ASYMPTOTIC EQUIVALENCE OF ORDINARY LEAST SQUARES
AND GENERALIZED LEAST SQUARES IN REGRESSIONS
WITH INTEGRATED REGRESSORS

BY

PETER C.B. PHILLIPS
AND
JOON Y. PARK

COWLES FOUNDATION PAPER No. 703

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY
Box 2125, Yale Station
New Haven, Connecticut 06520
1988
Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions With Integrated Regressors

P. C. B. PHILLIPS and JOON Y. PARK*

1. INTRODUCTION

For years, statisticians have been interested in conditions where least squares regression is efficient. In finite samples, necessary and sufficient conditions for ordinary least squares (OLS) and generalized least squares (GLS) equivalence are well known. They arose originally in work by Anderson (1948) and were independently studied by Kruskal (1968), Zyskind (1967), and Rao (1967). More recently, Kariya (1985) and Malley (1986) studied these important conditions. But they are more important in theory than in practice, since they are so seldom satisfied. This is particularly true for time series regressions.

For infinite samples, however, the situation is different. Grenander and Rosenblatt (1957) found a necessary and sufficient condition for OLS to be asymptotically efficient (relative to GLS) in a regression with fixed regressors and stationary errors, requiring that the spectrum of the error process be constant on the elements of the regression spectrum (in effect, those sets where the spectral mass of the regressors is concentrated). This condition is satisfied in many important time series cases, including regressions on polynomial and trigonometric functions of time. Thus one can defend a time series stationary around a polynomial trend (whose spectral mass is at the origin) by performing a least squares regression on a polynomial of time and by taking residuals. The resulting series may then be analyzed by traditional methods without any loss of (asymptotic) efficiency. This approach forms the basis of much applied work (see Anderson 1971).

Of frequent interest, however, are regressions involving stochastic regressors rather than deterministic functions of time. For example, in economics long-run regularities between various macroeconomic variables often suggest regression formulation in terms of the levels or log levels of the relevant time series. Such time series are typically slow moving and usually well represented by simple ARIMA models with a single unit root. Integrated autoregressive integrated moving average (ARIMA) regressors are nonstationary and nonergodic; the results of Grenander and Rosenblatt (1957) on least squares efficiency do not strictly apply. If the errors in a regression relating the time series are stationary, however, and if the regressors are integrated processes, then least squares still might be expected to be asymptotically efficient. An intuitive explanation is as follows: ARIMA processes with a single unit root all have spectra with a singularity (a pole) at the origin, so power is effectively concentrated at a single point; namely, the 0 frequency. If the error spectrum is continuous, it is necessarily constant on the elements of the regression spectrum (the origin) where the spectral power of the regressors is concentrated, suggesting that the Grenander–Rosenblatt condition continues to hold in this stochastic regressor case. This article shows that this intuition is correct—that the Grenander–Rosenblatt result does indeed extend to this type of regression.

Krämer (1986) discovered a simple example of this phe-

* P. C. B. Phillips is Stanley B. Resor Professor of Economics and Professor of Statistics, Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520-2125. Joon Y. Park is Graduate Student and Alfred P. Sloan Doctoral Dissertation Fellow, Department of Economics, Yale University, New Haven, CT 06520-2125. Phillips received support from National Science Foundation Grant SES 85-19595, and Park received support from the Sloan Foundation. The authors are grateful to two referees and an associate editor for helpful comments. Their thanks also go to Glenn Ames for her skill and effort in preparing the manuscript of this article.

Journal of the American Statistical Association
March 1988, Vol. 83, No. 401, Theory and Methods

111
nommenon. He studied a two-variable regression model driven by a stationary AR($p$) error process with a regressor generated by an ARIMA($r$, l, s) model, demonstrating the asymptotic equivalence of OLS and GLS in this regression. But his method of derivation does not easily generalize to multiple regressions.

This article deals directly with the multiple regression case. Our method of proof relies on the theory of weak convergence and yields generalizations of Krämer’s results in a very straightforward manner. (Proofs are given in the Appendix.)

2. EFFICIENCY OF OLS

Consider the regression model

$$y_t = x_t \beta + u_t, \quad t = 1, 2, \ldots ,$$  \hspace{1cm} (1)

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

$$x_t = x_{t-1} + u_t, \quad t = 1, 2, \ldots ,$$  \hspace{1cm} (2)

Assume that the innovation sequences \{u_t\} and \{v_t\} in (1) and (2) are statistically independent, so 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 require only that the partial sum process $S_t = \sum w_s$ satisfies a multivariate invariance principle. More specifically, if

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

then

$$X_T(r) \Rightarrow B(r) \text{ as } T \uparrow \infty.$$  \hspace{1cm} (3)

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

$$\Sigma = \lim_{T \to \infty} T^{-1} \mathbb{E}(S_T S_T') = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \Sigma_2 \end{bmatrix}$$ \hspace{1cm} (4)

Since \{u_t\} and \{v_t\} are independent, $B(r)' = (B_1(r), B_2(r))'$ where $B_1(r)$ and $B_2(r)$ are independent Brownian motions of dimension 1 and $m$, respectively, with variance matrices $\sigma_1^2$ and $\Sigma_2$.

Multivariate invariance principles of this type were proved by Eberlaffin (1986) and Phillips and Durlauf (1986). They apply for a very wide class of weakly dependent and possibly heterogeneous distributed innovation sequences \{w_t\}, including processes that are strong mixing under simple conditions on the mixing decay rates but not processes with long memory (e.g., fractionally integrated processes). Following Hall and Heyde (1980, p. 146), invariance principles such as (3) may also be shown to apply to a large class of linear processes, including those generated by all stationary and invertible autoregressive moving average (ARMA) models.

When \{w_t\} is stationary with spectral density matrix $f_{ww}(\lambda)$, then (4) may be written as

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

Throughout this article we assume that \{u_t\} is generated by the AR($p$) model

$$\rho_j u_{t-j} = e_t, \quad \rho_0 = 1,$$ \hspace{1cm} (5)

where \{e_t\} is iid($0, \sigma^2$) and the roots of $\sum_{j=0}^{p} \rho_j z^j = 0$ lie outside the unit circle. Then

$$\sigma_1^2 = 2\pi f_x(0) = \left( \sum_{j=0}^{p} \rho_j \right)^{-2} \sigma^2.$$  \hspace{1cm} (6)

For a sample of $T$ observations (1) is written in conventional matrix form as $y = X\beta + u$. The asymptotic distribution of the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ is easily obtained, this being a special case of a more general result of Phillips and Durlauf (1986, theorem 4.1).

**Lemma 2.1.** As $T \uparrow \infty$, \n
$$T(\hat{\beta} - \beta) \overset{d}{\to} \int_0^T B_1(r) dB_2(r) dr,$$  \hspace{1cm} (6)

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

In this lemma the asymptotic distribution of the OLS estimator is a simple functional of vector Brownian motion. The integral $\int_0^T B_1(r) dB_2(r)$ in (6) is interpreted as a vector of stochastic integrals with respect to the univariate Brownian motion $B_1(r)$. The matrix $\int_0^T B_1(r) dB_2(r) dr$ is a quadratic functional of the vector Brownian motion $B_2(r)$ and is nonsingular with probability 1.

The representation (6) is useful in what follows: It allows demonstration of the asymptotic efficiency of OLS in the model (1), and leads to some interesting consequences concerning the distribution of statistical tests (see Sec. 3). Finally, note from (6) that $\hat{\beta} = \beta + O_p(T^{-1/2})$, with $\hat{\beta}$ a consistent estimator of $\beta$.

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

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

**Theorem 2.3.** As $T \uparrow \infty$, (a) $(X'X)^{1/2}(\hat{\beta} - \beta) \Rightarrow N(0, \sigma^2I)$ and (b) $(X'\Omega^{-1}X)^{1/2}(\hat{\beta} - \beta) \Rightarrow N(0, \sigma^2I)$. Both (a) and (b) remain true if $\hat{\beta}$ is replaced with $\hat{\beta}$. The asymptotic normality of $(X'X)^{1/2}(\hat{\beta} - \beta)$—useful in formulating statistical tests—is obtained in the proof of
Theorem 2.3 through a simple conditioning argument. In particular, it implies that the usual $F$ statistic for testing a linear hypothesis in (1) has an asymptotic chi-squared distribution upon appropriate standardization. Note also the difference in the variances of the two limiting distributions in Theorem 2.3: It has some interesting consequences (see Sec. 4).

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 asymptotic distribution of Wald-type test statistics for testing $H_0$.

Theorem 3.1. Under the null hypothesis $H_0$ and as $T \to \infty$, (a) $(R\hat{\beta} - r)'[R(X'X)^{-1}R]'^{-1}(R\hat{\beta} - r)/\sigma^2 \Rightarrow \chi^2_q$ and (b) $(R\beta - r)'[R(X'\hat{\Omega}^{-1}X)'R]'^{-1}(R\beta - r)/\sigma^2 \Rightarrow \chi^2_q$. Both (a) and (b) remain true if $\hat{\beta}$ is replaced by $\beta$.

Using $\hat{\beta}$ rather than $\hat{\beta}$ in Theorem 3.1(b) gives

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

the Wald statistic for testing $H_0$ in the standard linear regression model with nonstochastic regressors and (known) error covariance matrix $\Sigma^2$. Interestingly, $W_i$ still has a limiting $\chi^2_q$ distribution even when $x_i$ is a rather general integrated process generated by (2), because of the strict exogeneity of $x_i$. When the innovation sequences $\{u_t\}$ and $\{w_t\}$ driving (1) and (2) are dependent, the limiting distributions of statistics such as $W_i$ are no longer $\chi^2$. [See Phillips and Durlauf (1986) for pertinent results.]

Consistent estimates of $\sigma^2$ and $\sigma^2$ are needed to make the tests in Theorem 3.1 operational for statistical inference. It is simple to show the following theorem:

Theorem 3.2. (a) $s^2 = T^{-1}(y - X\hat{\beta})'\hat{\Omega}^{-1}(y - X\hat{\beta})/\sigma^2$ and (b) $s^2_n = (\sum_{t=0}^{p} \rho_t)^{-2} \hat{\sigma}^2_n \Rightarrow \sigma^2$.

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

$$s^2 = T^{-1}(y - X\hat{\beta})'\hat{\Omega}^{-1}(y - X\hat{\beta})/\sigma^2.$$ 

and

$$s^2_n = \left(\sum_{t=0}^{p} \hat{\rho}_t\right)^{-2} \hat{\sigma}^2_n \Rightarrow \sigma^2.$$

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

$$W_2 = (R\hat{\beta} - r)'[R(X'X)^{-1}R]'^{-1}(R\hat{\beta} - r)/s^2 \Rightarrow \chi^2_q$$

(7)

and

$$W_3 = (R\hat{\beta} - r)'[R(X'\hat{\Omega}^{-1}X)'R]'^{-1}(R\hat{\beta} - r)/s^2 \Rightarrow \chi^2_q$$

(8)

where we employ the feasible GLS estimator

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

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

$$s^2 = T^{-1}(y - X\hat{\beta})'(y - X\hat{\beta})/\sigma^2 \Rightarrow \sigma^2 = E(u^2).$$

Here the usual Wald statistic for testing $H_0$ is

$$W_4 = (R\hat{\beta} - r)'[R(X'X)^{-1}R]'^{-1}(R\hat{\beta} - r)/s^2,$$

and from Theorem 3.1(b) $W_4 \Rightarrow (\sigma^2/\sigma^2) \chi^2_q$. Thus the conventional Wald statistic for testing $H_0$ based on an OLS regression has a limiting distribution proportional to a $\chi^2_q$. When $u_t$ is generated by (5), the constant of proportionality is $[\sum_{t=0}^{p} \rho_t]^2 \sigma^2$. 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 $W_4$ can be very different from the conventional $\chi^2_q$ when there is serial correlation.

4. ADDITIONAL REMARKS AND EXTENSIONS

The proofs of these results depend heavily on the theory of weak convergence. These methods seem to provide a convenient way of handling the complications resulting from 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 a simple way of demonstrating the validity of conventional asymptotic chi-squared theory for classical tests of linear hypotheses in multiple regression with integrated processes. Section 3 shows that conventional theory applies without modification for tests based on feasible GLS estimates of the coefficients [see (8)]. For OLS-based tests, it is sufficient to replace the usual error variance estimator (as in the definition of $W_d$) with a consistent estimator of the error spectrum at the origin, leading to $W_2$ [given in (7)]. With this simple modification conventional asymptotic chi-squared theory applies to the OLS-based statistic $W_2$.

Our theory has been developed for regressions without a fitted intercept. But all of our results continue to apply where (1) includes a constant or even a polynomial function of time, in addition to the integrated regressor $x_i$. The only modification to the asymptotic formulas that is required for these extensions is that the Brownian motion $B_2$ be replaced by the corresponding demeaned or detrended stochastic process. Thus when (1) involves a constant as well as $x_i$, formula (6) is simply replaced by

$$T(\hat{\beta} - \beta) \Rightarrow \left\{ \int_0^t \tilde{B}_2 dB_2 \right\}^{-1} \sum_{i=0}^{t} \tilde{B}_2 dB_2,$$

where $\tilde{B}_2(r) = B_2(r) - \int_0^r B_2(t) dt$. Here corresponding adjustments also apply to the stated formulas for the estimators and tests, so all variables are replaced by deviations from their sample means.

Results in this article do not apply to models with re-
so by the continuous mapping theorem (CMT) and Lemma 2.1

\[(X'Y)^{1/2} (\tilde{\beta} - \beta) \Rightarrow \int_0^\infty B_z(r) B_z(r)' \, dr \int_0^\infty B_z(r) dB_z(r),\]

Now suppose the \(n = m + 1\)-dimensional Brownian motion \(B(r)\) is defined on the probability space \((\Omega, F, P)\), and let \(F_t\) denote the sub-\(\sigma\)-field of \(F\) generated by \(\{B(r): 0 \leq r \leq t\}\). The symbol \(\cdot \mid F_t\) signifies the conditional distribution relative to \(F_t\). Since \(B(t)\) is Gaussian and independent of \(B(0)\),

\[\int_0^t B_z(r) dB_z(r) \mid_{F_t} = N(0, \sigma_I T)\]

Since the latter distribution does not depend on realizations of \(B(t)\), it is also the unconditional distribution. Part (a) of the theorem follows immediately.

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

\[\|X_t^*(r) - \phi X_t(r)\| = \max_{\|\phi\|_1} |X_t^*(r) - \phi X_t(r)| \to 0\]

Thus

\[\|X_t^*(r) - \phi X_t(r)\| \to 0\]

Therefore \(\tilde{\beta}\) is the OLS estimator of \(\beta\) in the transformed model, from Lemma 2.1

\[T(\tilde{\beta} - \beta) \Rightarrow \int_0^\infty B_z^*(r) dB_z^*(r)' \, dr \int_0^\infty B_z^*(r) dB_z^*(r)\]

\[B^*(r) = \phi B(r), \text{ where } \phi \text{ signifies equality in distribution.} \]

Asymptotic equivalence follows, since by cancellation of the scale factor \(\phi^2\)

\[\int_0^\infty B_z^*(r) dB_z^*(r)' \, dr \int_0^\infty B_z^*(r) dB_z^*(r)\]

as required.

**Proof of Theorem 2.3.** From lemma 3.1(b) of Phillips and Durlauf (1986),

\[T^{-1}X'X \Rightarrow \int_0^\infty B_z(r) B_z(r)' \, dr,\]  

(4.4)
Proof of Theorem 3.1. By the CMT and Lemma 2.1,
\[
\{R(T^{-1}X'X)^{-1}R'\}^{1/2} T(R\hat{\beta} - r) \\
\Rightarrow \left[ R\left( \int_0^\infty B_2(r)B_2(r)'dr \right)^{-1} R' \right]^{-1/2} \\
\times R\left( \int_0^\infty B_2(r)B_2(r)'dr \right)^{-1} \int_0^\infty B_2(r)dB_2(r) \\
= N(0, \sigma^2 I_q).
\]

The last line follows from the conditioning argument used 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 (A.7), but is otherwise entirely analogous. The invariance of the results to the replacement of \( \hat{\beta} \) by \( \hat{\beta} \) is also straightforward.

Proof of Theorem 3.2
\[
\hat{\sigma}^2 = T^{-1}(y - X\hat{\beta})'\Omega^{-1}(y - X\hat{\beta}) \\
= T^{-1}(y^* - X^*\hat{\beta})(y^* - X^*\hat{\beta}) \\
= T^{-1}u^*u^* - T^{-1}(T^{-1}u^*X^*)(T^{-1}X^*X)^{-1}(T^{-1}X^*u^*) \\
= T^{-1}u^*u^* + o_p(1) \\
\xrightarrow{p} \sigma^2
\]
as required for (a). Part (b) follows immediately.

[Received October 1980. Revised June 1987.]

REFERENCES