Discussion paper 411

ISSN 0316-5078

INSTITUTE FOR ECONOMIC RESEARCH

QUEEN'S UNIVERSITY

Kingston, Ontario, Canada K7L 3N6

TWO PAPERS ON LINEAR MODELS

Gordon Fisher, Allan W. Gregory, Michael McAleer
Department of Economics
Queen's University

Discussion Paper No. 411

December, 1980

AN INVARIANCE PROPERTY OF GENERALIZED CLASSICAL LINEAR ESTIMATORS

Gordon Fisher and Allan W. Gregory
Department of Economics
Queen's University

Discussion Paper No. 411

February, 1981

ABSTRACT

When coefficients of endogenous variables are known, it is demonstrated that two-stage least squares and instrumental variable estimators are invariant to the form in which these variables enter computations, as raw data or estimates. Exclusion of instruments and knowledge of coefficients are related to identifiability testing, and a test presented.

AN INVARIANCE PROPERTY OF GENERALIZED CLASSICAL LINEAR ESTIMATORS

- Gordon Fisher and Allan W. Gregory -

Consider the following illustrative specification of the demand-for-money function:

$$(1) \qquad m-p = \alpha + \beta r + \gamma y + \Sigma\lambda_i z_i + \varepsilon$$

in which the variables are cast in natural logarithms, α,β,γ and the λ_i are unknown coefficients, and ε is a zero mean spherical random error; the z_i are all predetermined while m,p,r and y may be endogenous: m is the nominal money stock, p is 'the' price level, r is 'the' nominal rate of interest, and y is aggregate real income. If we are interested in generalizing this equation, to test the homogeneity postulate for example, it might be written with dependent variable $(m - \theta p)$ or even more generally as:

$$(2) \qquad \phi m - \theta p - \beta r - \gamma y = \alpha + \Sigma\lambda_i z_i + \varepsilon \quad ,$$

in which normalization has to be decided. In the usual stock-adjustment formulation, for example, (m-p) is treated as the natural logarithm of the real money stock, r and y are treated as exogenous and there is only one z_i, namely the lagged money stock (see, for example, Goldfeld, 1973). In this formulation, estimation amounts to a standard application of regression theory to equation (1).

If, on the contrary, m,r,p and y are regarded as endogenous then, subject to knowledge of the rest of the system in which (2) is now presumed embedded, we must first check identification and agree upon normalization before we

proceed to estimation. If the whole system is estimated by full-information maximum-likelihood (FIML), or the equation is estimated by limited-information maximum-likelihood (LIML) or least variance ratio (LVR), then the estimates are invariant to normalization. This is a convenient property since it does not then matter whether we regard (2) as the money demand equation, for example, or as the price determining equation implicit in (2) after normalizing on the coefficient of p (see Walters, 1965 and 1967). Thus a difference in the causality presumed to be at work among the same set of variables does not affect empirical estimation of corresponding coefficients; and if such differences arise as a matter of economic theory, then estimation may play its rightful role as a means to scientific resolution, since it is not prejudiced toward any competing theory.

Notwithstanding these fundamentals, it is still very much part of current practice to apply generalized classical linear (GCL) methods (Basmann, 1957), like two-stage least squares (2SLS) or instrumental variables (IV), which are *not* invariant to normalization of the kind just described. Moreover, since such methods have intuitive appeal (while FIML, LIML and LVR are more remote in this respect), it is often difficult for a student to appreciate why invariance to normalization is a valuable property, or why normalization is an issue worth discussing in simultaneous equation models (see *e.g.* Aigner and Sawa, 1974). After all, normalization is not an issue in classical regression analysis and, presumably in consequence, lack of invariance to normalization is not usually presented as a special disadvantage of 2SLS. However, the intuitive appeal of a method like 2SLS, which comes directly from the algorithm that is used for its computation, can itself be misleading. For example, when

some of the endogenous regressors in a structural equation have known coefficients, the algorithm gives no clear-cut guide as to whether these variables should enter as they stand or as their reduced form estimates. As an illustration, suppose that the variables m,p,r and y in equation (2) are endogenous, that we agree to normalize on ϕ (=1), and that θ is known: how then should 2SLS be applied? The regression formulation would suggest constructing the dependent variable $(m-\theta p)$ and applying 2SLS in the usual way. However, this is not necessarily the obvious way to proceed because, when θ is unknown, p would enter as its corresponding unrestricted reduced form estimate $\hat{p}$. Should, then, the dependent variable be $(m-\theta\hat{p})$ or $(\hat{m}-\theta\hat{p})$ when θ is known? Moreover, if the dependent variable is changed in these ways, how might the 2SLS estimates of the remaining coefficients be affected?

In this note it is demonstrated that when some of the coefficients of the endogenous variables in a structural equation are known, then the 2SLS estimates of the remaining coefficients are invariant to the form in which the endogenous variables are entered, as raw data or as unrestricted reduced form estimates. The notation of a general model is first introduced and then the invariance property of 2SLS is developed by way of comparison with the invariance-to-normalization property of LIML and LVR, using the classical identifiability test statistic (Anderson and Rubin, 1949, 1950).

Let an equation, say the first, of an over-identifiable linear system, comprising M endogenous and K predetermined variables in n co-ordinate observations, be written as:

$$(3) \qquad y_1 = Y_1\beta_1 + X_1\gamma_1 + \varepsilon_1 \quad .$$

Y_1 comprises m, and X_1 comprises k, linearly independent columns and $[Y_1 \vdots X_1]$ has rank $(m+k) < K < n$; also, $\varepsilon_1 \sim N(0, I_n \sigma_{11})$ independently of all predetermined variables $X = [X_1 \vdots X_2]$ in the system, X having rank K. The coefficient vectors β_1 and γ_1 have m and k elements respectively. Corresponding to a normalized version of (2), equation (3) may be re-written

$$(4) \qquad y_1^* = Y_1^* \beta_1^* = y_1 - Y_1 \beta_1 = X_1 \gamma_1 + \varepsilon_1$$

in which $Y_1^* = [y_1 \vdots Y_1]$ and $\beta_1^{*T} = [1 \vdots -\beta_1^T]$. For known β_1^*, equation (4) may be regarded as a classical regression.

If y_1^* is expressed in terms of the reduced form corresponding to it, then

$$(5) \qquad y_1^* = X_1 \gamma_1 + X_2 \gamma_2 + \varepsilon_1 \quad ,$$

subject to the over-identifying hypothesis H: $\gamma_2 = 0$. This hypothesis may in principle be tested using the classical F-ratio:

$$(6) \qquad F = \frac{y_1^{*T}(P_\Omega - P_\omega) y_1^*}{y_1^{*T}(I_n - P_\Omega) y_1^*} \cdot \frac{n-k}{K-k} = \frac{\beta_1^{*T} Y_1^{*T}(P_\Omega - P_\omega) Y_1^* \beta_1^*}{\beta_1^{*T} Y_1^{*T}(I_n - P_\Omega) Y_1^* \beta_1^*} \cdot \frac{n-k}{K-k}$$

in which P_Ω denotes orthogonal projection on Ω, defined as the span of X, and P_ω has a corresponding meaning for ω, the span of X_1. Let $\nu(\beta_1^*)$ denote the ratio of quadratic forms in β_1^* on the right-hand side of (6). Note that $\omega \subset \Omega$, the dimension of Ω is K and the dimension of ω is k. Equivalent expressions for (6) under H are:

$$(7) \qquad F = \nu(\beta_1^*) \frac{n-K}{K-k} = \frac{\varepsilon_1^T (P_\Omega - P_\omega) \varepsilon_1}{\varepsilon_1^T (I_n - P_\Omega) \varepsilon_1} \cdot \frac{n-K}{K-k}$$

which, given the spherical normality of ε_1, is sufficient to demonstrate $F \sim F(K-k,n-K)$ exactly, when β_1^* is known (Anderson and Rubin, 1949, 1950).

With this notation and background we may state the following:

(i) LIML and LVR estimates $\hat{\beta}_1^*$ of β_1^* are obtained by minimizing $\nu(\beta_1^*)$, and hence F, with respect to the unknown elements of β_1^* (Basmann, 1960). When this is done, H has the best 'chance' of being declared 'acceptable' among all admissible estimates of β_1^*.

(ii) Re-normalization of (3) or (4) merely involves scalar multiplication of its coefficients (and the equation error). Such scalar multiplication cannot affect the minimization of $\nu(\beta_1^*)$ which establishes the invariance-to-normalization of LIML and LVR estimates of β_1^*.

(iii) 2SLS estimates b_1^* of β_1^* are obtained by minimization of the numerator of $\nu(\beta_1^*)$ with respect to the unknown elements of β_1^* (Basmann, 1960). Since the numerator of $\nu(\beta_1^*)$ is *not* invariant to scalar multiplication, b_1^* is *not* invariant to normalization.

(iv) Given the estimates $\hat{\beta}_1^*$ and b_1^*, corresponding estimates for γ_1 are obtained as least squares estimates from (4), replacing β_1^* by an estimate. For LIML see Koopmans and Hood (1953, pp. 167-9). For 2SLS, $b_1^* = [1 \vdots -b_1^T]$ where $b_1 = \{Y_1^T(P_\Omega - P_\omega)Y_1\}^{-1}Y_1^T(P_\Omega - P_\omega)y_1$. If the 2SLS estimate of γ_1 is c_1, then $c_1 = (X_1^TX_1)^{-1}X_1^T(y_1 - Y_1b_1)$. One way of recognizing the correctness of this formula for c_1 is by noting that if $y_1 - Y_1b_1 - X_1c_1 = e_1$, then $P_\omega e_1 = 0$ (Fisher, 1981); this implies $P_\omega(y_1 - Y_1b_1) = X_1c_1$ and hence the formula for c_1.

(v) Since $P_\omega e_1 = 0$, $\nu(b_1^*) = e_1^TP_\Omega e_1 / e_1^T(I_n - P_\Omega)e_1$ (Fisher, 1981). Basmann's (1960) modification of F in (7) is then $\{\nu(b_1^*)(n-K)\}/(K-k-m)$ which is approximately $F(K-k-m, n-K)$ under H, the degrees of freedom in the numerator recognizing the extra m restrictions introduced in estimating β_1^*.

The invariance property of 2SLS may now be developed and distinguished from the invariance-to-normalization property of LIML and LVR. Notice that the formulae in (*iv*) above for b_1 and c_1 may be re-written:

$$(8) \qquad \begin{bmatrix} b_1 \\ c_1 \end{bmatrix} = \begin{bmatrix} \{Y_1^T(P_\Omega - P_\omega)Y_1\}^{-1} Y_1^T(P_\Omega - P_\omega)P_\Omega y_1 \\ (X_1^T X_1)^{-1} X_1^T P_\Omega (y_1 - Y_1 b_1), \end{bmatrix} ,$$

since $(P_\Omega - P_\omega)P_\Omega = P_\Omega - P_\omega$ and $(P_\Omega X_1)^T = X_1^T$ by virtue of $\omega \subset \Omega$ and the properties of orthogonal projections. More precisely, $(P_\Omega - P_\omega)$ is the orthogonal projection on $\omega^\perp \cap \Omega$, *i.e.* $P_\Omega(I_n - P_\omega)$, and clearly $\omega^\perp \cap \Omega \subset \Omega$. The fact that P_Ω may be introduced to pre-multiply y_1 and $(y_1 - Y_1 b_1)$ in (8), without effect on the calculation of b_1 and c_1, immediately yields the following result: if any of the elements of β_1 is known, and hence need not be estimated, then the 2SLS estimates of the unknown coefficients are invariant to the choice of dependent variable between (a) the known linear combination of the endogenous observations; (b) the same linear combination of their corresponding least squares reduced form predictions; or (c) any combination of (a) and (b) consistent with (3). Specifically, if $Z_1 = [Y_1 \vdots X_1]$ and $\alpha_1^T = [\beta_1^T \vdots \gamma_1^T]$, then the following formulae are equivalent expressions for the 2SLS estimates of α_1:

$$(\hat{Z}_1^T \hat{Z}_1)^{-1} \hat{Z}_1^T y_1 = (Z_1^T P_\Omega Z_1)(P_\Omega Z_1)^T y_1 = (Z_1^T P_\Omega Z_1)^{-1} (P_\Omega Z_1)^T P_\Omega y_1 \quad ,$$

$\hat{Z}_1$ being Z_1 with Y_1 replaced by its reduced form estimate. Moreover, if Y_1 and β_1 are conformally partitioned as

$$Y_1\beta_1 = [Y_{11} \vdots Y_{12} \vdots Y_{13}] \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{31} \end{bmatrix}$$

comprising blocks of m_1, m_2 and m_3 variables ($m_1 + m_2 + m_3 = m$), and β_{11} and β_{21} are known while β_{31} is not, then

$$b_{31} = \{Y_{13}^T(P_\Omega - P_\omega)Y_{13}\}^{-1} Y_{13}^T(P_\Omega - P_\omega)(y_1 - Y_{11}\beta_{11} - Y_{12}\beta_{21}) .$$

Since, as already noted, $(P_\Omega - P_\omega)P_\Omega = P_\Omega - P_\omega$, it is obvious that it makes no difference to 2SLS estimation of β_{31} whether y_1, Y_{11} and Y_{12} or their corresponding least squares reduced from predictions are used; or indeed any proper combination of the two, like y_1, $\hat{Y}_{11}$, Y_{12}. Thus, in the example of equations (1) and (2), all of the possibilities suggested lead to the same 2SLS estimates of the unknown coefficients.

Moreover, the same property holds for IV estimation. Suppose a subset of $K^* < K$ of the exogenous variables, including X_1, replaces X as a set of instrumental variables. If the span of these variables is denoted by Λ, then $\omega \subset \Lambda \subset \Omega$, and consequently $(P_\Lambda - P_\omega)P_\Lambda = (P_\Lambda - P_\omega)$. Thus the same invariance property holds for IV estimation as holds for 2SLS estimation, so long as $K^* - k \geq m^*$ where m^* is the number of unknown coefficients of endogenous variables. This result makes clear the fact that, if the equation to be estimated is underidentifiable, then knowledge of the values of enough coefficients will justify estimation taking place. In effect, the known coefficients are a form of *a priori* information and these play the role of non-exclusion restrictions for purposes of estimation. The condition $(K^* - k) \geq m^*$ simply ensures that there are sufficient 'extra' exogenous variables available to estimate the unknown coefficients of endogenous variables.

Exclusion of exogenous variables from the instrumental set when in truth the excluded variables are really part of the model, represents a loss of infor-

mation. In these circumstances, it is important to check that the exclusions are statistically justifiable, *as an approximation*. In this regard, the identifiability test-statistic noted in (v) above may be applied, using the residuals e_1^* from the IV regression and the orthogonal projection P_Λ:

$$F = \frac{e_1^{*T} P_\Lambda e_1^*}{e_1^{*T}(I_n - P_\Lambda)e_1^*} \cdot \frac{n-K^*}{K^*-k-m^*} \quad .$$

This statistic may be used to test whether, given knowledge of some coefficients in an equation, the total number of exclusions may be regarded as satisfactory. Such exclusions relate only to the exogenous variables, of course, and may arise either from the theoretical structure of the model or as a matter of approximmation in defining a convenient instrumental set.

ACKNOWLEDGEMENTS

The authors are grateful to Michael McAleer for advice, criticism and general interest. The financial support of the Social Sciences and Humanities Research Council of Canada for the second author is gratefully acknowledged.

REFERENCES

Aigner, D.J. and T. Sawa (1974), "Identification and Normalization: A Note", *Journal of Econometrics*, 2, pp. 389-391.

Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, 20, pp. 46-63.

Anderson, T.W. and H. Rubin (1950), "The Asymptotic Properties of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, 21, pp. 570-582.

Basmann, R.L. (1957), "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation", *Econometrica*, 25, pp. 77-83.

Basmann, R.L. (1960), "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics", *Journal of the American Statistical Association*, 55, pp. 650-659.

Fisher, G.R. (1981), "Two Types of Residuals and the Classical Identifiability Test Statistic", in *Selected Econometric Papers in Memory of Stefan Valavanis*, Amsterdam: North-Holland, forthcoming.

Goldfeld, S. (1973), "The Demand for Money Revisited", *Brookings Papers on Economic Activity*, 3, pp. 575-646.

Koopmans, T.J. and W.C. Hood (1953), "The Estimation of Simultaneous Linear Economic Relationships", Chapter 6 of W.C. Hood and T.J. Koopmans (Eds.), *Studies in Econometric Method*, New York: Wiley, pp. 112-199.

Walters, A.A. (1965), "Professor Friedman on the Demand for Money", *Journal of Political Economy*, 73, pp. 545-551.

Walters, A.A. (1967), "The Demand for Money - The Dynamic Properties of the Multiplier", *Journal of Political Economy*, 75, pp. 293-298.

THE GEOMETRY OF SPECIFICATION ERROR

Gordon Fisher and Michael McAleer
Department of Economics
Queen's University

Discussion Paper No. 411

December, 1980

The financial support of the Canadian Commonwealth Scholarship and Fellowship Plan for the second author is gratefully acknowledged.

ABSTRACT

This paper discusses the traditional specification problem from a geometric viewpoint. While the traditional emphasis is on the properties of estimators, the geometric approach also allows an easy development of corresponding results for inference. Errors arising from artificial inclusion or exclusion of variables are considered in terms of augmentations or restrictions on a given maintained hypothesis, and this allows a corresponding interpretation of tests based upon the Wald and Lagrange Multiplier Principles. It is demonstrated that biases arising from incorrect exclusion of variables do not invalidate the traditional F-test.

THE GEOMETRY OF SPECIFICATION ERROR

- Gordon Fisher and Michael McAleer -

1. *Introduction*

For over two decades the approach pioneered by Theil (1957) has been the traditional framework for examining specification error involving artificial inclusion or exclusion of variables in a linear regression. Of course, since incorrectly excluding variables may be regarded as imposing restrictions on the 'true' model, and since the 'true' model corresponding to incorrect inclusion of variables may itself be regarded as the result of imposing restrictions on the false model, it is obviously possible to re-cast the specification problem in terms of restricted least squares. Evidently, this was first noted by Stewart (1976, pp. 96-100), and the more formal analysis of Riddell and Buse (1980) demonstrates some of the interesting didactical features that are often passed over in the traditional discussion. Nevertheless, one of the great drawbacks of examining the problem from the traditional or the restricted least squares viewpoints is that attention is concentrated on estimation because inferential matters get pushed into the background. Part of the reason for this is that, while the distributional aspects of regression are defined on observation space, the discussion of specification error is cast in terms of parameter space. Though this is the traditional way of tackling the problem, it does involve an awkward switch from the one space to the other, and this tends to direct attention toward bias and efficiency alone without giving natural vent to considerations of inference. Yet when the same problem is cast in terms of observation space, there emerges not only a direct link with the general problem of formulating and analyzing a linear hypothesis, but also an obvious route along which results on bias and efficiency may be developed concisely. Indeed, such an

approach has the distinct merit of emphasizing the natural interplay between estimation and testing in respect of specification error and other general problems that arise in connection with a linear hypothesis.

The concern of this paper is to develop an approach to specification error which emphasizes geometric aspects of the problem. This has the virtue of obtaining the standard results succinctly while drawing attention to important features which might otherwise be missed. Moreover, certain of the geometric properties we shall develop form a natural basis for discussing inferences about specification, and also allow us to draw a distinction between these sorts of inferences and inferences that might otherwise be made were the specification problem itself to be passed over unnoticed. In this and several other respects, a number of new results are developed. Errors committed by including or excluding variables incorrectly are analyzed in terms of augmentations or restrictions of a given maintained hypothesis, and this permits us to exploit the relations between tests based on the Wald and Lagrange Multiplier Principles. For example, while it is well known that biases are introduced by excluding variables incorrectly, it turns out that the appropriate F-test, though based upon biased estimates, is nevertheless valid.

The plan of the paper is as follows. There is first a general discussion of the linear hypothesis from the geometric viewpoint. This is justified not only because of its unfamiliarity, but also because it forms the basis of the analysis in the rest of the paper. There is then an illustration with a familiar example in Section 3. In Section 4, the same analysis is applied to specification error, and in Section 5 extensions of the results are discussed.

2. *The Linear Hypothesis*

It should be emphasized at the outset that mistakes made in formulating a linear hypothesis, whether by inclusion or exclusion of variables, are in prin-

ciple testable using standard methods. For this reason, it is helpful to begin with a general method of formulating and testing a linear hypothesis. The approach to be adopted is due to Kruskal (1961), with important extensions by Seber (1964a, 1964b, 1964c and 1966) and Kruskal (1968).

Consider the vector y which ranges over n-dimensional Euclidean space, $\mathcal{E}_n$, on which the inner product $(\cdot,\cdot)$ is given. The distribution of y is multivariate normal and its moments may be defined in terms of the given inner product. Thus for fixed x and z in $\mathcal{E}_n$, the unique vector expectation of y, say μ, is defined by the linear functional $E\{(x,y)\} = (x,\mu)$ while Σ_0, the unique linear transformation on $\mathcal{E}_n$ called the dispersion (or covariance) operator of y, is defined by interpreting $D\{(x,y),(z,y)\}$ as $(x,\Sigma_0 z)$, a proper inner product, except possibly non-negative definite. Without loss of generality, it is assumed that $\Sigma_0 = I_n\sigma^2$, whence $y \sim N(\mu, I_n\sigma^2)$ and there is no confusion in writing $E(y) = \mu$, $D(y) = I_n\sigma^2$. Notice that while vector expectation does not depend on the given inner product, the definition of dispersion does. Thus for some idempotent linear transformation G, for example, if $(\cdot,G\cdot) \equiv \langle\cdot,\cdot\rangle$ defines a new inner product, $E\{\langle x,y\rangle\} = \langle x,\mu\rangle$ but $D\{\langle x,y\rangle, \langle z,y\rangle\} = \langle x,\Sigma_0 G^T z\rangle$, *not* $\langle x,\Sigma_0 z\rangle$. This is helpful in understanding the discussion of components of specification error later on.

It is given that $\mu\epsilon\Omega$, a p-dimensional sub-space, but otherwide μ and σ^2 are unknown. It is then desired to test the linear hypothesis H: $\mu\epsilon\omega\subset\Omega$, ω being (p-r)-dimensional. Thus r is the number of linear restrictions on Ω to define ω.

The following notation will be adopted. Least squares estimators of parameters in Greek letters will be denoted by corresponding Roman letters, indexed according to the associated sub-space of estimation. For example, corresponding to Ω, the least squares estimates of μ and σ^2 are m_Ω and s^2_Ω, respec-

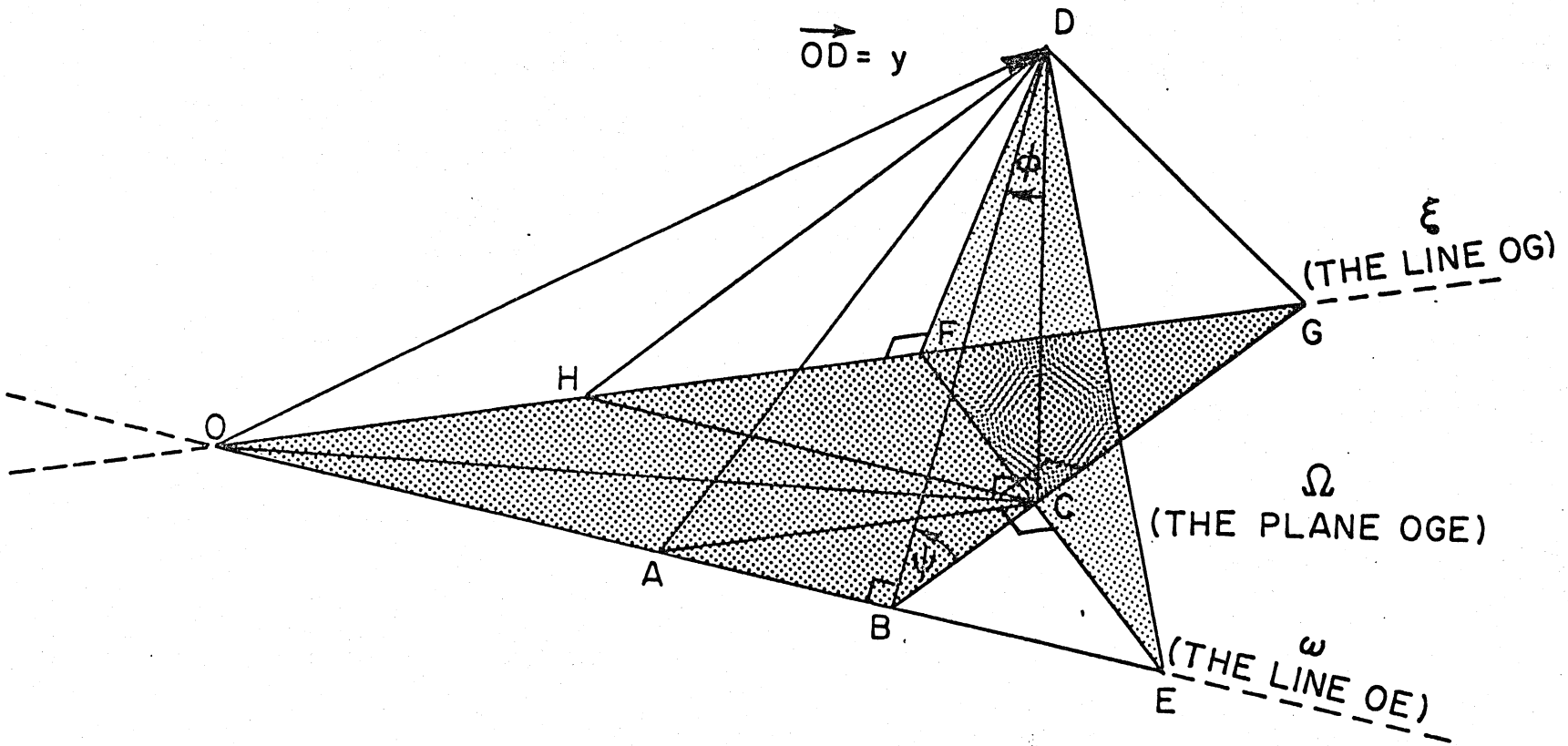
tively. Also, P will denote perpendicular projection, *i.e.* orthogonal projection relative to $(\cdot,\cdot)$, and this will be indexed according to the sub-space on which projection takes place; corresponding projection on the ortho-complement in $\mathcal{E}_n$ is denoted by Q, with the same indexing. For example, P_ω is on ω and $Q_\omega \equiv (I_n - P_\omega)$ is on $\omega^\perp$, and so on. Dimension is denoted by $d(\cdot)$ and length relative to $(\cdot,\cdot)$ by $\|\cdot\|$. With this notation, $m_\Omega = P_\Omega y$, $s_\Omega^2 = \{\|Q_\Omega y\|^2/d(\Omega^\perp)\}$, $m_\omega = P_\omega y$ and $s_\omega^2 = \{\|Q_\omega y\|^2/d(\omega^\perp)\}$. Of course, $d(\Omega^\perp) = n-p$ and $d(\omega^\perp) = n-p+r$. To complete preliminaries: $(P_\Omega - P_\omega) = P_\Omega Q_\omega = P_{\omega^\perp \cap \Omega}$; symmetry ensures $P_\Omega Q_\omega = Q_\omega P_\Omega$; and under H, $E(P_{\omega^\perp \cap \Omega} y) = P_{\omega^\perp \cap \Omega}\mu = 0$ since $\mu \in \omega$.

The standard test statistic for H is:

$$F = \frac{\|(P_\Omega - P_\omega)y\|^2/d(\omega^\perp \cap \Omega)}{\|Q_\Omega y\|^2/d(\Omega^\perp)} = \frac{y^T(P_\Omega - P_\omega)y/r}{y^T Q_\Omega y/(n-p)} \quad . \tag{1}$$

There are various expressions corresponding to equation (1), but the right-hand side of this equation is fundamental in the sense it comprises components of y in sub-spaces arising from estimation, and the dimensions of these sub-spaces. Given the ratio of dimensions, the formula has a natural geometric interpretation, namely, the square of the tangent of the angle of inclination of $Q_\omega y$ to $Q_\Omega y$ ($\tan^2\phi = \{BC/CD\}^2$ in Diagram 1), which clearly varies according as the distance $\|m_\Omega - m_\omega\|$ (BC in Diagram 1). The same point is made explicitly in the following alternative expression:

$$F = \frac{\|m_\Omega - m_\omega\|^2}{r s_\Omega^2} \quad . \tag{2}$$



Key:

$\vec{OB} = P_\omega y$	$\vec{OF} = P_\xi y$	$\vec{OC} = P_\Omega y$
$\vec{OA} = F_\omega y$	$\vec{OH} = F_\xi y$	$\vec{BC} = (P_\Omega - P_\omega)y = m_\Omega - m_\omega$
$\vec{AB} = (P_\omega - F_\omega)y$	$\vec{HF} = (P_\xi - F_\xi)y$	$\vec{CD} = Q_\Omega y$
$\vec{BD} = Q_\omega y$	$\vec{FD} = Q_\xi y$	
$\vec{AD} = (I_n - F_\omega)y$	$\vec{HD} = (I_n - F_\xi)y$	

DIAGRAM 1

Usually, in a practical problem, H is expressed in the form of r restricting relations $(a_i,\mu) = 0$ $(i = 1,2,\ldots,r)$ or as $A^T\mu = 0$, $\mu\epsilon\Omega$ where A is n x r of rank $r \leq p$ of known columns a_i. In this case, which is quite general, $\omega = \Omega \cap N[A^T]$ where $N[A^T] = \{\lambda: A^T\lambda = 0, \lambda\epsilon\mathbb{R}_n\}$. Corresponding to this formulation, the unique perpendicular projection on $\omega^{\perp} \cap \Omega$ may be written $P_\Omega A(A^T P_\Omega A)^{-1} A^T P_\Omega$, provided the span of A and $\Omega^{\perp}$ intersect only at the origin (Seber, 1964a, p. 262). It is then straightforward to demonstrate that $\mathbb{F}$ in (1) embodies the Wald (1943) Principle of testing restrictions using only unrestricted estimates, since:

$$\mathbb{F} = \frac{y^T(P_\Omega - P_\omega)y}{rs^2_\Omega} = \frac{(A^T m_\Omega)^T[D_\Omega(A^T m_\Omega)]^{-1}(A^T m_\Omega)}{r} , \qquad (3)$$

where $D_\Omega(\cdot)$ denotes estimate of dispersion corresponding to the space Ω, *i.e.* evaluated at $\sigma^2 = s^2_\Omega$. (While $D(\cdot)$ generally denotes dispersion, when D is indexed by a sub-space it denotes estimate of dispersion corresponding to estimation in that sub-space.)

Of course, $\mathbb{F}$ in (1), and hence in (2) and (3), has the central F(r,n-p) distribution under H. Further $r\mathbb{F}$ is a quadratic form based upon the unrestricted estimates m_Ω and s^2_Ω and the given restrictions only; upon replacing s^2_Ω with σ^2, it is seen to be a quadratic form in normal variates under H. Since also s^2_Ω is asymptotically equivalent to σ^2_Ω, the corresponding maximum-likelihood estimator of σ^2, it is obvious that $r\mathbb{F} \sim \chi^2(r)$ approximately for large n. This is the standard large-sample Wald test (see Seber, 1964a).

It is worth remarking that, contrary to Seber (1964a), neither (1) nor (2) may be construed as test-statistics based strictly upon the Lagrange Multiplier Principle (Silvey, 1959). The test-statistic corresponding to the Lagrange Multiplier Principle is:

$$M = \frac{y^T(P_\Omega - P_\omega)y}{y^T Q_\omega y} = \frac{y^T(P_\Omega - P_\omega)y}{y^T(P_\Omega - P_\omega)y + y^T Q_\Omega y} \quad , \tag{4}$$

which is exactly distributed as $\beta_1(\frac{r}{2}, \frac{n-p}{2})$, since the two components in the denominator of the expression in (4), each divided by σ^2, are independent chi-square variates with r and (n-p) degrees of freedom, respectively. However, this β_1-distribution is in one-for-one correspondence with the F(r,n-p) distribution. Hence there is no conflict in testing H *via* equations (1), (2) or (3) using the F(r,n-p) distribution or *via* (4) using the $\beta_1(\frac{r}{2}, \frac{n-p}{2})$ distribution (Fisher, 1978, 1980; Fisher and McAleer, 1980). Notice that in Diagram 1, $M = \sin^2\phi = \cos^2\Psi$.

3. An Example

In a practical regression problem, y is a vector of n co-ordinate observations and Ω is the span of an n x p matrix, say W, of rank p. Thus μ may be written $\mu = W\alpha$ for some $\alpha \epsilon \mathbb{R}_p$, and y as $W\alpha + \varepsilon$ with $\varepsilon \sim N(0, I_n\sigma^2)$. If as above, $A^T\mu = 0, \mu \epsilon \mathcal{S}[W]$ under H, where $\mathcal{S}[\cdot]$ denotes span, then there must also exist an $A^* = P_\Omega A$ which satisfies the same condition since $\Omega = \mathcal{S}[W]$ and hence $A^{*T}\mu = A^T P_\Omega \mu = A^T\mu = 0$. In view of this, there must also exist a matrix B of order p x r with its r columns in the span of W^T since obviously $W(W^TW)^{-1}W^TA = W(W^TW)^{-1}B$, in which $B = W^TA$, and hence

$$0 = A^T\mu = A^TW\alpha = A^{*T}W\alpha = A^TW(W^TW)^{-1}W^TW\alpha = B^T\alpha \quad .$$

Thus for every set of admissible restrictions $A^T\mu = 0$, $\mu = W\alpha$, there corresponds a B such that $B^T\alpha = 0$; moreover, if $B^T\alpha = 0$, then $0 = B^T(W^TW)^{-1}W^TW\alpha = A^{*T}W\alpha = A^TP_\Omega W\alpha = A^T\mu$. Generally, then, there are many A's to a given B, but corresponding to the A's there is a unique A*, given $\mu \epsilon \Omega = \mathcal{S}[W]$. Moreover,

there is no need, from a theoretical viewpoint, to consider non-homogeneous restrictions of the kind $B^T\alpha = \rho$ or $A^T\mu = \kappa$, since this case may readily be adjusted to the original one. Any solutions α^* or μ^* yield $B^T\alpha^* = \rho$ and $A^T\mu^* = \kappa$, whence $B^T(\alpha-\alpha^*) = 0$ and $A^T(\mu-\mu^*) = 0$. Hence, consideration of the regression $y - \mu^* = W(\alpha-\alpha^*) + \varepsilon$ accounts for the non-homogeneity without damage to the generality of the homogenous case as stated (see *e.g.* Seber, 1964a, 1966).

To help fix ideas still further, consider the above regression when $W = [X:Z]$: X has (p-r) columns, Z has r columns, B^T has r rows and $B^T = [0_{p-r}:I_r]$. If $\alpha^T = [\beta^T:\gamma^T]$ is conformably partitioned:

$$y = W\alpha + \varepsilon = X\beta + Z\gamma + \varepsilon \quad , \tag{5}$$

then $B^T(W^TW)^{-1}W^TW\alpha = A^{*T}\mu = 0$ and $\omega = \mathscr{L}[W] \cap N[A^{*T}] = \mathscr{L}[W] \cap \mathscr{L}[X]$, *i.e.* $\omega = \mathscr{L}[X]$. In these circumstances, equation (5) under H becomes:

$$y = X\beta + \varepsilon \quad , \tag{6}$$

and hence equation (1) specializes to:

$$F = \frac{y^T(W(W^TW)^{-1}W^T - X(X^TX)^{-1}X^T)y}{y^T(I_n - W(W^TW)^{-1}W^T)y} \cdot \frac{(n-p)}{r} \quad . \tag{7}$$

A useful alternative formula for equation (7) may be developed as follows. Let ξ denote $\mathscr{L}[Z]$; since the matrices W, X and Z each have linearly independent columns, it is clear that $\Omega = \omega \oplus \xi$. Let F denote an oblique projection relative to $(\cdot,\cdot)$: F_ω will denote the oblique projection on ω relative to $(\cdot,\cdot)$ which is the orthogonal projection relative to $(\cdot,Q_\xi\cdot)$; F_ξ is, similarly, oblique on ξ relative to $(\cdot,\cdot)$ and orthogonal relative to $(\cdot, Q_\omega\cdot)$. In terms of matrices: $F_\omega = X(X^TQ_\xi X)^{-1}X^TQ_\xi$ and $F_\xi = Z(Z^TQ_\omega Z)^{-1}Z^TQ_\omega$, which are, of course,

idempotent but not symmetric. It follows immediately that $P_\Omega = F_\omega + F_\xi$ and hence:

$$Q_\omega P_\Omega = (P_\Omega - P_\omega) = Q_\omega F_\omega + Q_\omega F_\xi \quad , \tag{8}$$

which, since $Q_\omega F_\omega = 0$, reduces to:

$$P_\Omega - P_\omega = Q_\omega F_\xi = Q_\omega Z(Z^T Q_\omega Z)^{-1} Z^T Q_\omega \quad . \tag{9}$$

In view of this, equation (7) reduces to the familiar expression:

$$F = \frac{y^T Q_\omega Z(Z^T Q_\omega Z)^{-1} Z^T Q_\omega y}{y^T (I_n - W(W^T W)^{-1} W^T) y} \cdot \frac{n-p}{r} \quad . \tag{10}$$

The discussion thus far has concentrated on testing a set of restrictions on a given maintained hypothesis, using estimates under that maintained hypothesis. We may characterize this way of looking at matters as testing the movement from a given maintained hypothesis toward an hypothesis H which is restricted in some specified way; or, more precisely, testing the movement from a given maintained hypothesis, call it H_Ω, toward a restricted version of H_Ω, denoted by H_ω, which will be termed 'the restricted hypothesis'. It is quite natural to base this test upon estimation under H_Ω (*i.e.* in Ω) simply because this is the hypothesis (*i.e.* space) from which movement is to be considered. This is precisely what the Wald Principle tells us to do. However, as equation (4) makes clear, a test may be formulated *as if* movement were the other way round, *i.e.* from estimation under H_ω toward estimation under H_Ω. Such a test is based upon the Lagrange Multiplier Principle.

For the purpose of analyzing specification error, it is useful to consider not just two *principles* of testing the same hypothesis, but rather two *classes* of hypotheses. Hypotheses of the first kind are essentially concerned

with *restrictions* on a given maintained hypothesis and these have already been considered. Hypotheses of the second kind are concerned with *augmentations* of a given maintained hypothesis, and tests of them may be described as testing a movement from a given maintained toward a specified *augmented* (or modified) hypothesis. In terms of linear hypotheses, if the maintained is H_Ω: $\mu \epsilon \Omega$ as before, this may be augmented to become H_Λ: $\mu \epsilon \Omega \oplus \nu \equiv \Lambda$, where ν is a sub-space in $\mathcal{E}_n$ of dimension q such that $d(\Lambda) = p+q = k \leqq n$. Thus ν might be $\mathcal{A}[U]$ where U is an n x q matrix of linearly independent columns such that [W:U] has rank k. Returning to specification error, incorrect inclusion of variables will be seen to correspond to an augmentation of the given maintained hypothesis, while incorrect exclusion of variables corresponds to a restriction of the maintained hypothesis.

Turning back to an hypothesis of the first kind and considering a movement from H_Ω to H_ω, the form of the test is to demonstrate whether the squared distance between efficient estimates of μ in Ω and ω is significantly different from zero; if it is, H_ω is rejected; if not, H_ω is not rejected. Since the hypothesis of departure is H_Ω, it is natural to base significance testing upon estimation of σ^2 in Ω (the Wald Principle), but the Lagrange Multiplier Principle may also be applied, whence σ^2 is estimated in ω. In the case of hypotheses of the second kind, a movement from H_Ω toward H_Λ is considered; but now $d(\Omega) < d(\Lambda)$, and movement is from the space of smaller to the space of larger dimension, not *vice-versa*. Again the test seeks to determine whether the squared distance from an efficient estimate of μ in Λ to the corresponding estimate in Ω, is, or is not, significantly different from zero: but this time if it is, H_Λ is not rejected, whereas if it is not significantly different, H_Λ is rejected. Moreover, H_Ω is again the hypothesis of departure and hence Ω is the natural space for estimation of σ^2; but since Ω is now the space of smaller

dimension, the natural procedure to apply is the Lagrange Multiplier Principle, not the Wald Principle, and hence to base significance testing on estimation of σ^2 in Ω. In short, in considering a movement from a maintained to a *restricted* hypothesis, the natural principle to apply is that of Wald; but in considering a movement from a maintained to an *augmented* hypothesis, the natural principle to apply is the Lagrange Multiplier Principle. However, the other principle may be applied in each case, subject to re-definition of the test-statistic and application of the appropriate distribution.

4. Specification Error

It is assumed that H_Ω: $\mu \epsilon \Omega$ applies throughout. In ignorance of this fact, μ is presumed to lie in Λ, that is, in k-dimensional rather than p-dimensional space, where $p < k$. Since $\Lambda = \Omega \oplus \nu$, it is natural to consider the decomposition:

$$P_\Lambda y = F_\Omega y + F_\nu y \quad ,$$

where F_Ω is orthogonal on Ω relative to $(\cdot, Q_\nu \cdot)$, and F_ν is orthogonal on ν relative to $(\cdot, Q_\Omega \cdot)$. Immediately we have:

(i) $P_\Lambda y$ is unbiased for μ, since $E(P_\Lambda y) = P_\Lambda \mu = \mu$, because $\mu \epsilon \Omega \subset \Lambda$.

(ii) $F_\nu y$ has expectation zero, since F_ν is orthogonal on ν relative to $(\cdot, Q_\Omega \cdot)$ and $Q_\Omega \mu = 0$, $\mu \epsilon \Omega$. Hence,

(iii) $E(F_\Omega y) = E(P_\Lambda y) = \mu$; or directly, $E(F_\Omega y) = F_\Omega \mu = \mu$, since F_Ω is on Ω. Moreover,

(iv) $D(P_\Lambda y) = P_\Lambda \sigma^2$, whereas the minimum variance linear unbiased estimator has dispersion $P_\Omega \sigma^2$. Since the difference between the two estimators is $(P_\Lambda - P_\Omega)y$ and $(P_\Lambda - P_\Omega)$ is itself a perpendicular projection, the dispersion arising from this difference is precisely the difference of the

individual dispersions, namely $(P_\Lambda - P_\Omega)\sigma^2$. This may be regarded as a representation of the total loss of efficiency due to specification error.

(v) $(P_\Lambda - P_\Omega)y$ may be decomposed into $(F_\Omega - P_\Omega)y + F_\nu y$ and these represent components of $(P_\Lambda - P_\Omega)y$ attributable to the spaces Ω and ν, respectively.

(vi) The dispersion corresponding to this decomposition is:

$$D\{(x,(P_\Lambda - P_\Omega)y),(z,(P_\Lambda - P_\Omega)y)\} = (x,(P_\Lambda - P_\Omega)z)\sigma^2$$
$$= (x,[(F_\Omega - P_\Omega)(F_\Omega^T - P_\Omega) + F_\nu F_\nu^T + (F_\Omega F_\nu^T + F_\nu F_\Omega^T)]z)\sigma^2,$$

for fixed x and z in $\mathcal{E}_n$, since $P_\Omega F_\nu^T = 0$. Thus the total dispersion $(P_\Lambda - P_\Omega)\sigma^2$ may be interpreted as arising from dispersion: (a) within Ω, $(F_\Omega - P_\Omega)(F_\Omega^T - P_\Omega)\sigma^2$; (b) within ν, $F_\nu F_\nu^T\sigma^2$; and (c) between the two spaces, *i.e.* as a result of correlation between the components of y in the two spaces, $(F_\Omega F_\nu^T + F_\nu F_\Omega^T)\sigma^2$. Thus:

(vii) within Ω, there can never be a gain in efficiency from presuming $\mu\epsilon\Lambda$, rather than $\mu\epsilon\Omega$, since $(F_\Omega - P_\Omega)(F_\Omega^T - P_\Omega)$ is non-negative definite.

(viii) However, when $\Omega \perp \nu$, $F_\Omega = P_\Omega$ and so the loss of efficiency from within Ω is zero; *pari passu*, $F_\Omega F_\nu^T = F_\nu F_\Omega^T = 0$. Nevertheless, under the same condition, $F_\nu F_\nu^T = P_\nu$ since $F_\nu = P_\nu$ and so the total loss of efficiency may be represented as $(P_\Lambda - P_\Omega)\sigma^2 = P_\nu\sigma^2$ which again is non-negative definite. More directly, when $\Omega \perp \nu$ the oblique projections become perpendicular and so $P_\Lambda = P_\Omega + P_\nu$, whence $(P_\Lambda - P_\Omega)\sigma^2 = P_\nu\sigma^2$, as before. Thus there can never be an advantage, from the point of view of efficient estimation, of estimating μ in Λ rather than μ in Ω, even though there may be no loss within Ω.

(ix) It is perhaps worth remarking that while $(F_\Omega - P_\Omega)$ is non-null, its square is null. This is because both F_Ω and P_Ω are on Ω, whence $F_\Omega(F_\Omega - P_\Omega) = F_\Omega - P_\Omega = P_\Omega(F_\Omega - P_\Omega)$, a result which implies $(F_\Omega - P_\Omega)^2 = 0$.

(x) In view of *(i)*, $E(s_\Lambda^2) = E\{||Q_\Lambda y||^2/d(\Lambda^\perp)\} = \sigma^2$, so that estimates of both μ and σ^2 are unbiased.

(xi) Of course, if the possibility of specification error is in mind, a test for the exclusion of the improperly included variables is possible. This naturally involves considering a movement from H_Λ: $\mu \epsilon \Lambda$ to H_Ω when we have available only estimates of μ and σ^2 in Λ. Since $\Omega \subset \Lambda$, the problem might be treated like the testing of a restricted hypothesis by applying the Wald Principle in the form:

$$F = \frac{y^T(P_\Lambda - P_\Omega)y}{y^T Q_\Lambda y} \cdot \frac{n-k}{q} \quad . \tag{11}$$

The test is valid despite the fact it presumes H_Λ holds, which is invalid. This is because *(i)* and *(x)* apply. Moreover, had we proceeded from a valid hypothesis, namely $\mu \epsilon \Omega$, then we would have been forced to apply the Lagrange Multiplier Principle using:

$$M = \frac{y^T(P_\Lambda - P_\Omega)y}{y^T Q_\Omega y} \quad , \tag{12}$$

and thus treating the problem as one of significance testing of an augmented hypothesis. However, in practice this makes no difference whatsoever so long as it is recognized that (12) is β_1-distributed; convenience then dictates that (11) be used. Finally,

(xii) notice that central to the points *(i)*-*(xi)* above is the difference between the two regressions $(P_\Lambda - P_\Omega)y$: this is distributed on $\mathcal{E}_n$ as $N(0,(P_\Lambda - P_\Omega)\sigma^2)$ and has squared length which forms the numerator of valid test statistics; its distribution is spherical on its own sub-space since, for x and z in $\Omega^\perp \cap \Lambda$, $D\{(x,(P_\Lambda - P_\Omega)y),(z,(P_\Lambda - P_\Omega)y)\} = (x,(P_\Lambda - P_\Omega)z)\sigma^2 = (x,z)\sigma^2$.

Turning now to the obverse case where again H_Ω applies but μ is presumed to lie in $\omega \subset \Omega$, $d(\omega) = (p-r)$; *i.e.* the model is estimated as if H_ω: $\mu \epsilon \omega$ is at work. Since $\Omega = \omega \oplus \xi$, $P_\Omega = F_\omega + F_\xi$. Let μ_ω be the component of μ in ω and μ_ξ be the component of μ in ξ: $\mu_\omega + \mu_\xi = \mu$. Then,

(xiii) $F_\omega y$ is unbiased for μ_ω: $E(F_\omega y) = F_\omega(\mu_\omega + \mu_\xi) = F_\omega\mu_\omega = \mu_\omega$. However, $P_\omega y$, the least squares estimator of μ under H_ω, is biased for μ_ω, save when ω and ξ are orthogonal: $E(P_\omega y) = P_\omega\mu = \mu_\omega + P_\omega\mu_\xi$; if $\omega \perp \xi$, $P_\omega\mu_\xi = 0$ and $E(P_\omega y) = \mu_\omega$. It follows from these results that:

(xiv) $P_\omega y$ is biased for μ, even when $\omega \perp \xi$, but except when $\mu_\xi \epsilon \omega$ or H_ω actually applies; both of these conditions are ruled out by hypothesis (but see Section 5).

(xv) $D(P_\omega y) = P_\omega\sigma^2$ whereas the dispersion of the minimum variance linear unbiased estimator of μ, $P_\Omega y$, is $P_\Omega\sigma^2$. Since $D\{(P_\Omega - P_\omega)y\} = (P_\Omega - P_\omega)\sigma^2$ is non-negative definite, it follows that $P_\omega y$ will never be less efficient than $P_\Omega y$, where efficiency is defined in respect of dispersion only and thereby permits comparison between biased and unbiased estimators. It is also the case that:

(xvi) $P_\Omega\sigma^2 = \{F_\omega F_\omega^T + F_\xi F_\xi^T + (F_\omega F_\xi^T + F_\xi F_\omega^T)\}\sigma^2$. Hence, so far as estimation of μ_ω is concerned, $P_\omega y$ is never less efficient (in the same sense as *(xv)* above) than the minimum variance linear unbiased estimator, since $(F_\omega F_\omega^T - P_\omega)\sigma^2 = (F_\omega - P_\omega)(F_\omega^T - P_\omega)\sigma^2$ is also non-negative definite (*c.f. (vi)* above). However, $(F_\omega - P_\omega) = 0$ when $\omega \perp \xi$.

(xvii) Notice the double danger inherent in *(xiv)*, *(xv)* and *(xvi)* as a result of presuming μ to lie in ω: (a) $P_\omega y$ is generally biased for μ_ω and for μ; and (b) its distribution generally has less 'spread' around the biased mean than the most efficient linear unbiased estimator, in each case.

(xviii) In view of *(xiii)* and *(xiv)*, $E(s_\omega^2) = E\{||Q_\omega y||^2/d(\omega^\perp)\} = \sigma^2 + \{||Q_\omega \mu_\xi||^2/d(\omega^\perp)\}$, and so s_ω^2 is biased upward for σ^2, even when $\omega \perp \xi$ (but see Section 5).

(xix) With regard to testing for specification error, the natural test would be based upon the Lagrange Multiplier Principle, since estimates of μ and σ^2 in ω are available. To be sure, ω does not contain Ω so that biases are present under H_ω, but this does *not* render the test invalid. By considering a movement from H_Ω to H_ω, it is readily seen that the Wald Principle will lead to a valid test, and this itself renders valid the corresponding small-sample specialization of the Lagrange Multiplier test. This situation is in contrast to the one examined in *(xi)* above, where Λ contains Ω and hence no biases are involved under H_Λ.

5. *Discussion*

It is of interest to enquire about the matrix differences $(P_\Lambda s_\Lambda^2 - P_\Omega s_\Omega^2)$ and $(P_\Omega s_\Omega^2 - P_\omega s_\omega^2)$ corresponding to the two cases of specification error examined above, since these are sample estimates of population measures of efficiency. The first estimate may be rewritten $\{P_\Lambda(s_\Lambda^2/s_\Omega^2) - P_\Omega\}s_\Omega^2$ and the second as $\{P_\Omega - P_\omega(s_\omega^2/s_\Omega^2)\}s_\Omega^2$. Since (s_Λ^2/s_Ω^2) may be less than unity and (s_ω^2/s_Ω^2) greater than unity, it should be clear that the non-negative definiteness of $(P_\Lambda - P_\Omega)\sigma^2$ and $(P_\Omega - P_\omega)\sigma^2$ may not be preserved when σ^2 is estimated in the various spaces. This result may hold even when orthogonality is present. However, since $(s_\Lambda^2/s_\Omega^2) \simeq 1$ when the extra variables are not significant, it is likely that $(P_\Lambda s_\Lambda^2 - P_\Omega s_\Omega^2)$ will be non-negative definite in this case. In the case of (s_ω^2/s_Ω^2), since the excluded variables ought to be included, the ratio may be

markedly greater than unity. Hence, it may not then be possible to say whether $(P_\Omega s_\Omega^2 - P_\omega s_\omega^2)$ will be non-negative definite.

It is well known that when variables are inadvertently excluded and these are orthogonal to the included ones, then $\mu_\omega \perp \mu_\xi$ and $P_\omega \mu_\xi = 0$ but $E(s_\omega^2) = \sigma^2 + \{||Q_\omega \mu_\xi||^2 / d(\omega^\perp)\} = \sigma^2 + \{||\mu_\xi||^2 / d(\omega^\perp)\} \geqq \sigma^2$, except when $\mu_\xi = 0$ (*i.e.* H_ω is valid). There are two points worth mentioning here. First, if $\xi \subset \omega$ (*i.e.* the two sub-spaces are linearly dependent with ξ the sub-space of smaller dimension), then $Q_\omega \mu_\xi = 0$ and $E(s_\omega^2) = \sigma^2$. In this case, the estimated error variance is unbiased for σ^2 and so is $P_\omega y$ for μ, since $\mu_\omega + P_\omega \mu_\xi = \mu_\omega + \mu_\xi = \mu$, but it is not possible to identify μ_ω and μ_ξ separately. Of course, the likelihood is that μ_ξ will not lie in ω but simply be 'close' to it; that is, there will be a high correlation between any column of Z and some or all of the columns of X. More generally, as the maintained hypothesis is augmented, so the chances of co-linearity between the explanatory variables are increased. This contrasts with restricting the maintained hypothesis, in which case the number of explanatory variables is decreased.

The second point concerns the presumption that the matrices X and Z (which define the sub-spaces ω and ξ, respectively) are such that the span of each has dimension precisely equal to the number of its columns. Suppose, however, that the rank of Z is smaller than the number of its columns. Then $\mathcal{S}[Z] = \xi$ has dimension less than r and $N[Z]$ is, in consequence, non-null. Hence, there will exist vectors $\gamma \neq 0$ in $N[Z]$ and for these $Z\gamma = 0$. Clearly, no bias in estimating μ with $P_\omega y$ can then arise, but $\gamma \neq 0$ and hence H_Ω still applies. This possibility seems to have gone unnoticed in the econometric literature since the assumption of full column rank eliminates it. However, it should be stressed that the rank assumption is made to ensure coefficients are *estimable*, while the case just examined is concerned with their *existence*.

REFERENCES

Fisher, G.R., 1978, The linear hypothesis and structural change, unpublished paper, Department of Economics, Queen's University.

Fisher, G.R., 1980, Propositions, principles and methods: The linear hypothesis and structural change, Discussion Paper No. 410, Institute for Economic Research, Queen's University.

Fisher, G.R. and M. McAleer, 1980, Principles and methods in the testing of alternative models, Discussion Paper No. 400, Institute for Economic Research, Queen's University. Paper to the North American meeting of the Econometric Society, Denver, Colorado, September 1980.

Kruskal, W.H., 1961, The coordinate-free approach to Gauss-Markov estimation and its application to missing and extra observations, Proceedings of the fourth Berkeley symposium on mathematical statistics and probability 1, 435-451.

Kruskal, W.H., 1968, When are Gauss-Markov and least squares estimators identical? A coordinate-free approach, Annals of Mathematical Statistics 39, 70-75.

Riddell, W.C. and A. Buse, 1980, An alternative approach to specification errors, Australian Economic Papers 19, 211-213.

Seber, G.A.F., 1964a, The linear hypothesis and idempotent matrices, Journal of the Royal Statistical Society B 26, 261-266.

Seber, G.A.F., 1964b, Linear hypotheses and induced tests, Biometrika 51, 41-48.

Seber, G.A.F., 1964c, The linear hypothesis and large sample theory, Annals of Mathematical Statistics 35, 773-779.

Seber, G.A.F., 1966, The linear hypothesis: A general theory (London: Griffin).

Silvey, S.D., 1959, The Lagrangian multiplier test, Annals of Mathematical Statistics 30, 389-407.

Stewart, J., 1976, Understanding econometrics (London: Hutchinson).

Theil, H., 1957, Specification errors and the estimation of economic relationships, Review of the International Statistical Institute 25, 41-51.

Wald, A., 1943, Tests of statistical hypotheses concerning several parameters when the number of observations is large, Transactions of the American Mathematical Society 54, 426-482.