $T \uparrow \infty$:

(a) $(X'X)^{1/2}(\hat{\beta}-\beta) \Rightarrow N(0, \sigma_1^2 I)$;

(b) $(X'\Omega^{-1}X)^{1/2}(\hat{\beta}-\beta) \Rightarrow N(0, \sigma^2 I)$.

Both (a) *and* (b) *remain true if* $\hat{\beta}$ *is replaced by* $\tilde{\beta}$.

The asymptotic normality of $(X'X)^{1/2}(\hat{\beta}-\beta)$ is obtained in the proof of Theorem 2.3 by a very simple conditioning argument. This normality is important and useful in the formulation of statistical tests. In particular, it implies that the conditional F-statistic for testing a linear hypothesis in (1) has an asymptotic chi squared distribution upon appropriate standardization, as is the case for the standard regression model. Also the difference in the variances of the two limiting distributions in Theorem 2.3 should be noted. This has some interesting consequences which will be elaborated below.

3. STATISTICAL TESTS

Suppose we wish to test the linear hypothesis

$$H_0 : R\beta = r$$

where R is $q \times m$ of rank $q < m$. The following Theorem gives the main results on the asymptotic distribution of Wald-type test statistics.

THEOREM 3.1. *Under the null hypothesis* H_0 *and as* $T \uparrow \infty$:

(a) $(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r)/\sigma_1^2 \Rightarrow \chi_q^2$

(b) $(R\hat{\beta}-r)'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\hat{\beta}-r)/\sigma^2 \Rightarrow \chi_q^2$.

Both (a) *and* (b) *remain true if* $\hat{\beta}$ *is replaced by* $\tilde{\beta}$.

Using $\tilde{\beta}$ rather than $\hat{\beta}$ in (b) of Theorem 3.1 we have

$$W_1 = (R\tilde{\beta}-r)'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\tilde{\beta}-r)/\sigma^2$$

which is the Wald statistic for testing H_0 in the standard linear regression model with nonstochastic regressors and (known) error covariance matrix $\sigma^2\Omega$. It is interesting to note that W_1 still has a limiting χ_q^2 distribution even when x_t is rather general integrated process generated by (2). As the proof of the Theorem makes clear, this result holds because of the strict exogeneity of x_t , without which the conditioning argument used in the proof does not go through. When the innovation sequences $\{u_t\}$ and $\{v_t\}$ that drive (1) and (2) are dependent the limiting distributions of statistics like W_1 are no longer χ^2 . The reader is referred to Phillips and Durlauf (1986) for results which apply in this case.

To make the tests in Theorem 3.1 operational for statistical inference we need a consistent estimator of σ^2 . It is simple to show:

THEOREM 3.2

(a) $\hat{\sigma}^2 = T^{-1}(y - X\tilde{\beta})'\Omega^{-1}(y - X\tilde{\beta}) \xrightarrow[p]{} \sigma^2$,

(b) $\hat{\sigma}_1^2 = (\Sigma_{j=0}^{p}\rho_j)^{-2}\hat{\sigma}^2 \xrightarrow[p]{} \sigma_1^2$.

These estimators depend on Ω and the AR coefficients ρ_j . When the order p of the autoregression for u_t is known the coefficients ρ_j may be consistently estimated by the usual two step procedure based on the OLS residuals. Call these consistent estimators $\hat{\rho}_j$ and write $\hat{\Omega} = \Omega(\hat{\rho})$. Then

$$s^2 = T^{-1}(y - X\tilde{\beta})'\hat{\Omega}^{-1}(y - X\tilde{\beta}) \xrightarrow[p]{} \sigma^2$$

and

$$s_1^2 = \left(\Sigma_{j=0}^{p}\hat{\rho}_j\right)^{-2}s^2 \xrightarrow[p]{} \sigma^2 .$$

These estimated error variances may now be used in statistical tests. We find that

$$(R\tilde{\tilde{\beta}} - r)'\left[R(X'\hat{\Omega}^{-1}X)^{-1}R'\right]^{-1}(R\tilde{\tilde{\beta}} - r)/s^2 \Rightarrow \chi_q^2$$

where we employ the feasible GLS estimator

$$\tilde{\tilde{\beta}} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}y) .$$

On the other hand, if (1) is estimated by OLS, the conventional error variance estimator is:

$$\hat{s}^2 = T^{-1}(y - X\hat{\beta})'(y - X\hat{\beta}) \xrightarrow[p]{} \sigma_u^2 = E(u_t^2) \ .$$

In this case the usual Wald statistic for testing H_0 is:

$$W_2 = (R\hat{\beta} - r)'\left[R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta} - r)/\hat{s}^2 \ ;$$

and we deduce from Theorem 3.1(a) that

$$W_2 \Rightarrow (\sigma_1^2/\sigma_u^2)\chi_q^2 \ .$$

Thus, the conventional Wald statistic for testing H_0 based on an OLS regression has a limiting distribution which is proportional to a χ_q^2. When u_t is generated by (5) the constant of proportionality is $\sigma^2/(\Sigma_{j=0}^p\rho_j)^2\sigma_u^2$. For spherical errors this is unity. For an AR(1) it is $(1 - \rho_1)/(1 + \rho_1)$, which shows that the asymptotic distribution of W_2 can be very different from the conventional χ_q^2 when there is serial correlation.

4. EXTENSIONS TO STATIONARY ERRORS

The results in the previous two sections apply when $\{u_t\}$ is generated by the stationary AR(p) (5). Since a stationary process whose spectral density is positive and continuous may be arbitrarily well approximated by an AR process of finite order it may be expected that our results hold asymptotically for a wider class of models with stationary error processes. The following theorem indicates that this is so and establishes the asymptotic efficiency of OLS for such a wider class of error processes.

THEOREM 4.1. *If* $\{u_t\}$ *is stationary with positive and continuous spectral density* $f_u(\lambda)$ *then as* $T \uparrow \infty$:

$$T^{-2}X'\Omega^{-1}X \Rightarrow \frac{\sigma^2}{2\pi f_u(0)} \int_0^1 B_2(r)B_2(r)'dr \ ;$$

$$T^{-2}(X'X)^{-1}(X'\Omega X)(X'X)^{-1} \Rightarrow \frac{2\pi f_u(0)}{\sigma^2}\left\{\int_0^1 B_2(r)B_2(r)'dr\right\}^{-1} \ ;$$

where $B_2(r)$ *is m-dimensional Brownian motion with covariance matrix* Σ_2 .

Note that given X the conditional covariance matrices of the OLS and GLS estimators in (1) are given by:

$$\mathrm{var}(\hat{\beta}|X) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \ ,$$

$$\mathrm{var}(\tilde{\beta}|X) = \sigma^2(X'\Omega^{-1}X)^{-1} \ .$$

Theorem 4.1 shows that asymptotically the conditional covariance matrices of $T(\hat{\beta}-\beta)$ and $T(\tilde{\beta}-\beta)$ are identical. These conditional covariance matrices are represented in terms of a matrix quadratic functional of a vector Brownian motion sample path. Upon integration the unconditional asymptotic covariance matrix of $T(\hat{\beta}-\beta)$ and $T(\tilde{\beta}-\beta)$ is seen to be

$$2\pi f_u(0)E\left\{\int_0^1 B_2(r)B_2(r)'dr\right\}^{-1} \ .$$

5. FINAL REMARKS

The proofs of our results depend heavily on the theory of weak convergence. These methods seem to provide a very convenient way of handling the complications that result from the presence of stochastic regressors generated by ARIMA models. Not only do they provide a means of establishing the asymptotic efficiency of OLS in regressions of this type. They also yield simple representations of the limiting distributions in terms of functionals of Brownian motion. Furthermore, the conditioning argument developed in the proofs of Theorems 2.3 and 3.1 gives us a simple way of demonstrating the validity of conventional asymptotic χ^2 theory for classical tests of linear hypotheses in multiple regression with integrated processes.

The results reported here do not apply in models where there is regressor-error correlation, as in econometric models of simultaneous equations. However, similar techniques may be brought to bear to analyze such regressions. The reader is referred to Phillips and Durlauf (1986) for details.

APPENDIX

Proof of Lemma 2.1. The result can be easily deduced from Theorem 4.1(a) of Phillips and Durlauf (1986).

Proof of Theorem 2.2. The matrix R may be chosen so that

$$x_t^* = \Sigma_{j=0}^p \rho_j x_{t-j} \ , \quad u_t^* = \Sigma_{j=0}^p \rho_j u_{t-j} \tag{A1}$$

for $t > p$. Since the end corrections leading to x_t^* and u_t^* for $1 \leq t \leq p$ do not affect asymptotic results we may assume, without loss of generality, that the transformation (A1) holds for all $t = 1, 2, \ldots$ with the convention that $x_{-p+1} = \ldots = x_0 = 0$. It follows that $v_t^* = x_t^* - x_{t-1}^* = \Sigma_{j=0}^p \rho_j v_{t-j}$ and

$$w_t^* = \Sigma_{j=0}^p \rho_j w_{t-j} \tag{A2}$$

where $w_t^{*\prime} = (u_t^*, v_t^{*\prime})$ and where, without loss of generality in what follows, we may set $w_{-p+1} = \ldots = w_0 = 0$.

The new process $\{w_t^*\}_1^\infty$ defined by (A2) has partial sums which satisfy the multivariate invariance principle (3). For instance, if $\{w_t\}_1^\infty$ is strong mixing with mixing numbers α_k that satisfy $\Sigma_1^\infty \alpha_k^{1-2/\delta} < \infty$ for some $\delta > 2$ then the same is true of the transformed sequence $\{w_t^*\}_1^\infty$ (see, for example, White (1984, p. 153)). In fact, using $X_T^*(r)$ to denote the random element constructed from partial sums of w_t^* we have, in place of (3),

$$X_T^*(r) \Rightarrow B^*(r)$$

where $B^*(r)$ is vector Brownian motion with covariance matrix $\Sigma^* = (\Sigma_{j=0}^{p}\rho_j)\Sigma$. When $\{w_t\}$ is stationary the new covariance matrix $\Sigma^* = 2\pi f_{w^*w^*}(0) = (\Sigma_{j=0}^{p}\rho_j)^2(2\pi f_{ww}(0))$ may be deduced quite simply from the action of the linear filter (A2). In the general case, we need only write:

$$\text{(A3)} \qquad S_T^* = \Sigma_1^T w_t^* = (\Sigma_{j=0}^{p}\rho_j)(\Sigma_1^T w_t) - \Sigma_{j=1}^{p}\rho_j \Sigma_{t=T-j+1}^{T} w_t ;$$

and it follows that:

$$\lim_{T\to\infty} T^{-1}E(S_T^* S_T^{*\prime}) = \varphi^2\Sigma = \Sigma^* .$$

where $\varphi = \Sigma_{j=0}^{p}\rho_j$.

Notice that the transformed model is driven by the new process $\{w_t^*\}_1^\infty$ in exactly the same way as the original model (1) is by $\{w_t\}$. Since $\tilde{\beta}$ is the OLS estimator of β in the transformed model we now deduce from Lemma 2.1 that

$$T(\tilde{\beta}-\beta) \Rightarrow \left[\int_0^1 B_2^*(r)B_2^*(r)'dr\right]^{-1}\left[\int_0^1 B_2^*(r)dB_1^*(r)\right] .$$

However, $B^*(r) \equiv \varphi B(r)$ where the symbol " $\equiv$ " signifies equality in distribution. Asymptotic equivalence now follows since, by cancellation of the scale factor φ^2 , we have:

$$\left[\int_0^1 B_2^*(r)B_2^*(r)'dr\right]^{-1}\left[\int_0^1 B_2^*(r)dB_1^*(r)\right] \equiv \left[\int_0^1 B_2(r)B_2(r)'dr\right]^{-1}\left[\int_0^1 B_2(r)dB_1(r)\right]$$

as required.

<u>Proof of Theorem 2.3</u>. From Lemma 3.1(b) of Phillips and Durlauf (1986) we have:

(A4) $$T^{-2}X'X \Rightarrow \int_0^1 B_2(r)B_2(r)'dr$$

and so by the continuous mapping theorem (cmt) and Lemma 2.1 we find:

$$(X'X)^{1/2}(\hat{\beta}-\beta) \Rightarrow \left[\int_0^1 B_2(r)B_2(r)'dr\right]^{-1/2} [\int_0^1 B_2(r)dB_1(r)] \ .$$

Now suppose the n (= m+1) dimensional Brownian motion $B(r)$ is defined on the probability space (Ω,F,P) and let F_2 denote the sub σ-field of F that is generated by $\{B_2(r) : 0 \le r \le 1\}$. We use the symbol " $\cdot|F_2$ " to signify the conditional distribution relative to F_2 in what follows. Since $B_1(r)$ is Gaussian and independent of $B_2(r)$ we deduce that:

$$\int_0^1 B_2(r)dB_1(r)\Big|_{F_2} \equiv N(0,\ \sigma_1^2\int_0^1 B_2(r)B_2(r)'dr)$$

and

$$\left[\int_0^1 B_2(r)B_2(r)'dr\right]^{-1/2} [\int_0^1 B_2(r)dB_2(r)]\Big|_{F_2} \equiv N(0,\ \sigma_1^2 I) \ .$$

However since the latter distribution does not depend on realizations of $B_2(r)$ it is also the unconditional distribution. Part (a) of the theorem follows immediately.

To prove part (b) we first show that as $T \uparrow \infty$

(A5) $$\|X_T^*(r) - \varphi X_T(r)\| = \max_i \sup_r |X_{Ti}^*(r) - \varphi X_{Ti}(r)| \xrightarrow[p]{} 0$$

where $\varphi = \Sigma_{j=0}^p \rho_j$. We note that for $(k-1)/T \le r < k/T$ we have

$$|X^*_{Ti}(r) - \varphi X_{Ti}(r)| = T^{-1/2}|\Sigma_1^{k-1}\Sigma_{s=0}^{p}\rho_s w_{i,t-s} - \varphi\Sigma_1^{k-1}w_{it}|$$

$$= T^{-1/2}|\Sigma_{s=0}^{p}\rho_{s+1}\Sigma_{k-1-s}^{k-1}w_{ir}|$$

$$\leq T^{-1/2}|\varphi|\Sigma_{k-1-p}^{k-1}|w_{ir}| \ .$$

Thus

$$\|X^*_T(r) - \varphi X_T(r)\| \leq T^{-1/2}p|\varphi|(\max_i \max_t|w_{it}|)$$

$$\xrightarrow[p]{} 0$$

proving (A5). It now follows that as $T \uparrow \infty$:

$$h(X^*_T(r)) - h(\varphi X_T(r)) \xrightarrow[p]{} 0$$

where h is any uniformly continuous functional on $D^n[0,1]$, the product space of n copies of $D[0,1]$. In particular,

$$\int_0^1 X^*_T(r)X^*_T(r)'dr - \varphi^2\int_0^1 X_T(r)X_T(r)'dr \xrightarrow[p]{} 0$$

and we deduce directly that:

(A6) $$T^{-2}X'\Omega^{-1}X - \varphi^2T^{-2}X'X = T^{-2}X*'X* - \varphi^2T^{-2}X'X \xrightarrow[p]{} 0 \ .$$

From (A6) and (A4) we obtain

(A7) $$T^{-2}X'\Omega^{-1}X \Rightarrow \varphi^2\int_0^1 B_2(r)B_2(r)'dr \ .$$

Part (b) of the theorem now follows since $\sigma^2 = \varphi^2\sigma_1^2$. Similar arguments show that parts (a) and (b) remain true when $\hat{\beta}$ is replaced by $\tilde{\beta}$.

Proof of Theorem 3.1. By the cmt and Lemma 2.1 we deduce that:

$$\left\{R(T^{-2}X'X)^{-1}R'\right\}^{-1/2} T(R\hat{\beta} - r)$$

$$\Rightarrow \left[R(\int_0^1 B_2(r)B_2(r)'dr)^{-1}R'\right]^{-1/2} R\left\{\int_0^1 B_2(r)B_2(r)'dr\right\}^{-1} \int_0^1 B_2(r)dB_1(r)$$

$$= N(0, \sigma_1^2 I_q) .$$

The last line follows from the same conditioning argument used earlier in the proof of Theorem 2.3(a). Part (a) of the Theorem now follows from a further application of the cmt. The proof of part (b) makes use of (A7) but is otherwise entirely analogous. The invariance of the results to the replacement of $\hat{\beta}$ by $\tilde{\beta}$ is also straightforward.

Proof of Theorem 3.2

$$\hat{\sigma}^2 = T^{-1}(y - X\tilde{\beta}')'\Omega^{-1}(y - X\tilde{\beta}) = T^{-1}(y^* - X^*\tilde{\beta})'(y^* - X^*\tilde{\beta})$$

$$= T^{-1}u^{*\prime}u^* - T^{-1}(T^{-1}u^{*\prime}X^*)(T^{-2}X^{*\prime}X^*)^{-1}(T^{-1}X^{*\prime}u^*)$$

$$= T^{-1}u^{*\prime}u^* + o_p(1)$$

$$\xrightarrow[p]{} \sigma^2$$

as required for (a). Part (b) follows immediately.

Proof of Theorem 4.1. We first approximate $f_u(\lambda)$ from above and below by the spectra of stationary, finite order autoregressive processes. Thus

$$f_1(\lambda) \le f_u(\lambda) \le f_2(\lambda)$$

where $f_j(\lambda) = (\sigma^2/2\pi)|\Sigma_{k=0}^{p_j} b_{jk} e^{i\lambda k}|^{-2}$, $(j = 1, 2)$. The corresponding covariance matrices satisfy:

$$\sigma^2\Omega_1 \leq \sigma^2\Omega \leq \sigma^2\Omega_2 ;$$

and similarly:

$$X'\Omega_2^{-1}X \leq X'\Omega^{-1}X \leq X'\Omega_1^{-1}X .$$

Now, as in the proof of Theorem 2.3 we obtain:

$$T^{-2}X'\Omega_j^{-1}X \Rightarrow \varphi_j^2\int_0^1 B_2(r)B_2(r)'dr$$

where

$$\varphi_j^2 = \left(\Sigma_{k=0}^{p_j} b_{jk}\right)^2 = \sigma^2/2\pi f_j(0) .$$

Using Skorohod's theorem (see, for example, Billingsley (1979, p. 337)) it is now possible to define a common probability space supporting the random matrices Y_{jT} , Y_j and Z_T for which

$$Y_{jT} \equiv T^{-2}X'\Omega_j^{-1}X$$

$$Y_j \equiv \varphi_j^2\int_0^1 B_2(r)B_2(r)'dr$$

$$Z_T \equiv T^{-2}X'\Omega^{-1}X$$

and such that:

(A8) $\qquad Y_{2T} \leq Z_T \leq Y_{1T}$ a.s.

and

(A9) $\quad Y_{jT} \to Y_j \quad$ a.s.

Take any ω for which (A8) and (A9) hold and let α be any m-vector. Then

$$\alpha' Y_2(\omega)\alpha \le \underline{\lim}_{T\to\infty},\ \overline{\lim}_{T\to\infty}\ \alpha' Z_T(\omega)\alpha \le \alpha' Y_1(\omega)\alpha \ .$$

However φ_1 and φ_2 may be made arbitrarily close to $\sigma^2/2\pi f_u(0)$. It follows that

$$Z_T(\omega) \to Z(\omega) = (\sigma^2/2\pi f_u(0)\varphi_1^2)Y_1(\omega)$$

for each ω . We deduce that

$$T^{-2}X'\Omega^{-1}X \Rightarrow \frac{\sigma^2}{2\pi f(0)} \int_0^1 B_2(r)B_2(r)'dr$$

proving part (a).

To prove (b) we first recall that from earlier arguments

$$T^{-2}X'X \Rightarrow \int_0^1 B_2(r)B_2(r)'dr \ .$$

To determine the limiting behavior of $T^{-2}X'\Omega X$ we now approximate $f_u(\lambda)$ from above and below by the spectra of stationary finite order moving average processes. Thus, let

$$f_1(\lambda) \le f_u(\lambda) \le f_2(\lambda)$$

where $f_j(\lambda) = (\sigma^2/2\pi)|\Sigma_{k=0}^{p_j} a_{jk}e^{i\lambda k}|^2$, $(j = 1, 2)$. Now we have

$$X'\Omega_1 X \le X'\Omega X \le X'\Omega_2 X$$

and proceeding in the same way as before we deduce that

$$T^{-2}X'\Omega X \Rightarrow (2\pi f_u(0)/\sigma^2)\int_0^1 B_2(r)B_2(r)'dr \ .$$

It follows that

$$\left(T^{-2}X'X\right)^{-1}(T^{-2}X'\Omega X)\left(T^{-2}X'X\right)^{-1} \Rightarrow (2\pi f_u(0)/\sigma^2)\left\{\int^1 B_2(r)B_2(r)'dr\right\}^{-1}$$

as required for part (b).

REFERENCES

Billingsley, P. (1979). *Probability and Measure.* Wiley: New York.

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

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

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

Krämer, W. (1986). "Least squares regression when the independent variable follows an ARIMA process," *Journal of the American Statistical Association*, 81, 150-154.

Kruskal, W. (1968). "When are Gauss-Markov and least squares estimators identical? A coordinate free approach," *Annals of Mathematical Statistics*, 39, 70-75.

Phillips, P. C. B. and S. Durlauf (1986). "Multiple time series regression with integrated processes," *Review of Economic Studies* (to appear).

Rao, C. R. (1967). "Least squares using an estimated dispersion matrix and its application to measurement of signals," *Fifth Berkeley Symposium in Mathematical Statistics and Probability*, 1, 355-372. University of California: Berkeley, CA.

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

Zyskind, G. (1967). "On canonical forms, non negative covariance matrices and best and simple least squares linear estimators in linear models," *Annals of Mathematical Statistics*, 38, 1092-1109.