

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 734R

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgement by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

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

Revised: August 1985

Forthcoming in Econometrica

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

By Donald W. K. Andrews¹

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 conditional distribution of the vector of errors is symmetric given the matrix of regressors. 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 the property of 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.

Revised: August 1985

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, spectral, and instrumental variable (IV) estimators. Of course, symmetry of the estimators implies median unbiasedness (that is, the probability of an overestimate equals that of an underestimate, see Lehmann [30]), and also mean unbiasedness if the estimators possess one or more moments. The result holds if the conditional distribution of the vector of errors is symmetric given the matrix of regressors, 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 also are origin (or shift) equivariant (see the definition given in equation (8) below). The property of origin equivariance of an estimator often is

desirable, because it implies that the estimator is appropriate for any choice of origin used to measure the variable (see Andrews [6]). As pointed out by Breusch [11], this property also is useful because it implies that the value of the estimand does not affect the distribution of the estimator, except by affecting the location of the estimator's distribution. Hence, the parameter dependence of the distribution of the estimator is of a very simple form, and Monte Carlo results performed for specific values of the estimand have much wider validity than often is recognized.

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 [12] and Prais-Winsten [39] procedures, (iii) Durbin's [15] estimator, (iv) Amemiya's [2] estimator, (v) Pierce's [38] estimator, (vi) Swamy's [44] estimator of the random coefficients model, and (vii) various estimators of the error components model (e.g., see Maddala [33]), (3) all ML and quasi-ML estimators (provided the specified conditional (quasi-) likelihood of the errors given the regressors is symmetric), (4) spectral and band spectral regression estimators (see Hannan [18], Duncan and Jones [14], and Engle [16]), (5) numerous robust estimators such as (i) Huber M-estimators (see Huber [23] and Yohai and Maronna [47]), (ii) bounded influence

M-estimators (see Krasker and Welsch [27] and Maronna and Yohai [36]), (iii) L-estimators (see Bickel [8]), (iv) R-estimators (see Adichie [1] and Jurečková [24]), (v) minimum distance estimators (see Koul and DeWet [26]), and (vi) GEM estimators (see Andrews [5]), (6) adaptive estimators (e.g., see Bickel [10]), (7) various one-step estimators that are equal to a Gauss-Newton step away from an initial estimator (e.g., see Bickel [8, 9]), and (8) IV estimators, including those of White [46] and Krasker and Welsch [28].

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 for certain estimators under certain conditions.² The generality of the result, however, has not been fully appreciated in the literature. For example, a large number of papers introduce feasible GLS procedures, or present Monte Carlo evidence for them, but do not discuss or exploit this result. Furthermore, in the literature on robust, adaptive, and spectral estimation the result has not received widespread attention.³ The purpose of this note is to illustrate the generality of the result, with respect to both the estimation procedure considered and the underlying errors assumptions, and to provide general conditions that eliminate the need for separate symmetry proofs for each estimator or modelling situation that arises.

The proof of symmetry is not difficult. It relies on a simple oddness result that has been known for a long time, undoubtedly: An odd function of a random vector with a symmetric distribution also has a symmetric distribution. For example, Hodges and Lehmann [22] used this result in showing that their estimator of location has a symmetric

distribution about the true location parameter. Kakwani [25] used this result in showing that Zellner's SUR estimator has a symmetric distribution. More recently, Breusch [11], Magnus [34], Harvey [19], and others (see footnote 2) have used the oddness result to show symmetry of various estimators.

The usual method of exploiting the oddness result (e.g., see Kakwani [25], Breusch [11], etc.) is to show that the estimator minus the estimand can be written in closed form as an odd function of the errors and some preliminary estimator that is an odd or even function of the errors. In this note, however, we consider estimators that are defined implicitly by a maximization problem or system of equations (which may depend on preliminary estimators). Hence, a different method must be used to find conditions under which the estimator minus the estimand is an odd function of the errors. The advantage of considering implicitly defined estimators is that it allows one to consider a very wide class of estimators, as indicated above.

2. SYMMETRY RESULTS

The linear model considered here is written as

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

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 may be fixed or random. The regressor matrix 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 \mathbb{R}^k .⁴ The distribution of the vector of errors u conditional on X is assumed to be symmetric about an n -vector of zeroes (for almost all X).⁵ That is, the conditional distributions of u and $-u$ given X are equivalent. This condition on the errors obviously holds if u and X are independent and u is symmetric. In addition, it allows for heteroscedasticity of the sort where the error variances are related to the values of the regressors. If instrumental variables (IV's) Z are used in estimation, then the symmetry assumption is assumed to hold conditional on Z (for almost all Z), where the matrix Z includes X .

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 assumptions regarding independence (over time), identical distributions, existence of moments, or normality of marginal distributions.

Before describing the general class of estimators for which the symmetry result holds, we consider a simple example that exemplifies the method of proof of the result for implicitly defined estimators. Consider an estimator $\hat{\beta}_s$ that is defined as the solution to the maximization of $r(y-X\beta)$ over $\beta \in \mathbb{R}^k$, where $r(\cdot)$ is a real-valued function that is an even function of its argument. For example, we might have $r(y-X\beta) = \rho((y-X\beta)'(y-X\beta))$, for some convex function $\rho(\cdot)$. Further, suppose $r(\cdot)$ is such that $\hat{\beta}_s$ is always defined and is unique. The claim is that $\hat{\beta}_s$ is symmetrically distributed about β_0 . Using the oddness result described in Section 1, it suffices to show that $\hat{\beta}_s - \beta_0$ is an

odd function of errors u .

Let $\hat{\beta}_s(u)$ and $\hat{\beta}_s(-u)$ denote the estimator $\hat{\beta}_s$ applied to the data $(y, X) = (u + X\beta_0, X)$ and $(\tilde{y}, X) \equiv (-u + X\beta_0, X)$, respectively. Let $\operatorname{argmax}_{\beta \in R^k} r(y - X\beta)$ denote the vector β in R^k that maximizes $r(y - X\beta)$.

The oddness of $\hat{\beta}_s(u) - \beta_0$ in u is established using the following change of variables manipulations:

$$\hat{\beta}(u) = \operatorname{argmax}_{\beta \in R^k} r(u + X[\beta_0 - \beta])$$

$$= - \operatorname{argmax}_{\tilde{\beta} \in \beta_0 - R^k} r(u + X\tilde{\beta}) + \beta_0,$$

by letting $\tilde{\beta} = \beta_0 - \beta$, where $\beta_0 - R^k \equiv \{\tilde{\beta} : \tilde{\beta} = \beta_0 - \beta, \text{ for some } \beta \in R^k\}$,

$$= - \operatorname{argmax}_{\tilde{\beta} \in R^k} r(-u + X[-\tilde{\beta}]) + \beta_0$$

(2) since $r(\cdot)$ is an even function and $\beta_0 - R^k = R^k$,

$$= - \operatorname{argmax}_{\beta \in R^k + \beta_0} r(-u + X[\beta_0 - \beta]) + 2\beta_0$$

by letting $\beta = \beta_0 + \tilde{\beta}$,

$$= -\hat{\beta}_s(-u) + 2\beta_0$$

since $R^k + \beta_0 = R^k$.

Thus, $\hat{\beta}_s(u) - \beta_0 = -(\hat{\beta}_s(-u) - \beta_0)$, as desired. This method of proof is the basis of the proof given below for the general class of estimators.

We now consider the definition of this class.

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; $\hat{\beta}$, to the maximization problem or system of equations. The second step determines a unique estimator $\hat{\beta}$ from the set $\hat{\beta}$, which may contain zero, one, or more elements.⁶

The least absolute deviations (LAD) estimator and M-estimators (see Maronna and Yohai [36]) exemplify the case where multiple solutions may exist for the maximization 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. 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 $\hat{\beta}$. More specifically, $\hat{\beta}$ consists of those values $\bar{\beta}$ in \mathcal{B} such that

$$(3) \quad (\bar{\beta}, \bar{\theta}_1) \text{ solves } \max_{\beta \in \mathcal{B}, \theta_1 \in \Theta_1} \tilde{r}(y, Z, \beta, \theta_1, \hat{\beta}_1, \hat{\theta}_2), \text{ or}$$

$$(4) \quad (\bar{\beta}, \bar{\theta}_1) \text{ solves } \tilde{r}(y, Z, \beta, \theta_1, \hat{\beta}_1, \hat{\theta}_2) = 0,$$

where Θ_1 is the parameter space of θ_1 . The set $\hat{\beta}$ is the null set if the relevant problem, viz., (3) or (4), 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 reason for

introducing instruments is discussed below.) The instrument Z may be random or non-random. No assumptions are placed on the distribution of the IV's (except the above assumption that u given Z is symmetric). The estimator $\hat{\beta}_1$, which appears as an argument or \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 [12], Prais and Winsten [39], Durbin [15], and Andrews [5]). The parameter θ_1 , on the other hand, is estimated jointly with β as indicated in (3) and (4) (e.g., see Beach and MacKinnon [7] and MaCurdy [32]). 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 [38]). 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 [46] and Krasker and Welsch [28].

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 (3) or equation (4). 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 $\hat{\beta}$. In order to establish the symmetry of $\hat{\beta}$ about β_0 , we need to make assumptions on $\tilde{r}(\cdot)$, $\hat{\beta}_1$, $\hat{\theta}_2$, and the tie-breaking rule that allow us to show that $\hat{\beta}$ is an odd function of the errors. We 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 (3), 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 (4), 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 (3) or (4) that have been proposed in the literature. For example, all of the estimators referred to in the Introduction satisfy this assumption. The class of Bayes estimators is the notable case where assumption A1 fails. This is not restrictive, however, since Bayes procedures that are based on proper prior distributions do not have distributions that are symmetric about the true parameter value.

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 conditional distributions of u and $-u$ are identical. Thus, any parameter of the conditional distribution of u is identical to that of the conditional distribution of $-u$. In consequence, most estimation procedures that estimate some parameter of the conditional 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 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 (3) or (4) with r^* replacing \tilde{r} . In this case, if r^* satisfies the assumptions listed here for \tilde{r} , then the result concerning $\hat{\theta}_1$ of Theorem 1 below establishes A3 for the

initial estimator $\hat{\theta}_2$. If the initial estimator $\hat{\theta}_2$ is defined in some other manner, often it still is 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 [36]) 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). Alternatively, $\hat{\beta}$ might be defined as that element of \hat{B} that is generated by a specific iterative computational algorithm that uses $\hat{\beta}_2$ as its starting value (e.g., see Harvey [20], also see Sielken and Hartley [42]).

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 $a \in \mathbb{R}^k$.

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 re-

quired 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 .⁷

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,

$$\text{(Equivariance)} \quad s(B+\xi, \beta+\xi) = s(B, \beta) + \xi, \quad \forall B \in \mathcal{B}; \forall \beta \in \mathcal{B} \cup \{\eta\}; \forall \xi \in \mathcal{L},$$

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

By definition, the set \mathcal{L} is the linear subspace that is parallel to \mathcal{B} . That is, $\mathcal{L} = \mathcal{B} - \mathcal{B} \equiv \{\xi \in \mathbb{R}^k : \xi = \beta_1 - \beta_2, \text{ for some } \beta_1, \beta_2 \in \mathcal{B}\}$. A discussion of the equivariance condition is given in Section 3 below. 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 the particular tie-breaking rule used does not effect the asymptotic properties of the estimators. The finite sample distributions of these estimators, however, depends on the form of the tie-breaking rule, and conditions such as those of A4 are needed to establish distributional symmetry and unbiasedness of the estimators. As Sielken and Hartley [42]

show, it is easy to construct estimators that are not symmetric about β_0 if no conditions are placed on the form of the tie-breaking rule. The necessity of such conditions indicates that tie-breaking rules probably warrant more extensive consideration than they have received thus far.

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

THEOREM 1: *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 .*

The proof of the Theorem is given in Section 4 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 $\bar{\theta}_1$ such that $(\bar{\beta}, \bar{\theta}_1)$ solves (3) or (4), 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 . No additional conditions are needed on this tie-breaking rule, because the random set $\hat{\Theta}$ is exactly the same for u and $-u$, and so, $v(\cdot)$ chooses the same value from $\hat{\Theta}$ whether the errors generating the data are given by u or $-u$.

It is interesting to note that no conditions are needed to restrict the way in which θ_1 and $\hat{\theta}_2$ enter the function $r(\cdot)$. This follows because the symmetry result relies on the evenness (or oddness) of $r(\cdot)$ as a function of u , and this property is not affected by the way in which the non-random argument θ_1 or the random argument $\hat{\theta}_2$ enter $r(\cdot)$.

provided $\hat{\theta}_2$ is an even function of u . Also, it is worth pointing out that $r(\cdot)$ can be an even or an odd function of its first three arguments in the systems of equations case (see A1), because the solutions to a system of equations are unchanged when any equation is multiplied by minus one.

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 $\hat{\beta}$ is said to be median unbiased if

$$(5) \quad P(a'(\hat{\beta} - \beta_0) \leq 0, \hat{\beta} \neq \eta) = P(a'(\hat{\beta} - \beta_0) \geq 0, \hat{\beta} \neq \eta), \quad \forall a \in R^k.$$

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

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

$$(6) \quad \int \hat{\beta}(\omega) 1_{[\hat{\beta}(\omega) \neq \eta]} dP(\omega) / P(\hat{\beta}(\omega) \neq \eta),$$

provided the integral is finite, where ω represents a specific sample realization of the random variables (y, Z) . An estimator $\hat{\beta}$ is said to be unbiased if $E\hat{\beta} = \beta_0$, 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 [43] establish the existence of the expectation of Zellner's estimator for the SUR model, under weak conditions on the distribution

of the errors. Also, Fuller and Battese [17] show that a particular feasible GLS estimator for the variance components model has finite expectation for certain error distributions. For the general class of estimators considered above, the existence of one or more moments is an open question. The work of Phillips [34], however, 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. ORIGIN EQUIVARIANCE RESULTS

In this section, we show that in the linear model most estimators of regression function parameters and error distribution parameters possess the properties of origin (or shift) equivariance and origin invariance, respectively. These properties are of interest for two reasons. First, origin equivariance can be motivated by the symmetries of the statistical model itself, and can replace the somewhat arbitrary assumption of unbiasedness in the formulation of the Gauss-Markov Theorem (see Lehmann [31] and Andrews [6]). Second, origin equivariance or origin invariance of estimators implies that the parameter dependence of the distributions of the estimators is of a very simple form, and in consequence, Monte Carlo results for such estimators are not parameter dependent.

The condition of origin equivariance is best motivated by considering the coordinate-free linear model, which is a generalization of the linear model considered above in the case of fixed regressors (see Kruskal [29], Malinvaud [35], Drygas [13], and Andrews [6]). In the coordinate-free model, the dependent variable vector y is viewed as an element of a vector space V . The regression function is specified not in terms of regressor variables and parameters, but in terms of an affine subspace (i.e., a translated linear

subspace), in which the expectation of the dependent variable vector must lie. That is, the model specifies $Ey \in L$, where L is an affine subspace of the vector space V that contains y . Often V is just R^n . In this framework, no coordinate system of V needs to be specified. The coordinate-free approach to the linear model has a number of advantages including those of generality, simplicity, geometrical interpretability, elegance, and computational flexibility (see the references above).

In Andrews [6], it is shown that the coordinate-free linear model also is origin invariant. That is, if we consider the linear subspace F that is parallel to the affine subspace L that contains Ey , then the same statistical model is generated no matter which point in F is chosen as the origin of the vector space V . When the origin of V is shifted between points in F , the estimand shifts in like manner. That is, the estimand is origin equivariant. Since the origin is arbitrary within F from a mathematical perspective, it seems reasonable to require an estimator to be appropriate for any origin in F . This is the case only if it shifts in like manner to the estimand and the origin, i.e., only if it also is origin equivariant.

In terms of model (1) above, an origin shift within F corresponds to the transformation

$$(7) \quad y = X\beta + u \mapsto y^* = X\beta^* + u,$$

where $y^* = y + X\xi$ and $\beta^* = \beta + \xi$, for some $\xi \in \mathcal{L}$, where \mathcal{L} is the linear subspace that is parallel to \mathcal{B} (i.e., $\mathcal{L} = \mathcal{B} - \mathcal{B}$). An estimator is origin equivariant then, if it satisfies

$$(8) \quad \hat{\beta}(y + X\xi, Z) = \hat{\beta}(y, Z) + \xi, \quad \forall \xi \in \mathcal{L},$$

where $\hat{\beta}(y, Z)$ denotes the estimator $\hat{\beta}$ when applied to the data (y, Z) .

As alluded to above, the GLS estimator is the best linear origin equivariant estimator, in the sense of minimum mean squared error uniformly over the parameter space \mathcal{B} and over the class of all error distributions with given finite, non-singular covariance matrix (see Lehmann [31] and Andrews [6]). In addition, if the errors have multivariate normal distribution with a given non-singular covariance matrix, then the GLS estimator is the best origin equivariant estimator in the sense of minimum risk, uniformly over the parameter space \mathcal{B} , for any convex loss function (see Lehmann [31]). Thus, origin equivariance is a property of some interest from a theoretical perspective.

In order to establish the origin equivariance of the generic estimator $\hat{\beta}$ introduced in Section 2, 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. That is,
- (9)
$$\hat{\theta}_2(y + X\xi, Z) = \hat{\theta}_2(y, Z), \quad \forall \xi \in \mathcal{L},$$

where $\hat{\theta}_2(y + X\xi, Z)$ and $\hat{\theta}_2(y, Z)$ denote the estimator $\hat{\theta}_2$ applied to the data $(y + X\xi, Z)$ 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 (3) or (4). Proceed by identifying $\hat{\theta}_2$ with the solution $\hat{\theta}_1$ of (3) or (4) for some function \tilde{r} , and then apply the origin invariance result for $\hat{\theta}_1$ given in the Theorem below.

THEOREM 2: 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 Theorem hold whether or not the errors u have symmetric distribution given Z .

Following Breusch [11], we point out that Theorem 2 and the definitions of origin equivariance and origin invariance, given in equations (8) and (9), respectively, imply that the distributions of $\hat{\beta} - \beta_0$ and $\hat{\theta}_1$ do not depend on β_0 . (They still depend, however, on the distribution of Z .) Since we are interested generally in the distributions of $\hat{\beta}$ and $\hat{\theta}_1$ for each true β_0 in \mathcal{B} , this result is extremely informative. It also is useful because it shows that Monte Carlo results for the estimators $\hat{\beta}$ and $\hat{\theta}_1$, generated for some fixed β_0 , are valid for any other β_0 in \mathcal{B} .

4. PROOFS

PROOF OF THEOREM 1: We only consider the case where $\hat{\beta}$ is defined as the solution of the maximization problem (3). 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 (4).

We condition on Z and show that the symmetry result holds for almost all Z . This implies that the symmetry result also holds unconditionally.

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. Let

$\text{argmax}_{(\beta, \theta_1) \in \mathcal{B} \times \theta_1} r(\cdot)$ denote the sets of solutions $(\hat{\beta}, \hat{\theta}_1)$ for β and θ_1 ,

respectively, for the maximization of $r(\cdot)$ over $\mathcal{B} \times \theta_1$. We have

$$\begin{aligned} (\hat{\beta}(u), \hat{\theta}_1(u)) &= \text{argmax}_{(\beta, \theta_1) \in \mathcal{B} \times \theta_1} r(u + X[\beta_0 - \beta], u + X[\beta_0 - \hat{\beta}_1(u)], \beta - \hat{\beta}_1(u), Z, \theta_1, \hat{\theta}_2(u)) \\ &= \text{argmax}_{(\xi, \theta_1) \in [\mathcal{B} - \beta_0] \times \theta_1} r(u - X\xi, u + X[\beta_0 - \hat{\beta}_1(u)], \xi - [\hat{\beta}_1(u) - \beta_0], Z, \theta_1, \hat{\theta}_2(u)) + (\beta_0, \underline{0}) \end{aligned}$$

by letting $\xi = \beta - \beta_0$, where $\mathcal{L} \equiv \mathcal{B} - \beta_0$ is a linear subspace,

$$= \text{argmax}_{(\xi, \theta_1) \in \mathcal{L} \times \theta_1} r(-u + X\xi, -u + X[\beta_0 - \hat{\beta}_1(-u)], -\xi - [\hat{\beta}_1(-u) - \beta_0], Z, \theta_1, \hat{\theta}_2(-u)) + (\beta_0, \underline{0})$$

(10) using the evenness of r in its first three arguments (A4), and assumptions A2 and A3,

$$= -\text{argmax}_{(\beta, \theta_1) \in [\beta_0 - \mathcal{L}] \times \theta_1} r(-u + X[\beta_0 - \beta], -u + X[\beta_0 - \hat{\beta}_1(-u)], \beta - \hat{\beta}_1(-u), Z, \theta_1, \hat{\theta}_2(-u)) + (2\beta_0, \underline{0})$$

by letting $\beta = \beta_0 - \xi$, where $\beta_0 - \mathcal{L} = \mathcal{B}$ because \mathcal{L} is a linear subspace,

$$= -(\hat{\beta}(-u), \hat{\theta}_1(-u)) + (2\beta_0, \underline{0}).$$

That is, $\hat{\beta}(u) - \beta_0 = -(\hat{\beta}(-u) - \beta_0)$, and $\hat{\theta}_1(u) = \hat{\theta}_1(-u)$. Hence,

$$(11) \quad \hat{\theta}_1(u) = v(\hat{\theta}_1(u)) = v(\hat{\theta}_1(-u)) = \hat{\theta}_1(-u), \text{ and}$$

$$\begin{aligned}
\hat{B}(-u) - \beta_0 &\equiv s(\hat{B}(-u), \hat{\beta}_2(-u)) - \beta_0 = s(\hat{B}(-u) - \beta_0, \hat{\beta}_2(-u) - \beta_0) \text{ by A4b,} \\
&= s(-[\hat{B}(u) - \beta_0], -[\hat{\beta}_2(u) - \beta_0]) \text{ by (10) and A4a,} \\
&= -s(\hat{B}(u) - \beta_0, \hat{\beta}_2(u) - \beta_0) \text{ by A4b,} \\
&= -(\hat{\beta}(u) - \beta_0) , \text{ by A4b again.}
\end{aligned}$$

Thus, $\hat{\beta} - \beta_0$ and $\hat{\theta}_1$ are odd and even functions of u , respectively, and $\hat{\beta}$ has a symmetric distribution about β_0 , conditional on Z . *Q.E.D.*

PROOF OF THEOREM 2: We append (y) or (y^*) to \hat{B} , $\hat{\beta}$, $\hat{\beta}_1$, and $\hat{\theta}_2$ to denote that these statistics are calculated using the data (y, Z) or $(y^*, Z) \equiv (y + X\xi, Z)$, for some $\xi \in \mathcal{L} \equiv \mathcal{B} - \mathcal{B}$. Using the argmax notation of the proof of Theorem 1, we have

$$\begin{aligned}
(\hat{B}(y^*), \hat{\theta}_1(y^*)) &= \operatorname{argmax}_{(\beta, \theta_1) \in \mathcal{B} \times \theta_1} r(y - X[\beta - \xi], y + X[\xi - \beta_1(y^*)], \beta - \hat{\beta}_1(y^*), Z, \theta_1, \hat{\theta}_2(y^*)) \\
&= \operatorname{argmax}_{(\beta, \theta_1) \in \mathcal{B} \times \theta_1} r(y - X[\beta - \xi], y - X\hat{\beta}_1(y), \beta - \xi - \hat{\beta}_1(y), Z, \theta_1, \hat{\theta}_2(y)) \\
&\quad \text{using assumptions B1 and B2,} \\
(12) \quad &= \operatorname{argmax}_{(\tilde{\beta}, \theta_1) \in \mathcal{B} \times \theta_1} r(y - X\tilde{\beta}, y - X\hat{\beta}_1(y), \tilde{\beta} - \hat{\beta}_1(y), Z, \theta_1, \hat{\theta}_2(y)) + (\xi, 0) \\
&\quad \text{by letting } \tilde{\beta} = \beta - \xi, \text{ and using } \mathcal{B} = \mathcal{B} - \xi, \\
&= (\hat{B}(y), \hat{\theta}_1(y)) + (\xi, 0) .
\end{aligned}$$

That is, $\hat{B}(y + X\xi) = \hat{B}(y) + \xi$, and $\hat{\theta}_1(y + X\xi) = \hat{\theta}_1(y)$. Hence, $\hat{\theta}_1 \equiv v(\hat{\theta}_1(y))$ is origin invariant. Also,

$$\begin{aligned}
 \hat{\beta}(y^*) &= s(\hat{B}(y^*), \hat{\beta}_2(y^*)) = s(\hat{B}(y) + \xi, \hat{\beta}_2(y) + \xi) && \text{by (12) and B1,} \\
 (13) \qquad &= s(\hat{B}(y), \hat{\beta}_2(y)) + \xi && \text{by A4,} \\
 &= \hat{\beta}(y) + \xi,
 \end{aligned}$$

and so, $\hat{\beta}$ is origin equivariant.

Q.E.D.

*Cowles Foundation, Department of Economics
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] BREUSCH, T. S.: "Useful Invariance Results for Generalized Regression Models," *Journal of Econometrics*, 13 (1980), 328-340.
- [12] 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.
- [13] DRYGAS, H.: *The Coordinate-Free Approach to Gauss-Markov Estimation*, Lecture Notes in Operations Research and Mathematical Systems, No. 40. New York: Springer-Verlag, 1970.
- [14] DUNCAN, D. B., and R. H. JONES: "Multiple Regression with Stationary Errors," *American Statistical Association Journal*, 61(1966), 917-928.
- [15] DURBIN, J.: "Estimation of Parameters in Time-Series Regression Models," *Journal of the Royal Statistical Society B*, 22(1960), 139-153.
- [16] ENGLE, R. F.: "Band Spectrum Regression," *International Economic Review*, 15(1974), 1-11.
- [17] FULLER, W. A. and G. E. BATTESE: "Transformations for Estimation of Linear Models with Nested-Error Structure," *Journal of the American Statistical Association*, 68 (1973), 626-632.
- [18] HANNAN, E. J.: "Regression for Time Series," in *Time Series Analysis*, ed. by M. Rosenblatt. New York: John Wiley & Sons, 1963.
- [19] HARVEY, A. C.: "On the Unbiasedness of Robust Regression Estimators," *Communications in Statistics-Theory and Methods*, A7-8 (1978), 779-783.
- [20] 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.

- [21] HENDRY, D. F. and F. SRBA: "The Properties of Autoregressive Instrumental Variables Estimators in Dynamic Systems," *Econometrica*, 45(1977), 969-990.
- [22] HODGES, J. L., Jr., and E. L. LEHMANN: "Estimates of Location Based on Rank Tests," *Annals of Mathematical Statistics*, 34(1963), 598-611.
- [23] HUBER, P. J.: "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Annals of Statistics*, 1(1973), 799-821.
- [24] JUREČKOVÁ, J.: "Nonparametric Estimates of Regression Coefficients," *Annals of Mathematical Statistics*, 42(1971), 1328-1338.
- [25] KAKWANI, N. C.: "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equation Estimators," *American Statistical Association Journal*, 62(1967), 141-142.
- [26] KOUL, H., and T. DeWET: "Minimum Distance Estimation in a Linear Regression Model," *Annals of Statistics*, 11(1983), 921-932.
- [27] KRASKER, W. S., and R. E. WELSCH: "Efficient Bounded-Influence Regression Estimation," *Journal of the American Statistical Association*, 77(1982), 595-604.
- [28] _____: "Resistant Estimation of Simultaneous-Equations Models Using Weighted Instrumental Variables," *Econometrica*, forthcoming, 1985.
- [29] KRUSKAL, W.: "The Coordinate-Free Approach to Gauss-Markov Estimation and its Application to Missing and Extra Observations," *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics, and Probability Theory*, I (1961), 435-461.
- [30] LEHMANN, E. L.: *Testing Statistical Hypotheses*. New York: John Wiley & Sons, 1959.
- [31] _____: *Theory of Point Estimation*. New York: John Wiley & Sons, 1983.

- [32] 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.
- [33] MADDALA, G. S.: "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39(1971), 341-358.
- [34] 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.
- [35] MALINVAUD, E.: *Statistical Methods of Econometrics*, 1st Edition. New York: North-Holland, 1966.
- [36] 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.
- [37] PHILLIPS, P. C. B.: "The Distribution of GLS," unpublished research note, Cowles Foundation, Yale University, 1984.
- [38] PIERCE, D. A.: "Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors," *Biometrika*, 58(1971), 299-312.
- [39] PRAIS, S. J., and C. B. WINSTEN: "Trend Estimators and Serial Correlation," Cowles Commission Discussion Paper No. 383, Chicago, Illinois, 1954.
- [40] ROTHENBERG, T. J.: "Approximate Normality of Generalized Least Squares Estimates," *Econometrica*, 52(1984), 811-825.
- [41] SARGAN, J. D.: "The Maximum Likelihood Estimation of Economic Relationships with Autoregressive Residuals," *Econometrica*, 29(1961), 414-426.

- [42] SIELKEN, R. L. and H. O. HARTLEY: "Two Linear Programming Algorithms for Unbiased Estimation of Linear Models," *Journal of the American Statistical Association*, 68(1973), 639-641.
- [43] 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.
- [44] SWAMY, P. A. V. B.: "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, 38(1970), 311-323.
- [45] TAYLOR, W. E.: "Small Sample Properties of a Class of Two Stage Aitken Estimators," *Econometrica*, 45(1977), 497-508.
- [46] WHITE, H.: "Instrumental Variables Analogues of Generalized Least Squares Estimators," forthcoming in *Advances in Statistical Analysis and Statistical Computing*. Volume 1, ed. by R. S. Mariano. Greenwich, CT: JAI Press, 1985.
- [47] YOHAI, V. J., and R. A. MARONNA: "Asymptotic Behaviour of M-Estimators for the Linear Model," *Annals of Statistics*, 7(1979), 258-268.

FOOTNOTES

¹I would like to thank J. L. Powell, P. C. B. Phillips, M. D. Shapiro, T. M. Stoker, and two referees for their helpful comments on this paper. Research support by the National Science Foundation, through grant number SES-8419789, is gratefully acknowledged.

²The symmetry result has been established by Adichie [1], Harvey [19] and Koul and DeWet [26], for certain robust estimators in models with independent, identically distributed (iid) errors. Kakwani [25] has shown that Zellner's SUR estimator is symmetrically distributed in models with iid error vectors. Taylor [45], has shown that a particular two-stage Aitken estimator is symmetrically distributed in a heteroskedastic error model. Maddala [33] has shown that his error components estimator is unbiased under the assumption of normality of the errors. Fuller and Battese [17] have proved unbiasedness of a class of estimators for the error components model. Breusch [11] and Magnus [34] have 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 [25], Fuller and Battese [17], Hendry and Srba [21], Breusch [11], Harvey and MacAvinchey [20], and Rothenberg [40]. Even these authors, however, make explicit references to unbiasedness only for estimators that possess closed form expressions given some covariance matrix estimator or some preliminary estimator of the regression function parameters. 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. The sufficient conditions given here not only encompass a wide range of estimation procedures and error characteristics (including autocorrelation and heteroskedasticity, they also permit nonhomogeneous linear restrictions on the regression parameter vector. In the literature, only Breusch [11] has considered linear restrictions.

³The paper by Harvey [19] shows symmetry for a particular class of robust estimators, viz., M-estimators, unfortunately however, this paper has not received widespread recognition in the literature.

⁴This 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.

⁵If the distribution of u given X 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 $\beta_0 + (c, 0, \dots, 0)'$. Hence, the estimators of the regression coefficients, excluding the constant term, have symmetric distributions about the true values.

⁶If one chooses to report a set-valued estimator, this second step is superfluous. A symmetry result, analogous to that established below for uniquely defined estimators, can be established for set-valued estimators. For set-valued estimators \hat{B} , we say that \hat{B} is symmetric about B_0 if the distribution of $\hat{B} - B_0$ is the same as that of $-(\hat{B} - B_0)$.

⁷If set-valued estimators \hat{B} are considered, then estimators need not be undefined in the case of less than full rank regressor matrices, and the symmetry property of footnote 6 still can be established.