COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 734

A NOTE ON THE UNBIASEDNESS OF FEASIBLE GLS, QUASI-MAXIMUM LIKELIHOOD, ROBUST, ADAPTIVE, AND SPECTRAL ESTIMATORS OF THE LINEAR MODEL

by Donald W. K. Andrews

December, 1984

A NOTE ON THE UNBIASEDNESS OF FEASIBLE GLS, QUASI-MAXIMUM LIKELIHOOD, ROBUST, ADAPTIVE, AND SPECTRAL ESTIMATORS OF THE LINEAR MODEL

By Donald W. K. Andrews[1]

HEADNOTE

This note presents a set of conditions on the defining functions of regression parameter estimators of the linear model. These conditions guarantee that the estimators are symmetrically distributed about the true parameter value, and hence, are median unbiased, provided the joint distribution of the errors is symmetric. The symmetry result holds even if the regression parameters are subject to linear restrictions. If the estimators possess one or more moments, then the symmetry result also implies mean unbiasedness. Similar conditions are provided that establish a property called origin (or shift) equivariance for the estimators. Common feasible GLS, quasi-ML, robust, adaptive, and spectral estimators are seen easily to satisfy the requisite conditions.

1. INTRODUCTION

This note presents a general result that establishes symmetry about the true parameter vector of the distributions of a wide class of estimators of regression function parameters in the linear model. Estimators covered by this result include feasible generalized least squares (GLS), quasi-maximum likelihood (ML), robust, adaptive, and spectral estimators. Of course, symmetry of the estimators implies median unbiasedness (that is, the probability of an overestimate equals that of an underestimate, see Lehmann [25]), and also mean unbiasedness if the estimators possess one or more moments. The result holds if the joint distribution of the errors is symmetric, the regressors and the errors are independent, and the defining function of the estimator in question satisfies a certain set of conditions, which is usually quite easy to verify.

The symmetry result is an important finite sample result both in itself, and because it is useful in simplifying the production and presentation of Monte Carlo results. For example, most Monte Carlo studies of feasible GLS estimators are less definitive and more complicated than necessary, since they do not exploit the theoretical unbiasedness of the estimators under investigation. Furthermore, the generality of the result obviates the duplication of proofs of symmetry for the myriad of estimators for which the result applies.

The conditions used to generate the symmetry result only need to be altered slightly to ensure that the estimators under consideration are also origin (or shift) equivariant (see the definition given in equation (5) below). The property of origin equivariance of an estimator may be desirable, because it implies that the estimator is appropriate for any choice of origin used to measure the variables (see Andrews [6]).

The model under consideration includes the standard linear regression model, the linear seemingly unrelated regressions (SUR) model, the multivariate linear regression model (in particular, the unrestricted reduced form of a simultaneous equations system), the random coefficients linear model, and the linear panel data model. The regression parameters may be subject to non-homogeneous linear restrictions, the regressors may be fixed or random, and the errors may be autocorrelated and/or heteroscedastic.

The symmetry and origin equivariance results apply to numerous estimators including: (1) least squares (LS) estimators, (2) feasible GLS estimators such as (i) weighted LS (e.g., see Amemiya [3]), (ii) Cochrane-Orcutt [11] and Prais-Winsten [33] procedures, (iii) Durbin's [13] estimator, (iv) Amemiya's [2] estimator, (v) Pierce's [32] estimator, (vi) Swamy's [37] estimator of the random coefficients model, and (vii) various estimators of the error components model (e.g., see Maddala [28]), (3) all ML and quasi-ML estimators (provided the specified (quasi-) likelihood of the errors is symmetric), (4) spectral and band spectral regression estimators (see Hannan [15], Duncan and Jones [12], and Engle [14]), (5) numerous robust estimators such as (i) Huber M-estimators (see Huber [19] and Yohai and Maronna [40]), (ii) bounded influence M-estimators (see Krasker and Welsch [23] and Maronna and Yohai [30]), (iii) L-estimators (see Bickel [8]), (iv) R-estimators (see Adichie [1] and Jurečková [20]), (v) minimum distance estimators (see Koul and DeWet [22]), and (vi) GEM estimators (see Andrews [5]), (6) adaptive estimators (e.g., see Bickel [10]), and (7) various one-step estimators that are equal to a Gauss-Newton step away from an initial estimator (e.g., see Bickel [8, 9]).

Since the estimators considered do not necessarily have closed form expressions, attention is paid to the possibilities of non-existence and non-uniqueness of the estimators.

The symmetry result presented here is already known by some--perhaps even by many--for certain estimators.[2] The large number of papers that introduce feasible GLS procedures or present Monte Carlo evidence for them, but do not discuss or exploit this result, indicates, however, that knowledge of the result and its generality is not sufficiently widespread. Furthermore, in the literature on robust, adaptive, and spectral estimation the result has not received widespread attention.

The proof of the symmetry result is not difficult. It relies on a simple result that, undoubtedly, has been known for a long time: An odd function of a random vector with a symmetric distribution also has a symmetric distribution. For example, Hodges and Lehmann [18] used this result in showing that their estimator of location has a symmetric distribution about the true location parameter. Kakwani [21] also used this result in showing that Zellner's SUR estimator has a symmetric distribution. More recently, Magnus [29] used this result to show symmetry of the distribution of the ML estimator for the linear model with normal errors and covariance matrix that depends on a finite number of parameters. Here we use the result in showing that a wide class of feasible GLS, quasi-ML, robust, adaptive, and spectral estimators have symmetric distributions.

2. SYMMETRY AND ORIGIN EQUIVARIANCE RESULTS

The linear model considered here is written as

(1) $$y = X\beta + u ,$$

where y and u are n-vectors of dependent random variables and errors, respectively, X is an $n \times k$ matrix of regressors, and β is a k-dimensional parameter vector.

The regressors X may be fixed or random, but are assumed to be independent of the errors u. In addition, X is assumed to have full column rank with some positive probability. The true value of β, denoted β_0, is unknown but is assumed to be known to lie in a parameter space $\mathcal{B}$ that is an affine subspace (i.e., a translated linear subspace) of R^k.[3] The joint distribution of the errors is assumed to be symmetric about an n-vector of zeroes.[4] That is, the distributions of u and $-u$ are equivalent.

The above assumptions are the only assumptions placed on the model, and they are sufficiently weak to incorporate all of the models listed in the Introduction. Note that the regressor matrix need not have full rank with probability one, nor do the errors need to satisfy any distributional, independence, or identical distribution assumptions.

The generic estimator under consideration, denoted $\hat{\beta}$, is taken to be the solution to either a maximization problem or a system of equations. Most of the estimators considered in the literature can be so defined (including those estimators that utilize zigzag iterative procedures, see the discussion below). Since it is possible that the solution is not unique or does not even exist, the estimator is defined in two steps. The first step defines the set of solutions, $\hat{B}$, to the maximization problem or system of equations. The second step determines a unique estimator from the set $\hat{B}$, which may contain zero, one, or many elements.

The least absolute deviations (LAD) estimator and M-estimators (see Maronna and Yohai [30]) exemplify the case where multiple solutions may exist for the minimization problem or system of equations. The problem of existence of a solution may arise when the parameter space for some nuisance parameter is not compact (since functions do not necessarily attain their infimum on non-compact sets). For example, a non-compact parameter space arises naturally if the errors are stationary, first-order autoregressive, since the set of all points that generate stationary errors (i.e., all points in (-1,1)) is open.

Let $\tilde{r}(y, Z, \beta, \theta_1, \hat{\beta}_1, \hat{\theta}_2)$ denote the optimand or system of equations whose solutions for $\beta \in \mathcal{B}$ yield the set $\hat{B}$. More specifically, $\hat{B}$ consists of those values $\overline{\beta}$ in $\mathcal{B}$ such that

$$(2) \qquad (\overline{\beta}, \overline{\theta}_1) \text{ solves } \max_{\beta\in\mathcal{B}, \theta_1\in\Theta_1} \tilde{r}(y, Z, \beta, \theta_1, \hat{\beta}_1, \hat{\theta}_2) \text{ , or}$$

$$(3) \qquad (\overline{\beta}, \overline{\theta}_1) \text{ solves } \tilde{r}(y, Z, \beta, \theta_1, \hat{\beta}_1, \hat{\theta}_2) = \underset{\sim}{0} \text{ ,}$$

where Θ_1 is the parameter space of θ_1 . The set $\hat{B}$ is the null set if the relevant problem, viz., (2) or (3), does not have a solution.

The defining function $\tilde{r}$ depends on a matrix of instrumental variables (IV's) Z that includes the regressor matrix X . The instruments, Z , may be random or non-random, but are assumed to be independent of the errors u . (The reason for introducing instruments is discussed below.) The estimator $\hat{\beta}_1$, which appears as an argument of $\tilde{r}$, is an initial estimator of β that may (or may not) be used in defining the optimand or system of equations. For example, estimators that are defined to be one Gauss-Newton step away from an initial estimator are of this form (see Bickel [8, 9]). The parameters θ_1 and θ_2 are (nuisance) parameters of the joint distribution of the errors that again may (or may not) be taken into account when estimating β . The parameter θ_2 is allowed to affect $\tilde{r}$ through an initial estimator $\hat{\theta}_2$. For example, θ_2 may consist of autoregressive or moving average coefficients of the errors, and $\hat{\theta}_2$ may be an estimator of these parameters (see Cochrane and Orcutt [11], Prais and Winsten [33], Durbin [13], and Andrews [5]). The parameter θ_1 , on the other hand, is estimated jointly with β as indicated in (2) and (3) (e.g., see Beach and McKinnon [7] and MaCurdy [27]). The parameters θ_1 and θ_2 may be infinite dimensional (as θ_2 is in Bickel [10]), and may have elements in common (as in Pierce [32]). Of course, the defining function of most estimators does not depend on all of the arguments $(y, Z, \beta, \theta_1, \hat{\beta}_1, \hat{\theta}_2)$ listed for $\tilde{r}$. But, different estimators depend on different arguments, so all of the arguments listed are needed in order to achieve general results.

We allow the estimation procedure to depend on IV's, even though the true model contains no endogenous variables, because the latter fact may be unknown, and IV procedures of one sort or another may be used as a safeguard. Given this possibility, it is useful to know the properties of the IV procedures when none of the regressors is endogenous. The results below apply to this situation. The IV estimators of interest include the standard IV estimator and the estimators of White [39] and Krasker and Welsch [24].

Zigzag iterative procedures for the estimation of β_0 involve alternating between estimating β and θ_2 , with each new estimate of β relying on the latest estimate of θ_2 , and vice versa. Each step of these procedures is usually based on solving a maximization problem or system of equations for β or θ_2. Hence, the last step in which β is estimated is of the desired form, viz., that of equation (2) or equation (3). The stopping rule used to determine the number of iterations performed can be incorporated into the definition of the estimator $\hat{\theta}_2$ (which equals the final iterated estimate of θ_2). For example, one could define $\hat{\theta}_2$ to be the first estimator from the infinite sequence of iterated estimators of θ_2 such that the difference between successive iterated estimates of θ_2 (or β, perhaps) is less than some prespecified constant. Using this approach, we see that most zigzag iterative estimators fit into the framework of the generic estimator described above.

We return to the description of the generic estimator, and make the following assumption about $\tilde{r}$:

A1) The defining function $\tilde{r}$ depends on y, β, and $\hat{\beta}_1$ only through $y - X\beta$, $y - X\hat{\beta}_1$, $\beta - \hat{\beta}_1$, or $\hat{\theta}_2$. That is, we can write

$$\tilde{r}(y, Z, \beta, \theta_1, \hat{\beta}_1, \hat{\theta}_2) = r(y - X\beta, y - X\hat{\beta}_1, \beta - \hat{\beta}_1, Z, \theta_1, \hat{\theta}_2) .$$

Further, in the case of the maximization problem (2), r is an even function of its first three arguments:

$$r(y - X\beta, y - X\hat{\beta}_1, \beta - \hat{\beta}_1, Z, \theta_1, \hat{\theta}_2) = r(-[y-X\beta], -[y-X\hat{\beta}_1], -[\beta-\hat{\beta}_1], Z, \theta_1, \hat{\theta}_2) .$$

In the case of a system of equation problem (3), r is either an odd or an even function of its first three arguments.

Assumption A1 is extremely easy to verify, and is almost universally met by those estimation procedures of type (2) or (3) that have been proposed in the literature. For example, all of the estimators referred to in the Introduction satisfy this assumption.

The second assumption we make is of import only if an initial estimator $\hat{\beta}_1$ of β is utilized by the generic estimator:

A2) The initial estimator $\hat{\beta}_1$ is such that $\hat{\beta}_1 - \beta_0$ is an odd function of the errors u. That is, viewed as a function of u, $\hat{\beta}_1$ satisfies

$$\hat{\beta}_1(u) - \beta_0 = -(\hat{\beta}_1(-u) - \beta_0) .$$

The verification of this assumption follows by the results of this note applied to the initial estimator $\hat{\beta}_1$ rather than to the generic estimator $\hat{\beta}$. If $\hat{\beta}_1$ depends on some other initial estimator, say $\hat{\beta}_2$, then we apply the present results to $\hat{\beta}_2$ and then to $\hat{\beta}_1$, etc. The number of such initial estimators that needs to be considered must be finite, otherwise $\hat{\beta}_1$ is not properly defined. The results of this note apply to, and are quite easy to verify for, most of the estimators that have been suggested in the literature as initial estimators. Hence, assumption A2 is relatively easy to verify, and is satisfied quite generally.

By assumption, the distributions of u and $-u$ are identical. Thus, any parameter of the distribution of u is identical to that of the distribution of $-u$. In consequence, most estimation procedures that estimate some parameter of the distribution of u are invariant to changes in the data from u to $-u$. We require this property to hold for the initial estimator $\hat{\theta}_2$ of θ_2 (if such an estimator is used in defining the generic estimator $\hat{\beta}$):

A3) The initial estimator $\hat{\theta}_2$ of θ_2 is an even function of u. That is, viewed as a function of u, $\hat{\theta}_2$ satisfies $\hat{\theta}_2(u) = \hat{\theta}_2(-u)$.

The verification of this assumption can be done in a number of ways. If $\hat{\theta}_2$ has a closed form expression, assumption A3 is usually straightforward to verify. If $\hat{\theta}_2$ is the solution to a maximization problem or system of equations based on a defining function $\tilde{\tilde{r}}$, one can use the results of the present note to verify A3, by identifying the initial estimator $\hat{\theta}_2$ as the estimator $\hat{\theta}_1$ derived from the solution to (2) or (3) with $\tilde{\tilde{r}}$ replacing $\tilde{r}$. In this case, if $\tilde{\tilde{r}}$ satisfies the assumptions listed here for $\tilde{r}$, then the result concerning $\hat{\theta}_1$ of the Theorem below establishes A3 for the initial estimator θ_2. If the initial estimator $\hat{\theta}_2$ is defined in some other manner, often it is still not difficult to establish A3 in some ad hoc fashion.

Next we consider the second step in defining the generic estimator $\hat{\beta}$. This step consists of defining a tie-breaking rule, call it s, that assigns to every set of solutions $\hat{B}$ a unique estimator $\hat{\beta}$. The tie-breaking rule is allowed to depend on an alternative estimator of β, say $\hat{\beta}_2$. Thus, $\hat{\beta} = s(\hat{B}, \hat{\beta}_2)$. For example, the rule might be to take $\hat{\beta}$ to be that element of $\hat{B}$ that is closest to $\hat{\beta}_2$ (with further rules specified to break remaining ties). More specifically, when carrying out M-estimation procedures (e.g., see Maronna and Yohai [30]) we might choose that solution to the defining system of equations that is closest to the least squares estimator or to the LAD estimator (if it is unique).

The alternative estimator $\hat{\beta}_2$ (which may equal $\hat{\beta}_1$) is assumed to be such that $\hat{\beta}_2 - \beta_0$ is an odd function of the errors. This assumption can be verified in the same manner as is A2. Whenever X is of full rank such an estimator $\hat{\beta}_2$ exists, since the LS estimator qualifies. When X is of less than full rank, no such estimator exists and $\hat{\beta}_2$ is set equal to η, where η is an abstract symbol that denotes that the estimator is not defined as an element of $\mathcal{B}$. Note that we define $\eta = -\eta$ and $\eta \pm a = \eta$, for all $a \in R^k$.

If the solution set $\hat{B}$ has a single element, then the tie-breaking rule sets $\hat{\beta}$ equal to that element. If $\hat{B}$ is empty, then $\hat{\beta}$ is defined to equal either η or $\hat{\beta}_2$. In other cases, the tie-breaking rule is required to satisfy the equivariance and oddness conditions stated below. Note that $\hat{B}$ is necessarily equal to η (i.e., essentially is undefined) when X has less than full rank. This follows because $\hat{B}$ is either empty or has multiple elements, and all potential alternative estimators $\hat{\beta}_2$ that satisfy the oddness condition must equal η .

The alternative estimator $\hat{\beta}_2$ and the tie-breaking rule s are assumed to satisfy:

A4a) The alternative estimator $\hat{\beta}_2$ is such that $\hat{\beta}_2 - \beta_0$ is an odd function of the errors u .

b) The tie-breaking function s takes values in $\mathcal{B} \cup \{\eta\}$, and is equivariant and odd. That is,

(Equivariance) $s(B+\tilde{\beta}, \beta+\tilde{\beta}) = s(B,\beta) + \tilde{\beta}$, $\forall B \subset \mathcal{B}$; $\forall \beta, \tilde{\beta} \in \mathcal{B} \cup \{\eta\}$,

(Oddness) $s(-B, -\beta) = - s(B,\beta)$, $\forall B \subset \mathcal{B}$; $\forall \beta \in \mathcal{B} \cup \{\eta\}$

As mentioned above, the verification of A4a parallels that of A2. The verification of A4b is straightforward.

It is interesting to note that tie-breaking rules are not discussed at any length in the literature. This may be due to the concentration on asymptotics in the literature, coupled with the common property of proposed estimators that no tie-breaking rule is needed for n sufficiently large with probability one. For finite samples, however, unique solutions to (2) and (3) often are not guaranteed. In consequence, tie-breaking rules are needed, and conditions such as those of A4 are needed to establish distributional symmetry and unbiasedness of the estimators under consideration. The necessity of such conditions indicates that tie-breaking rules probably warrant more extensive consideration than they have received thus far.

Before presenting the theorem that establishes the symmetry of $\hat{\beta}$, we discuss another property of the estimator $\hat{\beta}$ that is easily proved in the present framework. Consider the following transformation of the model (1):

$$(4) \qquad y = X\beta + u \mapsto y^* = X\beta^* + u \ ,$$

where $y^* = y + X\tilde{\beta}$ and $\beta^* = \beta + \tilde{\beta}$, for some $\tilde{\beta} \in \mathcal{B}$. This transformation occurs if the origin of the vector space that contains y is shifted by an amount $X\tilde{\beta}$. Since, in principle, the origin is arbitrary, we may be interested in estimators $\hat{\beta}$ that are appropriate for any choice of origin. Such estimators are called <u>origin</u> (or shift) <u>equivariant</u>, and satisfy

$$(5) \qquad \hat{\beta}(y + X\tilde{\beta},\ Z) = \hat{\beta}(y,Z) + \tilde{\beta}\ ,\quad \forall \tilde{\beta} \in \mathcal{B}\ ,$$

where $\hat{\beta}(y,Z)$ denotes the estimator $\hat{\beta}$ when applied to the data (y,Z). (See Andrews [6] for a more detailed motivation and discussion of this property. Also, see Lehmann [26].)

In order to establish the origin equivariance of the generic estimator $\hat{\beta}$ we need two assumptions in addition to those of A1 and A4, for the case where initial estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\theta}_2$ are used in the definition of $\hat{\beta}$:

B1) The initial estimator $\hat{\beta}_1$ and the tie-breaking alternative estimator $\hat{\beta}_2$ are origin equivariant.

B2) The initial estimator $\hat{\theta}_2$ is origin invariant (within $X\mathcal{B} - X\mathcal{B}$). That is,

(6) $$\hat{\theta}_2(y + X\tilde{\beta}, Z) = \hat{\theta}_2(y,Z) \ , \ \forall\tilde{\beta} \in \mathcal{B} \ ,$$

where $\hat{\theta}_2(y + X\tilde{\beta}, Z)$ and $\hat{\theta}_2(y,Z)$ denote the estimator $\hat{\theta}_2$ applied to the data $(y + X\tilde{\beta}, Z)$ and (y,Z) , respectively.

Just as assumption A2 can be established by using the symmetry results of this note applied to $\hat{\beta}_1$, assumption B1 can be established by applying the origin equivariance results below to $\hat{\beta}_1$ and $\hat{\beta}_2$. Similarly, assumption B2 can be established for initial estimators $\hat{\theta}_2$ that can be defined as solutions to maximization or systems of equations problems of the form (2) or (3). Proceed by identifying $\hat{\theta}_2$ with the solution $\hat{\theta}_1$ of (2) or (3) for some function $\approx{r}$, and then apply the origin invariance result for $\hat{\theta}_1$ given in the Theorem below.

The various results alluded to above are given now in the following theorem:

<u>THEOREM</u>: (a) *For the linear model* (1), *all estimators* $\hat{\beta}$ *that satisfy* A1, *and when applicable,* A2, A3, *and/or* A4, *have distributions that are symmetric about the true parameter* β_0 *(and are such that* $\hat{\beta} - \beta_0$ *is an odd function of the errors* u *). Further, if an estimator* $\hat{\theta}_1$ *is estimated simultaneously with* $\hat{\beta}$ *, then under the same assumptions,* $\hat{\theta}_1$ *is an even function of the errors* u .

(b) *All estimators* $\hat{\beta}$ *that satisfy* A1, *and when applicable,* B1, B2, *and/or* A4, *are origin equivariant. Further, if an estimator* $\hat{\theta}_1$ *is estimated simultaneously with* $\hat{\beta}$ *, then under the same assumptions,* $\hat{\theta}_1$ *is origin invariant. The results of this part of the Theorem hold whether or not the errors* u *have symmetric distribution.*

The proof of the Theorem is given in Section 3 below.

Note that the estimator $\hat{\theta}_1$ of the Theorem is precisely defined as follows: $\hat{\theta}_1 = v(\hat{\Theta})$, where $\hat{\Theta}$ is the set of all points $\overline{\theta}_1$ such that $(\overline{\beta}, \overline{\theta}_1)$ solves (2) or (3), and $v(\cdot)$ is a tie-breaking rule that chooses a unique element from $\hat{\Theta}$. If $\hat{\Theta}$ is empty, then $v(\hat{\Theta})$ is equal to η.

Numerous examples of estimators that are covered by the Theorem are given in the Introduction. Further examples can be found in the literature.

The estimators $\hat{\beta}$ considered in the Theorem may be undefined (i.e., equal to η) with positive probability. Hence, to even consider the properties of mean and median unbiasedness for the estimators $\hat{\beta}$, these properties must be defined appropriately. We adopt the following definitions: The estimator $\hat{\hat{\beta}}$ is said to be <u>median unbiased</u> if

(7) $$P(a'(\hat{\hat{\beta}} - \beta_0) \leq 0,\ \hat{\hat{\beta}} \neq \eta) = P(a'(\hat{\hat{\beta}} - \beta_0) \geq 0,\ \hat{\hat{\beta}} \neq \eta)\ ,\quad \forall a \in R^k\ .$$

With this definition, part a of the Theorem establishes median unbiasedness of all estimators $\hat{\beta}$ that satisfy its conditions.

The <u>expectation</u> of an estimator $\hat{\hat{\beta}}$, denoted $E\hat{\hat{\beta}}$, is defined to equal

(8) $$\int \hat{\hat{\beta}}(\omega) 1_{[\hat{\hat{\beta}}(\omega) \neq \eta]} dP(\omega) / P(\hat{\hat{\beta}}(\omega) \neq \eta)\ ,$$

provided the integral is finite, where ω represents a specific sample realization of the random variables (y,Z). An estimator $\hat{\hat{\beta}}$ is said to be <u>unbiased</u> if $E\hat{\hat{\beta}} = \beta_0$, for any true parameter β_0 in $\mathcal{B}$. Part a of the Theorem establishes mean unbiasedness of all estimators $\hat{\beta}$ that satisfy its conditions, provided their expectations are finite. Note that Srivastava and Raj [36] establish the existence of the expectation of Zellner's estimator for the SUR model, under weak conditions on the distribution of the errors. For the general class of estimators considered above, the existence of one or more moments is an open question. The work of Phillips [31], however, clearly is relevant to this question. Phillips derives an expression for the exact distribution of a class of feasible GLS estimators for the case of normally distributed errors.

3. PROOFS

PROOF OF THEOREM: We only consider the case where $\hat{\beta}$ is defined as the solution of the maximization problem (2). By altering the wording below from "maximizes the function r " to "sets the function r equal to a vector of zeroes," the same proof applies to estimators that solve the system of equations (3).

For notational convenience we append (u) or (-u) to $\hat{\beta}$, $\hat{\beta}_1$, $\hat{\theta}_2$, etc. to denote that these statistics are calculated using the data $(y,Z) \equiv (u + X\beta_0, Z)$ or $(-u + X\beta_0, Z)$, respectively. By definition, $\hat{B}(u)$ is the set of points β such that (β, θ_1) maximizes $r(u + X(\beta_0 - \beta), u + X(\beta_0 - \hat{\beta}_1(u)), \beta - \hat{\beta}_1(u), Z, \theta_1, \hat{\theta}_2(u))$ over $\mathcal{B} \times \Theta_1$.

Let $\hat{\Xi}(u)$ be the set of points ξ, and $\hat{\Theta}(u)$ be the set of points θ_1, such that (ξ, θ_1) maximizes

$$r(u - X\xi, u - X(\hat{\beta}_1(u) - \beta_0), \xi - (\hat{\beta}_1(u) - \beta_0), Z, \theta_1, \hat{\theta}_2(u)) \tag{9}$$

over $L \times \Theta_1$, where $L \equiv \mathcal{B} - \beta_0$. Note that $\hat{B}(u) - \beta_0 \equiv \hat{\Xi}(u)$. Now, by definition, $-\hat{\Xi}(u)$ and $\hat{\Theta}(u)$ are the sets of points ξ and θ_1, respectively, such that (ξ, θ_1) maximizes

$$r(u+X\xi, u-X(\hat{\beta}_1(u) - \beta_0), -\xi-(\hat{\beta}_1(u) - \beta_0), Z, \theta_1, \hat{\theta}_2(u)) \equiv D, \tag{10}$$

over $(\beta_0 - \mathcal{B}) \times \Theta_1$. Using A2 and A3, and then A1, we find

$$
\text{(11)} \quad \begin{aligned} D &= r(u+X\xi,\ u+X(\hat{\beta}_1(-u)-\beta_0),\ -\xi+(\hat{\beta}_1(-u)-\beta_0),\ Z,\ \theta_1,\ \hat{\theta}_2(-u)) \\ &= r(-u-X\xi,\ -u-X(\hat{\beta}_1(-u)-\beta_0),\ +\xi-(\hat{\beta}_1(-u)-\beta_0),\ Z,\ \theta_1,\ \hat{\theta}_2(-u)) \end{aligned}
$$

Hence, by the above definitions of $\hat{\Xi}(u)$ and $\hat{\Theta}(u)$, $\hat{\Xi}(-u)$ and $\hat{\Theta}(-u)$ are the sets of points ξ and θ_1 , respectively, such that (ξ, θ_1) maximizes D over $L \times \Theta_1$. Since $\mathcal{B}$ is an affine subspace, $\beta_0 - \mathcal{B}$ equals L , and so, $-\hat{\Xi}(u) = \hat{\Xi}(-u)$ and $\hat{\Theta}(u) = \hat{\Theta}(-u)$. This gives

$$
\text{(12)} \quad \hat{\theta}_1(u) = v(\hat{\Theta}(u)) = v(\hat{\Theta}(-u)) = \hat{\theta}_1(-u) \ , \quad \text{and}
$$

$$
\text{(13)} \quad \hat{B}(u) - \beta_0 \equiv \hat{\Xi}(u) = -\hat{\Xi}(-u) = -(\hat{B}(-u) - \beta_0) \ ,
$$

where $v(\cdot)$ is the tie-breaking rule that yields a unique estimator of $\hat{\theta}_1$ from the set $\hat{\Theta}$ of solutions to (2) or (3). Equation (12) shows that $\hat{\theta}_1$ is an even function of u . Equation (13) is used in showing the following:

$$
\begin{aligned} \hat{\beta}(-u) - \beta_0 &\equiv s(\hat{B}(-u), \hat{\beta}_2(-u)) - \beta_0 = s(\hat{B}(-u) - \beta_0, \hat{\beta}_2(-u) - \beta_0) \quad \text{by A4b,} \\ &= s(-[\hat{B}(u) - \beta_0], -[\hat{\beta}_2(u) - \beta_0]) \quad \text{by (13) and A4a,} \\ &= -s(\hat{B}(u) - \beta_0, \hat{\beta}_2(u) - \beta_0) \quad \text{by A4b,} \\ &= -(\hat{\beta}(u) - \beta_0) \ , \quad \text{by A4b again.} \end{aligned}
$$

Hence, $\hat{\beta} - \beta_0$ is an odd function of u , and $\hat{\beta}$ has a symmetric distribution about β_0 , conditional on Z . Since this holds for all realizations of Z , and Z and u are independent, $\hat{\beta}$ has unconditional distribution symmetric about β_0 .

Next, part b of the Theorem is established. For notational convenience we append (y,Z) or $(y+X\tilde{\beta}, Z)$ after $\hat{B}$, $\hat{\beta}$, $\hat{\beta}_1$, or $\hat{\theta}_2$ to denote that these statistics are calculated using the data (y,Z) or (y^*, Z) , respectively, where $y^* \equiv y+X\tilde{\beta}$. By definition and assumption A1, $\hat{B}(y+X\tilde{\beta}, Z)$ is the set of points β such that (β, θ_1) maximizes

$$r(y+X[\tilde{\beta}-\beta],\ y+X[\tilde{\beta}-\hat{\beta}_1(y+X\tilde{\beta},Z)],\ \beta-\hat{\beta}_1(y+X\tilde{\beta},Z),\ Z,\ \theta_1,\ \hat{\theta}_2(y+X\tilde{\beta},Z))$$
$$(14)\qquad = r(y-X[\beta-\tilde{\beta}],\ y-X\hat{\beta}_1(y,Z),\ \beta-\tilde{\beta}-\hat{\beta}_1(y,Z),\ Z,\ \theta_1,\ \hat{\theta}_2(y,Z)) \equiv D_2$$

over $\mathcal{B}\times\Theta_1$, where the equality uses assumptions B1 and B2.

Also by definition, the set of points $\beta-\tilde{\beta}$ such that $(\beta-\tilde{\beta}, \theta_1)$ maximizes D_2 with respect to $(\beta-\tilde{\beta}, \theta_1)$ and over $\mathcal{B}\times\Theta_1$ is $\hat{B}(y,Z)$. Hence, the set of points β such that (β, θ_1) maximizes D_2 with respect to (β, θ_1) and over $\mathcal{B}\times\Theta_1$ is $\hat{B}(y,Z)+\tilde{\beta}$. Using (14), this gives

$$(15)\qquad \hat{B}(y+X\tilde{\beta}, Z) = \hat{B}(y,Z) + \tilde{\beta}\ , \quad \forall\tilde{\beta}\in\mathcal{B}\ .$$

Now, we have

$$\hat{\beta}(y+X\tilde{\beta}, Z) \equiv s(\hat{B}(y+X\tilde{\beta}, Z),\ \hat{\beta}_2(y+X\tilde{\beta}, Z))$$
$$= s(\hat{B}(y,Z)+\tilde{\beta},\ \hat{\beta}_2(y,Z)+\tilde{\beta}) \quad \text{by (15) and B1,}$$
$$= s(\hat{B}(y,Z),\ \hat{\beta}_2(y,Z))+\tilde{\beta} \quad \text{by A4,}$$
$$(16)\qquad \equiv \hat{\beta}(y,Z)+\tilde{\beta}\ .$$

Since (16) holds for all $\tilde{\beta}\in\mathcal{B}$, $\hat{\beta}$ is origin equivariant.

A similar argument shows that $\hat{\theta}_1(y+X\tilde{\beta}) = \hat{\theta}_1(y)$, and so, $\hat{\theta}_1$ is origin invariant. *Q.E.D.*

Cowles Foundation, Yale University

REFERENCES

[1] ADICHIE, J. N.: "Estimates of Regression Parameters Based on Rank Tests," *Annals of Mathematical Statistics*, 38(1967), 894-904.

[2] AMEMIYA, T.: "Generalized Least Squares with an Estimated Autocovariance Matrix," *Econometrica*, 41(1973), 723-732.

[3] _____: "Regression Analysis When the Variance of the Dependent Variable Is Proportional to the Square of Its Expectation," *Journal of the American Statistical Association*, 68(1973), 928-934.

[4] ANDREWS, D. F.: "Robust Estimation for Multiple Linear Regression," Proceedings of the 39th Session of the International Statistical Institute,45, Book 1, 1973, 105-111.

[5] ANDREWS, D. W. K.: "Robust and Efficient Estimation of Nonlinear Regression Models with Dependent Errors," manuscript, Cowles Foundation, Yale University, 1983.

[6] _____: "Origin Equivariant Gauss-Markov Estimation," manuscript, Cowles Foundation, Yale University, 1984.

[7] BEACH, C. M. and J. G. MacKINNON: "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," *Econometrica*, 46(1978), 51-58.

[8] BICKEL, P. J.: "On Some Analogues to Linear Combinations of Order Statistics in the Linear Model," *Annals of Statistics*, 1(1973), 597-616.

[9] _____: "One-Step Huber Estimates in the Linear Model," *Journal of the American Statistical Association*, 70(1975), 428-434.

[10] _____: "On Adaptive Estimation," *Annals of Statistics*, 10(1982), 647-671.

[11] COCHRANE, D., and G. H. ORCUTT: "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," *American Statistical Association Journal*, 44(1949), 32-61.

[12] DUNCAN, D. B., and R. H. JONES: "Multiple Regression with Stationary Errors," *American Statistical Association Journal*, 61(1966), 917-928.

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

[14] ENGLE, R. F.: "Band Spectrum Regression," *International Economic Review*, 15(1974), 1-11.

[15] HANNAN, E. J.: "Regression for Time Series," in *Time Series Analysis*, ed. by M. Rosenblatt. New York: John Wiley & Sons, 1963.

[16] HARVEY, A. C., and I. D. McAVINCHEY: "The Small Sample Efficiency of Two-Step Estimators in Regression Models with Autoregressive Disturbances," Department of Economics Discussion Paper No. 78-10, University of British Columbia, 1978.

[17] HENDRY, D. F. and F. SRBA: "The Properties of Autoregressive Instrumental Variables Estimators in Dynamic Systems," *Econometrica*, 45(1977), 969-990.

[18] HODGES, J. L., Jr., and E. L. LEHMANN: "Estimates of Location Based on Rank Tests," *Annals of Mathematical Statistics*, 34(1963), 598-611.

[19] HUBER, P. J.: "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Annals of Statistics*, 1(1973), 799-821.

[20] JUREČKOVÁ, J.: "Nonparametric Estimates of Regression Coefficients," *Annals of Mathematical Statistics*, 42(1971), 1328-1338.

[21] KAKWANI, N. C.: "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equation Estimators," *American Statistical Association Journal*, 62(1967), 141-142.

[22] KOUL, H., and T. DeWET: "Minimum Distance Estimation in a Linear Regression Model," *Annals of Statistics*, 11(1983), 921-932.

[23] KRASKER, W. S., and R. E. WELSCH: "Efficient Bounded-Influence Regression Estimation," *Journal of the American Statistical Association*, 77(1982), 595-604.

[24] _____: "Resistant Estimation of Simultaneous-Equations Models Using Weighted Instrumental Variables, manuscript, Sloan School of Management, Massachusetts Institute of Technology, 1983.

[25] LEHMANN, E. L.: *Testing Statistical Hypotheses*. New York: John Wiley & Sons, 1959.

[26] _____: *Theory of Point Estimation*. New York: John Wiley & Sons, 1983.

[27] MaCURDY, T. E.: "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 18(1982), 83-114.

[28] MADDALA, G. S.: "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39(1971), 341-358.

[29] 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.

[30] MARONNA, R. A., and V. J. YOHAI: "Asumptotic Behaviour of General M-Estimates for Regression and Scale with Random Carriers," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 58(1981), 7-20.

[31] PHILLIPS, P. C. B.: "The Distribution of GLS," unpublished research note, Cowles Foundation, Yale University, 1984.

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

[33] PRAIS, S. J., and C. B. WINSTEN: "Trend Estimators and Serial Correlation," Cowles Commission Discussion Paper No. 383, Chicago, Illinois, 1954.

[34] ROTHENBERG, T. J.: "Approximate Normality of Generalized Least Squares Estimates," *Econometrica*, 52(1984), 811-825.

[35] SARGAN, J. D.: "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals," *Econometrica*, 29(1961), 414-426.

[36] SRIVASTAVA, V. K., and B. RAJ: "The Existence of the Mean of the Estimator in Seemingly Unrelated Regressions," *Communications in Statistics--Theory and Methods*, A8(7)(1979), 713-717.

[37] SWAMY, P. A. V. B.: "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, 38(1970), 311-323.

[38] TAYLOR, W. E.: "Small Sample Properties of a Class of Two Stage Aitken Estimators," *Econometrica*, 45(1977), 497-508.

[39] WHITE, H.: "Instrumental Variables Analogues of Generalized Least Squares Estimators," manuscript, University of California, San Diego, 1983.

[40] YOHAI, V. J., and R. A. MARONNA: "Asymptotic Behaviour of M-Estimators for the Linear Model," *Annals of Statistics*, 7(1979), 258-268.

FOOTNOTES

[1] I would like to thank J. L. Powell, P. C. B. Phillips, and M. D. Shapiro for their helpful comments on this paper.

[2] The symmetry result has been established by Adichie [1] and Koul and DeWet [22], for the specific robust estimators they propose. Kakwani [21] has shown that Zellner's SUR estimators is symmetrically distributed. Taylor [38] has shown that a particular two-stage Aitken estimator is symmetrically distributed in the special case of a heteroscedastic error model. For a more general model he has shown the weaker result that it is "unbiased to any order in probability." Maddala [28] has shown that his error components estimator is unbiased under the assumption of normality of the errors. Magnus [29] has shown symmetry of the distribution of the MLE for the linear model with normal errors and covariance matrix that depends on a finite number of parameters. Numerous others indicate their knowledge that the symmetry result applies to many more estimators than those just mentioned. See, for example, Kakwani [21], Hendry and Srba [17], Harvey and MacAvinchey [16], and Rothenberg [34]. Even these authors, however, make explicit references to unbiasedness only for estimators that possess closed form expressions given some covariance matrix estimator. Many of the estimators referred to in the Introduction cannot be written as such. For estimators defined implicitly, the problems of uniqueness and existence of the estimators must be addressed, as is done below. Furthermore, none of the above authors provide general sufficient conditions for the symmetry result to hold, as are given here. Note that the sufficient conditions given here not only encompass a wide range of estimation procedures, they also permit

linear restrictions on the regression parameter vector. Such restrictions have not been considered in any of the special cases treated in the literature.

[3]This assumption on the parameter space $\mathcal{B}$ is not made just for convenience. If $\mathcal{B}$ is not an affine subspace, e.g., if $\mathcal{B}$ is a compact set or an orthant, then the symmetry result below will not hold in general, see the proof of the Theorem below.

[4]If the distribution of u is symmetric about an n-vector of identical constants, c, not equal to 0, and the regression function contains a constant term, then a symmetry result still holds. The result of the Theorem below can be extended to show that the estimators considered are symmetric about $\beta_0 + (c, 0, \ldots, 0)'$. Hence, the estimators of the regression coefficients, excluding the constant term, have symmetric distributions about the true values.