Cowles Foundation Paper 703

Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions With Integrated Regressors

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

It is shown that in the multiple regression model $y_t = x_t'\beta + u_t$, where u_t is a stationary autoregressive process and x_t is an integrated m-vector process, the asymptotic distributions of the ordinary least squares (OLS) and generalized least squares (GLS) estimators of β are identical. This generalizes a result obtained by Kramer (1986) for two-variate regression and extends fixed regressor theory developed by Grenander and Rosenblatt (1957). Our approach uses a multivariate invariance principle and yields explicit representations of the asymptotic distributions in terms of functionals of vector Brownian motion. We also provide some useful asymptotic results for hypothesis tests of the model. Thus if x_t is generated by a vector (autoregressive integrated moving average) ARIMA(r, l, s) model and u_t is generated by an independent (autoregressive) AR(p) process, then $T(\hat{\beta} - \beta)$ and $T(\tilde{\beta} - \beta)$ have the same limiting distribution (where $\hat{\beta}$ and $\tilde{\beta}$ are the OLS and GLS estimators, respectively). This distribution is nonnormal and most conveniently represented in terms of a vector Brownian motion $(B_1(r), B_2(r)')$ as the functional $\{\int_0^1 B_2B_2'\}^{-1}\{\int_0^1 B_2\, dB_1\}$, where B_2 is an m-vector Brownian motion independent of B_1 Furthermore, if $X(T \times m)$ is a matrix of T observations of x_t, then $(X'X)^{1/2}(\hat{\beta} - \beta)$ and $(X'X)^{1/2}(\tilde{\beta} - \beta)$ have the same limiting normal distribution as $T \uparrow \infty$ But the variance of this normal distribution is given by $\sigma_1^2 = 2\pi f_u(0)$ [where $f_u(\lambda)$ is the spectral density of u_t] and not $\sigma^2 = \text{var}(u_t)$. Traditionally constructed asymptotic tests of significance are invalid in the present context; however, these tests may be made robust by simply replacing the usual estimators of σ^2 with consistent estimates of σ_1^2.

KEY WORDS: Asymptotic efficiency; Integrated process; Invariance principle; Multiple regression; Vector Brownian motion.

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 detrend a time series stationary about 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 nomenon. 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

$$y_t = x_t'\beta + u_t, \qquad t = 1, 2, \ldots, \tag{1}$$

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

$$x_t = x_{t-1} + v_t, \qquad t = 1, 2, \ldots. \tag{2}$$

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 $S_t = \sum_1^t w_k$ satisfies a multivariate invariance principle. More specifically, if

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

then

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

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

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

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 σ_1^2 and Σ_2.

Multivariate invariance principles of this type were proved by Eberlain (1986) and Phillips and Durlauf (1986). They

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

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

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

$$\sum_{j=0}^{p} \rho_j u_{t-j} = \varepsilon_t, \qquad \rho_0 = 1, \tag{5}$$

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

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

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 $T \uparrow \infty$,

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

where $B(r)' = (B_1(r), B_2'(r))$ is 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 $\int_0^1 B_2\,dB_1$ in (6) is interpreted as a vector of stochastic integrals with respect to the univariate Brownian motion $B_1(r)$. The matrix $\int_0^1 B_2B_2'\,dr$ 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 $\hat{\beta} = \beta + O_p(T^{-1})$, with $\hat{\beta}$ 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 $\mathrm{E}(uu') = \sigma^2\Omega$. 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^*, 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

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 $T \uparrow \infty$, (a) $(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)/\sigma_1^2 \Rightarrow \chi_q^2$ and (b) $(R\tilde{\beta} - r)'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\tilde{\beta} - r)/\sigma^2 \Rightarrow \chi_q^2$. Both (a) and (b) remain true if $\tilde{\beta}$ is replaced by $\tilde{\tilde{\beta}}$.

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

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

the Wald statistic for testing H_0 in the standard linear regression model with nonstochastic regressors and (known) error covariance matrix $\sigma^2\Omega$. Interestingly, W_1 still has a limiting χ_q^2 distribution even when x_t is a rather general integrated process generated by (2), because of the strict exogeneity of x_t. When the innovation sequences $\{u_t\}$ and $\{v_t\}$ driving (1) and (2) are dependent, the limiting distributions of statistics such as W_1 are no longer χ^2. [See Phillips and Durlauf (1986) for pertinent results.]

Consistent estimates of σ^2 and σ_1^2 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) $\hat{\sigma}^2 = T^{-1}(y - X\tilde{\beta})'\Omega^{-1}(y - X\tilde{\beta}) \underset{p}{\to} \sigma^2$ and (b) $\hat{\sigma}_1^2 = (\sum_{j=0}^{p} \rho_j)^{-2}\hat{\sigma}^2 \underset{p}{\to} \sigma_1^2$.

These estimators depend on Ω and the AR coefficients ρ_j. When 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

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

and

$$s_1^2 = \left(\sum_{j=0}^{p} \hat{\rho}_j\right)^{-2} s^2 \underset{p}{\to} \sigma_1^2,$$

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

$$\hat{s}^2 = T^{-1}(y - X\hat{\beta})'(y - X\hat{\beta}) \underset{p}{\to} \sigma_u^2 = \mathrm{E}(u_t^2).$$

Here the usual Wald statistic for testing H_0 is

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

and from Theorem 3.1(a) $W_4 \Rightarrow (\sigma_1^2/\sigma_u^2)\chi_q^2$. Thus the conventional Wald statistic for testing H_0 based on an OLS regression has a limiting distribution proportional to a χ_q^2. When u_t is generated by (5), the constant of proportionality is $[(\sum_{j=0}^{p} \rho_j)^2\sigma_u^2]^{-1}\sigma^2$. 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 χ_q^2 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_4) 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_t. 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 degressor-error correlation, such as econometric models of simultaneous equations. But similar techiques may be used to analyze such regressions (see Phillips and Durlauf 1986).

APPENDIX: PROOFS OF THE MAIN RESULTS

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 Q may be chosen so that

$$x_t^* = \sum_{j=0}^{p} \rho_j x_{t-j}, \qquad u_t^* = \sum_{j=0}^{p} \rho_j u_{t-j}, \tag{A.1}$$

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

$$w_t^* = \sum_{j=0}^{p} \rho_j w_{t-j}, \tag{A.2}$$

where $w_t^{*\prime} = (u_t^*, v_t^{*\prime})$ and (without loss of generality) $w_{-p+1} = \cdots = w_0 = 0$.

The new process $\{w_t^*\}$ defined by (A.2) has partial sums that satisfy the multivariate invariance principle (3). For instance, if $\{w_t\}$ is strong mixing with mixing numbers α_k that satisfy $\sum_1^\infty \alpha_k^{1-2/\delta} < \infty$ for some $\delta > 2$, then the same is true of the transformed sequence $\{w_t^*\}_1^\infty$ (see, e.g., White 1984, p. 153). In fact, using $X_T^*(r)$ to denote the random element constructed from partial sums of w_t^* gives $X_T^*(r) \Rightarrow B^*(r)$ in place of (3), where $B^*(r)$ is vector Brownian motion with covariance matrix $\Sigma^* = (\sum_{j=0}^{p} \rho_j)^2 \Sigma$. When $\{w_t\}$ is stationary, the new covariance matrix $\Sigma^* = 2\pi f_{w^*w^*}(0) = (\sum_{j=0}^{p} \rho_j)^2 (2\pi f_{ww}(0))$ may be deduced from the action of the linear filter (A.2). In the general case

$$S_T^* = \sum_1^T w_t^* = \left(\sum_{j=0}^{p} \rho_j\right)\left(\sum_1^T w_t\right) - \sum_{j=1}^{p} \rho_j \sum_{t=T-j+1}^{T} w_t, \tag{A.3}$$

and it follows that $\lim_{T\to\infty} T^{-1} \mathrm{E}(S_T^* S_T^{*\prime}) = \varphi^2 \Sigma = \Sigma^*$, where $\varphi = \sum_{j=0}^{p} \rho_j$.

Note that the transformed model is driven by the new process $\{w_t^*\}$ the same as the original model (1) is by $\{w_t\}$. Since $\tilde\beta$ is the OLS estimator of β in the transformed model, from Lemma 2.1

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

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

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

so by the continuous mapping theorem (CMT) and Lemma 2.1

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

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 generated by $\{B_2(r) : 0 \le r \le 1\}$. The symbol $\cdot|F_2$ signifies the conditional distribution relative to F_2. Since $B_1(r)$ is Gaussian and independent of $B_2(r)$,

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

and

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

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 $T \uparrow \infty$

$$\|X_T^*(r) - \varphi X_T(r)\| = \max_i \sup_r |X_{Ti}^*(r) - \varphi X_{Ti}(r)| \xrightarrow[p]{} 0, \tag{A.5}$$

where $\varphi = \sum_{j=0}^{p} \rho_j$. Note that for $(k - 1)/T \le r < k/T$

$$\begin{aligned} |X_{Ti}^*(r) - \varphi X_{Ti}(r)| &= T^{-1/2} \left| \sum_1^{k-1} \sum_{s=0}^{p} \rho_s w_{i,t-s} - \varphi \sum_1^{k-1} w_{it} \right| \\ &= T^{-1/2} \left| \sum_{s=0}^{p} \rho_{s+1} \sum_{k-1-s}^{k-1} w_{it} \right| \\ &\le T^{-1/2} |\varphi| \sum_{k-1-p}^{k-1} |w_{it}|. \end{aligned}$$

Thus

$$\|X_T^*(r) - \varphi X_T(r)\| \le T^{-1/2} p |\varphi| (\max_i \max_t |w_{it}|) \xrightarrow[p]{} 0,$$

proving (A.5). 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

Proof of Theorem 3.1. By the CMT and Lemma 2.1,

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

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

$$\times R\left\{\int_0^1 B_2(r)B_2(r)'\,dr\right\}^{-1} \int_0^1 B_2(r)\,dB_1(r)$$

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

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 $\tilde{\beta}$ is also straightforward.

Proof of Theorem 3.2

$$\begin{aligned}\hat{\sigma}^2 &= T^{-1}(y - X\tilde{\beta}')'\Omega^{-1}(y - X\tilde{\beta}) \\ &= T^{-1}(y^* - X^*\tilde{\beta})'(y^* - X^*\tilde{\beta}) \\ &= T^{-1}u^{*\prime}u^* - T^{-1}(T^{-1}u^{*\prime}X^*)(T^{-2}X^{*\prime}X^*)^{-1}(T^{-1}X^{*\prime}u^*) \\ &= T^{-1}u^{*\prime}u^* + o_p(1) \\ &\xrightarrow[p]{} \sigma^2\end{aligned}$$

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

[*Received October 1986. Revised June 1987.*]

REFERENCES

Anderson, T. W. (1948), "On the Theory of Testing for Serial Correlation," *Skandinavisk Aktuarietidskrift*, 31, 88–116.

——— (1971), *Statistical Analysis of Time Series*, New York: John Wiley.

Eberlain, E. (1986), "On Strong Invariance Principles Under Dependence Assumptions," *The Annals of Probability*, 14, 260–270.

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

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

Kariya, T. (1985), "Testing in the Multivariate General Linear Model," New York: Kinokuniya.

Kramer, 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.

Malley, J. D. (1986), "Optimal Unbiased Estimation of Variance Components," New York: Springer-Verlag.

Phillips, P. C. B., and Durlauf, S. (1986), "Multiple Time Series Regression With Integrated Processes," *Review of Economic Studies*, 53, 473–496.

Rao, C. R. (1967), "Least Squares Using an Estimated Dispersion Matrix and Its Application to Measurement of Signals," in *Fifth Berkeley Symposium in Mathematical Statistics and Probability* (Vol. 1), Berkeley: University of California Press, pp. 355–372.

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

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.