Tests for Specification Errors in Classical Linear Least-squares Regression Analysis

By J. B. RAMSEY

Michigan State University

[Received December 1967. Final revision February 1969]

SUMMARY

The effects on the distribution of least-squares residuals of a series of model mis-specifications are considered. It is shown that for a variety of specification errors the distributions of the least-squares residuals are normal, but with non-zero means. An alternative predictor of the disturbance vector is used in developing four procedures for testing for the presence of specification error. The specification errors considered are omitted variables, incorrect functional form, simultaneous equation problems and heteroskedasticity.

1. INTRODUCTION

THE objectives of this paper are two. The first is to derive the distributions of the classical linear least-squares residuals under a variety of specification errors. The errors considered are omitted variables, incorrect functional form, simultaneous equation problems and heteroskedasticity. It is assumed that the disturbance terms are independently and normally distributed. It will be shown that the effect of the specification errors considered above is, with the exception of the error of heteroskedasticity, to yield residuals which though normally distributed do not have zero means, so that the distribution of the squared residuals is non-central χ^2. The second objective is to derive procedures to test for the presence of the specification errors considered in the first part of the paper. The tests are developed by comparing the distribution of residuals under the hypothesis that the specification of the model is correct to the distribution of the residuals yielded under the alternative hypothesis that there is a specification error of one of the types considered in the first part of the paper.

As a preliminary step to deriving the test procedures the classical least-squares residual vector is transformed to a sub-vector which has more desirable properties for testing the null hypothesis that the specification of the model is correct. Also, under certain assumptions, with respect to the alternative hypothesis, it is shown that the mean vector of the residuals can be approximated by a linear sum of vectors $\mathbf{q}_j$, $j = 1, 2, \ldots$. The vectors $\mathbf{q}_j$, $j = 1, 2, \ldots$ are obtained from the least-squares estimates of the dependent variable. Four tests are developed, RESET, RASET, KOMSET and BAMSET. RESET tests the significance of a regression of the residuals on a linear function of vectors $\mathbf{q}_j$, $j = 1, 2, \ldots$. RASET is a rank correlation test between $\mathbf{q}_1$ and a ranking of the squared residuals. KOMSET is the application of the Kolmogorov test statistic to discriminate between the central and non-central F distributions. BAMSET is a modification of Bartlett's M test for the heterogeneity of variance.

Considering the extent to which least squares has been used as a tool of analysis for linear regression models one is surprised to note the lack of attention which has been given to the distribution of least-squares residuals. Recently, efforts have been made to redress this omission, chiefly by Anscombe, Tukey, Box, Cox, Tidwell and Theil in a short series of papers published since 1960.

As this paper is not concerned with problems of serial correlation, the excellent work of Wald, Anderson, Durbin, Watson and others on the analysis of residuals with respect to serial correlation problems is not included in this paper. The omission is not to be construed as a denigration of the importance of this work, but merely as the wish to omit any discussion not directly relevant to the topic of an already lengthy paper.

The papers by Anscombe (1961) and Anscombe and Tukey (1963) contain summaries of the work which had been done before the 1960's and a discussion of the recent contributions of Anscombe and Tukey. The topics examined by the authors were the assumptions of normality, heteroskedasticity, non-additivity and the problem of outliers. To test for normality Anscombe derived two estimators, g_1 and g_2, for the scale-invariant measures of skewness and kurtosis. The test for heteroskedasticity is to consider that the variance of the dependent variable about its conditional mean is proportional to $\exp[\gamma \mathbf{x}_i' \boldsymbol{\beta}]$, where γ is some small constant, $\mathbf{x}_i'$ is a vector of regressors and $\boldsymbol{\beta}$ is the regression coefficient vector. Anscombe derived an estimator for γ.

Box and Cox (1964), and in an earlier paper Box and Tidwell (1962), also consider the problems of non-normality, heteroskedasticity and removable non-additivity, but take a different approach. They wish to find that transformation of the variables such that the regression between the transformed variables satisfies the full ideal conditions.

Kuh and Meyer (1955), in comparing regressions between deflated and undeflated variables, showed that if the true model is in terms of deflated variables the use of undeflated variables leads to heteroskedastic residuals. Later, Goldfeld and Quandt (1965) devised two tests of the relationship between the residuals from a least-squares regression and the deflation index.

The work of this paper will be seen to be complementary to those briefly reviewed above. First, one of the main points of interest in the literature to date is the problem of non-normality. In this paper normality of the disturbance terms is assumed to hold even for the mis-specified models to be examined in the next section. Both Tukey (1959) and Box and Cox (1964) have looked at non-additivity which arises from the use of incorrect scale of measurement for the dependent variable and which can be removed by an appropriate power transformation. This problem is not discussed in this paper. On the other hand, attention is paid to a wide range of specification errors which are particularly relevant to econometric analysis. To some extent the approach taken in this paper generalizes the work of Anscombe and Tukey on heteroskedasticity and that of Goldfeld and Quandt on the effect of deflation indices.

2. The Analysis of the Effect of Certain Specification Errors on the Distribution of the Classical Least-squares Residuals

2.1. *The Distribution of Residuals under the Full Ideal Conditions*

In order to understand clearly what is implied by each of the mis-specifications discussed below one must define precisely what is meant by the correctly specified model. In this section and the next it is assumed that the correctly specified model satisfies the "full ideal conditions" as defined by Anscombe (1961, pp. 1, 2). Thus, it

is assumed that one has N observations on $K+1$ variables, one of which is the dependent variable denoted by y_i, and the other variables are regressor variables denoted by x_{ij}, $j = 1, ..., K$, $i = 1, 2, ..., N$. The relationship between the dependent variable and the regressors is given by (1a) or (1b):

$$y_i = \sum_j x_{ij}\beta_j + u_i \quad (i = 1, 2, ..., N), \tag{1a}$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1b}$$

where $\mathbf{y}$ is an $N \times 1$ vector, $\mathbf{X}$ an $N \times K$ matrix of rank K, $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients and $\mathbf{u}$ is an $N \times 1$ vector of disturbance terms. Equation (1b) may be rewritten as (2):

$$\mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{u}, \tag{2}$$

which stresses the fact that the distribution of the vector $\mathbf{u}$ is the distribution of the vector $\mathbf{y}$ about the mean $\mathbf{X}\boldsymbol{\beta}$ *conditional* on the observed matrix $\mathbf{X}$. It is assumed that the distribution of $\mathbf{u}$ is normal with null mean vector and covariance matrix given by $\sigma^2\mathbf{I}$, i.e. conditional on the matrix $\mathbf{X}$ one has a set of N independent observations on the variable y which has constant variance about its conditional mean.

The least-squares residual vector is defined by (3):

$$\hat{\mathbf{u}} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} = \mathbf{M}\mathbf{y}, \quad \text{or} \quad \hat{u}_i = \mathbf{m}_i'\mathbf{y}, \tag{3}$$

where $\mathbf{m}_i'$ is the ith row of $\mathbf{M}$. An estimator of σ^2 is

$$(N-K)^{-1}\hat{\mathbf{u}}'\hat{\mathbf{u}} = (N-K)^{-1}\mathbf{y}'\mathbf{M}\mathbf{y}. \tag{4}$$

Under the full ideal conditions as specified in the previous paragraph the distribution of $\hat{\mathbf{u}}$ is singular normal with null mean vector and covariance matrix given by $\sigma^2\mathbf{M}$. The matrix $\mathbf{M}$ is positive semi-definite, non-diagonal and idempotent of rank $N-K$ so that, although the u_i $(i = 1, ..., N)$ are independent and identically distributed as $N(0, \sigma^2)$, the least-squares residuals $\hat{u}_i$ are not independently and identically distributed. The distribution of $(\hat{\mathbf{u}}'\hat{\mathbf{u}})/\sigma^2$ is central χ^2 with $N-K$ degrees of freedom.

The basic approach followed in the first part of this paper is to consider only those mis-specifications which yield distributions of $\hat{\mathbf{u}}$ which, conditional on the observed matrix $\mathbf{X}$, are normal. The cases where mis-specification leads to a non-normal distribution of $\hat{\mathbf{u}}$ are to be discussed in a later paper. One should emphasize that in the analysis to follow the distributions of $\hat{\mathbf{u}}$ and of $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ are derived conditional on the observed matrix of regressors.

2.2. *Some Examples of Specification Errors and Their Effect upon the Distribution of Least-squares Residuals*

2.2.1. *The omitted variable case*

Consider the following model which satisfies the full ideal conditions:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \gamma\mathbf{z} + \mathbf{u}, \quad \mathbf{u} \sim N(\mathbf{0}, \sigma_u^2\mathbf{I}), \tag{5}$$

where $\mathbf{y}$ is the $N \times 1$ regressand vector, $\mathbf{X}$ is the $N \times K$ regressor matrix of rank K, $\mathbf{z}$ is the $N \times 1$ vector for the $K+1$st regressor, $\boldsymbol{\beta}$ is the $K \times 1$ vector of coefficients, γ is the coefficient for the $K+1$st regressor and $\mathbf{u}$ is the $N \times 1$ vector of disturbances.

The model considered for regression is given in (5a):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\delta} + \mathbf{v}, \tag{5a}$$

where $\mathbf{y}$ and $\mathbf{X}$ are as before, $\boldsymbol{\delta}$ is the $K \times 1$ vector of coefficients and $\mathbf{v}$ is the $N \times 1$ vector of disturbances. If (5) is the correctly specified model, then (5a) does not fulfil the full ideal conditions and $\mathbf{v} = \gamma\mathbf{z} + \mathbf{u}$. Assume that the N dimensional vector $\mathbf{z}$ is non-stochastic. As (5) meets the full ideal conditions the distribution of $\mathbf{u}$ is independent of $\mathbf{z}$.

The least-squares residuals for the model (5a) are

$$\hat{\mathbf{v}} = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{z}\gamma + \mathbf{M}\mathbf{u},$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Under the assumptions made above one sees that the mean vector of $\hat{\mathbf{v}}$ and the covariance matrix of $\hat{\mathbf{v}}$ are

$$E(\hat{\mathbf{v}} \mid \mathbf{X}) = \mathbf{M}\mathbf{z}\gamma, \quad E(\hat{\mathbf{v}}\hat{\mathbf{v}}' \mid \mathbf{X}) = \sigma_u^2 \mathbf{M}. \tag{6}$$

Further, one notes that the distribution of $\hat{\mathbf{v}}$ is singular normal under the assumptions given above. Also, as the characteristic roots of $\mathbf{M}$ are 1 of multiplicity $N-K$ and 0 of multiplicity K, $(\hat{\mathbf{v}}'\hat{\mathbf{v}})/\sigma_u^2$ is distributed as non-central χ^2 with $N-K$ degrees of freedom and non-central parameter given by $\lambda = \gamma^2(\mathbf{z}'\mathbf{M}\mathbf{z})/\sigma_u^2$. The distribution of each $\hat{v}_i^2/\sigma_u^2$, $i = 1, 2, \ldots, N$, is also non-central χ^2 with one degree of freedom and non-central parameter given by $\lambda_i = \gamma^2(\mathbf{z}'\mathbf{m}_i\mathbf{m}_i'\mathbf{z})/\sigma_u^2$.

2.2.2. *The case of mis-specification of the functional form*

In this paper attention will be restricted to mis-specification of the regressors only. Specifically, consider the model given in (7) which satisfies the full ideal conditions:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}, \quad \mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}), \tag{7}$$

where $\mathbf{y}$ is the $N \times 1$ regressand vector, $\mathbf{Z}$ is an $N \times K$ regressor matrix of rank K, $\boldsymbol{\gamma}$ is a $K \times 1$ vector of coefficients and $\mathbf{u}$ is the $N \times 1$ vector of disturbance terms. The model considered for regression analysis is that given in (8):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}, \tag{8}$$

where $\mathbf{y}$ is as in (7), $\mathbf{X}$ is an $N \times K$ regressor matrix of rank K, $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients and $\mathbf{v}$ is the $N \times 1$ vector of disturbance terms.

Assume that the relationship between each element in $\mathbf{Z}$ and each element in $\mathbf{X}$ is that shown in (9):

$$z_{ij} = g_j(x_{ij}) \quad (j = 1, 2, \ldots, K, \quad i = 1, 2, \ldots, N), \tag{9}$$

where g_j is some non-linear function with constant parameters, e.g. g_j might represent some power function, logarithmic function, *et cetera*. One could also consider for each i, $i = 1, 2, \ldots, N$, z_{ij}, as a function of several x_{ij}, $j = 1, 2, \ldots, K$. As we have assumed in the correctly specified model, (7), that the distribution of $\mathbf{u}$ is independent of $\mathbf{Z}$, it follows that the distribution of $\mathbf{u}$ is independent of $\mathbf{X}$, whose elements are functions of the elements of $\mathbf{Z}$ alone.

The least-squares residuals for the model given in (8) are

$$\hat{\mathbf{v}} = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{Z}\gamma + \mathbf{M}\mathbf{u}, \tag{10}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Conditional on the observed $\mathbf{X}$ matrix one sees that

$$E(\hat{\mathbf{v}} \mid \mathbf{X}) = \mathbf{M}\mathbf{Z}\boldsymbol{\gamma}, \quad E(\hat{\mathbf{v}}\hat{\mathbf{v}}' \mid \mathbf{X}) = \sigma_u^2 \mathbf{M}, \tag{11}$$

where $\mathbf{Z}$ is defined by substituting the elements of $\mathbf{X}$ into equation (9). On the basis of the specification of the model given in (7), one sees that the distribution of $\hat{\mathbf{v}}$ conditional on $\mathbf{X}$ is singular normal with mean vector and covariance matrix given by (11). Further, as $\mathbf{M}$ is idempotent of rank $N-K$, the distribution of $(\hat{\mathbf{v}}'\hat{\mathbf{v}})/\sigma_u^2$ is non-central χ^2 with $N-K$ degrees of freedom and non-central parameter given by $\lambda = (\boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}\mathbf{Z}\boldsymbol{\gamma})/\sigma_u^2$. The distribution of each $\hat{v}_i^2/\sigma_u^2$, $i = 1, 2, \ldots, N$, is also non-central χ^2 with one degree of freedom and non-central parameter $\lambda_i = (\boldsymbol{\gamma}'\mathbf{Z}'\mathbf{m}_i\mathbf{m}_i'\mathbf{Z}\boldsymbol{\gamma})/\sigma_u^2$.

2.2.3. *The simultaneous equation case*

The "simultaneous equation problem" can be easily formulated under the framework used so far. Thus, consider equation (12) which summarizes the simultaneous equation system at some point in time, t:

$$\mathbf{y}_t'\boldsymbol{\Gamma} + \mathbf{z}_t'\boldsymbol{\Delta} + \mathbf{u}_t' = \mathbf{0}, \tag{12}$$

where $\mathbf{y}_t$ is an $M \times 1$ vector of endogenous variables, $\mathbf{z}_t$ is a $K \times 1$ vector of non-stochastic exogenous variables, $\mathbf{u}_t$ is an $M \times 1$ vector of stochastic error terms, $\boldsymbol{\Gamma}$ is an $M \times M$ matrix of the coefficients relating the M endogenous variables in each of the M relationships, and $\boldsymbol{\Delta}$ is a $K \times M$ matrix of coefficients relating the exogenous variables in each of the M relationships. In the following discussion it is convenient to drop the subscript t.

The mth structural equation of the system given in (12) may be written as

$$y = \mathbf{y}_m'\boldsymbol{\gamma}_m + \mathbf{z}_m'\boldsymbol{\delta}_m + u_m, \tag{13}$$

where u_m is the mth element of $\mathbf{u}$, and where, without loss of generality, one assumes that the transpose of the mth column of $\boldsymbol{\Gamma}$ is $(-1, \boldsymbol{\gamma}_m, \mathbf{0})$ where $\boldsymbol{\gamma}_m$ contains only non-zero elements and that the transpose of the mth column of $\boldsymbol{\Delta}$ is $(\boldsymbol{\delta}_m, \mathbf{0})$ where $\boldsymbol{\delta}_m$ contains only non-zero elements. y is that variable which corresponds to the first element in the mth column of $\boldsymbol{\Gamma}$ and $\mathbf{y}_m, \mathbf{z}_m$ are the appropriately dimensioned subvectors of $\mathbf{y}_t$ and $\mathbf{z}_t$ respectively which contain those elements of $\mathbf{y}_t, \mathbf{z}_t$ which correspond to the coefficients in $\boldsymbol{\gamma}_m$ and $\boldsymbol{\delta}_m$. Let M_m and K_m denote the number of endogenous and exogenous variables respectively in the mth equation which have non-zero coefficients and define the $(M_m - 1) + K_m$ dimensional vector $\mathbf{x}_m'$ by

$$\mathbf{x}_m' = (\mathbf{y}_m', \mathbf{z}_m'), \tag{14}$$

and the transpose of the corresponding coefficient vector $\boldsymbol{\beta}_m$ by

$$\boldsymbol{\beta}_m' = (\boldsymbol{\gamma}_m', \boldsymbol{\delta}_m'), \tag{14'}$$

so that (13) becomes

$$y = \mathbf{x}'\boldsymbol{\beta} + u, \tag{13'}$$

where one drops the subscript m for convenience. Suppose that the joint distribution for the M_m variables $y, \mathbf{y}_m$ is denoted by f, then

$$\begin{aligned} f(y, \mathbf{y}_m | \mathbf{z}_m) \xrightarrow{T} f(u, \mathbf{y}_m | \mathbf{z}_m) &= g(u | \mathbf{y}_m, \mathbf{z}_m)\, h(\mathbf{y}_m | \mathbf{z}_m) \\ &= g(u | \mathbf{x})\, h(\mathbf{y}_m | \mathbf{z}_m), \end{aligned} \tag{15}$$

where T denotes the transformation:

$$T\colon u = y - \mathbf{x}'\boldsymbol{\beta}, \quad \mathbf{x} = \mathbf{x}, \tag{16}$$

the Jacobian of which is 1. The function g denotes the distribution of u conditional on $\mathbf{x}$ and h is the distribution of $\mathbf{y}_m$ given $\mathbf{z}_m$. The "simultaneous equation problem" is now easily stated. By Kendall (1951, p. 12), one can obtain unbiased estimators of all the regression parameters of the model given in (13) from knowledge of the marginal distribution of u if, and only if, the joint distribution of u and $\mathbf{y}_m$ factors into the product of the marginal distribution of u and the conditional distribution of $\mathbf{y}_m$ given $\mathbf{z}_m$, i.e. if and only if

$$f(u, \mathbf{y}_m | \mathbf{z}_m) = g(u) h(\mathbf{y}_m | \mathbf{z}_m), \tag{17}$$

where $g(u)$ is the marginal distribution of u. But, under the simultaneous equation framework assumed in equation (12), $g(u|\mathbf{x})$ will not in general be equal to $g(u)$. In short, the condition given in (17) will not be met.

To proceed further and more explicitly, assume that $g(u|\mathbf{x}) = g(u|\mathbf{y}_m, \mathbf{z}_m)$ is a normal density function with mean $\eta_{u|y_m}$ and variance $\sigma^2_{u|y_m}$. If $\boldsymbol{\Sigma}$ denotes the covariance matrix for the joint distribution of u and $\mathbf{y}_m$, also assumed to be normal, then

$$\eta_{u|y_m} = \eta_u + \boldsymbol{\Sigma}_{uy} \boldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y}_m - \boldsymbol{\eta}_{y_m}), \quad \sigma^2_{u|y_m} = \sigma^2_u + \boldsymbol{\Sigma}_{uy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yu}, \quad \text{for all vectors } \mathbf{y}_m, \tag{18}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{uu} & \boldsymbol{\Sigma}_{uy} \\ \boldsymbol{\Sigma}_{yu} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}$$

denotes the appropriate partitions of the joint covariance matrix for u and $\mathbf{y}_m$. $\eta_u = 0$ is the unconditional mean of u and $\boldsymbol{\eta}_{y_m}$ is the $(M_m - 1)$ dimensional mean vector for the vector $\mathbf{y}_m$.

Assume now that one has N independent observations on the M_m joint conditional distribution of y and $\mathbf{y}_m$, namely $f(y, \mathbf{y}_m | \mathbf{z}_m)$, which has been assumed to be multivariate normal. Under these assumptions the conditional mean vector of the least-squares residuals is given by

$$E\{\hat{\mathbf{u}} | \mathbf{X}\} = E\{\mathbf{M}\mathbf{y} | \mathbf{X}\} = \mathbf{M}\boldsymbol{\eta}_{u|y_m}, \tag{19}$$

where $\mathbf{X}$ is the $N \times \{(M_m - 1) + K_m\}$ matrix of observations on the "regressors" for the mth structural relation, $\mathbf{y}$ is the $N \times 1$ vector of observations on the "dependent" variable, $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\boldsymbol{\eta}_{u|y_m}$ is the $N \times 1$ dimensional mean vector of the error vector $\mathbf{u}$ conditional on $\mathbf{X}$. It is assumed that $\boldsymbol{\eta}_{u|y_m}$ is not orthogonal to the matrix $\mathbf{M}$. Note that the matrix $\mathbf{X}$ is defined by

$$\mathbf{X} = (\mathbf{Y}_m, \mathbf{Z}_m), \tag{20}$$

where $\mathbf{Y}_m$ is an $N \times (M_m - 1)$ matrix of observations on the endogenous variables and $\mathbf{Z}_m$ is an $N \times K_m$ matrix of observations on the non-stochastic exogenous variables.

The conditional covariance matrix for the least-squares estimator of the regression coefficient vector is given by

$$E\{\mathbf{M}(\mathbf{u} - \boldsymbol{\eta}_{u|y_m})(\mathbf{u} - \boldsymbol{\eta}_{u|y_m})'\mathbf{M}'\} = \sigma^2_{u|y_m}\mathbf{M}, \tag{21}$$

where $\mathbf{M}$ is as defined above and $\sigma^2_{u|y_m}$ is the variance of the error term conditional on $\mathbf{X}$. Note that the above result assumes that the conditional variance is constant over i ($i = 1, 2, \ldots, N$).

Further, remembering that $\mathbf{M}$ is idempotent of rank $(N - M_m + 1 - K_m)$ and that the distribution of $\hat{\mathbf{u}}$ is singular normal with conditional mean vector $\mathbf{M}\boldsymbol{\eta}_{u|y_m}$, then the

distribution of the quadratic form $\hat{\mathbf{u}}'\hat{\mathbf{u}}/\sigma^2_{u|y_m}$ is under these assumptions non-central χ^2 with $(N-M_m+1-K_m)$ degrees of freedom and non-central parameter

$$\lambda = \boldsymbol{\eta}'_{u|y_m}\mathbf{M}\boldsymbol{\eta}_{u|y_m}/\sigma^2_{u|y_m},$$

as $\mathbf{M}$ is idempotent of rank $(N-M_m+1-K_m)$ and providing $\boldsymbol{\eta}_{u|y_m}$ is not orthogonal to $\mathbf{M}$.

2.2.4. *The case of simple heteroskedasticity*

Assume that the model to be estimated is that specified in (22):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\mathbf{u}, \quad \mathbf{u}\sim N(\mathbf{0},\boldsymbol{\Sigma}), \tag{22}$$

where $\mathbf{y}$ is the $N\times 1$ regressand vector, $\mathbf{X}$ is the $N\times K$ regressor matrix of rank K, $\boldsymbol{\beta}$ is the $K\times 1$ vector of regression coefficients, $\mathbf{u}$ is the $N\times 1$ vector of disturbance terms and $\boldsymbol{\Sigma}$ is the $N\times N$ diagonal covariance matrix of $\mathbf{u}$ with σ_i^2, $i=1,2,\ldots,N$, as diagonal elements. The least-squares residuals are

$$\hat{\mathbf{u}} = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{u} \quad \hat{u}_i = \mathbf{m}'_i\mathbf{u} \quad (i=1,2,\ldots,N), \tag{22'}$$

where $\mathbf{M} = \mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{m}'_i$ is the ith row of $\mathbf{M}$. The mean vector and covariance matrix of $\hat{\mathbf{u}}$, which is distributed as singular normal, are given in (23):

$$E(\hat{\mathbf{u}}\,|\,\mathbf{X}) = \mathbf{0}, \quad E(\hat{\mathbf{u}}\hat{\mathbf{u}}'\,|\,\mathbf{X} = \mathbf{M}\boldsymbol{\Sigma}\mathbf{M}. \tag{23}$$

Denoting the non-zero characteristic roots of $\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}$ by γ_j, $j=1,2,\ldots,N-K$, one notes that the conditional distribution of $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ is a weighted sum of $N-K$ central χ^2 variables with one degree of freedom each, where the weights are proportional to the roots γ_j, $j=1,2,\ldots,N-K$. The distribution of $\hat{u}_i$, $i=1,2,\ldots,N$, conditional on $\mathbf{X}$ is normal with zero mean and variance

$$\bar{\sigma}_i^2 = \sum_{j=1}^{N} m_{ij}^2\,\sigma_j^2,$$

where m_{ij} is the (i,j)th element of $\mathbf{M}$. Thus, the distribution of each $\hat{u}_i^2/\bar{\sigma}_i^2$, $i=1,2,\ldots,N$, is central χ^2 with one degree of freedom.

2.3. *Summary of the Analysis of Residuals for Mis-specified Regression Models*

The above few pages have demonstrated that under a variety of specification errors the distribution of the least-squares residual vector is singular normal with non-null mean vector of the form $\mathbf{M}\boldsymbol{\xi}$, where $\mathbf{M} = \mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{X}$ being the observed $N\times K$ regressor matrix and where $\boldsymbol{\xi}$ is a vector which summarizes the effect of the specification error on the mean of the disturbance vector in the mis-specified model. The precise formulation of the vector $\boldsymbol{\xi}$ depends upon the exact specification of the error. The exception to the above result is the case of simple heteroskedasticity where the mean of the least-squares residual vector in the mis-specified model was found to be null.

Looking at the distributions of the squared residuals it was found that for all but one of the specification errors considered the distribution of each $\hat{u}_i^2/\sigma_u^2$, $i=1,2,\ldots,N$, is non-central χ^2 with one degree of freedom and non-central parameter λ_i, where $\lambda_i \propto \boldsymbol{\xi}'\mathbf{m}_i\mathbf{m}'_i\boldsymbol{\xi}$, $\mathbf{m}'_i$ is the ith row of $\mathbf{M}$ and $\mathbf{M}\boldsymbol{\xi}$ is the vector discussed in the previous paragraph. The exceptional instance is that of simple heteroskedasticity in which case the distribution of each $\hat{u}_i^2$, $i=1,2,\ldots,N$, is scaled central χ^2 with one degree of freedom and scaling factor $\bar{\sigma}_i^2 = \sum_{j=1}^{N} m_{ij}^2\,\sigma_j^2$, where σ_j^2 are the diagonal elements of the covariance matrix of the disturbance vector.

3. An Optimal Residual Vector

One of the difficulties facing anyone trying to devise tests on the distribution of the least-squares residuals is that though the vector of residuals is normally distributed with zero mean vector under the full ideal conditions, the covariance matrix of the least-squares residual vector is given by $\sigma^2\mathbf{M}$, where $\mathbf{M}=\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. As is well known $\mathbf{M}$ is non-diagonal, singular and in general does not have equal diagonal elements, i.e. the residuals are heteroskedastic and are not distributed independently. Anscombe and Tukey (1963) and Goldfeld and Quandt (1965) recognized this problem and overcame it with varying degrees of success by suitable definition of the functions of the least-squares residuals which were chosen as test statistics. Theil (1965) approached the problem from a different point of view by developing a set of residuals which have more desirable properties under the full ideal conditions. For $\mathbf{X}$, an $N\times K$ regressor matrix of rank K, $\mathbf{M}$ is $N\times N$ of rank $N-K$, so that the maximum number of independent residuals which can be obtained is $N-K$. One wishes, therefore, to obtain an $(N-K)\times 1$ residual vector whose distribution under the full ideal conditions is non-singular normal with null mean vector and covariance matrix given by $\sigma^2\mathbf{I}_{N-K}$. (Throughout the remainder of the paper, identity matrices will carry a suffix indicating their dimension.) Theil's residual vector, which is discussed below, fulfils the above requirements.

3.1. *The Optimal Residual Vector*

Theil (1965) considers an $N-K$ dimensional residual vector defined by

$$\tilde{\mathbf{u}}=\mathbf{A}'\mathbf{y}, \tag{24}$$

where $\mathbf{A}$ is an $N\times(N-K)$ matrix which satisfies the following conditions:

$$\left.\begin{aligned} &\text{(i)} && \mathbf{A}'\mathbf{X}=\mathbf{0};\\ &\text{(ii)} && \mathbf{A}'\mathbf{A}=\mathbf{I}_{N-K};\end{aligned}\right\} \tag{25}$$

(iii) Subject to the constraints (i) and (ii), $\mathbf{A}$ minimizes the expected value of the sum of squares of the "errors of prediction", i.e. $\mathbf{A}$ is a solution to (26):

$$\begin{aligned}\min_{\mathbf{A}} E\{\mathbf{u}'(\mathbf{A}-\mathbf{J})(\mathbf{A}-\mathbf{J})'\mathbf{u}\} &= \min_{\mathbf{A}} \sigma^2 \operatorname{tr}(\mathbf{A}-\mathbf{J})(\mathbf{A}-\mathbf{J})' \\ &= 2\sigma^2 \min_{\mathbf{A}}(N-K-\operatorname{tr}\mathbf{A}'\mathbf{J}),\end{aligned} \tag{26}$$

where $(\mathbf{A}-\mathbf{J})'\mathbf{u}$ denotes the "error in predicting" $\mathbf{u}$, $\mathbf{J}$ is an $N\times(N-K)$ matrix which is obtained by deleting K columns from $\mathbf{I}_N$. Theil calls his vector, $\tilde{\mathbf{u}}=\mathbf{A}'\mathbf{y}$, the best linear unbiased scalar covariance matrix residual vector. (Note that in Theil (1965) the vector $\tilde{\mathbf{u}}$ is called an "estimator" of $\mathbf{u}$, which is an unfortunate usage of the term. This terminology has since been corrected in Theil (1968).)

To obtain an explicit solution suppose that one has decided not to predict the first K error terms, i.e. $\mathbf{J}$ is obtained by deleting the first K columns of $\mathbf{I}_N$ (Theil, 1965, pp. 1070–1071). Partition the $\mathbf{X}$ matrix into its first K rows, $\mathbf{X}_0$, and its last $N-K$ rows, $\mathbf{X}_1$. The only restriction required at this point is that the $K\times K$ sub-matrix $\mathbf{X}_0$

be non-singular. Consider conformable partitions of the **M** matrix and of the **A** matrix:

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{00} & \mathbf{M}_{01} \\ \mathbf{M}_{10} & \mathbf{M}_{11} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \end{pmatrix}. \tag{27}$$

Let **P** be the $(N-K)\times(N-K)$ matrix such that

$$\mathbf{P}'\mathbf{M}_{11}\mathbf{P} = \mathbf{D}, \tag{28}$$

where **D** is the diagonal matrix of eigenvalues of $\mathbf{M}_{11}$. From these matrices Theil (1965, pp. 1071–1072) defines the **A** matrix by

$$\mathbf{A}_1 = \mathbf{P}\mathbf{D}^{\frac{1}{2}}\mathbf{P}', \quad \mathbf{A}_0 = -(\mathbf{X}_1\mathbf{X}_0^{-1})'\mathbf{A}_1. \tag{29}$$

It is easy to verify that $\mathbf{A}'\mathbf{X} = \mathbf{0}$, $\mathbf{A}'\mathbf{A} = \mathbf{I}_{N-K}$, that **A** satisfies the three conditions given in equations (25) and (26), and that $\mathbf{A}\mathbf{A}' = \mathbf{M}$. The distribution of $\mathbf{u}$ under the full ideal conditions is non-singular normal with null mean vector and covariance matrix $\sigma^2\mathbf{I}_{N-K}$.

3.2. *The Distribution of* $\tilde{\mathbf{u}}$ *in Mis-specified Models*

In the previous section it was shown that for some of the specification errors considered there the mean vector of the least-squares estimator was of the form $\mathbf{M}\boldsymbol{\xi}$, where $\mathbf{M} = \mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, **X** being the $N\times K$ matrix of observations on the included regressors and $\boldsymbol{\xi}$ being a non-stochastic vector (conditional on **X**), whose precise definition depends upon the nature of the mis-specification. For Theil's residuals $\tilde{\mathbf{u}} = \mathbf{A}'\mathbf{y}$, the mean vector is $\mathbf{A}'\boldsymbol{\xi}$, i.e. $\mathbf{A}'$ replaces **M** in all the results obtained in Section 2. The sum of squared residuals is exactly the same:

$$\tilde{\mathbf{u}}'\tilde{\mathbf{u}} = \mathbf{y}'\mathbf{A}\mathbf{A}'\mathbf{y} = \mathbf{y}'\mathbf{M}\mathbf{y} = \hat{\mathbf{u}}'\hat{\mathbf{u}}. \tag{30}$$

Further, the non-central parameter λ is invariant to choice of the $K\times K$ matrix $\mathbf{X}_0$ as is shown in (31):

$$\lambda = \boldsymbol{\xi}'\mathbf{A}\mathbf{A}'\boldsymbol{\xi}/\tilde{\sigma}^2 = \boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi}/\tilde{\sigma}^2, \tag{31}$$

for any choice of $\mathbf{X}_0$ subject to $\mathbf{X}_0$ being non-singular and where $\tilde{\sigma}^2$ represents the appropriate scaling factor. The exact definitions of both $\tilde{\sigma}^2$ and $\boldsymbol{\xi}$ depend upon the nature of the mis-specification being considered.

In order to see what difference the choice of $\mathbf{X}_0$ makes consider the following restatement of the problem addressed by Theil. For a given matrix **X** of order $N\times K$ one wishes to find a matrix **B** of order $(N-K)\times N$ such that

(i) $\mathbf{B}\mathbf{X} = \mathbf{0}$, i.e. the rows of **B** are contained in the row kernel of **X**;
(ii) The rows of **B** are orthogonal and normalized;
(iii) Subject to conditions (i) and (ii) the rows of **B** maximize the sum of the cosines of the angles between the rows of **B** and an $(N-K)$ dimensional subset of a Euclidean basis, E_N.

These three conditions correspond to the three conditions considered by Theil. The correspondence for the first two is obvious. For the third, note that the matrix representation of E_N is the N-dimensional identity matrix so that the matrix **J** corresponds to an $(N-K)$-dimensional subset of E_N. Thus Theil's third condition is seen to be mathematically equivalent to condition (iii) above. Partitioning **B** into its first

K columns and remaining $N-K$ columns conformably to the row partition of $\mathbf{X}$ one obtains from condition (i):

$$\mathbf{B}_1\mathbf{X}_1 = -\mathbf{B}_0\mathbf{X}_0. \tag{32}$$

Written in this way one notes that for any choice of $\mathbf{B}_0$ such that $\mathbf{B}_0\mathbf{X}_0$ is non-null one can obtain a matrix $\mathbf{B}$ satisfying conditions (i) and (ii) above. Further, one sees that a permutation of the rows of $\mathbf{X}$ makes no difference to the solution of the system (32). Thus, for a given $(N-K)$-dimensional subset of E_N, i.e. for a given choice of $\mathbf{J}$, condition (iii) ensures a unique solution for the elements of $\mathbf{B}$.

The problem of trying to obtain a unique solution stems from the arbitrariness of the choice for $\mathbf{J}$, which choice can be made in $\binom{N}{k}$ different ways. Theil does not provide a criterion for choosing the appropriate $\mathbf{J}$, or, in his formulation of the problem, the appropriate $\mathbf{X}_0$. (There are, however, a few brief comments on this problem in which Theil indicates that the choice of $\mathbf{X}_0$ may affect the power of tests based on the residual vector $\tilde{\mathbf{u}}$, (1965, pp. 1076–1077).) The notational correspondence between the two formulations is

$$\mathbf{B}_1 = \mathbf{A}_1', \quad \mathbf{B}_0 = -\mathbf{A}_1'\mathbf{X}_1\mathbf{X}_0^{-1} = \mathbf{A}_0', \tag{33}$$

As for any choice of $\mathbf{X}_0$, or, rather, for any choice of $\mathbf{J}$, $\mathbf{B}$ is an orthonormal basis for the space orthogonal to the space spanned by the columns of $\mathbf{X}$, one notes that distance and angles between vectors in the space orthogonal to $\mathbf{X}$ are invariant to the choice of $\mathbf{J}$. Indeed, the choice of $\mathbf{J}$ merely rotates the basis for the space spanned by $\mathbf{B}$. Thus, the choice as to which rows of $\mathbf{X}$ are to be included in the sub-matrix $\mathbf{X}_0$ can be easily settled. From the above discussion, one notes that one chooses $\mathbf{B}$ so as to maximize $\operatorname{tr}\mathbf{BJ}$ subject to the constraints $\mathbf{BX} = \mathbf{0}$ and $\mathbf{BB}' = \mathbf{I}_{N-K}$ for given $\mathbf{J}$. But, $\operatorname{tr}\mathbf{BJ} = \operatorname{tr}\mathbf{D}^{\frac{1}{2}}$, where one takes the positive square roots. However, $\operatorname{tr}\mathbf{D} = \operatorname{tr}\mathbf{M}_{11}$, so that the choice of $\mathbf{J}$ which maximizes $\operatorname{tr}\mathbf{BJ}$ is determined simply from the $(N-K)$ largest elements on the diagonal of $\mathbf{M}$, or equivalently the K smallest elements on the diagonal of $\mathbf{M}$ indicate the rows of $\mathbf{X}$ to be put into the sub-matrix $\mathbf{X}_0$.

Briefly stated, the procedure is as follows. Choose the K smallest elements on the diagonal of $\mathbf{M}$. The subscripts of the K smallest elements correspond to the rows of $\mathbf{X}$ to be included in $\mathbf{X}_0$. Having chosen the sub-matrices $\mathbf{X}_0$ and $\mathbf{X}_1$, one can calculate the matrix $\mathbf{A}$ by means of the equations given in (27), (28) and (29) after having permuted the rows of $\mathbf{X}$ so that $\mathbf{X}_0$ occupies the first K rows.

3.3. *Derivation of a Function to Approximate the Mean of $\tilde{u}_i$ under the Alternative Hypotheses*

The first two tests to be discussed in the next sub-section attempt to discriminate between the null hypothesis of no mis-specification and the alternative that the means of the $\tilde{u}_i$, $i = 1, 2, \ldots, N-K$, can be approximated by some function of the included regressors. It is the purpose of this sub-section to derive the appropriate function.

The basic criteria for such a function are that, first the statistics defined by the function should be distributed independently of the residuals under the null hypothesis, second the function should be useful under a wide range of alternatives with respect to the specific types of mis-specification one can make, and last it should be easily evaluated. An appropriate function which meets these criteria is that defined by $q_{1i} = \mathbf{a}_i'\hat{\mathbf{y}}^{(2)}$, $i = 1, 2, \ldots, N-K$, where $\mathbf{a}_i'$ is the ith row of $\mathbf{A}$; and $\hat{\mathbf{y}}^{(2)}$ is the $N \times 1$ vector

formed from the squares of the elements of the vector of least-squares estimates of the dependent variable.

With respect to the first criterion one can easily show that $\tilde{u}_i$ is distributed independently of $\hat{y}_i^2$, $i = 1, 2, \ldots, N$, under the null hypothesis, so that each of $\tilde{u}_i, \tilde{u}_i^2$ is distributed independently of the $(N-K)$ vector $\mathbf{q}_1 = \mathbf{A}'\hat{\mathbf{y}}^{(2)}$.

If one restricts attention to the specification errors considered in Sections 2.2.1 to 2.2.3, that is the errors of omitted variables, incorrect functional form and simultaneous equation problems, then it would seem reasonable to assume that ξ_i, the mean of the ith residual, can be expressed as a linear function of the moments about the origin of $\hat{y}_i$, the least-squares estimator of the conditional mean of y_i. Thus, one assumes that

$$\xi_i = \alpha_0 + \alpha_1 \mu_{01i} + \alpha_2 \mu_{02i} + \alpha_3 \mu_{03i} + \ldots \quad (i = 1, 2, \ldots, N), \tag{34}$$

where μ_{0ki}, $k = 1, 2, \ldots$, are the moments about the origin of $\hat{y}_i$ given $\mathbf{X}$. For example, consider the problem of an omitted variable where the $\mathbf{X}$ matrix reduces to a vector. Under the same conditions as were listed in 2.2.1, the conditional expectation of $\hat{y}_i$ is $x_i\{\beta + \gamma(\sum xz)/(\sum x^2)\}$, $i = 1, 2, \ldots, N$, where x_i is the ith observation on the regressor, z_i is the ith value of the omitted variable and β and γ are the corresponding regression coefficients. Thus, the second moment about the origin of $\hat{y}_i$ is

$$x_i^2\{\beta + \gamma(\sum xz)/(\sum x^2)\}^2 + \sigma^2,$$

where σ^2 is the variance of the error terms, so that if $z_i = x_i^2$ say, then the quantities ζ_i defined by $\zeta_i = z_i\gamma$ are equal to

$$-\gamma\sigma^2\left\{\beta + \gamma\frac{\sum xz}{\sum x^2}\right\}^{-2} + \gamma\left\{\beta + \gamma\frac{\sum xz}{\sum x^2}\right\}^{-2} E(\hat{y}_i^2 | \mathbf{x}).$$

The main point is that, provided $\boldsymbol{\zeta}$ is not orthogonal to the regressor matrix, one will in general be able to approximate the variable ζ_i in terms of a polynomial expansion of the conditional moments about the origin of $\hat{y}_i$ as illustrated.

It has been shown above that the conditional mean of $\tilde{u}_i$ under the specification errors discussed in Sections 2.2.1 to 2.2.3 is

$$E\{\tilde{u}_i | \mathbf{X}\} = \mathbf{a}_i'\boldsymbol{\xi} = \zeta_i \quad (i = 1, 2, \ldots, N-K). \tag{35}$$

Using equation (34), letting $\mathbf{i}' = (1, 1, \ldots, 1)$, and defining $\boldsymbol{\mu}_{0k}' = (\mu_{0k1}, \ldots, \mu_{0kN})$, one obtains the result that

$$\begin{aligned} \boldsymbol{\zeta} &= \mathbf{A}'[\alpha_0\mathbf{i} + \alpha_1\boldsymbol{\mu}_{01} + \alpha_2\boldsymbol{\mu}_{02} + \ldots] \\ &= \alpha_2\mathbf{A}'\boldsymbol{\mu}_{02} + \alpha_3\mathbf{A}'\boldsymbol{\mu}_{03} + \ldots, \end{aligned} \tag{35'}$$

since $\mathbf{A}'$ is orthogonal to the matrix $\mathbf{X}$. If the regression is not in deviation terms and a constant term is not included in $\mathbf{X}$, then equation (35′) becomes

$$\boldsymbol{\zeta} = \alpha_0\mathbf{A}'\mathbf{i} + \alpha_2\mathbf{A}'\boldsymbol{\mu}_{02} + \alpha_3\mathbf{A}'\boldsymbol{\mu}_{03} + \ldots. \tag{36}$$

Thus, using (35′), one has the result that

$$\boldsymbol{\zeta} = E\{\alpha_2\mathbf{A}'\hat{\mathbf{y}}^{(2)} + \alpha_3\mathbf{A}'\hat{\mathbf{y}}^{(3)} + \alpha_4\mathbf{A}'\hat{\mathbf{y}}^{(4)} + \ldots\}, \tag{37}$$

where the $N \times 1$ dimensional vectors $\hat{\mathbf{y}}^{(j)}$ are defined by $\hat{\mathbf{y}}^{(j)'} = (\hat{y}_1^j, \hat{y}_2^j, \ldots, \hat{y}_N^j)$.

Defining q_{ij} by $q_{ij} = \mathbf{a}_i'\hat{\mathbf{y}}^{(j+1)}$, $i = 1, 2, \ldots, N-K$, $j = 1, 2, \ldots$, one notes that the vector $\boldsymbol{\zeta}$ can be approximated by an equation in the expectations of the $(N-K)$

dimensional vectors $\mathbf{q}_j$. This result provides the basis for the regression test to be discussed in the next section. One notes that the regression model:

$$\tilde{\mathbf{u}} = \alpha_0 \mathbf{i} + \alpha_1 \mathbf{q}_1 + \alpha_2 \mathbf{q}_2 + \ldots + \mathbf{e}, \tag{38}$$

is an errors in the variables type of regression model. Indeed, it is the stochastic independence of $\tilde{\mathbf{u}}$ and $\hat{\mathbf{y}}$ which enables one to interpret equation (38) as an errors in the variables model, for to interpret it otherwise the regression model relative to (38) would suffer from the specification error that $\mathbf{e}$ would not be statistically independent of the regressors. Thus, although the least-squares estimators of the regression coefficients $\alpha_0, \alpha_1, \alpha_2, \ldots$ will be biased downwards, the estimator

$$\hat{\alpha}_0 \mathbf{i} + \hat{\alpha}_1 \mathbf{q}_1 + \hat{\alpha}_2 \mathbf{q}_2 + \ldots$$

will be an unbiased estimator of the conditional mean of $\tilde{\mathbf{u}}$, provided ζ_i can be approximated by a polynomial in the conditional moments about the origin of $\hat{y}_i$: see, for example, Lindley's comments on errors in the variables models (Lindley, 1947, p. 232). However, one should note that cases can arise for which one cannot express ζ_i in terms of a polynomial in the conditional moments.

If one were to assume further that the relationship between $E(\tilde{u}_i^2 | \mathbf{X})$ and $\mathbf{a}_i' \boldsymbol{\mu}_{02}$, $i = 1, 2, \ldots, N-K$, is monotonic, then one has the basis for a rank test between the ranking of the squared residuals and that of the q_{i1}, $i = 1, 2, \ldots, N-K$.

4. Discussion of the Test Procedures

The previous section discussed how to obtain a suitable residual vector. The important requirement which is met by those residuals is that the joint distribution under the null hypothesis is normal, with null mean vector and covariance matrix given by $\sigma^2 \mathbf{I}_{N-K}$. This section discusses how to use the residuals developed in the previous section in testing for specification errors.

4.1. *The Regression Specification Error Test (RESET)*

Under the null hypothesis that the full ideal conditions hold, $\tilde{\mathbf{u}}$ is distributed as normal with null mean vector and covariance matrix $\sigma^2 \mathbf{I}_{N-K}$. Under the alternative hypothesis, $\tilde{\mathbf{u}}$ is distributed as normal, but with mean vector given by $\mathbf{A}'\boldsymbol{\xi}$ and the covariance matrix by $\bar{\sigma}^2 \mathbf{I}_{N-K}$. (It should be noted that attention is being restricted here to the cases where u_i is distributed normally with non-zero mean, i.e. the cases corresponding to the errors discussed in Sections 2.2.1 to 2.2.3.) In the omitted variables case the vector $\mathbf{A}'\boldsymbol{\xi}$ can be regarded as the net contribution of the omitted variable to the conditional mean of the dependent variable, where "net" means that one is considering the effect of the omitted variable given the other regressors. Thus, under the alternative hypothesis, one may write $\tilde{\mathbf{u}}$ as

$$\tilde{\mathbf{u}} = \mathbf{A}'\boldsymbol{\xi} + \mathbf{e}, \tag{39}$$

where the $(N-K) \times 1$ vector $\mathbf{e}$ is normally distributed with null mean vector.

The model formulated in equation (39) suggests the possibility of testing for specification error under these assumptions by regressing $\tilde{\mathbf{u}}$ on $\boldsymbol{\zeta} = \mathbf{A}'\boldsymbol{\xi}$ and performing an F test on the estimated regression. However, $\boldsymbol{\xi}$ is unknown. Instead, as shown in the last section, one may, under the assumptions discussed, approximate the regressor vector $\boldsymbol{\zeta}$ by a linear sum of the vectors $\mathbf{q}_j$, $j = 1, 2, \ldots$. Thus, one considers the regression:

$$\tilde{u}_i = \alpha_0 + \alpha_1 q_{i1} + \alpha_2 q_{i2} + \ldots + e_i \quad (i = 1, 2, \ldots, N-K), \tag{40}$$

where e_i is assumed to be normal with zero mean, and one applies an F test on the joint significance of the α's. How many vectors $\mathbf{q}_j$ are needed will depend upon the particular circumstances so that no general rule can be specified. The author has found, however, that using $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$ has been sufficient.

Under the null hypothesis the regression coefficient estimators have zero expected values and the covariance matrix of $\mathbf{e}$ is $\sigma^2\mathbf{I}_{N-K}$. Under the alternative hypothesis that the mean vector of $\tilde{\mathbf{u}}$ is $\mathbf{A}'\boldsymbol{\xi}$, one or more of the coefficient estimates will have non-zero means, i.e. the multiple correlation coefficient is non-zero. The covariance matrix of $\mathbf{e}$ under the alternative hypotheses considered in Sections 2.2.1 to 2.2.3 will be given by $\bar{\sigma}^2\mathbf{I}_{N-K}$, where $\bar{\sigma}^2$ is the variance of $\tilde{u}_i$ about its conditional mean.

If the alternative hypothesis is that $\tilde{\mathbf{u}}$ is distributed as $N(\mathbf{0},\boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is a diagonal matrix with unequal elements, then one cannot use the regression test to discriminate against such an alternative. Further, if the alternative hypothesis is that the distribution of $\tilde{\mathbf{u}}$ is not normal then the regression test is no longer well defined.

4.2. *The Rank Specification Error Test (RASET)*

The following tests all use the squared residuals as independent estimates of the error variance. Under the null hypothesis, $\tilde{u}_i^2$, $i = 1, 2, \ldots, N-K$, are distributed as $\sigma^2\chi_1^2$, whereas under the first three alternative hypotheses, which were discussed in Sections 2.2.1 to 2.2.3, $\tilde{u}_1^2/\bar{\sigma}^2$, $i = 1, 2, \ldots, N-K$, are distributed as $\chi^2(1, \lambda_i)$, where $\bar{\sigma}^2$ is the variance of $\tilde{u}_i$ about its conditional mean $\mathbf{a}_i'\boldsymbol{\xi}$, $\lambda_i = \boldsymbol{\xi}'\mathbf{a}_i\mathbf{a}_i'\boldsymbol{\xi}/\bar{\sigma}^2$. One notes that the larger λ_i, $i = 1, 2, \ldots, N-K$, the greater the power of tests discriminating between central and non-central χ^2 distributions. Thus, it is the size of $\boldsymbol{\xi}'\mathbf{a}_i\mathbf{a}_i'\boldsymbol{\xi}$ *relative* to $\bar{\sigma}^2$ which is important. It can be shown that for the case of an omitted variable, $\sum_{i=1}^{N-K}\lambda_i = E\{\Delta\text{RSS}\}/\bar{\sigma}^2 - 1$, where ΔRSS is the extra regression sum of squares due to the omitted variable and $\bar{\sigma}^2$ is the variance of the error terms.

The previous test considered regressing $\tilde{u}_i$ on the vectors $\{\mathbf{q}_j\}$, $j = 1, 2, \ldots$. If, however, one can assume that the second moment of $\tilde{u}_i$, i.e. $\boldsymbol{\xi}'\mathbf{a}_i\mathbf{a}_i'\boldsymbol{\xi} + \bar{\sigma}^2$ varies monotonically with q_{i1}, $i = 1, 2, \ldots, N-K$, where $q_{i1} = \mathbf{a}_i'\hat{\mathbf{y}}^{(2)}$, then one can consider the relationship between monotonic transformations of $\tilde{u}_i^2$ and q_{i1}. Consequently, one is led to consider the relationship between the rankings of the $\tilde{u}_i^2$ and of the q_{i1}. One should note that $\tilde{u}_i^2$ is an estimate of $\boldsymbol{\xi}'\mathbf{a}_i\mathbf{a}_i'\boldsymbol{\xi} + \bar{\sigma}^2$ so that the ranking of $\tilde{u}_i^2$, $i = 1, 2, \ldots, N-K$ is an estimate of the ranking of $\boldsymbol{\xi}'\mathbf{a}_i\mathbf{a}_i'\boldsymbol{\xi}$.

First, one rearranges the q_{i1} into ascending order permuting the elements of the vector $\tilde{\mathbf{u}}$ conformably and one transforms $q_{i1} \to j$, $j = 1, 2, \ldots, N-K$, where j denotes the index of the jth largest q_{i1}. One now assigns to each $\tilde{u}_i^2$ an integer r_i such that the value of r_i indicates that $\tilde{u}_i^2$ is the rth largest variance estimate. In short, the set of $N-K$ numbers $\{r_i\}$ is a permutation of the integers $1, 2, 3, \ldots, N-K$. The test statistic to be used is Spearman's rank correlation coefficient, R_s, which is defined by

$$R_s = 1 - \frac{6}{(N-K)\,[(N-K)^2-1]} \sum_{i=1}^{N-K} (r_i - i)^2. \tag{41}$$

For values of $(N-K) \leqslant 10$, tables are readily available. For larger values one uses the asymptotic result that the statistic t_R defined by

$$t_R = \left(\frac{(N-K-2)\,R_s^2}{(1-R_s^2)}\right)^{\frac{1}{2}} \tag{42}$$

is distributed as "Student's" t with $(N-K-2)$ degrees of freedom under the null hypothesis. Kendall and Stuart (1961, Vol. 2, p. 477) point out that even for $(N-K)$ as small as 10 the approximation is quite good.

Under the null hypothesis $\tilde{u}_i^2$ is distributed independently of q_{i1}, so that the expected value of R_s is zero and its variance is $(N-K-2)^{-1}$. Under the alternative hypothesis that $\tilde{u}_i^2$ is distributed as $\bar{\sigma}^2\chi^2(1, \lambda_i)$, $i = 1, 2, \ldots, N-K$, the power of the test depends upon the assumption of a monotonic relationship between the second moment of $\tilde{u}_i$ and q_{i1}.

Spearman's rank correlation coefficient was chosen in preference to other possible rank tests of independence on two grounds. First, the test is simple. Second, as Stuart (1956, p. 287) has shown, Spearman's coefficient has maximum asymptotic relative efficiency when compared with a number of other randomness tests against normal alternatives. Stuart (1954) also considered the alternative of calculating the linear regression between the rankings which test had an asymptotic relative efficiency of 1 compared to the 0·98 for Spearman's coefficient. However, the rank test would seem to be more robust to non-normality.

One notes that Spearman's coefficient does not depend upon the assumption of normality. If the alternative hypothesis is that of simple heteroskedasticity, the second moment of $\tilde{u}_i$ is σ_i^2, $i = 1, 2, \ldots, N-K$, and $\tilde{u}_i^2$ is distributed as $\sigma_i^2\chi_1^2$, $i = 1, 2, \ldots, N-K$. If σ_i^2 varies monotonically with q_{i1} over $i = 1, 2, \ldots, N-K$, then the rank test is applicable.

4.3. *The Kolmogorov Specification Error Test (KOMSET)*

One of the disadvantages of testing on the squared residuals $\tilde{u}_i^2$ is that their distribution depends upon the unknown scale factor $\bar{\sigma}^2$. The rank test is not affected by lack of knowledge of $\bar{\sigma}^2$ as the test is invariant to scale changes. An alternative approach is to consider ratios of the squared residuals so as to eliminate the effect of the common scale factor. One therefore considers the distribution of $w_i = \tilde{u}_j^2/\tilde{u}_k^2$, $j \neq k$, where the j, k are chosen so that the w_i are independent.

4.3.1. *The distribution of ratios of $\tilde{u}_i^2$ under the null and alternative hypotheses*

Under the null hypothesis that the full ideal conditions hold,

$$\tilde{u}_j^2/\bar{\sigma}^2, \quad j = 1, 2, \ldots, N-K,$$

is distributed as central χ^2 with one degree of freedom. For the moment assume $N-K$ is even, so that one has $l = (N-K)/2$ statistics w_i. Thus, the w_i, $i = 1, 2, \ldots, l$, form a set of l independent variates from the central F distribution with $(1, 1)$ degrees of freedom. Under the first three alternative hypotheses considered in Sections 2.2.1 to 2.2.3, $\tilde{u}_j^2/\bar{\sigma}^2$ is distributed as non-central χ^2 with one degree of freedom, so that w_i, $i = 1, 2, \ldots, l$, form a set of l independent variates from double non-central F distributions with $(1, 1)$ degrees of freedom and non-central parameters λ_1, λ_2, where $\lambda_1 = (\boldsymbol{\xi}'\mathbf{A}_1\mathbf{A}_1\boldsymbol{\xi})/\bar{\sigma}^2$ and $\lambda_2 = (\boldsymbol{\xi}'\mathbf{A}_2\mathbf{A}_2'\boldsymbol{\xi})/\bar{\sigma}^2$. For the purposes of the analysis to follow it is convenient to approximate the non-central distributions with central distributions.

4.3.2. *Approximation of the non-central F by central F through reparametrization*

The relationship between the non-central and central χ^2 distributions is easily achieved by equating the first two moments of both distributions. The

approximation is, in fact, quite good (Kendall and Stuart, 1961, Vol. 2, p. 229). The correspondence is

$$\chi^2(\nu, \lambda) \simeq a\chi^2_b, \tag{43}$$

where $a = (\nu+2\lambda)/(\nu+\lambda)$, $b = (\nu+\lambda)^2/(\nu+2\lambda)$, ν is the degrees of freedom and λ is the non-central parameter. Thus, letting $g_1 = \hat{u}_1^2/\bar{\sigma}^2$, $g_2 = \hat{u}_2^2/\bar{\sigma}^2$, one sees that g_1 is distributed as $a_1\chi^2_{b_1}$ and g_2 as $a_2\chi^2_{b_2}$, or that g_1/a_1 is distributed as $\chi^2_{b_1}$ and g_2/a_2 as $\chi^2_{b_2}$. If one defines z by $a_2 g_1/a_1 g_2$, then under the alternative hypothesis that $\hat{u}_i^2$ is distributed as $\bar{\sigma}^2\chi^2(1, \lambda_i)$, $i = 1, 2, \ldots, N-K$, one sees that z is distributed as F with b_1 and b_2 degrees of freedom. However, the a_i are unknown and in general not equal so one must consider instead the distribution of $w = g_1/g_2$. The distribution of w is obtained from that of z by transformation of variables with a Jacobian of a_2/a_1. Thus, one has

$$f_1(w) = (a_2/a_1)F_{b_1,b_2}, \tag{44}$$

a scaled central F with (b_1, b_2) degrees of freedom. Compare this with the distribution of w under the null hypothesis which is central F with $(1, 1)$ degrees of freedom:

$$f_0(w) = \frac{w^{-\frac{1}{2}}(1+w)^{-1}}{B(\frac{1}{2}, \frac{1}{2})} = F_{1,1}. \tag{45}$$

Some consideration must be given to the effect of the scaling factor a_2/a_1. As shown in (46), the scaling factor is a function of the degrees of freedom and of the non-central parameters of the component χ^2 variables:

$$\frac{a_2}{a_1} = \frac{(\nu_2+2\lambda_2)(\nu_1+\lambda_1)}{(\nu_2+\lambda_2)(\nu_1+2\lambda_1)}. \tag{46}$$

Obviously, if $\lambda_1 = \lambda_2$ and $\nu_1 = \nu_2 = 1$, the factor reduces to 1. The effect of the ratio appears, therefore, only when $\lambda_1 \neq \lambda_2$. Thus, one notes that

$$\lim_{\substack{\lambda_2\to\infty \\ \lambda_1\to 0}} (a_2/a_1) = 2 \quad \text{and} \quad \lim_{\substack{\lambda_1\to\infty \\ \lambda_2\to 0}} (a_2/a_1) = \tfrac{1}{2},$$

which are the maximum and minimum values respectively for the scale factor. Finally, note that $b_i = (1+\lambda_i)^2/(1+2\lambda_i) \geqslant 1$, the equality holding only if λ_i is zero.

Looking at the transformed distributions, one notes that under the null hypothesis there is no change, w is distributed as central F with $(1, 1)$ degrees of freedom. Under the alternative hypothesis one has w distributed as scaled central F with (b_1, b_2) degrees of freedom. To illustrate the analysis above consider Fig. 1a which shows the density functions of F under the null and alternative hypotheses. The diagrams and the above analysis suggest that a suitable test would be a non-parametric test on the cumulative distribution function.

4.3.3. *The Kolmogorov statistic*

Let Λ represent the scaling factor, so that under the alternative hypothesis w is distributed as $\Lambda F_{b_1,b_2}$. One considers the cumulative distribution functions, G_0, G_1, under the null and alternative hypotheses respectively, defined by

$$G_0(W) = \int_0^W F_{1,1}(w)\,dw, \quad G_1(W) = \int_0^W \Lambda F_{b_1,b_2}(\Lambda w)\,dw. \tag{47}$$

The Kolmogorov test is based on the sample cumulative distribution function $S_n(x)$ defined by

$$S_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ r/n, & x_{(r)} \leqslant x < x_{(r+1)}, \\ 1, & x_{(n)} \leqslant x, \end{cases} \tag{48}$$

where the $x_{(i)}$ are the order statistics. $S_n(x)$ is simply the proportion of observations

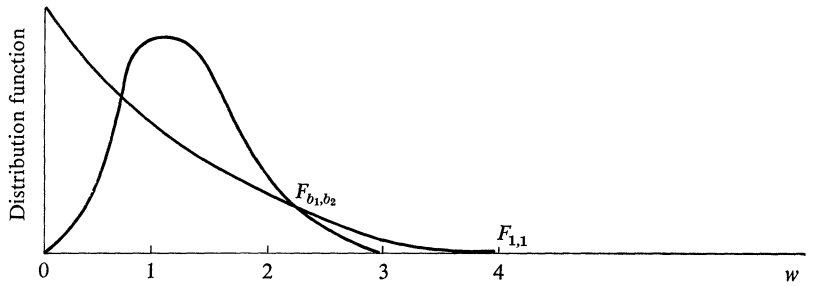

FIG. 1a. The density functions, $F_{1,1}$ and F_{b_1,b_2}, of the statistic w under the null and alternative hypotheses respectively.

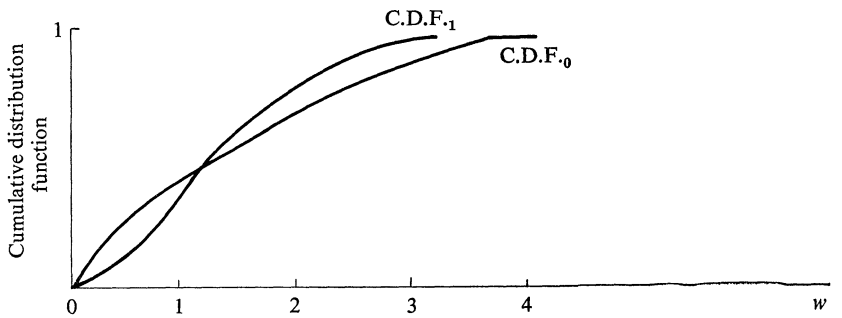

FIG. 1b. The cumulative distribution functions, G_0 and G_1, of the statistic w under the null and alternative hypotheses respectively.

not exceeding x. If $G_0(x)$ is the true cumulative distribution function, then by the Strong Law of Large Numbers,

$$\lim_{n\to\infty} \Pr\{S_n(x) = G_0(x)\} = 1. \tag{49}$$

The Kolmogorov statistic, D_n, is defined by

$$D_n = \sup_x |S_n(x) - G_0(x)|. \tag{50}$$

The distribution of D_n has been calculated exactly and listed in tables for n up to 100 (Birnbaum, 1952). For large n the asymptotic distribution has been obtained and tabulated (Smirnov, 1948). In the simple hypothesis testing case the distribution of D_n is completely distribution free under the null hypothesis, the distribution depending only on n. Further, whatever the true cumulative distribution function $G(x)$, one has for a size α test and critical value d_α:

$$\mathrm{P}\{D_n = \sup_x |S_n(x) - G(x)| > d_\alpha\} = \alpha. \tag{51}$$

Birnbaum (1952) has listed values of d_α for various α levels and n from 1 to 100. The probability statement in (51) can be inverted to yield a confidence interval for the *whole* distribution, thus:

$$\mathrm{P}\{S_n(x) - d_\alpha \leqslant G(x) \leqslant S_n(x) + d_\alpha, \text{ for all } x:\} = 1 - \alpha. \tag{52}$$

Indeed, (52) states one of the most useful properties of the Kolmogorov statistic.

For any n and α size of test, lower and upper bounds can be calculated at each i, $i = 1, 2, \ldots, n$, where i refers to the order of the order statistics. Define:

$$L(i) = \max(0, i/n - d_\alpha), \quad U(i) = \min(1, i/n + d_\alpha). \tag{53}$$

A very useful way of carrying out the test is a graphical method. This procedure is particularly useful if the calculations and graphing are all performed by a computer. One plots $L(i)$, $U(i)$ and $G_0(x)$ for $x \in [x_{(i)}, x_{(i+1)}]$ against $x_{(i)}$. If G_0, the null hypothesis, is to be accepted, the entire graph of G_0 must be contained within the contours plotted by $L(i)$ and $U(i)$. Tables are available for evaluating $G_0(x_{(i)})$ for the case under consideration in this paper. Alternatively, one may use the tables of the incomplete beta distribution and transform to the F.

4.3.4. *Discussion of the appropriate way of choosing the ratios of* $\tilde{u}_i^2$

Two points need to be settled before the Kolmogorov statistic can be applied in practice. First, it has been assumed that the optimal choice of null hypothesis was $F_{1,1}$. This point must be demonstrated. Second, one must settle the question of optimal choice of how to take the ratios of the $\tilde{u}_i^2$ to obtain the variables w distributed as F.

Obviously one wants as many estimates w_i as is possible in order to maximize the power of the Kolmogorov test. This requirement would indicate the use of $F_{1,1}$ as the null hypothesis. Further, the distinction between the distributions under the two hypotheses depends upon the difference between ν_i, the degrees of freedom under the null hypothesis, and $b_i = (\nu_i + \lambda_i)^2/(\nu_i + 2\lambda_i)$, i.e. the greater $(b_i - \nu_i)$ the more powerful is the test. As $(b_i - \nu_i) = \lambda_i^2/(\nu_i + 2\lambda_i)$, one sees immediately that the optimal degrees of freedom under the alternative hypothesis are $(1, 1)$. One, therefore, restricts attention to $w_i = \tilde{u}_j^2/\tilde{u}_k^2$, $i = 1, 2, \ldots, l$.

The choice of j and k can now be settled. The only requirement is that the w_i be independent under the null hypothesis so that no two w_i, $i = 1, 2, \ldots, l$, should have a common $\tilde{u}_j^2$, $j = 1, 2, \ldots, N - K$. Thus, a workable solution to the problem of how to calculate the w_i, $i = 1, 2, \ldots, l$, is given by taking ratios in pairs to give the sequence:

$$w_1 = \tilde{u}_2^2/\tilde{u}_1^2, \quad w_2 = \tilde{u}_4^2/\tilde{u}_3^2, \quad \ldots, \quad w_l = \tilde{u}_{2l}^2/\tilde{u}_{2l-1}^2, \tag{54}$$

where $\tilde{u}_j^2$ is the jth observed residual. As the residuals are distributed independently of the order of observation, then the variables w_i, $i = 1, 2, \ldots, l$, are distributed independently of the observed order.

4.3.5. *Comparison of the Kolmogorov test to some alternative non-parametric tests*

The advantages of the Kolmogorov test over other non-parametric tests, e.g. Pearson's χ^2 test, the Neyman–Barton "smooth" test and the Smirnov–von Mises test, are that D_n uses ungrouped data, its distribution is known for finite sample sizes and it provides a confidence interval for the true cumulative distribution. With respect to the use of Pearson's χ^2 test the work of Williams (1950) and of Massey (1951) show that the Kolmogorov test can detect differences between true and hypothesized cumulative distributions about one-half the size detectable by Pearson's χ^2 test. Further, as sample size increases D_n becomes even more efficient relative to the χ^2 test.

Another advantage is that the Kolmogorov test as applied above is robust to non-normality. If the original error terms are not normal then the distribution of w is no longer F. A final point is to look at the behaviour of the statistic D_n if the alternative hypothesis is simple heteroskedasticity. The distribution of w under the alternative hypothesis is now a scaled central F with (1, 1) degrees of freedom, where the effect of the scaling factor can be inferred from equation (44) by substituting σ_1^2/σ_2^2 for a_1/a_2. It would appear that this test is not very sensitive to such a mis-specification.

4.4. *Bartlett's M Specification Error Test (BAMSET)*

It has been seen that all of the tests proposed so far do not seem to be very sensitive to simple heteroskedasticity. Such a situation is at least potentially an advantage, for one would like to be able to discriminate between specification errors.

However, if the alternative hypothesis is that $\tilde{u}_i^2$, $i = 1, 2, \ldots, N-K$, are distributed as $\sigma_i^2 \chi_1^2$, one can consider as a possible test of heterogeneity, a modification of Bartlett's M test for the non-equality of variances (Kendall and Stuart, 1961, Vol. 2, pp. 234–236). The residuals $\tilde{u}_i^2$, $i = 1, 2, \ldots, N-K$, provide one with $N-K$ independent estimates of the variances σ_i^2, $i = 1, 2, \ldots, N-K$. Bartlett's M test is a modification of Pearson's likelihood-ratio test for heterogeneity. Under the null hypothesis, H_0, the σ_i^2 are all equal and under the alternative hypothesis H_1, the σ_i^2 are all different. Pearson, using the likelihood-ratio method, derived the statistic l defined by

$$l = \prod_{i=1}^{k} (s_i^2/s^2)^{n_i/2}, \tag{55}$$

where $s_i^2 = n_i^{-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$, $s^2 = n^{-1} \sum_{i=1}^{k} n_i s_i^2$, $n = \sum_{i=1}^{k} n_i$. Pearson assumed that the observations x_{ij} were drawings from k independent samples of size n_i each and that each sample was drawn from a normal population with mean μ_i and variance σ_i^2. Bartlett's modification was to replace the n_i by the corresponding degrees of freedom $\nu_i = (n_i - 1)$ and where $\nu = \sum_{i=1}^{k} \nu_i$ replaces n. Thus, one has l^* defined

$$l^* = \prod_{i=1}^{k} (s_i^2/s^2)^{\nu_i/2}. \tag{56}$$

The advantage of l^* over l is that for any n_i one has an unbiased test (Kendall and Stuart, 1961, Vol. 2, pp. 244–245). By transforming l^* to

$$-2 \log l^* = \nu \log s^2 - \sum_{i=1}^{k} \nu_i \log s_i^2, \tag{57}$$

one has a statistic which is asymptotically distributed as central χ^2 with $k-1$ degrees of freedom. The approximation is improved if one divides (57) by the scaling constant:

$$1+\tfrac{1}{3}(k-1)^{-1}\sum_{i=1}^{k}\left(\frac{1}{\nu_i}-\frac{1}{\nu}\right). \tag{58}$$

The analogue to the regression case would appear to obvious; $\tilde{u}_i^2$ corresponds to s_i^2, $(N-K)^{-1}\sum_1^{N-K}\tilde{u}_i^2$ corresponds to s^2, $\nu_i = 1$, and $\sum_1^k \nu_i = \nu = N-K$, where $k = N-K$. However, defining l^* in this way, while formally correct, is not very useful as $-2\log l^*$ will not be asymptotically distributed as χ^2 with $(k-1)$ degrees of freedom under the null hypothesis. This is because the smallest value of ν_i for which the asymptotic result will hold is 2 as may be easily verified. However, the following modification enables one to obtain a statistic which is asymptotically distributed as χ^2. Let

$$s_i^2 = \nu_i^{-1}\sum_{j=1}^{\nu_i}\tilde{u}_j^2, \quad i = 1,2,3;$$

$$s^2 = \nu^{-1}\sum_{i=1}^{k}\nu_i s_i^2 = (N-K)^{-1}\sum_{j=1}^{N-K}\tilde{u}_j^2; \quad \sum_{i=1}^{k}\nu_i = \nu = N-K; \quad k = 3.$$

In short, one divides the set of $(N-K)$ residuals into three non-intersecting sub-sets in order to form the statistics s_i^2, $i = 1,2,3$.

The distribution of $-2\log l^*$ defined in (57) with respect to s_i^2, s^2, ν_i, ν and k is, under the null hypothesis, asymptotically distributed as central χ^2 with two degrees of freedom. The null hypothesis, one recalls, is that $\tilde{u}_i^2$ is distributed as

$$\sigma^2\chi_1^2, \quad i = 1,2,\ldots,N-K.$$

Under the alternative hypothesis that $\tilde{u}_i^2$ is distributed as $\sigma_i^2\chi_1^2$, the $(N-K)$ σ_i^2 not being all equal, the distributions of s_i^2 and s^2 are weighted sums of central χ^2 distributions. In the original statement of the problem by Pearson and Bartlett only s^2 is distributed as a weighted sum of χ^2 distributions under the alternative hypothesis.

The value of k was chosen to be 3 for the following reason. The larger the values of ν_i, $i = 1,2,\ldots,k$, the closer is the approximation between the actual distribution of $-2\log l^*$ and χ_2^2 under the null hypothesis. On the other hand, in order to test for heterogeneity of variance, one needs several variance estimators. $k = 3$ was chosen as a useful compromise. In place of additional information about the distributions of $\tilde{u}_i^2, i = 1,2,\ldots,N-K$, the sub-sets are most easily chosen by setting

$$\left.\begin{aligned} s_1^2 &= (1/r_1)\sum_{i=1}^{r_1}\tilde{u}_i^2, \quad r_1 = [(N-K)/3], \\ s_2^2 &= (1/r_1)\sum_{i=r_1+1}^{r_2}\tilde{u}_i^2, \quad r_2 = 2r_1, \\ s_3^2 &= (1/r_3)\sum_{i=r_2+1}^{N-K}\tilde{u}_i^2, \quad r_3 = N-K-2r_1, \end{aligned}\right\} \tag{59}$$

where r_1 is the integer obtained from the integer division of $(N-K)$ by 3.

Bartlett's M test is obviously not defined for the cases considered in Sections 2.2.1 to 2.2.3, i.e. when the $\tilde{u}_i^2$ are distributed as non-central χ^2. However, the test is sensitive to the alternative hypothesis of simple heteroskedasticity. The chief disadvantage of the test is that it is not robust to non-normality.

4.5. *The Relative Merits of the Four Test Procedures*

The above few pages have considered the use of four tests, the regression specification error test (RESET), the rank specification error test (RASET), the Kolmogorov specification error test (KOMSET) and a modification of Bartlett's M specification error test (BAMSET). The first three were considered as tests for the first three specification errors considered in Section 2 and the last is a test for simple heteroskedasticity.

The four tests are to a considerable extent complementary rather than substitutes for each other. RESET is able to detect cases where RASET's power is reduced, i.e. where the regression of $\hat{u}_i^2$ on q_{i1} is non-monotonic. The power of RESET, however, depends upon the full ideal conditions holding for the regression of $\hat{u}_i$ on a linear sum of the vectors $\mathbf{q}_j, j = 1, 2, \ldots$. KOMSET'S power, on the other hand, is not reduced by non-monotonicity of the relationship between $\hat{u}_i^2$ and q_{i1}, nor does it need the assumptions required by RESET. BAMSET, of course, is not designed to detect specification errors which yield residuals distributed as non-central χ^2.

If the alternative hypothesis is that of non-normality of the error terms, then RESET and BAMSET are no longer well defined. RASET would seem to be insensitive to such a mis-specification. However, KOMSET is still well defined under these circumstances and with sufficient degrees of freedom should be able to detect the presence of such a mis-specification.

Moving to the third alternative, that is, the problem of heteroskedasticity, it has been shown that RESET is completely insensitive to such a departure from the full ideal conditions. RASET and KOMSET can discriminate against such alternatives in certain circumstances, but are presumably not as efficient as BAMSET which is a parametric test designed to detect this very problem.

If not one, but several, mis-specifications are made, the situation is not so clear. On the one hand, if the mis-specifications reinforce each other the tests become even more powerful. However, if they offset each other the tests will tend to accept the null hypothesis. If the net effect on the distribution of the residuals is to leave them normal with zero mean, then the tests discussed above, RASET, RESET and KOMSET, will not detect any specification error. However, BAMSET may indicate the presence of heteroskedasticity.

5. RESEARCH IN PROGRESS

With respect to the further research which needs to be done and which is in fact under way, the most important problem to be handled is the application of Monte Carlo techniques to examine the small sample properties of the four tests.† In addition, the power functions of the tests are to be examined and some attention is to be paid to the asymptotic relative efficiency of the tests in certain cases.

On the theoretical side two major areas of examination are: first, the cases where mis-specification leads to non-normality of the error terms in the mis-specified model, and, second, the problems raised by auto-correlation of the error terms. If the tests are robust to auto-correlation the usefulness of the procedures described in this paper will be greatly enhanced as much econometric work is concerned with time-series data.

† A computer program has been designed to carry out the entire set of test procedures. This program has been run successfully on a Control Data Corporation 3600 computer. The results obtained from some thirty runs with small sample problems drawn from the economic literature have been encouraging. Readers interested in obtaining a copy of the Fortran source deck with accompanying write-up should send their request to the author.

ACKNOWLEDGEMENT

The author would like to thank Arnold Zellner, Jan Kmenta and Harold Watts for their cogent criticisms of earlier drafts of this paper. Further, the author is most grateful to the referee for many comments which improved the paper considerably.

REFERENCES AND OTHER RELATED PUBLICATIONS

ANSCOMBE, F. J. (1961). Examination of residuals. *Proc. Fourth Berkeley Symp. Math. Statist. and Prob.*, **1**, 1–36.

ANSCOMBE, F. J. and TUKEY, J. W. (1963). The examination and analysis of residuals. *Technometrics*, **5**, 141–160.

BANCROFT, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Statist.*, **15**, 190–204.

BARTLETT, M. S. (1949). Fitting a straight line when both variables are subject to error. *Biometrics*, **5**, 207–212.

BIRNBAUM, Z. W. (1952). Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. *J. Amer. Statist. Ass.*, **47**, 425–441.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations. *J. R. Statist. Soc.* B, **26**, 211–243.

BOX, G. E. P. and TIAO, G. C. (1964). A note on criterion robustness and inference robustness. *Biometrika*, **51**, 169–173.

BOX, G. E. P. and TIDWELL, P. W. (1962). Transformation of the independent variables. *Technometrics*, **4**, 531–550.

CHRIST, C. F. (1960). Simultaneous equation estimation: any verdict yet? *Econometrica*, **28**, 835–845.

DURBIN, J. (1954). Errors in variables. *Rev. Int. Statist. Inst.*, **22**, 23–32.

DURBIN, J. and WATSON, G. S. (1950). Testing for serial correlation in least squares regression, I. *Biometrika*, **37**, 409–428.

—— (1951). Testing for serial correlation in least squares regression, II. *Biometrika*, **38**, 159–178.

FISHER, F. M. (1961). On the cost of approximate specification in simultaneous equation estimation. *Econometrica*, **29**, 139–170.

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

GOLDFELD, S. M. and QUANDT, R. E. (1965). Some tests for homoscedasticity. *J. Amer. Statist. Ass.*, **60**, 539–547.

GRILICHES, Z. (1957). Specification bias in estimates of production functions. *Farm Economics*, **39**, 8–20.

HAAVELMO, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, **11**, 1–12.

HILDRETH, C. (1960). Simultaneous equations: any verdict yet? *Econometrica*, **28**, 846–854.

KAC, M., KIEFER, J. and WOLFOWITZ, J. (1955). On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist.*, **26**, 189–211.

KENDALL, M. G. (1951). Regression, structure and functional relationships, Part I. *Biometrika*, **38**, 11–25.

—— (1952). Regression, structure and functional relationships, Part II. *Biometrika*, **39**, 96–108.

KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics*, Vols. 1, 2. London: Griffin.

KUH, E. and MEYER, J. R. (1955). Correlation and regression estimates when the data are ratios. *Econometrica*, **23**, 400–416.

LARSON, J. J. and BANCROFT, T. A. (1963). Biases in prediction by regression for certain incompletely specified models. *Biometrika*, **50**, 391–402.

—— (1963). Sequential model building for prediction in regression analysis, I. *Ann. Math. Statist.*, **34**, 462–479.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. New York: Wiley.

LINDLEY, D. V. (1947). Regression lines and the linear functional relationship. *J. R. Statist. Soc. Suppl.* **9**, 218–244.

LIU, T. C. (1960). Underidentification, structural estimation, and forecasting. *Econometrica*, **28**, 855–865.

MADANSKY, A. (1959). The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Ass.*, **54**, 173–205.

MASSEY, F. J., Jr (1951). The Kolmogorov–Smirnov test of goodness of fit. *J. Amer. Statist. Ass.*, **46**, 68–78.
NAIR, K. R. and SHRIVASTAVA, M. P. (1942). On a simple method of curve fitting. *Sankhyā*, **6**, 121–132.
PRAIS, S. J. and HOUTHAKKER, H. S. (1955). *The Analysis of Family Budgets.* Cambridge: University Press.
PEARSON, E. S. and HARTLEY, H. O. (1962). *Biometrika Tables for Statisticians*, Vol. 1. Cambridge: University Press.
RAFF, M. S. (1956). On approximating the point binomial. *J. Amer. Statist. Ass.*, **51**, 293–303.
RAMSEY, J. B. (1965). An expository note on errors in digital computers: their effect on the results of econometric analysis. *Systems Formulation, Methodology, and Policy Workshop*, Paper No. 6515. University of Wisconsin, Madison.
RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research.* New York: Wiley.
SMIRNOV, N. V. (1948). Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.*, **19**, 279–281.
STUART, A. (1954). Asymptotic relative efficiencies of distribution-free tests of randomness against normal alternatives. *J. Amer. Statist. Ass.*, **49**, 147–157.
—— (1956). The efficiencies of tests of randomness against normal alternatives. *J. Amer. Statist. Ass.*, **51**, 285–287.
THEIL, H. (1961). *Economic Forecasts and Policy*, 2nd ed. Amsterdam: North Holland.
—— (1965). The analysis of disturbances in regression analysis. *J. Amer. Statist. Ass.*, **60**, 1067–1079.
—— (1968). A simplification of the BLUS procedure for analysing regression disturbances. *J. Amer. Statist. Ass.*, **63**, 242–251.
THEIL, H. and GOLDBERGER, A. S. (1961). On pure and mixed statistical estimation in economics. *Int. Econ. Rev.*, **2**, 65–78.
TUKEY, J. W. (1959). One degree of freedom for non-additivity. *Biometrics*, **5**, 232–242.
WALD, A. (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Statist.*, **11**, 284–300.
WILLIAMS, C. A., Jr (1950). On the choice of the number and width of classes for the chi-square goodness of fit. *J. Amer. Statist. Ass.*, **45**, 77–86.
ZELLNER, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Ass.*, **57**, 348–368.
ZELLNER, A. and THEIL, H. (1962). Three-stage least squares: simultaneous estimation of simultaneous equations. *Econometrica*, **30**, 54–78.