EXACT FGLS ASYMPTOTICS FOR MA ERRORS

David M. Mandy
and
Sandor Fridli

David M. Mandy
Department of Economics
University of Missouri
118 Professional Building
Columbia, MO 65211 USA
Voice 573-882-1763; Fax 573-882-2697
e-mail MandyD@missouri.edu

Sandor Fridli
Department of Numerical Analysis
Eötvös L. University
Pázmány P. sétány 1/D
Budapest, H-1117 Hungary
Voice 463-3918; Fax 439-1648
e-mail fridli@ludens.elte.hu

June 29, 2001

Abstract. We show under very parsimonious assumptions that FGLS and GLS are asymptotically equivalent when errors follow an invertible MA(1) process. Although the linear regression model with MA errors has been studied for many years, asymptotic equivalence of FGLS and GLS has never been established for this model. We do not require anything beyond a finite second moment of the conditional white noise, uniformly bounded fourth moments and independence of the regressor vectors, consistency of the estimator for the MA parameter, and a finite nonsingular probability limit for the (transformed) averages of the regressors. These assumptions are analogous to assumptions typically used to prove asymptotic equivalence of FGLS and GLS in SUR models, models with AR(p) errors, and models of parametric heteroscedasticity.

Keywords: Moving Average, Generalized Least Squares, Asymptotic Distribution.

Suppose we want to perform Feasible Generalized Least Squares (FGLS) estimation of the linear structural equation

$$y_t = x_t'\beta + v_t; \quad t = 1, \ldots, T; \quad T = 1, 2, \ldots; \tag{1}$$

where y_t and x_t are observable random variables and independent $k \times 1$ random vectors, respectively; β is an unobservable $k \times 1$ vector of parameters to be estimated; v_t are unobservable random variables that follow the invertible first order moving average (MA) process

$$v_t = u_t + \alpha u_{t-1}; \quad t = 1, \ldots, T; \tag{2}$$

$\alpha \in (-1, 1)$ is an unobservable nuisance parameter; and $u_t|x_t \sim iid(0, \sigma^2 < \infty)$ for $t = 0, 1, \ldots$ are unobservable conditional white noise. As pointed out by West and Wilcox (1996), this statistical model arises naturally from certain economic models. Stacking the observations in the usual way with $Y_T = (\, y_1 \quad \ldots \quad y_T \,)'$, $X_T = (\, x_1 \quad \ldots \quad x_T \,)'$, and $V_T = (\, v_1 \quad \ldots \quad v_T \,)'$; the well-known covariance of V_T is

$$E[V_T V_T'] = \sigma^2 \Omega_T(\alpha), \text{ where } \Omega_T(\alpha) = \begin{pmatrix} 1+\alpha^2 & \alpha & 0 & \ldots & \ldots & 0 \\ \alpha & 1+\alpha^2 & \alpha & \ddots & \ldots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \alpha \\ 0 & \ldots & \ldots & 0 & \alpha & 1+\alpha^2 \end{pmatrix}; \tag{3}$$

and the GLS class of estimators for β is defined by the function

$$\hat{\beta}_T(\gamma) = (X_T'\Omega_T(\gamma)^{-1}X_T)^{-1}X_T'\Omega_T(\gamma)^{-1}Y_T. \tag{4}$$

When $\gamma = \alpha$ we have the (infeasible) GLS estimator, and when $\gamma = \hat{\alpha}_T$ is some feasible estimator of α we have an FGLS estimator. $\hat{\beta}_T(\alpha)$ has the Gauss-Markov property and

$$\sqrt{T}(\hat{\beta}_T(\alpha) - \beta) \xrightarrow{d} N\left(0, \sigma^2 \operatorname*{plim}_{T\to\infty}\left(\frac{X_T'\Omega_T(\alpha)^{-1}X_T}{T}\right)^{-1}\right), \tag{5}$$

so as usual in GLS estimation problems we can establish asymptotic efficiency of a feasible estimator within the class of linear unbiased estimators and also provide an asymptotic basis for inference if we can show

$$\sqrt{T}(\hat{\beta}_T(\hat{\alpha}_T) - \hat{\beta}_T(\alpha)) \xrightarrow{p} 0 \tag{6}$$

for some feasible $\hat{\alpha}_T$. This paper establishes (6) for any consistent $\hat{\alpha}_T$. We assume throughout that $\operatorname*{plim}_{T\to\infty} \frac{X_T'\Omega_T(\alpha)^{-1}X_T}{T}$ exists, is finite, and is nonsingular; and that the elements of x_t have uniformly bounded fourth moments.

1. Existing Literature on the Model (1)-(2)

Establishing $\sqrt{T}$ asymptotic equivalence of GLS and FGLS estimators is an old problem that was solved many years ago for some particular error structures. Even in modern treatments of this subject, these solutions are sometimes generalized to a claim that GLS and FGLS are always $\sqrt{T}$ asymptotically equivalent provided the number of nuisance parameters is bounded as $T \to \infty$, the covariance of the errors is a "well behaved" function of the nuisance parameters, and the nuisance parameters are estimated consistently.[1] This claim is incorrect. Schmidt (1976, p. 69) provided a counter-example showing that the claimed convergence can fail even when the inverse covariance has only $O(T)$ nonzero elements, there is only one nuisance parameter, and the estimator of the nuisance parameter is consistent and independent of the errors in the regression equation. Thus, the convergence must be addressed on a case-by-case basis for each particular error structure, or at least with some additional general assumptions about the structure of the error covariance.

The standard approach applied to the regression model with MA errors (1)-(2) is to prove the jointly sufficient conditions (see, for example, White (1984) p. 163)

$$\frac{1}{T} X_T' \left(\Omega_T(\hat{\alpha}_T)^{-1} - \Omega_T(\alpha)^{-1} \right) X_T \xrightarrow{p} 0 \tag{7a}$$

$$\frac{1}{\sqrt{T}} X_T' \left(\Omega_T(\hat{\alpha}_T)^{-1} - \Omega_T(\alpha)^{-1} \right) V_T \xrightarrow{p} 0. \tag{7b}$$

For many error structures that arise in GLS estimation problems the inverse error covariance has $O(T)$ nonzero elements and those elements consist of a fixed number of distinct functions of a fixed number of nuisance parameters. Under these circumstances, conditions analogous to (7) can be established using familiar arguments.[2] The main difficulty in proving (6) is that the MA error structure has an inverse covariance with neither of these properties. An exact expression for the (t, τ) element of $\Omega_T(\gamma)^{-1}$ is given by Whittle (1983, p. 75):

$$\omega_{t\tau T}(\gamma) = \frac{(-1)^{t+\tau} \gamma^{\tau-t} (\gamma^{2t} - 1)(\gamma^{2(T-\tau+1)} - 1)}{(\gamma^2 - 1)(\gamma^{2(T+1)} - 1)} \text{ for } 1 \le t \le \tau \le T. \tag{8}$$

To the authors' knowledge, asymptotic equivalence of GLS and FGLS using only the assumptions stated in

[1] Maddala (1971) is an early example of such a claim. Davidson and MacKinnon (1993, pp. 300-301) recently repeated the claim. Fuller (1976, p. 425) claims without proof that the convergence with MA errors can be established in a manner similar to the standard proof for AR errors, but withdrew the claim in the second edition (1996, p. 522).

[2] See Mandy and Martins-Filho (1994 and 1997). Note that the number of functions of the nuisance parameter in Schmidt's counter-example approaches ∞ with T.

the introduction has never been established when the elements of the inverse error covariance are given by (8).

There have been other approaches to feasible estimation of a regression model with MA errors. These approaches usually involve an assumption that u_t is normally distributed, or at least that u_t has a symmetric distribution, and that $\sqrt{T}(\hat{\alpha}_T - \alpha)$ is asymptotically normal. A relatively recent example is Zinde-Walsh and Galbraith (1991). We prove (6) in the next section without any of these properties. In particular, we allow $\hat{\alpha}_T$ to converge to α as slowly as $o_p(1)$. Note that such $\hat{\alpha}_T$ are readily available with no assumptions beyond those mentioned in the introduction (see, for example, Judge at al. (1985) p. 304). Our result shows that the estimator of Zinde-Walsh and Galbraith is asymptotically BLUE even when errors are not Gaussian (for the case of MA(1) errors).

In addition to the assumptions used by Zinde-Walsh and Galbraith, earlier approaches relied on approximations to the inverse covariance or on numerical solutions. Shaman (1969 and 1975) provided some approximations to (8), and Pierce (1971) used both approximations and numerical optimization. Harvey and Phillips (1979) avoided approximations, but relied on potentially troublesome iterative numerical techniques.

An exception is Amemiya (1973), who proved existence of a feasible estimator that is asymptotically equivalent to GLS when errors are ARMA, but the technique is not true FGLS because the estimator of the error covariance is nonparametric. This nonparametric estimation is difficult to implement because it uses a subsample of $N < T$ observations and requires that $N \to \infty$ with T, but does not specify how N should be chosen for any particular T.

2. Asymptotic Equivalence of GLS and FGLS

This section presents a proof of (6). Let $\hat{\alpha}_T$ be any consistent estimator of α. By adopting the convention $0^0 = 1$, we automatically include the $\gamma = 0$ case in (8). Since $\Omega_T(\gamma)^{-1}$ is symmetric, the elements for $t > \tau$ are obtained by interchanging the roles of t and τ in (8). The denominator in (8) cancels in the GLS rule (4) because it is independent of (t, τ), so for the purpose of establishing (6) we can assume

$$\omega_{t\tau T}(\gamma) = (-1)^{t+\tau}\left[\gamma^{2(T+1)+t-\tau} - \gamma^{t+\tau} - \gamma^{2(T+1)-t-\tau} + \gamma^{\tau-t}\right] \text{ for } 1 \leq t \leq \tau \leq T. \tag{8'}$$

Conditions (7) are established elementwise so we lose no generality by assuming $k = 1$. Then, substituting

(2) and (8') into (7) and using symmetry of $\Omega_T(\gamma)$, for (6) it suffices to show

$$\sum_{t=1}^{T}\sum_{\tau=t}^{T} h_{tT}(\tau)\left[\hat{\alpha}_T^{j_{tT}(\tau)} - \alpha^{j_{tT}(\tau)}\right] \xrightarrow{p} 0; \tag{9}$$

where $h_{tT}(\tau)$ is either $\frac{1}{T}x_t x_\tau(-1)^{t+\tau}$, $\frac{1}{\sqrt{T}}x_t u_\tau(-1)^{t+\tau}$, or $\frac{1}{\sqrt{T}}\alpha x_t u_{\tau-1}(-1)^{t+\tau}$; and $j_{tT}(\tau)$ is one of the entries in the first column of Table 1. Performing a change of variable from τ to $\ell = j_{tT}(\tau)$ on the inside sum in (9) and then reversing the order of summation yields

$$\begin{aligned}\sum_{t=1}^{T}\sum_{\tau=t}^{T} h_{tT}(\tau)\left[\hat{\alpha}_T^{j_{tT}(\tau)} - \alpha^{j_{tT}(\tau)}\right] &= \sum_{t=1}^{T}\sum_{\ell=a_T(t)}^{b_T(t)} h_{tT}\left(j_{tT}^{-1}(\ell)\right)\left[\hat{\alpha}_T^{\ell} - \alpha^{\ell}\right] \\ &= \sum_{\ell=c_T}^{d_T}\left[\left[\hat{\alpha}_T^{\ell} - \alpha^{\ell}\right]\sum_{t=p_T(\ell)}^{q_T(\ell)} h_{tT}\left(j_{tT}^{-1}(\ell)\right)\right],\end{aligned} \tag{10}$$

where the various limits of summation are given in Table 1.

$\ell = j_{tT}(\tau)$	$a_T(t)$	$b_T(t)$	c_T	d_T	$p_T(\ell)$	$q_T(\ell)$
$2(T+1)+t-\tau$	$T+t+2$	$2(T+1)$	$T+3$	$2(T+1)$	1	$\ell-T-2$
$t+\tau$	$2t$	$T+t$	2	$2T$	$\max\{1,\ell-T\}$	$\lfloor \ell/2 \rfloor$
$2(T+1)-t-\tau$	$T+2-t$	$2(T+1-t)$	2	$2T$	$\max\{1,T+2-\ell\}$	$\lfloor T+1-\ell/2 \rfloor$
$\tau-t$	1	$T-t$	1	$T-1$	1	$T-\ell$

TABLE 1. $\lfloor c \rfloor$ denotes the integer part of c. The last entry in the $a_T(t)$ column is set to 1 rather than 0 because $\hat{\alpha}_T^0 - \alpha^0 = 0$, so this term can be dropped from (10).

By the mean value theorem there exists a random variable $\gamma_{\ell T}$ between $\hat{\alpha}_T$ and α such that

$$\hat{\alpha}_T^{\ell} - \alpha^{\ell} = \ell\gamma_{\ell T}^{\ell-1}[\hat{\alpha}_T - \alpha]. \tag{11}$$

Substituting (11) into (10) and recalling that $\hat{\alpha}_T - \alpha \xrightarrow{p} 0$, it suffices to show $g_T = O_P(1)$ where

$$g_T = \sum_{\ell=c_T}^{d_T}\left[\ell\gamma_{\ell T}^{\ell-1}\sum_{t=p_T(\ell)}^{q_T(\ell)} h_{tT}\left(j_{tT}^{-1}(\ell)\right)\right]. \tag{12}$$

Let S denote the sample space, $s \in S$ a point in the sample space, P the probability measure, and μ_i the uniform bound on the i^{th} noncentral moments of the x_t sequence. Since $\gamma_{\ell T}(s)$ is between $\hat{\alpha}_T(s)$ and α,

$$|\gamma_{\ell T}(s)| \leq |\alpha| + |\gamma_{\ell T}(s) - \alpha| \leq |\alpha| + |\hat{\alpha}_T(s) - \alpha| \ \forall s \in S. \tag{13}$$

As $|\alpha| < 1$ there exists $m \in (|\alpha|, 1)$. Let $A_T = \{s \in S\colon |\alpha| + |\hat{\alpha}_T(s) - \alpha| \leq m\}$ and note that $P(A_T) \to 1$. That is, for any $\delta > 0$ there exists T_δ such that $T > T_\delta \Rightarrow P(A_T) > 1 - \frac{\delta}{2}$. From (13),

$$|\gamma_{\ell T}(s)| \leq m \ \forall s \in A_T, \text{ for every } \ell, T. \tag{14}$$

We first treat the case $h_{tT}(\tau) = \frac{1}{T} x_t x_\tau (-1)^{t+\tau}$. Define

$$M_{\ell\delta} = \sqrt{\frac{2\mu_4}{\delta[1-m]m^{\ell-1}}} \text{ and } C_{\ell T\delta} = \left\{ s \in S \colon \left| \frac{1}{T} \sum_{t=p_T(\ell)}^{q_T(\ell)} (-1)^{t+j_{tT}^{-1}(\ell)} x_t(s) x_{j_{tT}^{-1}(\ell)}(s) \right| \geq M_{\ell\delta} \right\}.$$

Using Chebyshev's Inequality,

$$P(C_{\ell T\delta}) \leq \frac{[q_T(\ell) - p_T(\ell)]^2 \mu_4}{T^2 M_{\ell\delta}^2}.$$

It is easy to check from Table 1 that

$$q_T(\ell) \leq \max\{q_T(\ell) \colon c_T \leq \ell \leq d_T\} \leq \max\{q_T(c_T), q_T(d_T)\} \leq T.$$

Hence $0 \leq q_T(\ell) - p_T(\ell) \leq T$ and

$$P(C_{\ell T\delta}) \leq \frac{\mu_4}{M_{\ell\delta}^2} = \frac{\delta[1-m]m^{\ell-1}}{2}.$$

So for $C_{T\delta} = \cup_{\ell=c_T}^{d_T} C_{\ell T\delta}$,

$$P(C_{T\delta}) \leq \sum_{\ell=c_T}^{d_T} P(C_{\ell T\delta}) \leq \frac{\delta[1-m]}{2} \sum_{\ell=c_T}^{d_T} m^{\ell-1} \leq \frac{\delta[1-m]}{2} \sum_{\ell=1}^{\infty} m^{\ell-1} = \frac{\delta}{2}.$$

Therefore $P(C_{T\delta}^c) > 1 - \frac{\delta}{2}$, where

$$C_{T\delta}^c = \bigcap_{\ell=c_T}^{d_T} \left\{ s \in S \colon \left| \frac{1}{T} \sum_{t=p_T(\ell)}^{q_T(\ell)} (-1)^{t+j_{tT}^{-1}(\ell)} x_t(s) x_{j_{tT}^{-1}(\ell)}(s) \right| < M_{\ell\delta} \right\}. \tag{15}$$

So $T > T_\delta \Rightarrow P(A_T \cap C_{T\delta}^c) > 1 - \delta$. Substituting (14) and (15) into (12), for $s \in A_T \cap C_{T\delta}^c$ we have

$$\begin{aligned} |g_T(s)| \leq \sum_{\ell=c_T}^{d_T} \ell m^{\ell-1} M_{\ell\delta} &= \sqrt{\frac{2\mu_4}{\delta[1-m]}} \sum_{\ell=c_T}^{d_T} \ell \left(\sqrt{m}\right)^{\ell-1} \\ &\leq \sqrt{\frac{2\mu_4}{\delta[1-m]}} \sum_{\ell=1}^{\infty} \ell \left(\sqrt{m}\right)^{\ell-1} \\ &= \sqrt{\frac{2\mu_4}{\delta[1-m]}} \frac{1}{[1-\sqrt{m}]^2}. \end{aligned}$$

Hence

$$T > T_\delta \Rightarrow P\left(\left\{ s \in S \colon |g_T(s)| \leq \sqrt{\frac{2\mu_4}{\delta[1-m]}} \frac{1}{[1-\sqrt{m}]^2} \right\}\right) > 1 - \delta. \tag{16}$$

That is, $g_T = O_P(1)$.

The case $h_{tT}(\tau) = \frac{1}{\sqrt{T}} x_t u_\tau (-1)^{t+\tau}$ is similar. Define

$$M_{\ell\delta} = \sqrt{\frac{2\sigma^2\mu_2}{\delta[1-m]m^{\ell-1}}} \text{ and } C_{\ell T\delta} = \left\{ s \in S \colon \left| \frac{1}{\sqrt{T}} \sum_{t=p_T(\ell)}^{q_T(\ell)} (-1)^{t+j_{tT}^{-1}(\ell)} x_t(s) u_{j_{tT}^{-1}(\ell)}(s) \right| \geq M_{\ell\delta} \right\}.$$

Then, using independence of the x_t's and $u_t|x_t \sim iid(0,\sigma^2)$, Chebyshev's Inequality yields

$$P(C_{\ell T\delta}) \leq \frac{[q_T(\ell) - p_T(\ell)]\sigma^2\mu_2}{TM_{\ell\delta}^2}.$$

The rest of the argument proceeds exactly as above, culminating in the conclusion

$$T > T_\delta \Rightarrow P\left(\left\{s \in S\colon |g_T(s)| \leq \sqrt{\frac{2\sigma^2\mu_2}{\delta[1-m]}\frac{1}{[1-\sqrt{m}]^2}}\right\}\right) > 1 - \delta. \tag{16'}$$

Since the u_τ's are *iid* and α is a constant, the case $h_{tT}(\tau) = \frac{1}{\sqrt{T}}\alpha x_t u_{\tau-1}(-1)^{t+\tau}$ is identical.

3. Discussion

The factoring of $\hat{\alpha}_T^\ell - \alpha^\ell$ from $\sum_t h_{tT}(\tau)$ in (10) is crucial to the proof. The potential for dependence between $\hat{\alpha}_T$ and $h_{tT}(\tau)$ means (9) cannot be proven by any method that attempts to directly use independence of the u_t's.

To see this, consider a special case in which $x_t(s) \equiv 1\ \forall t$, $\alpha = 0$, $h_{tT}(\tau) = \frac{1}{\sqrt{T}}x_t u_\tau(-1)^{t+\tau}$, and $j_{tT}(\tau) = 2(T+1)+t-\tau$ (this value of j contains the dominant term as $T \to \infty$ among the four possibilities in Table 1). Then, from (9) we must show $f_T(s) \xrightarrow{p} 0$, where

$$f_T(s) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\sum_{\tau=t+1}^{T} u_\tau(s)(-1)^{t+\tau}\hat{\alpha}_T(s)^{2(T+1)+t-\tau}. \tag{17}$$

Now suppose $\hat{\alpha}_T$ and u_τ are dependent in the following very simple way. Let $S = [0,1]$; P be Lebesgue measure;

$$\hat{\alpha}_T(s) = \begin{cases} 2, & s \in [0, 2^{-T}) \\ 0, & \text{otherwise} \end{cases};$$

and $u_\tau(s)$ be the Rademacher functions

$$u_1(s) = \begin{cases} 1, & 0 \leq s < 1/2 \\ -1, & 1/2 \leq s < 1 \end{cases}$$

extended to $s \in [0,\infty)$ by periodicity of period 1, and $u_\tau(s) = u_1(2^{\tau-1}s)$ for $\tau = 1, 2, \ldots$. Clearly $\hat{\alpha}_T \xrightarrow{p} \alpha = 0$, and recall that the Rademacher functions are orthonormal and identically distributed on $[0,1]$. Thus all of our assumptions are satisfied by this setup.

As $u_\tau(s) = 1$ on $s \in [0, 2^{-T})$ for $\tau = 1, \ldots, T$; (17) is

$$f_T(s) = \begin{cases} \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\sum_{\tau=t+1}^{T}(-1)^{t+\tau}2^{2(T+1)+t-\tau}, & s \in [0, 2^{-T}) \\ 0, & \text{otherwise} \end{cases}; \tag{17'}$$

and we see immediately that $P(\{s \in S \colon |f_T(s)| > 0\}) \leq 2^{-T} \to 0$, as required.

However, the terms $r_t = \sum_{\tau=t+1}^{T} (-1)^{t+\tau} 2^{2(T+1)+t-\tau} \chi_{[0,2^{-T})}$ are all constant multiples of $\chi_{[0,2^{-T})}$, the characteristic function of $[0, 2^{-T})$. Consequently any argument based on a certain degree of independence or orthogonality of the r_t's must fail. In particular, Chebyshev's Inequality applied directly to (17') is

$$P(\{s \in S \colon |f_T(s)| \geq M\}) \leq \frac{E\left(f_T^2\right)}{M^2} \text{ for arbitrary } M > 0. \tag{18}$$

By repeated applications of the formula for the partial sums of geometric series we obtain

$$E\left(f_T^2\right) = \frac{1}{T2^T} \left[\sum_{t=1}^{T-1} \sum_{\tau=t+1}^{T} (-1)^{t+\tau} 2^{2(T+1)+t-\tau} \right]^2 = \frac{16}{81} \frac{2^{3T}}{T} \left[2 - 3T - \frac{2}{(-2)^T} \right]^2,$$

so $\lim_{T\to\infty} E\left(f_T^2\right) = \infty$. Therefore we cannot conclude from (18) that $P(\{s \in S \colon |f_T(s)| \geq M\})$ tends to zero for arbitrary $M > 0$.

The direct approach to (9) fails in this example despite boundedness of $x_t(s)$ and $u_\tau(s)$.

4. Conclusion

We have shown under very parsimonious assumptions that FGLS and GLS are asymptotically equivalent when errors follow an invertible MA(1) process. We do not require anything beyond a finite second moment of the conditional white noise, uniformly bounded fourth moments and independence of the regressor vectors, consistency of the nuisance estimator, and a finite nonsingular probability limit for the (transformed) averages of the regressors. These assumptions are analogous to assumptions typically used to prove asymptotic equivalence of FGLS and GLS in SUR models, models with AR(p) errors, and models of parametric heteroscedasticity.

References

Amemiya, T., *Generalized Least Squares with an Estimated Autocovariance Matrix*, Econometrica **41/4** (July, 1973), 723-732.

Davidson, R. and J. G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993.

Fuller, W. A., *Introduction to Statistical Time Series*, Wiley, New York, 1976.

Fuller, W. A., *Introduction to Statistical Time Series*, Second ed., Wiley, New York, 1996.

Harvey, A. C. and G. D. A. Phillips, *Maximum Likelihood Estimation of a Regression Model with Autoregressive-Moving Average Disturbances*, Biometrika **66** (1979), 49-58.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T-C. Lee, *The Theory and Practice of Econometrics*, 2nd ed., Wiley, New York, 1985.

Maddala, G. S., *Generalized Least Squares with an Estimated Variance Covariance Matrix*, Econometrica **39** (1971), 23-33.

Mandy, D. M. and C. Martins-Filho, *A Unified Approach to Asymptotic Equivalence of Aitken and Feasible Aitken Instrumental Variables Estimators*, International Economic Review **35** (1994), 957-979.

Mandy, D. M. and C. Martins-Filho, *A Note on A Unified Approach to Asymptotic Equivalence of Aitken and Feasible Aitken Instrumental Variables Estimators*, International Economic Review **38** (1997), 479.

Pierce, D. A., *Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors*, Biometrika **58** (1971), 299-312.

Schmidt, P., *Econometrics*, Dekker, New York, 1976.

Shaman, P., *On the Inverse of the Covariance Matrix of a First Order Moving Average*, Biometrika **56/3** (1969), 595-600.

Shaman, P., *An Approximate Inverse for the Covariance Matrix of Moving Average and Autoregressive Processes*, The Annals of Statistics **3/2** (1975), 532-538.

West, K. and D. W. Wilcox, *A Comparison of Alternative Instrumental Variables Estimators of a Dynamic Linear Model*, Journal of the American Statistical Association **14** (1996), 281-293.

White, H., *Asymptotic Theory for Econometricians*, Academic Press, Orlando, 1984.

Whittle, P., *Prediction and Regulation by Linear Least-Square Methods*, Second ed., University of Minnesota Press, Minneapolis, 1983.

Zinde-Walsh, V. and J. W. Galbraith, *Estimation of a Linear Regression Model with Stationary ARMA(p,q) Errors*, Journal of Econometrics **47** (1991), 333-357.