INTERNATIONAL ECONOMIC REVIEW
Vol. 35, No. 4, November 1994

A UNIFIED APPROACH TO ASYMPTOTIC EQUIVALENCE OF AITKEN AND FEASIBLE AITKEN INSTRUMENTAL VARIABLES ESTIMATORS*

BY DAVID M. MANDY AND CARLOS MARTINS-FILHO[1]

Asymptotic equivalence of Aitken and feasible Aitken estimators in linear models with nonscalar identity error covariance matrices is usually established in a tedious case-by-case manner. Some general sufficient conditions for this equivalence exist, but there are problems with the extant conditions. These problems are discussed, and new widely applicable sufficient conditions are presented and applied to a variety of error structures.

1. INTRODUCTION

Consider the linear statistical model $y_n = X_n\beta + u_n$, where y_n is an $(n \times 1)$ stochastic vector, X_n is an $(n \times k)$ (possibly stochastic) almost everywhere full column rank regressor matrix, β is a $(k \times 1)$ vector of unknown nonstochastic parameters, and u_n is an $(n \times 1)$ error vector with $E(u_n) = 0$ and $E(u_n u_n') = \Omega_n$, an arbitrary symmetric positive definite $(n \times n)$ matrix. If Z_n is an $(n \times k)$ almost everywhere full column rank instrument matrix then an instrumental variables (IV) version of the Aitken (1935) estimator is

$$\hat{\beta}_n = (Z_n'\Omega_n^{-1}X_n)^{-1}Z_n'\Omega_n^{-1}y_n,$$

which is henceforth called an Aitken IV estimator. When Ω_n is unknown a feasible estimator must be constructed, usually entailing a parameterization that assumes Ω_n is a known function of an unknown $(r \times 1)$ vector θ with true value θ^0, so that $E(u_n u_n') = \Omega_n(\theta^0)$. Then, an IV version of the feasible Aitken estimator for β based on an estimator $\hat{\theta}^n$ for θ^0 is

$$\tilde{\beta}_n = (Z_n'\Omega_n(\hat{\theta}^n)^{-1}X_n)^{-1}Z_n'\Omega_n(\hat{\theta}^n)^{-1}y_n,$$

which is henceforth called a feasible Aitken IV estimator.

The small sample properties of $\tilde{\beta}_n$ are often unknown, so many estimation problems involve searching for an estimator $\hat{\theta}^n$ such that at least asymptotic normality of $\sqrt{n}(\tilde{\beta}_n - \beta)$ can be established. One frequently employed approach is to show that $\sqrt{n}\tilde{\beta}_n$ based on an estimator $\hat{\theta}^n$ converges in probability to $\sqrt{n}$ $\hat{\beta}_n$ since, in many applications, known sufficient conditions for consistency and asymptotic normality of $\hat{\beta}_n$ can be assumed. $\hat{\beta}_n$ also may possess certain asymptotic efficiency properties. Thus, under the sufficient conditions, if $\sqrt{n}(\hat{\beta}_n - \tilde{\beta}_n)$ converges

* Manuscript received April 1992; revised February 1994.

[1] We gratefully acknowledge helpful comments from Richard Baillie, Gregory Duncan, and an anonymous referee, none of whom bear any responsibility for remaining errors.

in probability to zero for a particular estimator $\hat{\theta}_n$ then the feasible estimator $\tilde{\beta}_n$ inherits the asymptotic properties of $\hat{\beta}_n$.

Proving convergence in probability in specific cases is often a tedious prospect, so it is useful to establish general sufficient conditions on the structure of $\Omega_n(\theta)$ that apply in many cases and are relatively easy to check. Two sets of sufficient conditions have appeared in the literature, but neither offers a complete solution to the problem. Section 2 reviews these extant results and discusses their shortcomings. We show that the asymptotic properties of certain widely-used estimators that were heretofore thought to be known are in fact unresolved issues. Section 3 develops new sufficient conditions that summarize the properties of some familiar models that are important in establishing asymptotic equivalence of Aitken and feasible Aitken IV estimators. Section 4 demonstrates how the sufficient conditions of Section 3 are easily applied to familiar models of heteroscedasticity, autocorrelation, seemingly unrelated regressions, and three-stage least squares. Some results for models that have not previously appeared in the literature are also established via the new sufficient conditions, as well as some of the unresolved properties from Section 2. Hence, the new conditions provide a convenient method for establishing asymptotic properties in existing and as yet unexplored models.

2. EXISTING CONDITIONS FOR ASYMPTOTIC EQUIVALENCE OF $\hat{\beta}_n$ AND $\tilde{\beta}_n$

One set of sufficient conditions for asymptotic equivalence of $\hat{\beta}_n$ and $\tilde{\beta}_n$ builds on sufficient conditions for asymptotic normality of $\sqrt{n}(\hat{\beta}_n - \beta)$. Schmidt (1976, p. 101) provides conditions for this asymptotic normality (and hence consistency of $\hat{\beta}_n$) when $\Omega_n(\theta^0) = I_n$. Restating his conditions for the case of a general $\Omega_n(\theta^0)$ matrix yields

(A1) $\quad \dfrac{1}{\sqrt{n}} Z'_n \Omega_n(\theta^0)^{-1} u_n \xrightarrow{d} N(0, Q)$ for some symmetric $(k \times k)$ matrix Q

(A2) $\quad \operatorname*{plim}_{n\to\infty} \dfrac{1}{n} Z'_n \Omega_n(\theta^0)^{-1} X_n = Q_{ZX}$, a finite positive definite matrix.

The proof that these conditions imply $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, Q_{ZX}^{-1} Q Q_{ZX}^{-1})$ is straightforward and parallels the proof provided by Schmidt. Theil (1971, p. 399) states additional conditions that, along with (A1) and (A2), are jointly sufficient for $\operatorname{plim}_{n\to\infty} \sqrt{n}(\hat{\beta}_n - \tilde{\beta}_n) = 0$ for a standard GLS model, in which X_n is nonstochastic and $Z_n = X_n$.[2] White (1984, p. 163) notes that Theil's theorem may be extended to the case of general instrumental variables with little difficulty. In particular, (A1) and (A2) combine with

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} Z'_n (\Omega_n(\hat{\theta}^n)^{-1} - \Omega_n(\theta^0)^{-1}) X_n = 0 \tag{1}$$

[2] Theil actually uses conditions that imply (A1) and (A2), but this does not affect the conclusion.

$$\text{(2)} \qquad \operatorname*{plim}_{n\to\infty} \frac{1}{\sqrt{n}} Z_n'(\Omega_n(\hat{\theta}^n)^{-1} - \Omega_n(\theta^0)^{-1})u_n = 0$$

to provide sufficient conditions for $\operatorname{plim}_{n\to\infty} \sqrt{n}(\hat{\beta}_n - \tilde{\beta}_n) = 0$.[3]

There are two main problems to be overcome in verifying equations (1) and (2). First, consistency of $\hat{\theta}^n$ *cannot* be combined with Slutsky's Theorem (as argued, for example, by Parks 1967, pp. 505–506) to establish equation (1) because $(1/n)Z_n'\Omega_n(\hat{\theta}^n)^{-1}X_n$ is a function of $\hat{\theta}^n$ that directly depends on n, while Slutsky's Theorem assumes the function is independent of the sample size. This problem arises in equation (2) as well, but a second difficulty with equation (2) is that Chebyschev's Inequality cannot be easily applied because the potential statistical dependence between $\hat{\theta}^n$ and u_n results in a complicated second moment matrix even if Z_n is nonstochastic, as in a standard GLS model. This is a serious problem since $\hat{\theta}^n$ and u_n are statistically dependent in most applications. Hence, consistency of $\hat{\theta}^n$ is sufficient for neither equation (1) nor equation (2), as demonstrated conclusively by Schmidt's (1976, p. 69) counter-example. Nevertheless, an approach like Theil's theorem can be successfully employed in a number of well-known cases with a consistent estimator $\hat{\theta}^n$ (for example, seemingly unrelated regressions Zellner 1962; finite order autoregressive errors Fuller 1976, pp. 424–425; and grouped heteroscedasticity Taylor 1977), and White applies his extension on a case-by-case to IV estimation of seemingly unrelated regressions with and without particular forms of heteroscedastic and autoregressive errors.

Since equations (1) and (2) involve only slightly simpler probability limits than the original problem, more easily verified sufficient conditions would be useful. Fuller and Battese (1973, Theorem 3, hereafter FB3) propose a second set of sufficient conditions for the standard GLS model. Their conditions are relatively easy to apply because only ordinary limits, differentiability, continuity, and the order in probability of $\hat{\theta}^n - \theta^0$ are involved. FB3 is used to establish asymptotic equivalence of Aitken and feasible Aitken estimators in standard GLS models by several authors in addition to Fuller and Battese (Swamy and Mehta 1977; Magnus 1978; Raj, Srivastava, and Ullah 1980; and Scott and Holt 1982).

Unfortunately, FB3 overlooks both of the aforementioned problems encountered in verifying equations (1) and (2). Crockett (1985, p. 203) notices that the proof of FB3 neglects the potential dependence between $\hat{\theta}^n$ and u_n, but does not demonstrate that the conclusion of FB3 is incorrect. In the Appendix, we establish conclusively that FB3 is incorrect by constructing a counter-example in which $\hat{\theta}^n$ is nonstochastic, so that the statistical dependence issue does not arise, but $(1/n)X_n'\Omega_n(\hat{\theta}^n)^{-1}X_n$ and $(1/\sqrt{n})X_n'\Omega_n(\hat{\theta}^n)^{-1}u_n$ still fail to converge to $(1/n)X_n'\Omega_n(\theta^0)^{-1}X_n$ and $(1/\sqrt{n})X_n'\Omega_n(\theta^0)^{-1}u_n$, respectively, because of their

[3] White relies on different conditions for the asymptotic normality than (A1) and (A2), and lists three conditions rather than equations (1) and (2) above. This is because he expresses his instruments as $Z_n' = X_n'\Omega_n(\hat{\theta}^n)^{-1}\bar{Z}_n(\bar{Z}_n'\Omega_n(\hat{\theta}^n)^{-1}\bar{Z}_n')^{-1}\bar{Z}_n'$ for some observable $(n \times k)$ matrix $\bar{Z}_n$, which allows White to state conditions in terms of the components X_n, Ω_n, and $\bar{Z}_n$ of Z_n. With this choice of instruments White's three conditions are sufficient for equations (1) and (2) above, which in turn are sufficient for $\operatorname{plim}_{n\to\infty} \sqrt{n}(\hat{\beta}_n - \tilde{\beta}_n) = 0$ (assuming (A1) and (A2)). Hence, under (A1) and (A2), equations (1) and (2) are at least as general as White's three conditions.

dependence on n. Hence, the conditions of FB3 do not assure that either of the difficulties of equations (1) and (2) are solved, and this deficiency means that more convenient conditions than equations (1) and (2) remain unknown. The breakdown of FB3 also means that estimators whose properties are established via FB3 may not possess the advertised properties. This is particularly troublesome for the panel estimation techniques of Swamy and Mehta (1977) and the two-step estimator of Magnus (1978) since these procedures are widely cited. Cragg (1992, p. 181) also overlooks the problems in verifying (1) and (2), as the example in the Appendix satisfies his assumptions but not his conclusions.

3. NEW SUFFICIENT CONDITIONS FOR ASYMPTOTIC EQUIVALENCE OF $\hat{\beta}_n$ AND $\tilde{\beta}_n$

Despite the known insufficiency of $\mathrm{plim}_{n\to\infty}\ \hat{\theta}^n = \theta^0$ for $\mathrm{plim}_{n\to\infty} \sqrt{n}(\hat{\beta}_n - \tilde{\beta}_n) = 0$, Theil's approach and White's extension are successful in many particular models when $\hat{\theta}^n$ is consistent. Our conditions clarify why $\mathrm{plim}_{n\to\infty} \sqrt{n}(\hat{\beta}_n - \tilde{\beta}_n) = 0$ in many familiar special cases of $\Omega_n(\theta)$ when $\hat{\theta}^n$ is consistent by summarizing the structure of these special cases that is important in proving the asymptotic equivalence.

The dependence on n can be partially addressed through an extension of Slutsky's Theorem provided by Amemiya (1985, Theorem 4.1.5) to the case of a sequence of functions that depend on n. The key to this extension is *uniform* convergence of the sequence as $n \to \infty$. Uniform convergence of individual elements of $\Omega_n(\theta)^{-1}$ holds in many models because these functions frequently do not even depend on n. For future reference, Amemiya's theorem for our context is restated here without proof.

LEMMA 1 (Amemiya 1985, Theorem 4.1.5). *Let $\varphi_n(\theta) \to \varphi(\theta)$ uniformly on an open set S containing θ^0, where φ is real-valued and continuous at θ^0. If* $\mathrm{plim}_{n\to\infty}$ $\hat{\theta}^n = \theta^0$, *then* $\mathrm{plim}_{n\to\infty}\ \varphi_n(\hat{\theta}^n) = \varphi(\theta^0)$.

The dependence on n is not completely resolved by this lemma because the dimensions of Z_n, $\Omega_n(\theta)^{-1}$, X_n, and u_n still depend on n. This leads to summations with n^2 terms in equations (1) and (2), which is problematic since the denominator of equation (1) in only $O(n)$ and the denominator of equation (2) is only $O(n^{1/2})$. In many models the number of nonzero elements in $\Omega_n(\theta)^{-1}$ is $O(n)$ so that the sums in equations (1) and (2) really have only $O(n)$ terms, but even with $O(n)$ terms the denominators of (1) and (2) may be inadequate for convergence unless the terms in the sums converge in some uniform manner. Often there are a finite number of distinct elements in $\Omega_n(\theta)^{-1}$ that simply repeat as $n \to \infty$. This imposes the needed uniformity provided the individual elements of $\Omega_n(\theta)^{-1}$ also satisfy Lemma 1, and the other random variables in the sums satisfy some routinely encountered "regularity" and "stability" conditions. In particular, equation (1) relies on certain fourth moments between the regressors and instruments possessing a uniform upper bound. Equation (2) relies on uniform boundedness of the second absolute moments of each instrument, error covariances conditional on the instruments that are uniformly bounded relative to the corresponding unconditional covariances, and absolutely convergent column-wise sums of the inverse covariance. This last condition is sometimes implied by stability of a stochastic process, and the conditional covariance condition is always true for a standard GLS model.

The statistical dependence between $\hat{\theta}^n$ and u_n is manageable if terms involving elements of $\Omega_n(\hat{\theta}^n)^{-1} - \Omega_n(\theta^0)^{-1}$ can be factored from the summations in equation (2), which include elements of u_n. This factoring occurs in most models as a consequence of the finite number of distinct functions in $\Omega_n(\theta)^{-1}$. Hence, a finite number of distinct functions in $\Omega_n(\theta)^{-1}$ can assist in two problems: the dependence between $\hat{\theta}^n$ and u_n, and the need for uniformity in the convergence of the terms of the sums.

Sufficiency of these properties for equations (1) and (2) is verified by the following theorem.

THEOREM 1. *In addition to* (A1) *and* (A2), *assume the following*:

(A3) $\Omega_n(\theta)^{-1}$ *has at most* $W < \infty$ *distinct nonzero elements for every* n, *denoted by* $g_{wn}(\theta)$ *for* $w = 1, \ldots, W$. *That is, there are* $n^2 - W$ *elements that are either zero or duplicates of other nonzero elements in* $\Omega_n(\theta)^{-1}$. *For each* w, $g_{wn}(\theta)$ *converges uniformly as* $n \to \infty$ *to a real-valued function* $g_w(\theta)$ *on an open set* S *containing* θ^0, *where* g_w *is continuous at* θ^0.

(A4) *The number of nonzero elements in each column (and row) of* $\Omega_n(\theta)^{-1}$ *is uniformly bounded by* $N < \infty$ *as* $n \to \infty$.

(A5) *There exists* $C < \infty$ *such that* $\sum_{i=1}^{n} |\omega_{ij}| \leq C$ *for every* $n = 1, 2, \ldots$ *and* $j = 1, \ldots, n$, *where* ω_{ij} *is the* (i, j) *element of* $\Omega_n(\theta^0)$.

(A6) *There exists* $B < \infty$ *such that*

(i) $E(z_{jh}x_{tq}z_{ih}x_{\tau q}) \leq B$ *for* $i, j, t, \tau = 1, 2, \ldots$ *and* $h, q = 1, \ldots, k$

(ii) $E(|z_{jh}z_{ih}|) \leq B$ *for* $i, j = 1, 2, \ldots$ *and* $h = 1, \ldots, k$

(iii) $|E(u_t u_\tau | z_{jh}, z_{ih})| \leq B|\omega_{t\tau}|$ *for* $i, j, t, \tau = 1, 2, \ldots$; $h = 1, \ldots, k$; *and almost every realization of* (z_{jh}, z_{ih}),

where z_{jh} *is the* (j, h) *element of* Z_n, x_{tq} *is the* (t, q) *element of* X_n, *and* u_t *is the* t*th element of* u_n. *Under* (A1) *through* (A6), *if* $\operatorname{plim}_{n\to\infty} \hat{\theta}^n = \theta^0$ *then* $\sqrt{n}(\tilde{\beta}_n - \beta) \xrightarrow{d} N(0, Q_{ZX}^{-1} Q Q_{ZX}^{-1})$. *That is,* (A1) *through* (A6) *are sufficient for the feasible Aitken estimator based on a consistent estimator of* θ^0 *to be asymptotically equivalent to the Aitken* IV *estimator (and hence consistent and asymptotically normal under* (A1) *and* (A2)).

PROOF. Since (A1) and (A2) are holding, White's result shows that it is sufficient to verify equations (1) and (2). Denote the (i, j) element of $\Omega_n(\theta)^{-1}$ by $f_{ijn}(\theta)$ and let $g_0(\theta) \equiv 0$ be the zero function. We must show a_{hq} and α_h are $o_p(1)$ for arbitrary (h, q), where a_{hq} is the (h, q) element of the left side of (1) and α_h is the hth element of the left side of (2). Letting $I_{iwn} \equiv \{j = 1, 2, \ldots, n : f_{ijn}(\theta) \equiv g_{wn}(\theta)\}$ be the index set of elements in row i of $\Omega_n(\theta)^{-1}$ that are equal to $g_{wn}(\theta)$, we have

$$a_{hq} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} z_{ih}x_{jq}(f_{ijn}(\hat{\theta}^n) - f_{ijn}(\theta^0))$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{w=1}^{W}\sum_{j\in I_{iwn}} z_{ih}x_{jq}(f_{ijn}(\hat{\theta}^n) - f_{ijn}(\theta^0)) + \sum_{j\notin \cup_{w=1}^{W} I_{iwn}} z_{ih}x_{jq}(f_{ijn}(\hat{\theta}^n) - f_{ijn}(\theta^0))\right]$$

$$= \frac{1}{n}\sum_{w=1}^{W}\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}x_{jq}(f_{ijn}(\hat{\theta}^n) - f_{ijn}(\theta^0)) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\notin \cup_{w=1}^{W} I_{iwn}} z_{ih}x_{jq}(f_{ijn}(\hat{\theta}^n) - f_{ijn}(\theta^0))$$

$$= \frac{1}{n}\sum_{w=1}^{W}\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}x_{jq}(g_{wn}(\hat{\theta}^n) - g_{wn}(\theta^0)) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\notin \cup_{w=1}^{W} I_{iwn}} z_{ih}x_{jq}(g_0(\hat{\theta}^n) - g_0(\theta^0))$$

$$= \frac{1}{n}\sum_{w=1}^{W}\left[(g_{wn}(\hat{\theta}^n) - g_{wn}(\theta^0))\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}x_{jq}\right].$$

By Lemma 1 and (A3), $g_{wn}(\hat{\theta}^n) - g_{wn}(\theta^0) = (g_{wn}(\hat{\theta}^n) - g_w(\theta^0)) + (g_w(\theta^0) - g_{wn}(\theta^0)) = o_p(1)$ for every $w = 1, \dots, W$. Since W is fixed as $n \to \infty$, we need only show that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}x_{jq} = O_p(1)$$

for arbitrary w. Taking the expectation of the square yields

$$E\left[\frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}x_{jq}\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t\in I_{jwn}}\sum_{\tau\in I_{iwn}} E(z_{ih}z_{jh}x_{tq}x_{\tau q})$$

$$\le \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t\in I_{jwn}}\sum_{\tau\in I_{iwn}} B \text{ by (A6(i))}$$

$$\le N^2 B \text{ by (A4)},$$

since each index set I_{jwn} and I_{iwn} contains no more than N elements.

Since the second moment bound is independent of n, applying Chebyschev's Inequality centered at 0 shows that $(1/n)\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}x_{jq} = O_p(1)$. The same manipulations used above show that

$$\alpha_h = \frac{1}{\sqrt{n}}\sum_{w=1}^{W}\left[(g_{wn}(\hat{\theta}^n) - g_{wn}(\theta^0))\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}u_j\right],$$

so we need only show that $(1/\sqrt{n})\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}u_j = O_p(1)$ for arbitrary w. Taking the expectation of the square yields

$$\begin{aligned}
E\left[\frac{1}{n}\left(\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}u_j\right)^2\right] &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t\in I_{jwn}}\sum_{\tau\in I_{iwn}} E(z_{jh}z_{ih}u_t u_\tau)\\
&\le \frac{1}{n}\sum_{j=1}^{n}\sum_{t\in I_{jwn}}\left[\sum_{i=1}^{n}\sum_{\tau\in I_{iwn}} |E(z_{jh}z_{ih}E(u_t u_\tau | z_{jh}z_{ih}))|\right]\\
&\le \frac{1}{n}\sum_{j=1}^{n}\sum_{t\in I_{jwn}}\left[\sum_{i=1}^{n}\sum_{\tau\in I_{iwn}} E(|z_{jh}z_{ih}||E(u_t u_\tau | z_{jh}z_{ih})|)\right]\\
&\le \frac{B}{n}\sum_{j=1}^{n}\sum_{t\in I_{jwn}}\left[\sum_{i=1}^{n}\sum_{\tau\in I_{iwn}} E(|z_{jh}z_{ih}|)|\omega_{t\tau}|\right] \text{ by (A6(iii))}\\
&\le \frac{B^2}{n}\sum_{j=1}^{n}\sum_{t\in I_{jwn}}\left[\sum_{i=1}^{n}\sum_{\tau\in I_{iwn}} |\omega_{t\tau}|\right] \text{ by (A6(ii))}.
\end{aligned}$$

By (A4), each τ belongs to at most N different index sets I_{iwn} for given w and n because the τth column of $\Omega(\theta^0)^{-1}$ has no more than N nonzero elements. Thus,

$$\sum_{i=1}^{n}\sum_{\tau\in I_{iwn}} |\omega_{t\tau}| \le N\sum_{\tau=1}^{n}|\omega_{t\tau}| \le NC \text{ by (A5)},$$

which implies

$$E\left[\frac{1}{n}\left(\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}u_j\right)^2\right] \le \frac{B^2}{n}\sum_{j=1}^{n}\sum_{t\in I_{jwn}} NC \le \frac{B^2}{n}\sum_{j=1}^{n} N^2C = B^2N^2C.$$

Since the second moment bound is independent of n, applying Chebyschev's Inequality centered at 0 shows that $(1/\sqrt{n})\sum_{i=1}^{n}\sum_{j\in I_{iwn}} z_{ih}u_j = O_p(1)$. □

REMARK. Asymptotic equivalence also means that (A1) through (A6) are sufficient for the feasible Aitken IV estimator to possess the asymptotic efficiency of the Aitken IV estimator, whether this be among: (i) linear unbiased estimators (even though the feasible estimator may be neither linear nor unbiased), as in a standard GLS model with (potentially) nonnormal u_n; (ii) consistent uniformly asymptotically normal estimators, as when $\hat{\beta}_n$ is asymptotically equivalent to the maximum likelihood estimator; (iii) the class of IV estimators, as in White (1984, Theorem 4.57); or (iv) some other class of estimators.

The example in the Appendix satisfies (A1) through (A6) for the choice of instruments $Z_n = X_n$, with the exception of (A3). This verifies the importance of (A3). To appreciate the importance of (A4), consider the following example. Let θ be a scalar and the (i, j) element of $\Omega_n(\theta)^{-1}$ be $f_{ij}(\theta) = \theta$ for $i \neq j$ and $f_{ii}(\theta) = 1 + \theta$, for $n = 1, 2, \dots$ and $i, j = 1, \dots, n$. Suppose X_n is a vector of ones and $\theta^0 = 0$. Since $\Omega_n(\theta^0) = I_n$, (A1), (A2), (A5), and (A6) are easily satisfied with instruments $Z_n = X_n$ provided the elements of u_n satisfy a central limit theorem. Since there are only two distinct f_{ij} functions, which are uniformly continuous on bounded sets and independent of n, (A3) is also satisfied. However, (A4) fails since $\Omega_n(\theta)^{-1}$ has n^2 nonzero elements. This leads to $(1/n)Z_n'(\Omega_n(\hat{\theta}^n)^{-1} - \Omega_n(\theta^0)^{-1})X_n = n\hat{\theta}^n$, which does not converge unless $\hat{\theta}^n = o_p(n^{-1})$. In particular, equation (1) can fail for consistent $\hat{\theta}^n$, such as $\hat{\theta}^n = (1/\sqrt{n})$.

FB3 uses a Taylor series expansion to approach the problem of establishing asymptotic equivalence of Aitken and feasible Aitken estimators. The Taylor series approach can be used to prove a variant of Theorem 1 provided that most of the boundedness assumptions of Theorem 1 still hold. Writing the elements of $\Omega_n(\theta)^{-1}$ as Taylor series about θ^0 essentially provides another method to accomplish the factoring of terms involving elements of $\Omega_n(\hat{\theta}^n)^{-1} - \Omega_n(\theta^0)^{-1}$ from the sums, thereby obtaining the needed uniformity and separation of $\hat{\theta}^n$ and u_n. One property needed to obtain the uniformity that assures convergence of the Taylor expansions is uniformly bounded derivatives of the elements in $\Omega_n(\theta)^{-1}$ at θ^0 as $n \to \infty$. This holds for most models, but even with bounded derivatives the Taylor residuals depend on $\hat{\theta}^n$. Thus, the Taylor residuals are statistically dependent on u_n, and this confounds attempts to use Chebyschev's Inequality. The only way to completely avoid the dependence problem is to factor *all* terms involving $\hat{\theta}^n$ from the sums involving u_n. The finite number of distinct elements assumed in (A3) facilitates this factoring, but without (A3) the Taylor series will accomplish this only if the Taylor residual disappears. That is, the elements of $\Omega_n(\theta)^{-1}$ must be proportional to *polynomials* in θ. Formally, Theorem 1 holds if (A3) is replaced by

(A3′) The (i, j) element of $\Omega_n(\theta)^{-1}$ can be expressed $g_n(\theta)f_{ijn}(\theta)$, where

(i) g_n is a sequence of functions, independent of (i, j), that converges uniformly to a real-valued function $g(\theta)$ as $n \to \infty$ on an open set S containing θ^0, where g is continuous at θ^0.
(ii) f_{ijn} is a polynomial of degree less than nz (z an integer), whose partial derivatives $D_{i_1}, \dots, {}_{i_l}f_{ijn}(\theta^0)$ at θ^0 ($l \leq nz$) are all bounded by W for every i, j, n (including $l = 0$. That is, including the function itself).

We will not repeat here the proof of Theorem 1 under this alternate assumption—a detailed proof is available from the authors on request. The alternate version of Theorem 1 may be useful in models where the number of distinct elements in $\Omega_n(\theta)^{-1}$ is not fixed.

4. APPLICATIONS

This section demonstrates that Theorem 1 applies to many familiar models and also most of the models that heretofore relied on FB3. Assumption (A3) is usually easier to verify and more widely applicable than (A3′), so we focus on applying the version of Theorem 1 proven above and simply note here that (A3′) may also be used in some of the models. Most of the asymptotic properties established in this section are not new, although some are, and the method of proof using Theorem 1 is new and demonstrates that the theorem is widely applicable precisely because many familiar models have the basic structure embodied in assumptions (A3) through (A5). This structure forms the basis for the previous case-by-case proofs of the known asymptotic properties.

To check the assumptions of Theorem 1, first recall that whenever asymptotic normality is investigated in a linear model (A1) must be established. This is usually accomplished by imposing assumptions more primitive than (A1) on the model and then utilizing a central limit theorem (CLT). For example, in a standard GLS model in which the transformed errors are IID, the Lindberg-Feller CLT applies. Alternatively, if the transformed errors are independent with bounded second and absolute third moments, use can be made of Liapunov's CLT. Schmidt (1976, p. 99) uses Schöenfeld's (1971) CLT to establish (A1) in the autoregressive model, while Campos (1986) uses Hannan's (1976) CLT to establish (A1) in a simultaneous equations model with lagged endogenous variables and VARMA errors. White (1984, Sections V.4 and V.5) discusses useful extensions of CLT's that rely on mixing distributions and martingale difference sequences. These extensions are particularly useful for IV estimators. Since (A1) is not our primary concern and the approach to (A1) varies depending on the model, we assume throughout this section that (A1) holds. Assumption (A2) is standard, while the regularity/stability assumption (A6) involves only moments and often cannot be verified or refuted. Therefore we assume throughout this section that assumptions (A1), (A2), and (A6) (hereafter called the maintained assumptions) hold, although there are well-known cases in which they fail. For instance, both (A2) and (A6) can fail if a time trend is included in the X_n matrix, while (A1) can fail if X_n is used for instruments when it is stochastic. In any particular context assumptions sufficient for the maintained assumptions must be in place before proceeding to apply Theorem 1. Given the maintained assumptions we must confirm (A3) through (A5) for any given covariance structure $\Omega_n(\theta)$.

The applications center around a seemingly unrelated regressions (SUR) model of m equations: $y_i = X_i\beta + u_i$ for $i = 1, \ldots, n$, where y_i and u_i are $(m \times 1)$ vectors of dependents and errors, respectively, for observation i on all m equations;

$$X_i = \begin{bmatrix} x'_{i1} & 0' & \cdots & 0' \\ 0' & x'_{i2} & \cdots & 0' \\ \vdots & \vdots & \ddots & \vdots \\ 0' & \cdots & 0' & x'_{im} \end{bmatrix}$$

is the $(m \times k)$ (possibly stochastic) matrix of regressors for observation i on all m equations (x_{ij} is the $(k_i \times 1)$ vector of regressors for observation i on equation j, the zero vectors are of conformable dimensions, and $k = \sum_{i=1}^{m} k_i$); and $\beta = (\beta'_1, \dots, \beta'_n)'$ is the $(k \times 1)$ vector of unknown parameters composed of the m $(k_i \times 1)$ subvectors β_i for each equation. Stacking the n observations yields the usual expression $y_n = X_n\beta + u_n$, where y_n and u_n are $(nm \times 1)$ and X_n is $(nm \times k)$, and $E(u_n) = 0$ with $E(u_n u'_n) = \Omega_n(\theta^0)$.

4.1. *Autoregressive Errors.* Suppose u_i follows the pth order vector autoregressive (VAR(p)) process

$$u_i = R_1 u_{i-1} + R_2 u_{i-2} + \cdots + R_p u_{i-p} + v_i \text{ for } i = 0, \pm 1, \pm 2, \dots,$$

where each R_j is an unknown $(m \times m)$ parameter matrix and $v_i \sim IID\,(0, \Sigma)$ for a symmetric positive definite $(m \times m)$ matrix Σ. Verification of (A3) and (A4) requires knowledge of the structure of $\Omega_n(\theta)^{-1}$, where θ consists of the distinct elements of Σ and each R_j. Following Guilkey and Schmidt (1973) and Judge et al. (1985, Section 12.3), since Σ is nonsingular Cholesky decomposition may be used to choose a lower triangular $(pm \times pm)$ matrix A such that $AE([u'_1 \dots u'_p]'[u'_1 \dots u'_p])A' = I_p \otimes \Sigma$. It is straightforward to verify that the $(mn \times mn)$ transformation matrix

$$P_n = \left[\begin{array}{ccc|cccccc}
 & A & & & & & & & \\
\hdashline
-R_p & \cdots & -R_1 & I_m & & & & & \\
 & \ddots & & & \ddots & & & & \\
 & & \ddots & & & \ddots & & & \\
 & & & & & & \ddots & & \\
 & & & \ddots & & & & \ddots & \\
 & & & & -R_p & \cdots & -R_1 & & I_m
\end{array}\right]$$

yields transformed errors $P_n u_n$ satisfying $E(P_n u_n (P_n u_n)') = I_n \otimes \Sigma$, so $\Omega_n(\theta)^{-1} = P'_n(I_n \otimes \Sigma^{-1})P_n$.[4] Carrying out this multiplication reveals that the upper left $(pm \times pm)$ block of $\Omega_n(\theta)^{-1}$ is $A'(I_p \otimes \Sigma^{-1})A + B'(I_p \otimes \Sigma^{-1})B$ and the lower right $(pm \times pm)$ block is $T'(I_p \otimes \Sigma^{-1})T$, where

[4] For the $m = 1$ and $p = 2$ case the (2, 1) element of A disagrees with the transformation provided by Fomby et al. (1988, p. 218) and Greene (1993, p. 429) but agrees with the transformation given by Fuller (1976, p. 423) and Judge et al. (1985, p. 294).

$$B = \begin{bmatrix} R_p & R_{p-1} & \cdots & \cdots & R_1 \\ 0 & \ddots & \ddots & \ddots & R_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & R_{p-1} \\ 0 & \cdots & \cdots & 0 & R_p \end{bmatrix}, \quad T = \begin{bmatrix} R_0 & 0 & \cdots & \cdots & 0 \\ R_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ R_{p-1} & \cdots & \cdots & R_1 & R_0 \end{bmatrix},$$

and we have defined $R_0 = -I_m$ for notational convenience. The $(m \times m)$ block of $\Omega_n(\theta)^{-1}$ in the (i, j) position (in terms of $(m \times m)$ blocks) is $\sum_{t=0}^{p-|i-j|} R'_t \Sigma^{-1} R_{t+|i-j|}$ for $p < \max\{i, j\}$ and $\min\{i, j\} \leq n - p$, where it is understood that this sum is zero when $p - |i - j| < 0$. Hence all elements of $\Omega_n(\theta)^{-1}$ that lie more than p $(m \times m)$ blocks off the main diagonal are zero, verifying (A4) with $N = m(2p + 1)$. Since the blocks of $\Omega_n(\theta)^{-1}$ are independent of n there are at most $W = pm(pm + 1) + (p + 1)m^2$ distinct functions in $\Omega_n(\theta)^{-1}$, all of which are independent of n for $n \geq W$ (implying uniform convergence trivially) and continuous at θ^0 because the operations involved in obtaining the blocks from θ are continuous at θ^0 when Σ is nonsingular. This verifies (A3).

All that remains is to verify (A5). Following Anderson (1971, p. 177) the m-dimensional VAR(p) process may be rewritten as a pm-dimensional VAR(1) process $e_i = Re_{i-1} + \varepsilon_i$, where $e_i = (u'_{i-p+1} \dots u'_i)'$ and $\varepsilon_i = (0' \dots 0'v'_i)'$ are $(pm \times 1)$ vectors, and

$$R = \begin{bmatrix} 0 & I_m & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & I_m \\ R_p & R_{p-1} & \cdots & \cdots & \cdots & R_1 \end{bmatrix}$$

is an $(mp \times mp)$ matrix. If the process is covariance stationary then the absolute eigenvalues of R are less than one, and also $E(e_i e'_j) = R^{|i-j|} E(e_t e'_t)$ for arbitrary t (Anderson 1971, p. 182). From the definition of e_i, the sum $\sum_{i=1}^n |E(u_i u'_j)|$ is the lower right $(m \times m)$ block of $\sum_{i=1}^n |E(e_i e'_j)|$, where $|\cdot|$ denotes the element-wise absolute value. So, if the elements of the latter sum are bounded independently of j as $n \to \infty$ then (A5) holds. Since

$$\sum_{i=1}^{n} |E(e_i e'_j)| \leq 2 \sum_{i=0}^{n} |E(e_i e'_0)| \leq 2 \left(\sum_{i=0}^{n} |R^i| \right) |E(e_0 e'_0)|,$$

rewriting $|R^i|$ in the Jordan Canonical Form yields

$$\sum_{i=1}^{n} |E(e_i e_j')| \leq 2|J| \left(\sum_{i=0}^{n} |\Lambda^i| \right) |J^{-1}| \; |E(e_0 e_0')|,$$

where Λ is a block diagonal matrix involving the eigenvalues of R and J is a fixed matrix depending only on R. Since the absolute eigenvalues are less than one $\sum_{i=0}^{\infty} |\Lambda^i|$ converges, which verifies (A5).

We have shown that, if the VAR(p) error process is covariance stationary and the maintained assumptions hold, then Theorem 1 applies and any feasible Aitken IV estimator based on consistent estimators of the contemporaneous covariance Σ and the VAR parameters R_i possesses the same asymptotic properties as the Aitken IV estimator. To our knowledge, this result has not previously appeared in the literature at the level of generality presented here. Consistent estimators for the VAR(1) case are discussed by Guilkey and Schmidt (1973) and Judge et al. (1985, section 12.3), and Theorem 1 verifies the asymptotic distribution of the feasible Aitken estimator for this case. Zellner's (1962) estimator and its variants (Zellner and Huang 1962) are all consistent, so Theorem 1 establishes Zellner's result for his special case of $p = 0$. Parks (1967) (whose results are suspect due to the reliance on Slutsky's Theorem to establish equations (1) and (2)) considers an SUR model with AR(1) errors. All of these models are special cases of the error structure considered here and are in the standard GLS rather than the IV context, as are typical treatments of single equation models with AR errors (for example, Fuller 1976, Theorem 9.7.1, who provides a consistent estimator of θ^0 for the AR(p) case). Anderson (1971, section 10.3.2) provides consistent estimators of the covariances of a stationary error process, but his approach does not assume a VAR(p) structure and also assumes the standard GLS context. Moreover, Anderson does not investigate feasible Aitken estimation of β. One unresolved problem with the general VAR(p) case is that, as far as we know preliminary consistent estimators of Σ and the R_j's for use in a two-step estimator for β have not been proposed.

The exact inverse $\Omega_n(\theta)^{-1}$ derived by Uppuluri and Carpenter (1969) when u_n follows an MA(1) process satisfies neither (A3) nor (A4), so Theorem 1 does not generally apply to VARMA error processes. To our knowledge, asymptotic equivalence of Aitken and feasible Aitken estimators has not been established for a general consistent estimator $\hat{\theta}^n$ when u_n is VARMA without a distributional assumption on u_n, even for the single equation nonstochastic regressor context with MA(1) errors.[5] Zinde-Walsh and Galbraith (1991) consider general ARMA errors but only do so for one equation and compare the feasible Aitken and maximum likelihood estimators with normal u_n, rather than the Aitken and feasible Aitken estimators with no particular error distribution. Amemiya (1973a) considers ARMA errors in a standard GLS model, but only establishes results for a specific estimator $\hat{\theta}^n$. Campos (1986) considers VARMA errors in a general simultaneous equations model, but does not examine feasible Aitken estimation of β.

Note finally that some approaches to estimation with AR(p) errors unnecessarily rely on normality to obtain $\Omega_n(\theta)^{-1}$ (Siddiqui 1958, Galbraith and Galbraith 1974,

[5] However, Fuller (1976, p. 425) mentions this problem.

Shaman 1975), or on approximations (Shaman 1975), and some approaches give alternative estimators for β that are consistent and asymptotically normal when errors are AR(p) or VAR(p) (Durbin 1960, Pierce 1971, White 1984 section VII.4, Campos 1986), but do not establish asymptotic equivalence of feasible Aitken and Aitken estimators, particularly for an arbitrary consistent estimator $\hat{\theta}^n$. All of the previous approaches to autocorrelation mentioned in this subsection except Guilkey and Schmidt, Parks, White, and Campos are for the single equation context, and all but White and Campos are for the standard GLS rather than the IV context.

4.2. *Heteroscedasticity.* Suppose now that u_i is heteroscedastic with $E(u_i u_i') = \Sigma_i$ for symmetric positive definite $(m \times m)$ matrices Σ_i and $E(u_i u_j') = 0$ for $i \neq j$. Then $\Omega_n(\theta)$ and $\Omega_n(\theta)^{-1}$ are block diagonal with diagonal blocks Σ_i and Σ_i^{-1}, respectively, so (A4) holds with $N = m$ and (A5) holds provided the contemporaneous covariances have an upper bound as $n \to \infty$. This is an assumption that is likely to be placed on any heteroscedastic model and may be implied by other assumptions for any given heteroscedastic structure, leaving only assumption (A3) as a serious problem in heteroscedastic models. White (1984, section VII.3) extends Taylor's (1977) model of constant variances within subgroups of observations to the SUR IV context. In this model θ consists of the distinct elements of a fixed number of distinct Σ_i matrices. This yields a fixed number of distinct elements in $\Omega_n(\theta)^{-1}$, none of which depend on n and all of which are continuous at θ^0 since each Σ_i is nonsingular. Hence (A3) is satisfied, and any consistent estimator of the contemporaneous covariance matrices in this model yields a consistent and asymptotically normal feasible Aitken IV estimator of β under the maintained assumptions. Taylor provides a consistent estimator of θ^0 for the single equation standard GLS case, and White provides a consistent estimator for the SUR IV context. Note that $n \to \infty$ in this model by letting the size of one or more subgroups of observations grow while keeping the number of subgroups fixed. Depending on the estimator $\hat{\theta}^n$, it may be necessary to let the size of all subgroups approach ∞ in order to establish consistency of $\hat{\theta}^n$.

Alternatively, suppose all n Σ_i matrices are distinct. Models of this type are frequently estimated by assuming some parametric structure for the heteroscedasticity of the form $\sigma_{hq}^i = g(\theta' v_{hq}^i)$, where σ_{hq}^i is the (h, q) element of Σ_i and v_{hq}^i is a vector of nonstochastic explanatory variables (for single equation models, see Goldberger 1964, Park 1966, Glejser 1969, Goldfeld and Quandt 1972, Amemiya 1973b, Harvey 1976, and Amemiya 1977; while multiple equation models are discussed by Singh and Ullah 1974 and Mandy and Martins-Filho 1993, who also discuss panel data models). In general, (A3) fails in these models because the number of distinct elements in $\Omega_n(\theta)^{-1}$ increases with n. However, this is partially a product of the way the limiting process is envisioned. If the limiting process on

$$V_n' \equiv [v_{11}^{1\prime} \cdots v_{mm}^{1\prime} v_{11}^{2\prime} \cdots v_{mm}^{2\prime} \cdots v_{11}^{2\prime} \cdots v_{mm}^{n\prime}]'$$

is "constant in repeated samples" (CRS; see Theil 1971, pp. 364–365) then the sample size is pn, where p is the number of repeated samples, and the sample size tends to infinity by letting $p \to \infty$. Hence, as $p \to \infty$ we can set $W = m(m + 1)n/2$

and the model behaves like Taylor's "constant within subgroups" model asymptotically. By this argument Theorem 1 applies to any model with a block diagonal covariance matrix if the maintained assumptions hold and the blocks repeat as $n \to \infty$. Therefore, in parametric heteroscedastic models under CRS we need only obtain a consistent estimator of θ^0 in order to immediately establish the asymptotic properties of feasible Aitken IV estimators via Theorem 1. Consistent estimators of θ^0 are readily available (Judge et al. 1985, pp. 431–441 provides a summary of single equation estimators, while consistent multiple equation and panel data estimators are discussed by Singh and Ullah 1974 and Mandy and Martins-Filho 1993).

Note that the Hildreth-Houck (1968) random coefficients model is included in this conclusion as a special case in which there is one equation, $g(\theta' v_{11}^{i}) = \theta' v_{11}^{i}$, and $v_{11}^{i} = (x_{i1}^2 \dots x_{ik}^2)$, even though previous proofs of asymptotic normality of the Hildreth-Houck estimator have required an estimator $\hat{\theta}^n$ of θ^0 satisfying $\hat{\theta}^n - \theta^0 = O_p((pn)^{-1/2})$ (Crockett 1985, Mandy and Martins-Filho 1993). Thus, asymptotic normality of the Hildreth-Houck estimator persists with weaker convergence of $\hat{\theta}^n$ provided the limiting process is CRS. This conclusion holds *only* because of the CRS assumption. Without CRS (A3) fails in the Hildreth-Houck model and therefore Theorem 1 is inapplicable. Hence, Theorem 1 does not provide a stronger result than Crockett or Mandy and Martins-Filho. Rather, Theorem 1 permits substitution of the CRS assumption for the stronger convergence of $\hat{\theta}^n$ required by these authors. These authors obtain the result without the CRS assumption because their stronger convergence, working through the theorem by Carroll and Ruppert (1982), provides an alternative to the fixed number of distinct elements used by Theorem 1 to enable application of Chebyschev's Inequality in overcoming the statistical dependence problem of equation (2). Carroll and Ruppert construct an application of Chebyschev's Inequality for equation (2) under heteroscedasticity when there is not a fixed number of distinct elements in $\Omega_n(\theta)^{-1}$, but only with convergence assumptions that require $\hat{\theta}^n = O_p((pn)^{-1/2})$ rather than the weaker consistency of $\hat{\theta}^n$.

The CRS assumption could help prove asymptotic equivalence with other, not necessarily heteroscedastic, error structures. In general, if this assumption fixes the number of distinct elements in the inverse covariance matrix as $n \to \infty$, then it will assist in establishing (A3).

Note also that, with the exception of White, the heteroscedastic models considered in this subsection were originally developed in a standard GLS context, and many have only been proposed in single equation models, but the asymptotic properties from Theorem 1 hold in any multiple equation IV context with heteroscedastic errors provided a consistent estimator of $\hat{\theta}^n$ is available and the maintained assumptions are satisfied as well as (A3). The main difference between the results for heteroscedastic models presented here and those of White is our emphasis on the structure of $\Omega_n(\theta)$ to establish the asymptotic distribution of $\sqrt{n}(\bar{\beta}_n - \beta)$.

4.3. *Special Cases*. The three stage least squares estimator for a simultaneous equations model with or without lagged endogenous variables is a feasible Aitken IV estimator in an SUR system; with the particular choice of instruments $Z_n = (D(D'D)^{-1}D' \otimes I_m)X_n$, where D is the $(n \times T)$ matrix of $T \leq k$ predetermined variables in the system; and the particular consistent estimator for the (h, q) element of the contemporaneous correlation matrix $(1/n) \sum_{i=1}^{n} (y_{ih} - x'_{ih}\bar{\beta}_h)(y_{iq} - x'_{iq}\bar{\beta}_q)$, where y_{ih} is the hth element of y_i and $\bar{\beta}_h$ is the two stage least squares estimator of β_h. Hence the proof of consistency and asymptotic normality of the three stage least squares estimator can be viewed as a special case of Theorem 1. Note that (A2) is satisfied in this context if all equations are identified by exclusion restrictions (Schmidt 1976, p. 205), while (A6) is usually unverifiable and therefore a pure assumption. However, some conditions on the instruments are needed for a general result like Theorem 1 even though they may not ordinarily be assumed when examining properties of three stage least squares. For example, White (1984, p. 171) notes that his results apply to three stage least squares with lagged endogenous variables provided his unverifiable assumptions on the instruments hold, and Campos (1986, Appendix B) also imposes assumptions on the instruments in a dynamic simultaneous equations model.

More importantly, no additional effort is required to conclude from Theorem 1 that the error variance structure in a simultaneous equations model need not be restricted to contemporaneous correlation in order to obtain a consistent and asymptotically normal system estimator for β via feasible Aitken IV estimation. All that is required is that instruments satisfying the maintained assumptions be available; the error structure satisfy (A3) through (A5), as is the case with the VAR(p) and heteroscedastic structures considered above; and a consistent estimator for the covariance matrix parameters be available. Note, however, that there may be additional identification considerations with such a model (see Harvey 1990, pp. 347–348). Similarly, two stage least squares modified to accommodate autoregressive or heteroscedastic errors in the equation is a consistent and asymptotically normal estimator for a single equation provided the maintained assumptions hold and a consistent estimator $\hat{\theta}^n$ is available.

The discussion of SUR models in the previous subsections also shows that the asymptotic distribution of the feasible Aitken IV estimator in any stochastic regressor model can be established via Theorem 1 when errors are autoregressive or heteroscedastic, provided instruments satisfying the maintained assumptions are available. This proviso may be difficult to overcome, however, as discussed by Schmidt (1976, Chapter 3). As a special case, the standard single equation autoregressive model (Schmidt 1976, section 3.2) is included in this conclusion.

Another area of application for Theorem 1 is models that heretofore relied upon FB3. All of these models were originally treated in a standard GLS context, so when Theorem 1 applies we have a generalization of the original results to the IV context. The nested-error models considered by Fuller and Battese (1973) and the two-stage sampling model of Scott and Holt (1982) both have block diagonal covariance matrices. Thus, as discussed in the subsection on heteroscedasticity, Theorem 1 applies provided the blocks repeat as $n \to \infty$. For Fuller and Battese, this means that n must approach ∞ by letting the number of "individuals" approach ∞ while the numbers of "measurements" and "determinations" are bounded. For Scott and Holt, the number of "clusters" must approach ∞ while the size of the clusters is bounded. Neither paper mentions this type of restriction because both rely on FB3, but the general need for a restriction of this type is not surprising in light of the discussion by Anderson and Hsiao (1982). Fuller and Battese propose consistent estimators for the parameters of their covariance matrix. Swamy and Mehta (1977) use FB3 on a covariance matrix that has $n(n+1)/2$ distinct elements, so none of the assumptions (A3) through (A5) necessarily hold. Since Raj et al. (1980) rely on Swamy and Mehta, the asymptotic properties of the estimators considered in both of these papers remain unknown unless a further restriction is added. The CRS assumption added to either of these models establishes the asymptotic properties, as discussed in the previous subsection, and both papers provide consistent estimators for the covariance matrix parameters. Unfortunately, Magnus (1978) has little assumed structure on the covariance matrix, so Theorem 1 does not reestablish the properties of his two-step estimator unless further restrictions are placed on the model. The general covariance structure considered by Magnus must be assumed to satisfy assumptions (A3) through (A5) in order for Theorem 1 to reestablish the asymptotic properties of his estimator. In general, Theorem 1 does not apply whenever the number of nonzero elements in the inverse covariance matrix may not be $O(n)$. The proof of Theorem 1 suggests that there is little hope for establishing asymptotic equivalence when the number of nonzero elements in the inverse covariance matrix is not $O(n)$ unless some additional convergence property is available to further restrict the convergence in probability of the terms in the sums in equations (1) and (2).

5. SUMMARY

Since the small sample properties of feasible Aitken IV estimators are usually too complicated to establish, only asymptotic properties are known for most estimators under nonscalar identity error covariance structures. These properties are usually established through a tedious case-by-case process of showing convergence in probability of the feasible Aitken IV estimator to the Aitken IV estimator. This process is further complicated by the presence of incorrect extant sufficient conditions for convergence of the two estimators. We introduce new sufficient conditions that greatly ease the task of demonstrating convergence in probability. These new sufficient conditions arise from the observation that in many models where asymptotic properties are known, the properties follow from three basic conditions: consistency of the estimator for the parameters of the covariance matrix, $O(n)$ nonzero elements composed of a fixed number of distinct elements in the inverse covariance matrix (or elements that are proportional to a polynomial in the covariance matrix parameters), and absolutely convergent column-wise sums of the covariance matrix. Since the new sufficient conditions are derived from existing models, they provide a unified method for proving asymptotic equivalence of Aitken and feasible Aitken estimators in many familiar models. Since the new conditions apply in IV contexts rather than only standard GLS contexts, they extend many GLS results to the IV framework. Moreover, the new conditions have the potential to quickly demonstrate asymptotic equivalence of Aitken and feasible Aitken estimators in linear models with heretofore unexplored error structures.

University of Missouri and University of Tennessee, U.S.A.
Oregon State University, U.S.A.

APPENDIX

THEOREM (Fuller and Battese 1973, p. 629, Theorem 3). *Assume the three "regularity conditions:"*

(R1) *For every n, the elements of $\Omega_n(\theta)^{-1}$ are continuously differentiable with respect to each component of θ in an open sphere S containing θ^0. Denote the matrix of derivatives with respect to the* ith *component of θ by $G_{ni}(\theta)$ for $i = 1, \dots, r$.*

(R2) $\lim_{n\to\infty} (1/n)X'_n\Omega_n(\theta)^{-1}X_n$ *is a nonsingular matrix for every $\theta \in S$, and the elements of*

$$\lim_{n\to\infty} \frac{1}{n} X'_n G_{ni}(\theta) X_n$$

are continuous functions of θ for $i = 1, \dots, r$.

(R3) *For every n, $\Omega_n(\hat{\theta}^n)$ is nonsingular, and there exists $\delta > 0$ such that $\hat{\theta}^n - \theta^0 = O_p(n^{-\delta})$.*

Then, $\tilde{\beta}_n$ possesses the same asymptotic distribution as $\hat{\beta}_n$.

Fuller and Battese seek to prove this theorem by showing that $\tilde{\beta}_n - \hat{\beta}_n = O_p(n^{-1/2-\delta})$. Unfortunately, this order in probability does not follow, as the following counter-example demonstrates.

A.1. *Example.* Let X_n be an $(n \times 1)$ vector of ones and assume for this example that u_n is normal with

$$\Omega_n(\theta) = \operatorname{diag}\left\{\frac{1}{f_1(\theta)}, \frac{1}{f_2(\theta)}, \dots, \frac{1}{f_n(\theta)}\right\},$$

where θ is a scalar and

$$f_t(\theta) = \begin{cases} 1, & \theta \in \left(-\infty, \dfrac{1}{t+1}\right] \cup \left(\dfrac{1}{t+1} + \dfrac{4}{c_t}, \infty\right) \\ \dfrac{t}{2} c_t^2\left(\theta - \dfrac{1}{t+1}\right)^2 + 1, & \theta \in \left(\dfrac{1}{t+1}, \dfrac{1}{t+1} + \dfrac{1}{c_t}\right] \\ t - \dfrac{t}{2} c_t^2\left(\theta - \left(\dfrac{1}{t+1} + \dfrac{2}{c_t}\right)\right)^2 + 1, & \theta \in \left(\dfrac{1}{t+1} + \dfrac{1}{c_t}, \dfrac{1}{t+1} + \dfrac{3}{c_t}\right] \\ \dfrac{t}{2} c_t^2\left(\theta - \left(\dfrac{1}{t+1} + \dfrac{4}{c_t}\right)\right)^2 + 1, & \theta \in \left(\dfrac{1}{t+1} + \dfrac{3}{c_t}, \dfrac{1}{t+1} + \dfrac{4}{c_t}\right], \end{cases}$$

$$c_t = 2t(t+1).$$

A graph of $f_t(\theta)$ is given in Figure 1. Assume $\theta^0 = 0$ and let $\hat{\theta}^n = (1/n)$ be our estimator for θ^0.

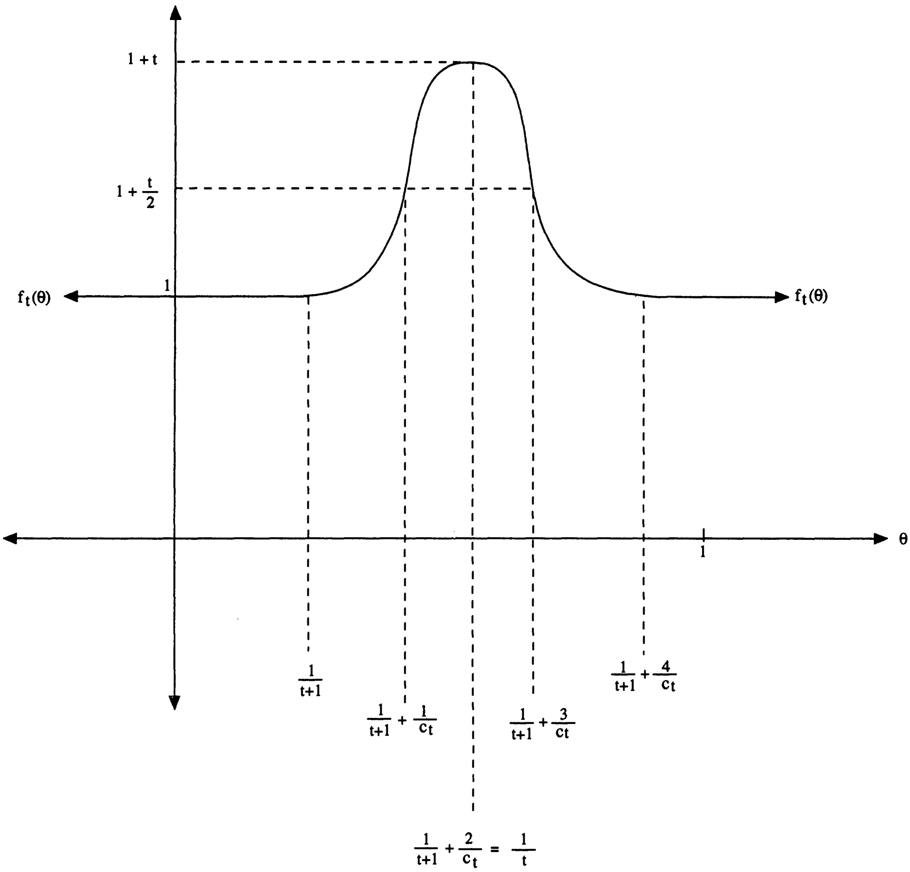

FIGURE 1

A.1.1. *Remarks.*

1. $f_t(\theta)$ is constructed from parabolas whose slopes are equal where they join. Hence, $f_t(\theta)$ is continuously differentiable on $\Re$ for $t = 1, 2, \ldots$.
2. As t increases, the height of the "spike" increases at a rate of t and the spike moves left and narrows.
3. $1/(t+1) + 4/c_t < 1/(t-1)$ for $t > 1$. Hence, the entire spike occurs between $1/(t+1)$ and $1/(t-1)$, and is centered at $1/t$.
4. $X'_n \Omega_n(\theta)^{-1} X_n = \sum_{t=1}^{n} f_t(\theta)$.
5. $X'_n G_n(\theta) X_n = \sum_{t=1}^{n} f'_t(\theta)$, which exists by remark 1.

A.1.2. *Verification of Regularity Conditions* (R1) *through* (R3).

(R1) This holds for any θ^0 and $S = \Re$, by remark 1.
(R2) $f_t(\theta) = 1$ for $\theta \in (-\infty, 0] \cup [(3/2), \infty)$ for every t. Thus, for θ in this part of the domain

$$\lim_{n\to\infty} \frac{1}{n} X'_n \Omega_n(\theta)^{-1} X_n = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} f_t(\theta) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} 1 = 1.$$

For $\theta \in (0, (3/2))$, let $n_\theta = \max\{n : \theta < 1/(n-1)\}$ where 1/0 is to be regarded as ∞ so that $n_\theta = 1$ for $\theta \in [1, (3/2))$. Note that n_θ exists and is finite for every $\theta \in (0, (3/2))$, and that $f_t(\theta) = 1$ for $t > n_\theta$. Thus, for $\theta \in (0, (3/2))$

$$\lim_{n\to\infty} \frac{1}{n} X'_n \Omega_n(\theta)^{-1} X_n = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} f_t(\theta) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n_\theta} f_t(\theta) + \lim_{n\to\infty} \frac{1}{n} \sum_{t=n_\theta+1}^{n} f_t(\theta)$$

$$\leq \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n_\theta} (1+t) + \lim_{n\to\infty} \frac{1}{n} \sum_{t=n_\theta+1}^{n} 1 \text{ since}$$

$$1 + t \text{ is the maximum of } f_t(\theta) \text{ on } \Re$$

$$\leq \lim_{n\to\infty} \frac{n_\theta(1+n_\theta)}{n} + \lim_{n\to\infty} \frac{n-n_\theta}{n} = 0 + 1 = 1.$$

Similarly,

$$\lim_{n\to\infty} \frac{1}{n} X'_n \Omega_n(\theta)^{-1} X_n \geq \lim_{n\to\infty} \frac{1}{n} \sum_{t=n_\theta+1}^{n} f_t(\theta) \text{ since } f_t(\theta) > 0 \text{ on } \Re \text{ for every } t.$$

$$= \lim_{n\to\infty} \frac{n-n_\theta}{n} = 1.$$

Thus,

$$\lim_{n\to\infty} \frac{1}{n} X'_n \Omega_n(\theta)^{-1} X_n = 1 \quad \forall \quad \theta \in \Re,$$

and 1 is invertible, so the first part of condition (R2) is verified. To verify the second part, note that if $\theta \in (-\infty, 0] \cup [(3/2), \infty)$ then $f'_t(\theta) = 0$ for every t. Thus, for θ in this part of the domain

$$\lim_{n\to\infty} \frac{1}{n} X'_n G_n(\theta) X_n = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} f'_t(\theta) = 0.$$

For $\theta \in (0, (3/2))$, note that $f'_t(\theta) = 0$ for $t > n_\theta$, so

$$\lim_{n\to\infty}\left|\frac{1}{n}X'_nG_n(\theta)X_n\right| = \lim_{n\to\infty}\left|\frac{1}{n}\sum_{t=1}^{n} f'_t(\theta)\right| \le \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}|f'_t(\theta)|$$

$$\le \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} tc_t \text{ since } tc_t \text{ is the maximum of } |f'_t(\theta)| \text{ on } \Re$$

$$\le \lim_{n\to\infty}\frac{1}{n}(n_\theta^2 c_{n_\theta}) = 0.$$

Thus, $\lim_{n\to\infty}(1/n)X'_nG_n(\theta)X_n = 0$ for every $\theta \in \Re$, which is a continuous function of θ on $\Re$, so the second part of condition (R2) is verified.

(R3) Clearly $\hat{\theta}^n - \theta^0 = O_p(n^{-1})$, so $\delta = 1$. In fact, $\hat{\theta}^n$ is nonstochastic so that $\hat{\theta}^n - \theta^0 = O(n^{-1})$. Moreover, $\Omega_n(\hat{\theta}^n)^{-1}$ exists for all n since $f_t(\theta) > 0$ for every $\theta \in \Re$ and $t = 1, 2, \ldots$.

A.1.3. *Demonstration that* $\hat{\beta}_n - \tilde{\beta}_n \neq O_p(n^{-1/2-\delta})$. Rather than writing a Taylor expansion as in Fuller and Battese, we shall examine directly the difference $\tilde{\beta}_n - \hat{\theta}_n$:

$$\begin{aligned}\tilde{\beta}_n - \hat{\beta}_n &= (X'_n\Omega_n(\hat{\theta}^n)^{-1}X_n)^{-1}X'_n\Omega_n(\hat{\theta}^n)^{-1}u_n \\ &\quad - (X'_n\Omega_n(\theta^0)^{-1}X_n)^{-1}X'_n\Omega_n(\theta^0)^{-1}u_n \\ &= \frac{\sum_{t=1}^{n} f_t(\hat{\theta}^n)u_t}{\sum_{t=1}^{n} f_t(\hat{\theta}^n)} - \frac{\sum_{t=1}^{n} f_t(\theta^0)u_t}{\sum_{t=1}^{n} f_t(\theta^0)} \text{ where } u_t \text{ is the } t\text{th element of } u_n \\ &= \frac{\sum_{t=1}^{n} f_t(\hat{\theta}^n)u_t}{\sum_{t=1}^{n} f_t(\hat{\theta}^n)} - \frac{1}{n}\sum_{t=1}^{n} u_t \text{ since } f_t(\theta^0) = f_t(0) = 1 \text{ for every } t \\ &= \sum_{t=1}^{n}\left(\frac{f_t\left(\frac{1}{n}\right)}{\sum_{i=1}^{n} f_i\left(\frac{1}{n}\right)} - \frac{1}{n}\right)u_t\end{aligned}$$

$$= \left(\frac{f_n\left(\frac{1}{n}\right)}{(n-1) + f_n\left(\frac{1}{n}\right)} - \frac{1}{n} \right) u_n + \sum_{t=1}^{n-1} \left(\frac{1}{(n-1) + f_n\left(\frac{1}{n}\right)} - \frac{1}{n} \right) u_t$$

since $f_t(1/n) = 1$ for $t < n$ by remark 3,

$$= \left(\frac{(1+n)}{(n-1) + (1+n)} - \frac{1}{n} \right) u_n + \sum_{t=1}^{n-1} \left(\frac{1}{(n-1) + (1+n)} - \frac{1}{n} \right) u_t$$

$$= \frac{1}{2} u_n - \frac{1}{2n} \sum_{t=1}^{n} u_t .$$

Therefore,

$$\sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n) = \frac{\sqrt{n}}{2} u_n - \frac{1}{2\sqrt{n}} \sum_{t=1}^{n} u_t .$$

Now, $E(1/(2\sqrt{n}) \sum_{t=1}^{n} u_t) = 0$ and

$$\operatorname{Var}\left(\frac{1}{2\sqrt{n}} \sum_{t=1}^{n} u_t \right) = \frac{1}{4n} \sum_{t=1}^{n} E(u_t^2) = \frac{1}{4n} \sum_{t=1}^{n} \frac{1}{f_t(\theta^0)} = \frac{1}{4n} \sum_{t=1}^{n} \frac{1}{f_t(0)} = \frac{1}{4},$$

where we have used the zero covariance of the u_t's and the fact that $f_t(0) = 1$ for every t. Thus, by Chebyschev's Inequality $(1/(2\sqrt{n})) \sum_{t=1}^{n} u_t$ is bounded in probability. But

$$\begin{aligned} \left| \frac{\sqrt{n}}{2} u_n \right| &= \left| \sqrt{n}(\hat{\beta}_n - \tilde{\beta}_n) + \frac{1}{2\sqrt{n}} \sum_{t=1}^{n} u_t \right| \\ &\leq \sqrt{n} |\hat{\beta}_n - \tilde{\beta}_n| + \left| \frac{1}{2\sqrt{n}} \sum_{t=1}^{n} u_t \right| \\ &= \sqrt{n} |\hat{\beta}_n - \tilde{\beta}_n| + O_p(1), \end{aligned}$$

so if $(\sqrt{n}/2)u_n$ is unbounded in probability then $\sqrt{n}(\tilde{\beta}_n - \hat{\beta}_n)$ is also unbounded in probability. But $P((\sqrt{n}/2)|u_n| < B)$ is the area under a $N(0, 1)$ density between $-(2B/\sqrt{n})$ and $(2B/\sqrt{n})$ for any $B > 0$, so $\lim_{n\to\infty} P((\sqrt{n}/2)|u_n| < B) = 0$ and $(\sqrt{n}/2)u_n$ is unbounded in probability. Therefore, $\tilde{\beta}_n - \hat{\beta}_n$ is not $O_p(n^{-1/2})$ and is also not $O_p(n^{-1/2-\delta}) = O_p(n^{-3/2})$. Thus, the conclusion that $\tilde{\beta}_n$ possesses the same asymptotic distribution as $\hat{\theta}_n$ cannot be established through reliance on $\tilde{\beta}_n - \hat{\theta}_n = O_p(n^{-1/2-\delta})$. In fact, in the present example $\sqrt{n}(\hat{\theta}_n - \beta) \sim N(0, 1)$ for every

n while $\sqrt{n}(\tilde{\beta}_n - \beta) \sim N(0, (n + 3)/4)$, which does not converge in distribution to $N(0, 1)$. □

REFERENCES

AITKEN, A. C., "On Least Squares and Linear Combinations of Observations," *Proceedings of the Royal Society of Edinburgh* 55 (1935), 42–48.

AMEMIYA, T., "Generalized Least Squares with an Estimated Autocovariance Matrix," *Econometrica* 41 (1973a), 723–732.

———, "Regression Analysis When the Variance of the Dependent Variable Is Proportional to the Square of the Expectation," *Journal of the American Statistical Association* 68 (1973b), 928–934.

———, "A Note on a Heteroscedastic Model," *Journal of Econometrics* 6 (1977), 365–370.

———, *Advanced Econometrics* (Cambridge: Harvard University Press, 1985).

ANDERSON, T. W., *The Statistical Analysis of Time Series* (New York: Wiley, 1971).

——— AND C. HSIAO, "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics* 18 (1982), 47–82.

CAMPOS, J., "Instrumental Variables Estimation of Dynamic Simultaneous Systems with ARMA Errors," *Review of Economic Studies* 53 (1986), 125–138.

CARROLL, R. J. AND D. RUPPERT, "Robust Estimation in Heteroscedastic Linear Models," *The Annals of Statistics* 10 (1982), 429–441.

CRAGG, J. G., "Quasi-Aitken Estimation for Heteroscedasticity of Unknown Form," *Journal of Econometrics* 54 (1992), 179–201.

CROCKETT, P. W., "Asymptotic Distribution of the Hildreth-Houck Estimator," *Journal of the American Statistical Association* 80 (1985), 202–204.

DURBIN, J., "Estimation of Parameters in Time Series Regression Models," *Journal of the Royal Statistical Society B* 22 (1960), 139–153.

FOMBY, T. B., R. C. HILL, AND S. R. JOHNSON, *Advanced Econometric Methods*, corrected softcover printing (New York: Springer-Verlag, 1988).

FULLER, W. A., *Introduction to Statistical Time Series* (New York: Wiley, 1976).

——— AND G. E. BATTESE, "Transformations for Estimation of Linear Models with Nested-Error Structure," *Journal of the American Statistical Association* 68 (1973), 626–632.

GALBRAITH, R. F. AND J. I. GALBRAITH, "On the Inverses of Some Patterned Matrices Arising in the Theory of Stationary Time Series," *Journal of Applied Probability* 11 (1974), 63–71.

GLEJSER, H., "A New Test for Heteroscedasticity," *Journal of the American Statistical Association* 64 (1969), 316–323.

GOLDBERGER, A. S., *Econometric Theory* (New York: Wiley, 1964).

GOLDFELD, S. M. AND R. E. QUANDT, *Nonlinear Methods in Econometrics* (Amsterdam: North-Holland, 1972).

GREENE, W. H., *Econometric Analysis*, 2nd ed. (New York: MacMillan, 1993).

GUILKEY, D. K. AND P. SCHMIDT, "Estimation of Seemingly Unrelated Regressions with Vector Autoregressive Errors," *Journal of the American Statistical Association* 68 (1973), 642–647.

HANNAN, E. J., "The Asymptotic Distribution of Serial Covariances," *The Annals of Statistics* 4 (1976), 396–399.

HARVEY, A. C., "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica* 44 (1976), 461–465.

———, *The Econometric Analysis of Time Series*, 2nd ed. (New York: Philip Allan, 1990).

HILDRETH, C. AND J. P. HOUCK, "Some Estimates for a Linear Model with Random Coefficients," *Journal of the American Statistical Association* 63 (1968), 584–595.

JUDGE, G. G., W. E. GRIFFITHS, R. C. HILL, H. LÜTKEPOHL, AND T-C. LEE, *The Theory and Practice of Econometrics*, 2nd ed. (New York: Wiley, 1985).

MAGNUS, J. R., "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix," *Journal of Econometrics* 7 (1978), 281–312.

MANDY, D. M. AND C. MARTINS-FILHO, "Seemingly Unrelated Regressions under Additive Heteroscedasticity: Theory and Share Equation Applications," *Journal of Econometrics* 58 (1993), 315–346.

PARK, R. E., "Estimation with Heteroscedastic Error Terms," *Econometrica* 34 (1966), 888.

PARKS, R. W., "Efficient Estimation of a System of Regression Equations When Disturbances Are Both Serially and Contemporaneously Correlated," *Journal of the American Statistical Association* 62 (1967), 500–509.

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

RAJ, B., V. K. SRIVASTAVA, AND A. ULLAH, "Generalized Two Stage Least Squares Estimators for a Structural Equation for both Fixed and Random Coefficients," *International Economic Review* 21 (1980), 171–183.

SCHMIDT, P., *Econometrics* (New York: Dekker, 1976).

SCHÖENFELD, P., "A Useful Central Limit Theorem for *m*-Dependent Random Variables," *Metrika* 17 (1971), 116–128.

SCOTT, A. J. AND D. HOLT, "The Effects of Two Stage Sampling on Ordinary Least Squares Methods," *Journal of the American Statistical Association* 77 (1982), 848–854.

SHAMAN, P., "An Approximate Inverse for the Covariance Matrix of Moving Average and Autoregressive Processes," *The Annals of Statistics* 3 (1975), 532–538.

SIDDIQUI, M. M., "On the Inversion of the Sample Covariance Matrix in a Stationary Autoregressive Process," *Annals of Mathematical Statistics* 29 (1958), 585–588.

SINGH, B. AND A. ULLAH, "Estimation of Seemingly Unrelated Regressions with Random Coefficients," *Journal of the American Statistical Association* 69 (1974), 191–195.

SWAMY, P. A. V. B. AND J. S. MEHTA, "Estimation of Linear Models with Time and Cross-Sectionally Varying Components," *Journal of the American Statistical Association* 72 (1977), 890–898.

TAYLOR, W., "Small Sample Properties of a Class of Two Stage Aitken Estimators," *Econometrica* 45 (1977), 497–508.

THEIL, H., *Principles of Econometrics* (New York: Wiley, 1971).

UPPULURI, V. R. R. AND J. A. CARPENTER, "The Inverse of a Matrix Occurring in First Order Moving Average Models," *Sankhyā* A31 (1969), 79–82.

WHITE, H., *Asymptotic Theory for Econometricians* (Orlando: Academic Press, 1984).

ZELLNER, A., "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association* 57 (1962), 348–368.

——— AND D. S. HUANG, "Further Properties of Efficient Estimators for Seemingly Unrelated Regression Equations," *International Economic Review* 3 (1962), 300–313.

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.