MICHAEL A. MAGDALINOS - SPYRIDON D. SYMEONIDES

ALTERNATIVE METHODS OF SIZE CORRECTION FOR SOME GLS TEST STATISTICS: THE CASE OF HETEROSCEDASTICITY

ABSTRACT

In this paper we consider the comparison of alternative size corrections of the t and F tests in the normal linear model with unknown error covariance matrix. Rothenberg (1984b) has derived corrected critical values for the F test from a chi-square Edgeworth approximation. Similar corrected critical values for the t test can be derived from a normal Edgeworth approximation. Alternative critical values can be obtained by using Edgeworth approximations based on the F or t distributions respectively. These corrections are locally exact, i.e. they reduce to the exact critical values when the error covariance matrix is known up to a multiplicative factor. Moreover, instead of correcting the critical values, we may use a Cornish-Fisher correction of the test statistic. Thus we avoid the problem of negative "probabilities" in the tails of an Edgeworth "distribution". The relative performance of these corrections is examined in the linear regression model with heteroscedastic errors. A simulation study supports the theoretical considerations in favour of the locally exact Cornish-Fisher corrections. Due to their moderate computational requirements and the simplicity of their use, the Cornish-Fisher corrections can be a useful tool in applied statistical and econometric work.

This research was supported by Center of Economic Research Grant E104. We wish to thank Professors L. Gills, E. Panas, and G. Phillips for their helpful comments. They are not responsible for any remaining errors or shortcomings.

PLAN

1. INTRODUCTION

When the normal linear regression model with nonscalar error covariance matrix is estimated by generalized least squares (GLS), any linear hypothesis is usually tested by the conventional F (or t) test, with the implicit assumption that the sample size is large enough to permit the chi-square (or normal) approximation. In finite samples, however, there is considerable discrepancy between the actual and the nominal size which night lead to erroneous inferences and to incorrect structural specification. Also, the well-known conflict among the classical testing procedures is mainly due to the fact that the Wald, likelihood ratio, and Lagrange multiplier tests have different sizes. The size correction should be adequate to eliminate most of the probability of conflict, as the differences between the actual and nominal size are large compared with the differences in power (see e.g. Rothenberg (1982), p. 529).

For a general class of models that includes most of the usual econometric specifications, Rothenberg (1984b) has derived general formulae for the Edgeworth corrected critical values for the Wald (F) test. Empirical evidence in favour of these corrections has been brought up by Magee (1989), in the context of the linear regression model with AR (1) errors. In both cases the Edgeworth expansion used was in terms of the chi-square distribution. Using similar methods, we may derive corrected critical values for the t test from a normal Edgeworth expansion.

Instead of using the "asymptotic" form of the test it seems preferable to make "degrees of freedom adjustments" and to derive expansions in terms of the F (or t) distribution instead of the chi-square (or normal) distribution. Although both approximations have an error of the same order of magnitude, the degrees of freedom adjustments seem to improve the approximation in finite samples.

The practice of using degrees of freedom adjustments in cases where only asymptotic results are available, has been critisized on the grounds that there is nothing in large sample asymptotic theory to justify these corrections (see, e.g. Dhrymes (1969), p. 220). But, there is no reason to restrict ourselves to information from the asymptotic method only. In small samples there are strong arguments in favour of the degrees of freedom adjustments.

If the error covariance matrix is known up to a multiplicative factor, then the degrees of freedom adjusted Wald test statistic is exactly distributed as an F variable. When the covariance matrix is estimated from the data, then the exact distribution is not known, and we switch to the asymptotic (without degrees of freedom adjustments) test. It is quite surprising to note that the latter test is more stringent than the former, i.e. the asymptotic test rejects the null hypothesis more often than the exact test. This is against all intuition, since it implies that, although we have introduced a new source of noise from the estimation of the error covariance matrix, the concentration of the GLS estimator around the true values has increased. Furthermore, the differences between the two tests are quite striking. Under realistic conditions, the acceptance region of the asymptotic test can be less than 50% of the acceptance region of the exact test (see Magdalinos (1983), p. 138-140). The difference tends to zero as the sample size increases, but not as fast as one might expect, especially when the number of the structural parameters is large.

Since the accuracy of an asymptotic approximation depends upon the nature of the first term in the expansion, the situation is essentially the same when refined asymptotic approximations are used. Thus, in the context of instrumental variables estimation, Kunimoto et al. (1983) and Morimune and Tsukuda (1984) report some cases where the approximations based on degrees of freedom adjusted distributions perform better than the approximations based on unadjusted distributions. This can only be explained by the fact that, when the model is sufficiently simplified, the degrees of freedom adjusted approximations reduce to the (known) exact distribution. This instigated some alternative approximations in terms of the t and F distributions (see Magdalinos (1985)). The same idea can be extended in the context of the GLS estimation. Suppose that in the normal linear regression model the error covariance matrix, up to a multiplicative factor, is known to belong to a ball of radius δ. An Edgeworth approximation is said to be locally exact if it reduces to the (known) exact distribution as $\delta \to 0$.

In other words, there is a limitless number of possible approximations that can be generated by changing the distribution function in terms of which the approximation is made. The choice does not affect the order of magnitude of the error term, but it is crucial for the small sample accuracy of the approximation. The traditional choice of the limiting (normal or chi-square) distribution, though analytically convenient, is unlikely to provide approximations with optimal small sample properties. A better choice can be based on the requirement of "local exactness", i.e. we select a distribution function which provides an asymptotic series that reduces to the exact formula when sufficient information allows to transform the model, so as to create a scalar error covariance matrix.

Moreover, one might use a Cornish-Fisher expansion (see, e.g. Cornish and Fisher (1937), Fisher and Cornish (1960), Hill and Davis (1968), Magdalinos (1985)), to correct the test statistic, instead of using an Edgeworth expansion to correct the critical values. The two corrections are asymptotically equivalent to the order of the required accuracy, but the former has two important advantages.

First, the Cornish-Fisher expansion avoids the well-known problems of the Edgeworth expansion in the tails of the approximated distribution. Since the Cornish-Fisher expansion is simply the inversion of the Edgeworth correction of the critical values, it can be expected to have very similar properties in the main body of the approximated distribution. But, in hypothesis testing we are interested in approximating the tail probabilities, and in the tail area the properties are quite different. The Edgeworth approximation is not a proper distribution function, and often assigns negative "probabilities" in the tails of the distribution. On the other hand, the Cornish-Fisher corrected statistic is a proper random variable and the tails of its distribution are well behaved.

Second, in applied econometric research it is more convenient to use the Cornish-Fisher correction of the test statistic rather than the Edgeworth correction of the critical values. The same Cornish-Fisher corrected statistic can be used for testing at any level of significance, whereas different Edgeworth corrected critical values have to be calculated for testing at different levels of significance. Also, once a Cornish-Fisher correction is given, the evaluation of the significance level (p-value) for a given realization of the test statistic is straightforward, since it requires the integration of standard density functions.

The property of "local exactness" can be easily extended to Cornish-Fisher expansions. We shall say that a Cornish-Fisher correction is locally exact if it reduces to a statistic whose exact distribution is known as $\delta\to 0$, where δ is the radius defined above.

All these correction methods lead to tests with sizes differing from the nominal size by an error of the same order of magnitude. Therefore, a Monte-Carlo study is needed to evaluate the relative performance of these corrections in small samples.

The structure of this paper is as follows. In Section 2 the general notation and assumptions are presented. Analytic formulae for the locally exact Edgeworth and Cornish-Fisher second order corrections of the t and F tests are presented in Sections 3 and 4 respectively. In Section 5 these formulae are specialized for the heteroscedastic specification of the linear model. We simplify considerably the original formulae in order to increase computational efficiency. The performance of the alternative size corrections considered in this paper is evaluated by a Monte-Carlo study, the findings of which are discussed in Section 6. Some general comments and concluding remarks are included in Section 7. All proofs are gathered in the Appendix.

2. NOTATION AND ASSUMPTIONS

Consider the equation

$$y=X\beta+\sigma u, \tag{2.1}$$

where y is the $T\times 1$ vector of observations on the dependent variable, X is the $T\times n$ matrix of the exogenous regressors, β is a $n\times 1$ vector of unknown parameters, and σu ($\sigma>0$) is the $T\times 1$ vector of unobserved errors. The random vector u is distributed as $N(0, \Omega^{-1})$, where the elements of the $T\times T$ matrix Ω are known functions of the unknown $k\times 1$ parameter vector γ and, possibly, of a

$T\times m$ matrix Z of observations on a set of exogenous variables, some of which might be regressors too. The vector γ belongs to the parameter space Θ, some open subset of the k-dimentional Euclidean space.

Let $\hat{\gamma}$ be a consistent estimator of γ. For any function $f=f(\gamma)$, we write $\hat{f} = f(\hat{\gamma})$. The feasible GLS estimators of β and σ^2 are

$$\hat{\beta} = (X'\hat{\Omega}X)^{-1}X'\hat{\Omega}y, \tag{2.2}$$

$$\hat{\sigma}^2 = (y-X\hat{\beta})'\hat{\Omega}(y-X\hat{\beta})/(T-n). \tag{2.3}$$

We write Ω_i, Ω_{ij}, etc. for the $T\times T$ martices of first, second, etc. order derivatives of Ω with respect to the elements of γ, and we define the $(k+1)\times 1$ vector δ with elements

$$\delta_0 = (\hat{\sigma}^2-\sigma^2)/\tau\sigma^2, \qquad \delta_i = (\hat{\gamma}_i-\gamma_i)/\tau \qquad (i=1,\dots,k), \tag{2.4}$$

where $\tau=1/\sqrt{T}$ is the "asymptotic scale" of our expansions.

We assume that the following regularity conditions are satisfied:

(i) The elements of Ω and Ω^{-1} are bounded for all T and all $\gamma\in\Theta$, and the matrices

$$A=X'\Omega X/T, \quad F=X'X/T \tag{2.5}$$

converge to non-singular limits as $T\to\infty$.

(ii) Up to the fourth order, the partial derivatives of the elements of Ω with respect to the elements of γ are bounded for all T and all $\gamma\in\Theta$.

(iii) The estimator $\hat{\gamma}$ is an even function of u, and it is functionally unrelated to the β parameters, i.e. it can be written as a function of X, Z, and σu only.

(iv) The vector δ admits a stochastic expansion of the form

$$\delta = d_1 + \tau d_2 + \omega(\tau^2), \tag{2.6}$$

where ω is an order of magnitude defined in the Appendix, and the expectations

$E(d_1 d_1'), \qquad E(\sqrt{T}d_1 + d_2)$

exist and have finite limits as $T\to\infty$.

The first two conditions imply that the matrices

$$A_i = X'\Omega_i X/T, \qquad A_{ij} = X'\Omega_{ij}X/T, \qquad A^*_{ij} = X'\Omega_i\Omega^{-1}\Omega_j X/T \tag{2.7}$$

are bounded, so that, the Taylor expansion of $\hat{\beta}$ is a stochastic expansion (Magdalinos (1986)). Provided that the parameters β and γ are functionally unrelated, the assumption (iii) is satisfied for a wide class of estimators of γ, that includes the maximum likelihood (ML) estimators, and the simple or iterative estimators based on the regression residuals (see Breusch (1980), Rothenberg (1984a)). Also, it can be shown that the condition (iv) is satisfied for the same class of estimators of γ, provided that a condition similar to (i) is satisfied by the Z matrix. It must be noted that we *do not* assume that the estimator $\hat{\gamma}$ is asymptotically efficient.

We define the scalars λ_0 and μ_0, the $k\times 1$ vectors λ and μ, and the $k\times k$ matrix Λ, from the equations

$$\begin{bmatrix} l_0 & l' \\ l & \Lambda \end{bmatrix} = \lim_{T\to\infty} E(d_1 d_1'), \qquad \begin{bmatrix} m_0 \\ m \end{bmatrix} = \lim_{T\to\infty} E(\sqrt{T} d_1 + d_2). \tag{2.8}$$

We denote any $n\times m$ matrix L with elements l_{ij} as

$$L = [(l_{ij})_{\ i=1,\dots,n;\ j=1,\dots,m}] \tag{2.9}$$

with the obvious modifications for vectors and square matrices. If l_{ij} are $n_i\times m_j$ matrices, then the notation (2.9) means that L is the $\left(\sum n_i\right)\times\left(\sum m_j\right)$ partitioned matrix with submatrices l_{ij}. Finally we use the tr, vec, $\otimes$, and matrix differentiation notation as defined in Dhrymes (1978), p. 518-540. Throughout this paper P_x, $\overline{P}_x$ stand for the orthogonal projectors into the space spanned by the columns of the X matrix, and its orthogonal complement respectively.

3. THE t TEST

Let e_0 be a known scalar and e a known $n\times 1$ vector. A test of the null hypothesis

$$e'\beta = e_0 \tag{3.1}$$

against one-sided altenatives can be based on the statistic

$$t = (e'\beta - e_0) / [\hat{\sigma}^2 e' (X'\hat{\Omega} X)^{-1} e]^{1/2}. \tag{3.2}$$

We define the $k\times 1$ vector l, and the $k\times k$ matrix L as

$$l = [(l_i)_{\ i=1,\dots,k}], \qquad L = [(l_{ij})_{\ i,j=1,\dots,k}], \tag{3.3}$$

where

$$l_i = e'GA_iGe/e'Ge, \qquad l_{ij} = e'GC_{ij}Ge/e'Ge, \tag{3.4}$$

$$G = (X'\Omega X/T)^{-1}, \qquad C_{ij} = A^*_{ij} - 2A_iGA_j + A_{ij}/2. \tag{3.5}$$

Lemma 1: Under the null hypothesis (3.1), the distribution function of the statistic (3.2) admits the Edgeworth expansion

$$\Pr(t \le x) = I(x) - \frac{\tau^2}{2}\left[\left(p_1 + \frac{1}{2}\right) + \left(p_2 + \frac{1}{2}\right)x^2\right]x\ i(x) + 0(\tau^3), \tag{3.6}$$

where

$$p_1 = \mathrm{tr}(\Lambda L) + l'\Lambda l/4 + l'(\mu + \lambda/2) - \mu_0 + (\lambda_0 - 2)/4,$$
$$p_2 = (l'\Lambda l - 2l'\lambda + \lambda_0 - 2)/4, \tag{3.7}$$

and $I(x)$, $i(x)$ are the standard normal distribution and density functions respectively.

Corollary 1: The Edgeworth corrected α% critical value of the statistic (3.2) is

$$\xi_\alpha = x_\alpha + \frac{\tau^2}{2}\left[\left(p_1 + \frac{1}{2}\right) + \left(p_2 + \frac{1}{2}\right)x_\alpha^2\right]x_\alpha, \tag{3.8}$$

where x_α is the α% significant point of the standard normal distribution.

Lemma 1 implies the following

Lemma 2: Under the null hypothesis (3.1), the distribution function of the statistic (3.2) admits the Edgeworth expansion

$$\Pr(t \le x) = I_{T-n}(x) - \frac{\tau^2}{2}(p_1 + p_2x^2)x\ i_{T-n}(x) + O(\tau^3), \tag{3.9}$$

where the quantities p_1 and p_2 are defined in (3.7), and $I_{T-n}(x)$, $i_{T-n}(x)$ are the distribution and density functions respectively of a t variable with T-n degrees of freedom. Moreover, the approximation is locally exact, i.e. if γ is known to belong to a ball of radius δ, then the approximation becomes exact as $\delta \to 0$.

Corollary 2: The Edgeworth corrected α% critical value of the statistic (3.2) is

$$\xi_\alpha = x_\alpha + \frac{\tau^2}{2}\left(p_1 + p_2 x_\alpha^2\right) x_\alpha , \tag{3.10}$$

where x_α is the α% significant point ot the t distribution with T-n degrees of freedom.

Using Lemma 2 we can easily prove the following

Theorem 1: Under the null hypothesis (3.1), and provided that the regularity conditions are satisfied, the Cornish-Fisher corrected statistic

$$\hat{t} = t - \frac{\tau^2}{2}(p_1 + p_2 t^2) t \tag{3.11}$$

is distributed, with an error of order $O(\tau^3)$, as a t variable with T-n degrees of freedom. Moreover, the approximation is locally exact, i.e. if γ is known to belong to a ball of radius δ, then the approximation becomes exact as $\delta \to 0$.

Since the parameters p_1 and p_2 are functions of the unknown parameter γ, in practice one has to substitute the estimates

$$\hat{p}_i = p_i(\hat{\gamma}) = p_i + \omega(\tau) \qquad (i=1,2) \tag{3.12}$$

for the parameters p_i in (3.8), (3.10), and (3.11) in order to obtain operational formulae.

In applied research, the statistic (3.11) can be used exactly as the corresponding t statistic in the classical linear model to test the "significance" of the structural parameters, or to test one-sided linear hypotheses on the elemernts of β, etc. Also, the significance level (p-value) of a given realization, t_0 say, of the statistic (3.2) can be obtained by comparing the transformed realization $\hat{t}_0$ with the tables of the t distribution, i.e. if $I_{T-n}(x)$ is the t distribution function with T-n degrees of freedom, then it is easy to show that

$$\begin{aligned} \Pr(t<t_0) &= I_{T-n}(\hat{t}_0) + O(\tau^3), \\ \Pr(t>t_0) &= 1 - I_{T-n}(\hat{t}_0) + O(\tau^3). \end{aligned} \tag{3.13}$$

For the two-sided test of significance of the k–th structural parameter β_k, we have that e has 1 in the k–th position and 0's elsewere, so that the elements of $\hat{l}$ and $\hat{L}$ are

$$\hat{l}_i = \hat{g}_k' \hat{A}_i \hat{g}_k / \hat{g}_{kk}, \qquad \hat{l}_{ij} = \hat{g}_k' \hat{C}_{ij} \hat{g}_k / \hat{g}_{kk} \tag{3.14}$$

respectively, where $\hat{g}_k$ is the k–th column, and $\hat{g}_{kk}$ the k–th diagonal element of the matrix $\hat{G} = (X'\hat{\Omega} X/T)^{-1}$.

4. THE F TEST

Let H be a $r\times n$ known matrix of rank r, and h a known $r\times 1$ vector. A test of the null hypothesis

$$H\beta - h = 0 \tag{4.1}$$

can be based on the Wald statistic

$$w = (H\hat{\beta} - h)' [H(X'\hat{\Omega} X)^{-1}H']^{-1}(H\hat{\beta} - h)/\hat{\sigma}^2. \tag{4.2}$$

We define the $k\times 1$ vector c, and the $k\times k$ matrices C and D as

$$\begin{aligned} c &= [(\mathrm{tr}A_iP)_{i=1,\dots,k}], \qquad C = [(\mathrm{tr}C_{ij}P)_{i,j=1,\dots,k}], \\ D &= [(\mathrm{tr}D_{ij}P)_{i,j=1,\dots,k}], \end{aligned} \tag{4.3}$$

where the matrices A_i and C_{ij} are defined in (2.7) and (3.5), and

$$P = GQG, \quad Q = H'(HGH')^{-1}H, \quad D_{ij} = A_iPA_j/2. \tag{4.4}$$

Lemma 3: Under the null hypothesis (4.1), the distribution function of the statistic (4.2) admits the Edgeworth expansion

$$\Pr(w \le x) = F_r(x) - \tau^2\left(h_1 + h_2\frac{x}{r+2}\right)\frac{x}{r} f_r(x) + O(\tau^3), \tag{4.5}$$

where

$$h_1 = \mathrm{tr}[\Lambda(C+D)] - c'\Lambda c/4 + c'\mu + r[c'\lambda/2 - \mu_0 - (r-2)\lambda_0/4],$$

$$h_2 = \mathrm{tr}(\Lambda D) + [c'\Lambda c - (r+2)(2c'\lambda - r\lambda_0)]/4, \tag{4.6}$$

and $F_r(x)$, $f_r(x)$ are the distribution and density functions respectively of a chi-square variable with r degrees of freedom.

Corollary 3: The Edgeworth corrected α% critical value of the statistic (4.2) is

$$\xi_\alpha = x_\alpha + \tau^2[(h_1/r) + (h_2/r(r+2))x_\alpha]x_\alpha, \tag{4.7}$$

where x_α is the α% significant point of the chi-square distribution with r degrees of freedom.

The parameters h_1 and h_2, being functions of the unknown parameter γ, are not known. So, in practice one uses their estimates

$$\hat{h}_i = h_i(\hat{\gamma}) = h_i + \omega(\tau) \qquad (i = 1, 2) \tag{4.8}$$

in order to render (4.7) an operational formula.

As the exact distribution of the statistic (4.2) has not been tabulated even for the case where γ is known, it is preferable to "correct" for the nominator degrees of freedom, thus obtaining the statistic

$$v = \left(H\hat{\beta} - h\right)'\left[H\left(X'\hat{\Omega}X\right)^{-1}H'\right]^{-1}\left(H\hat{\beta} - h\right)/r\hat{\sigma}^2. \tag{4.9}$$

The statistic (4.9) is the exact analogue of the familiar F statistic in the classical linear model, and it is exactly distributed as an F variable when the parameter γ is known.

From Lemma 3 we derive the following

Lemma 4: Under the null hypothesis (4.1), the distribution function of the statistic (4.9) admits the Edgeworth expansion

$$\Pr(v \le x) = F^r_{T-n}(x) - \tau^2(q_1 + q_2 x)x f^r_{T-n}(x) + O(\tau^3), \tag{4.10}$$

where

$$\begin{aligned} q_1 &= h_1/r + (r-2)/2, \\ q_2 &= h_2/(r+2) - r/2, \end{aligned} \tag{4.11}$$

and $F^r_{T-n}(x)$, $f^r_{T-n}(x)$ are the distribution and density functions respectively of an F variable with r and T–n degrees of freedom. Moreover, the approximation is locally exact, i.e. if γ is known to belong to a ball of radius δ, then the approximation becomes exact as $\delta \to 0$.

Corollary 4: The Edgeworth corrected α% critical value of the statistic (4.9) is

$$\xi_\alpha = x_\alpha + \tau^2 (q_1 + q_2 x_\alpha) x_\alpha, \tag{4.12}$$

where x_α is the α% significant point of the F distribution with r and T-n degrees of freedom.

Using Lemma 4 it is easy to prove the following

Theorem 2: Under the null hypothesis (4.1), and provided that the regularity conditions are satisfied, the Cornish-Fisher corrected statistic

$$\hat{v} = v - \tau^2 (q_1 + q_2 v) v \tag{4.13}$$

is distributed, with an error of order $O(\tau^3)$, as an F variable with r and T-n degrees of freedom. Moreover, the approximation is locally exact, i.e. if γ is known to belong to a ball of radius δ, then the approximation becomes exact as $\delta \to 0$.

The parameters q_1 and q_2 are functions of the unknown parameters h_1 and h_2, so that in practice one has to use the estimates

$$\hat{q}_1 = \hat{h}_1 / r + (r-2)/2, \qquad \hat{q}_2 = \hat{h}_2 / (r+2) - r/2, \tag{4.14}$$

in (4.12), and (4.13) in order to obtain operational formulae.

In applied research, the statistic (4.13) can be used exactly as the corresponding F statistic in the classical linear model to test the "significance" of the fitted equation, or to test two-sided linear hypotheses on the elements of β, etc. Also, the significance level (p-value) of a given realization, v_0 say, of the statistic (4.9) can be obtained by comparing the transformed realization $\hat{v}_0$ with the tables of the F distribution, i.e. if $F^r_{T-n}(x)$ is the F distribution function with r and T–n degrees of freedom, then it is easy to show that

$$\Pr(v > v_0) = 1 - F^r_{T-n}(\hat{v}_0) + O(\tau^3). \tag{4.15}$$

For a test of the joint significance of all the structural parameters, we have that $r=n$, and H is equal to the $n \times n$ identity matrix. Therefore, the elements of c, D, and $C^* = C + D$ are

$$c_i = \text{tr}A_iG, \quad d_{ij} = \text{tr}A_iGA_jG/2, \quad c_{ij}^* = \text{tr}\left(A_{ij}^* + A_{ij}/2\right)G - 3d_{ij} \tag{4.16}$$

and the coefficients of the Cornish-Fisher expansion are

$$\begin{aligned} q_1 &= [\text{tr}\Lambda C^* - c'\Lambda c/4 + c'\mu]/n + c'\lambda/2 - \mu_0 - (n-2)\,(\lambda_0-2)/4, \\ q_2 &= \frac{1}{n+2}\left[\text{tr}\Lambda D + \frac{1}{4}[c'\Lambda c - (n+2)(2c'\lambda - n(\lambda_0 - 2))]\right]. \end{aligned} \tag{4.17}$$

5. HETEROSCEDASTICITY AND RANDOM COEFFICIENTS

A model that includes many of the heteroscedastic and random coefficient specifications occurs when the error variance is assumed to be a linear function of a set of exogenous variables (see, e.g. Hildreth and Houck (1968), Goldfeld and Quandt (1972), Amemiya (1977)). In this case, the disturbances of equation (2.1) are assumed to be independent normal variables with zero mean and variances

$$\sigma_t^2 = z_t'\gamma \qquad (t = 1, \dots, T),$$

where (5.1)

$$z_t' = (1, z_{t2}, \dots, z_{tk})$$

is the t-th observation on k exogenous variables, and $\gamma = (\gamma_1, \dots, \gamma_k)'$ is a non-zero unknown parameter vector. We define the $T \times n$ matrix X, and the $T \times k$ matrix Z with rows x_t' and z_t' respectively.

Our assumptions imply that, without further restrictions, the parameters σ and γ are not simultaneously identified. A reasonable restriction is to set $\sigma = 1$, and this specification is used in the computation of the size corrections for the asymptotic chi-square (or normal) test. Under this restriction, however, the transformed model is supposed to have disturbances with identity variance matrix. This is true only if we can estimate the parameters γ exactly, i.e. for infinite sample size. In small samples the estimate of the variance of the transformed model usually differs from unity (variances as small as 0.5 and as large as 2.3 were observed in our experiments), indicating that the variance matrix of the disturbances in the transformed model is considerably different from the identity matrix. A reasonable method to account for this, is to estimate the variance of the transformed residuals and use the traditional formulae for the t and F test statistics. The resulting tests are locally exact, and, as we shall show in our simulation experiments, they have size closer to the nominal than the asymptotic tests.

In applied research, some of the most frequently used estimators of γ are:

(i) The Goldfeld-Quandt (GQ) estimator

$$\hat{\gamma}_{GQ}=\left[\sum_{t=1}^{T} z_t z_t'\right]^{-1}\sum_{t=1}^{T} z_t\left(y_t-x_t'\tilde{\beta}\right)^2, \tag{5.2}$$

where $\tilde{\beta}$ is the ordinary least squares (OLS) estimator of β.

(ii) The Amemiya (A) estimator

$$\hat{\gamma}_{A}=\left[\sum_{t=1}^{T}\left(z_t'\hat{\gamma}_{GQ}\right)^{-2} z_t z_t'\right]^{-1}\sum_{t=1}^{T}\left(z_t'\hat{\gamma}_{GQ}\right)^{-2} z_t\left(y_t-x_t'\tilde{\beta}\right)^2. \tag{5.3}$$

(iii) The iterative Amemiya (IA) estimator

$$\hat{\gamma}_{\alpha}=\left[\sum_{t=1}^{T}\left(z_t'\hat{\gamma}_{\alpha-1}\right)^{-2} z_t z_t'\right]^{-1}\sum_{t=1}^{T}\left(z_t'\hat{\gamma}_{\alpha-1}\right)^{-2} z_t\left(y_t-x_t'\tilde{\beta}_{\alpha-1}\right)^2, \tag{5.4}$$

where $\alpha = 2, 3, \ldots$ and $\hat{\gamma}_{\alpha-1}$, $\hat{\beta}_{\alpha-1}$ are the estimator of γ and the corresponding feasible GLS estimator of β from the previous iteration. Of course $\hat{\gamma}_1=\hat{\gamma}_A$.

(iv) The maximum likelihood (ML) estimator, which maximazes the function

$$L(\beta,\gamma)=-\frac{1}{2}\sum_{t=1}^{T}\log\left(z_t'\gamma\right)-\frac{1}{2}\sum_{t=1}^{T}\left(y_t-x_t'\beta\right)^2/\left(z_t'\gamma\right). \tag{5.5}$$

We can show that the regularity conditions of Section 2 are satisfied for all these estimators provided that the quantities $|z_{ti}|$, $|\sigma_t^2|$, $|\sigma_t^{-2}|$ are bounted, and that the matrices

$$A=\sum_{t=1}^{T}\sigma_t^{-2}x_t x_t'/T, \qquad \bar{A}=\sum_{t=1}^{T}\sigma_t^{-4}z_t z_t'/T,$$

$$F=\sum_{t=1}^{T}x_t x_t'/T, \qquad \bar{F}=\sum_{t=1}^{T}z_t z_t'/T \tag{5.6}$$

converge to non-singular limits as $T\to\infty$. Also, we define the matrices

$$\Gamma = \sum_{t=1}^{T}\sigma_t^2 x_t x_t' /T, \qquad \bar{\Gamma} = \sum_{t=1}^{T}\sigma_t^4 z_t z_t' /T,$$

$$G = A^{-1}, \quad \bar{G} = \bar{A}^{-1}, \quad B = F^{-1}, \quad \bar{B} = \bar{F}^{-1}. \tag{5.7}$$

It is well-known that the estimators A, IA, and ML are asymptotically efficient, whereas the estimator GQ is not. Moreover, it is easy to show that the IA estimator converges to the ML estimator as $\alpha\to\infty$.

The computation of the vectors l and c, and of the matrices L, C, and D from the definitions (3.3) and (4.3) is straightforward, but it can be computationally expensive when k is large. Then it is preferable to use the following computational procedure:

Given an arbitrary $n\times 1$ vector h, we define the matrices

$$L(h) = \sum_{t=1}^{T}\sigma_t^{-6}\left(x_t'h\right)^2 z_t z_t' /T, \qquad C(h) = \sum_{t=1}^{T}\sigma_t^{-4}\left(x_t'h\right)z_t x_t' /T \tag{5.8}$$

which are the moment matrices of the tranformed variables

$$z_t^* = z_t\left(x_t'h\right)/\sigma_t^3, \quad x_t^* = x/\sigma_t$$

so they can be computed using standard procedures.

Proposition 1: The vector ℓ and the matrix L in (3.3) can be computed form

$$l = -L(h)\gamma, \quad L = 2\,[L(h) - C(h)\,GC'(h)], \tag{5.9}$$

where $h=Ge/(e'Ge)^{1/2}$. Similarly, the vector c, and the matrices C and D in (4.3) can be computed from

$$c = -C_1\gamma, \quad C = 2(C_1 - C_2), \quad D = C_3/2,$$

where (5.10)

$$C_1 = \sum_{i=1}^{r}L(h_i), \quad C_2 = \sum_{i=1}^{r}C(h_i)GC'(h_i), \quad C_3 = \sum_{i=1}^{r}C(h_i)PC'(h_i)$$

and $h_i = \lambda_i^{1/2} w_i$, where λ_i (i= 1, … ,r) are the r non-zero eigenvalues of P and w_i are the corresponding orthonormalised eigenvectors of P. The difference betwe-

en the original definitions (3.3), (4.3) and the formulae (5.9), (5.10) are of order $O(\tau^2)$, i.e. it is negligible to the order of our approximation.

It remains to find the values of the parameters (2.8), which can be expressed in terms of the matrices (5.6), (5.7), and the vectors

$$\xi = \sum_{t=1}^{T} v_t z_t / T, \qquad \xi_1 = \sum_{t=1}^{T} \sigma_t^{-4} v_t z_t / T, \qquad \xi_2 = \sum_{t=1}^{T} \sigma_t^{-4} x_t' G x_t z_t / T,$$

where (5.11)

$$v_t = 2\sigma_t^2 x_t' B x_t - x_t' B \Gamma B x_t .$$

Proposition 2: Under the assumptions of this section, the parameters (2.8) for the GQ estimator of γ can be estimated as

$$\Lambda = 2\overline{B\Gamma B}, \quad \mu = -\overline{B}\xi, \quad \lambda = 2\gamma - \Lambda\overline{A}\gamma,$$

$$\lambda_0 = 2\left(1 - \gamma'\overline{A}\gamma\right) - \gamma'\overline{A}\lambda, \quad \mu_0 = \mathrm{tr}\left(\Lambda\overline{A}\right) - 2k - \gamma'\overline{A}\mu, \tag{5.12}$$

and for the A estimator of γ they can be estimated as

$$\Lambda = 2\overline{G}, \quad \lambda = 0, \quad \lambda_0 = 2\left(1 - \gamma'\overline{A}\gamma\right), \quad \mu_0 = -\gamma'\overline{A}\mu,$$

$$\mu = -\overline{G}\xi_1 - 4\sum_{i=1}^{k} \overline{G}\left[\overline{A}_i \overline{g}_i - \left(Z'\Omega_i\Omega^{-1}Z/T\right)\overline{b}_i\right]. \tag{5.13}$$

For the IA and ML estimators of γ, the parameters (2.8) can be estimated by using (5.13) with

$$\mu = -\overline{G}\xi_2, \tag{5.14}$$

where $\overline{g}_i$ is the i-th coloumn of the matrix $\overline{G}$, $\overline{b}_i$ is the i-th coloumn of the matrix $\overline{B}$, and $\overline{A}_i = (Z'\Omega\Omega_i Z/T)$.

Using Propositions 1 and 2, it is easy to compute the parameters (3.7) and (4.11) of the Cornish-Fisher expansions. Note that all the matrices appearing in Propositions 1 and 2 are $n \times n$ or $k \times k$, and, apart from $\overline{A}_i$ and $(Z'\Omega_i\Omega^{-1} Z/T)$, they are calculated in the estimation process. Furthermore, the trace of the matrix $\Lambda\bar{A}$ can be computed very easily. From a computational point of view, the formulae given above represent a considerable improvement of the original formulae (see Rothenberg (1984b)), and they simplify the calculation of the Edgeworth and Cornish-Fisher corrections. Of course, in the operational formulae, γ has to be substituted by the corresponding consistent estimator.

6. THE SIMULATION EXPERIMENTS

In the introduction we gave some heuristic arguments in favour of using locally exact rather than asymptotic approximations, and we proposed the use of the Cornish-Fisher correction of the test statistic instead of the Edgeworth correction of the critical values. As all these methods have an error of the same order of magnitude, we shall use Monte-Carlo experiments to compare their performance.

For the simulation we assumed the four parameter linear model

$$y_t = \sum_{j=1}^{4} x_{tj}\beta_j + u_t, \qquad u_t \sim N\left(0,\sigma_t^2\right), \qquad (t=1,\dots,20),$$

$$\sigma_t^2 = z_t'\gamma, \qquad z_t' = (1, x_{t2}, x_{t3}, z_{t4}). \tag{6.1}$$

Six values of γ were investigated: $\gamma_{(1)}' = (\gamma_1, 0, 0, 0)$, $\gamma_{(2)}' = (\gamma_1, 1, 0, 0)$, $\gamma_{(3)}' = (\gamma_1, 0, 0, 1)$, $\gamma_{(4)}' = (\gamma_1, 1, 1, 0)$, $\gamma_{(5)}' = (\gamma_1, 1, 0, 1)$, $\gamma_{(6)}' = (\gamma_1, 1, 1, 1)$.

In applied research one often faces not only heteroscedastic disturbances, but also multicollinear regressors. Therefore, we decided to use multicollinear explanatory variables. Following McDonald and Galarneau (1975) we computed the regressors by

$$x_{tj} = 1 \qquad (t=1,\dots,20 \text{ and } j=1),$$

$$x_{tj} = (1-A)^{1/2}\,\zeta_{t(j-1)} + A^{1/2}\,\zeta_{t4} \qquad (t=1,\dots,20 \text{ and } j=2,3,4), \tag{6.2}$$

where ζ_{tj} $(j=1,2,3,4)$ are independent $N(0,1)$ pseudo-random numbers, and A is the correlation coefficient between any two explanatory variables. Four values of A were considered: 0.0, 0.1, 0.5, 0.9.

Without loss of generality we confined ourselves to the case with $\sigma_t^2 \geq 1$. (Cases with $0 \leq \sigma_t^2 \leq 1$ are tackled by using the inverse of σ_t^2 instead of σ_t^2). So, we created the vectors $z_t' = (1, x_{t2}, x_{t3}, z_{t4})$, where z_{t4} (t=1,..., 20) are independent N(0,1) pseudo-random numbers, and the theoretical variances, $\sigma_t^2 = z_t'\gamma$, under the restriction that $\sigma_t^2 \geq 1$. To do so, we calculated γ_1 so that $\sigma_t^2 = z_t'\gamma \geq 1$ for all z_t'in our sample. Then, using the square roots of the variances, σ_t^2, we computed the heteroscedastic disturbances, u_t, from

$$u_t = \sigma_t \varepsilon_t, \quad \sigma_t = \sqrt{\sigma_t^2}, \tag{6.3}$$

where ε_t (t=1,..., 20) are independent N(0,1) pseudo-random numbers.

Following Breusch ((1980), Theorem 5, p. 336) and bearing in mind that the t and F test statistics can be derived from the Wald statistic, we deduce that the distributions of these statistics do not depend upon the values of the parameters β_j (j=1,..., 4) in the model (6.1) when the null hypothesis is correct. So, the results of the experiments concerning the actual size of the t and F tests are independent of the values of β_j (j=1,...,4). For the sake of simplicity we assumed $\beta_j = 0$ (j=1,...,4), thus calculating the y_t (t=1,..., 20) as

$$y_t = u_t, \qquad u_t \sim N(0, \sigma_t^2), \qquad \sigma_t^2 = z_t'\gamma \geq 1. \tag{6.4}$$

For the t test we considered four hypotheses of the form (3.1)

$$\beta_1 = 0, \quad \beta_2 = 0, \quad \beta_3 = 0, \quad \beta_4 = 0, \tag{6.5}$$

and for the F test we considered one hypothesis of the form (4.1) with

$$H = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad h = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{6.6}$$

For any combination of the values of γ and A a matrix of regressors, X, was created according to (6.2), so that the experimental results do not depend on a particular realization of the X matrix. For any realization of the X matrix a matrix Z, with rows z_t', was created, and 500 vectors y were used to implement 500 replications of the procedure discribed bellow:

The parameters β_j (j=1,..., 4) of the model

$$y_t = \sum_{j=1}^{4} x_{tj}\beta_j + u_t \tag{6.7}$$

were estimated via the OLS method, and the OLS residuals, $\tilde{u}_t$, were used to compute the OLS (Goldfeld-Quandt) estimate of γ (see (5.2)). Using the OLS residuals, $\tilde{u}_t$, and $\hat{\gamma}_{GQ}$ we computed the GLS (Amemiya) estimate of γ (see (5.3)). We used the Amemiya estimator because it is the simpler of all the asymptotically efficient estimators of γ discussed in Section 5, which are expected to give similar results. We used the calculated value of $\hat{\gamma}_A$ in order to estimate the model (6.7) via the GLS method. The GLS estimate of $\beta' = (\beta_1, \beta_2, \beta_3, \beta_4)$

$$\hat{\beta} = \left[\sum_{t=1}^{T}\left(z_t'\hat{\gamma}_A\right)^{-1} x_t x_t'\right]^{-1} \sum_{t=1}^{T}\left(z_t'\hat{\gamma}_A\right)^{-1} x_t y_t, \tag{6.8}$$

were used to compute the statistics (3.2), (3.11), (4.2), (4.9), and (4.13) and the Edgeworth corrected critical values (3.8), (3.10), (4.7), and (4.12) for the t and F tests.

Having completed 500 replications of the procedure discribed above, for each combination of the values of γ and A, we calculated the actual size of the different methods of conducting the t and F tests.

The order of execution of the experiments for the different couples (γ, A) was randomly determined, thus making the results independent of the sequential quality of the pseudo-random numbers needed for the creation of the X and Z matrices, and the y vectors. These pseudo-random numbers were generated by use of the I.M.S.L. library.

The results of the experiments are presented in Tables 1 and 2, and in Figures 1 through 6. Since the t test of each of the null hypotheses (6.5) against two-sided alternatives is a special case of the F test, we decided to examine the performance of the t test with one-sided alternatives. For this reason the results for the t test had to be evaluated separately for the cases of positive and negative t statistics.

TABLE 1
Null rejection probability estimates (positive t statistics)

H_0		β_3														
Size:		1%					5%					10%				
Test:		N	NE	T	TE	CFT	N	NE	T	TE	CFT	N	NE	T	TE	CFT
γ	A			%					%					%		
	.0	5.6	4.0	3.0	2.0	1.0	12.4	8.8	9.6	7.4	6.6	18.0	13.2	14.0	12.0	10.8
	.1	5.0	2.8	2.2	1.2	0.6	11.0	7.0	7.4	5.8	4.4	15.8	10.8	12.4	8.8	7.8
$\gamma_{(1)}$	.5	5.6	4.6	2.6	3.0	2.2	13.0	9.4	9.0	7.0	6.2	18.8	14.4	15.0	13.0	10.8
	.9	5.0	3.4	2.6	2.4	1.6	14.0	8.6	8.4	5.6	4.6	19.4	13.4	14.4	10.8	9.2
	.0	5.2	4.6	2.8	3.0	2.2	11.8	9.4	7.6	7.2	4.8	17.4	14.4	13.6	12.0	10.4
	.1	5.6	3.4	2.0	1.6	1.4	11.0	8.2	8.4	6.6	5.0	17.2	13.0	13.8	11.6	11.0
$\gamma_{(2)}$	.5	6.4	2.8	2.2	1.8	1.4	14.6	8.8	9.8	7.0	5.8	21.4	15.2	17.0	12.4	11.2
	.9	4.0	2.8	1.8	1.8	1.2	10.0	6.4	6.6	5.0	4.2	15.0	10.4	11.0	9.0	7.2
	.0	5.6	4.6	2.6	1.8	1.4	11.8	8.6	7.6	6.4	5.6	20.0	12.6	14.2	10.4	9.6
	.1	5.4	4.2	3.0	2.6	2.0	11.2	7.6	8.8	6.0	4.8	16.4	11.8	12.4	11.0	9.2
$\gamma_{(3)}$	.5	6.4	4.6	3.6	2.4	2.4	12.0	8.4	8.4	6.6	5.8	16.8	12.6	13.6	11.0	9.8
	.9	4.2	4.0	1.4	2.6	2.0	12.0	8.4	7.2	6.0	5.6	18.6	14.2	14.6	11.6	10.2
	.0	5.6	5.6	3.4	3.8	2.2	12.0	8.0	8.0	7.2	6.4	16.6	13.8	13.6	11.8	9.8
	.1	4.4	3.4	1.8	1.8	1.2	11.0	8.4	6.8	6.0	5.2	15.8	12.2	12.6	11.0	10.6
$\gamma_{(4)}$	.5	6.6	5.4	3.4	3.4	2.6	12.6	10.2	10.0	8.4	5.8	17.8	15.2	15.0	13.6	11.0
	.9	2.8	4.0	1.6	2.4	1.8	8.4	6.6	5.4	5.4	4.8	12.0	10.6	9.2	8.4	7.8
	.0	5.0	3.6	2.8	1.6	0.8	13.6	8.8	8.4	6.2	4.8	18.6	16.2	15.2	12.2	11.0
	.1	4.6	3.8	1.6	2.4	1.8	10.6	8.0	7.0	5.8	4.8	16.4	12.0	11.8	9.4	8.4
$\gamma_{(5)}$	.5	4.2	4.2	2.2	2.6	1.2	11.6	8.6	7.6	6.0	5.2	16.8	13.4	12.8	10.8	9.6
	.9	4.0	3.0	2.0	1.6	1.6	13.2	8.6	7.8	6.4	5.2	19.0	13.2	14.8	10.4	9.6
	.0	5.8	3.8	3.0	1.6	1.4	11.2	8.4	7.2	6.2	5.4	15.4	11.6	11.8	9.4	8.4
	.1	4.8	3.4	1.6	2.0	1.6	11.4	8.8	7.8	5.8	5.4	16.8	13.6	14.2	11.0	10.0
$\gamma_{(6)}$	.5	7.6	5.2	2.4	2.2	1.0	16.0	11.0	10.8	8.0	4.6	21.4	15.8	18.2	13.4	10.4
	.9	4.6	3.6	2.0	1.8	0.4	11.8	8.0	8.0	6.2	5.2	17.4	12.6	13.8	10.4	10.0

TABLE 2
Null rejection probability estimates
(F statistics)

H_0		$\beta_2, \beta_3, \beta_4$														
Size:		1%					5%					10%				
Test:		X2	X2E	F	FE	CFF	X2	X2E	F	FE	CFF	X2	X2E	F	FE	CFF
γ	A	%					%					%				
	.0	22.2	16.2	9.2	6.4	1.8	34.2	23.8	19.0	15.8	7.0	41.2	29.8	30.4	21.8	12.0
	.1	23.2	15.6	8.4	5.4	2.2	34.6	24.0	10.2	14.2	7.0	41.4	30.4	29.6	20.2	11.4
$\gamma_{(1)}$	.5	20.0	15.6	8.0	6.4	3.2	34.0	22.4	17.8	13.8	7.4	41.0	31.0	27.6	20.0	12.8
	.9	19.0	14.2	9.8	5.8	3.0	32.4	20.0	17.6	13.0	6.8	39.4	29.0	25.8	18.6	9.8
	.0	20.4	14.6	8.0	6.8	3.2	32.6	22.6	16.8	13.0	8.4	43.2	28.8	26.0	19.4	13.2
	.1	20.4	14.6	7.6	6.2	3.6	31.4	22.4	17.4	12.6	7.4	40.2	27.6	24.0	19.6	13.0
$\gamma_{(2)}$	.5	24.6	17.8	10.6	7.8	1.6	34.8	25.2	19.6	16.4	8.8	43.4	30.6	27.0	21.0	14.8
	.9	22.8	16.6	8.8	5.2	2.6	35.4	25.8	20.4	15.4	8.6	44.4	32.0	29.4	21.8	12.6
	.0	23.4	17.2	8.4	6.8	1.6	35.4	26.2	19.2	16.0	9.6	44.8	32.6	27.4	21.8	16.2
	.1	22.4	18.0	9.8	8.4	3.4	33.2	23.8	20.0	17.4	9.4	42.2	29.6	26.8	22.0	15.4
$\gamma_{(3)}$	.5	24.4	18.8	9.0	9.0	5.2	39.8	27.4	21.4	17.0	11.0	47.2	36.0	32.0	26.0	18.0
	.9	22.2	16.8	8.2	7.2	2.0	34.6	24.6	20.0	15.6	8.0	41.8	31.6	28.2	22.2	13.4
	.0	23.4	19.2	9.0	7.6	3.2	35.6	28.0	20.0	16.2	10.0	44.4	35.2	29.2	23.2	14.4
	.1	19.0	15.4	8.2	8.0	3.6	34.2	24.8	16.6	15.2	8.8	43.8	30.8	25.2	21.8	16.0
$\gamma_{(4)}$	.5	25.2	18.0	10.6	8.6	4.4	36.0	27.6	22.0	18.2	10.8	42.6	33.8	30.6	24.8	17.4
	.9	22.2	17.6	8.2	8.4	3.8	34.4	27.4	18.8	16.2	9.4	43.4	32.2	27.0	22.0	14.6
	.0	24.0	20.0	7.6	8.0	4.0	36.2	28.2	19.6	15.2	8.6	45.0	35.8	28.4	22.8	14.2
	.1	20.8	16.4	6.4	6.0	2.4	34.8	24.6	16.8	12.8	7.2	43.6	31.4	27.2	20.8	14.8
$\gamma_{(5)}$	.5	21.6	16.6	8.6	7.2	2.8	36.4	25.2	19.4	16.2	9.0	44.2	34.2	28.4	23.0	13.6
	.9	24.2	16.6	8.2	6.6	2.6	38.2	27.8	21.8	15.4	8.6	46.0	32.8	31.2	22.4	15.4
	.0	26.0	18.6	11.0	6.6	2.0	38.0	27.8	22.4	17.2	7.0	45.2	34.4	31.8	24.2	12.8
	.1	17.8	14.6	7.8	5.8	2.8	28.6	20.6	15.8	12.6	7.2	34.6	26.6	23.4	17.8	10.0
$\gamma_{(6)}$	.5	21.8	17.0	7.0	7.0	3.8	34.6	25.2	19.8	15.6	7.4	44.4	31.8	27.6	22.2	12.2
	.9	23.6	17.0	7.0	6.0	1.8	34.2	26.4	21.6	14.6	7.4	44.2	33.6	29.6	24.0	13.2

In Table 1 we give the null rejection probability estimates for the case of the positive t statistics for the test of significance of the parameter β_3. The results for the other t tests examined in our experiments are similar and omitted. The following notation is used: N, T are the t test procedures implemented by the statistic (3.2) using the normal and t distributions respectively, NE, TE are the t tests based on Edgeworth approximations from the normal and t distributions respectively, and CFT is the Cornish-Fisher corrected t test.

The comparison of the various t test procedures must be done in terms of the difference between the actual and nominal size of the corresponding testing methods. The various methods can be ranked in such a way, so that the smaller the difference between the actual and nominal size, the better is the corresponding test procedure.

Table 1 shows that the CFT test is, almost uniformly, the best followed by TE, and T, being the second and third best respectively. The worst performance is that of the N test. Inspection of the table reveals that the asymptotic tests (N, NE) are, in general, worse than the exact tests (T, TE, CFT) with the exception of some cases in which the NE correction pereforms better than the uncorrected T method. There are few cases with N better than NE, and/or T, TE better than CFT.

The null rejection probability estimates of the F test of "significance" of the fitted equation, i.e. the joint significance of the parameters β_2, β_3, and β_4, are presented in Table 2. The following notation is used: X2, F are the F test procedures implemented by the statistics (4.2), (4.9) using the chi-square and F distributions respectively, X2E, FE are the F tests based on Edgeworth approximations from the chi-square and F distributions respectively, and CFF is the Cornish-Fisher corrected F test.

As in the case of the t test, the various methods used to implement the F test can be ranked according to the difference between their actual and nominal size.

Table 2 shows that, in every case, i.e. for every combination of the values of γ and A, the CFF test performs better than the tests FE and F, being the second and thirt best respectively. The next best is the X2E test, whereas the X2 test is the worst of all. Therefore, using information of Table 2 the various F testing procedures can be ranked in decreasing order of performance as CFF, FE, F, X2E, X2. It should be emphasized that the uncorrected F tests is better not only from the chi-square test, but in most cases it is better than the Edgeworth-corrected chi-square test.

Also, the alternative methods of size correction can be compared diagrammatically. In a square diagram the actual size (i.e. the null rejection probability estimate at a given nominal size) of the t or F test is plotted against the nominal size. Under ideal conditions, the actual and nominal sizes are identical, and, consiquently, all the observations should be located at the 45 degrees diagonal of the diagram. This line is the cumulative distribution of a U(0,1) random variable. Of course, if we had a random sample from the U(0,1) distribution, then, because of the sampling error, the observations would not be located exactly at the 45 degrees line. However, they should be inside a symmetric band, around the 45 degrees line, and with width given in the table of the one-sample Kolmogorov-Smirnov distribution (see J. Durbin (1973)). That is, if a particular curve is inside the band, then its deviations from the 45 degrees line can be attributed to sampling error, i.e. they are statistically insignificant. On the other hand, if the curve intersects the band bountary, then we refuse the hypothesis that its deviations from the 45 degrees line are due to random sampling. It is clear that the band represents a kind of confidence interval which facilitates the interpretation of the information conveyed. In the diagrams the band corresponds to the 95% confidence interval. It should be noted that, as far as the diagrams for the t test are concerned, the lower left corner refers to the positive t statistics, whereas the upper right corner refers to the negative t statistics.

Here, we present only six of the one hundrend and twenty diagrams produced during the execution of our experiments. The chosen diagrams are representative of all the others, and were selected on the ground of their clarity. Figures 1, 2, and 3 portray the performance of the various t test procedures, whereas the performance of the F testing methods are illustrated in Figures 4, 5, and 6. Figures 1 and 4 show the performance of the various size corrections in those cases where the error term in the population is homoscedastic. Figures 1 through 6 have an advantage over the Tables 1 and 2 in that they give us information about the performance of the various test procedures at any level of the nominal size.

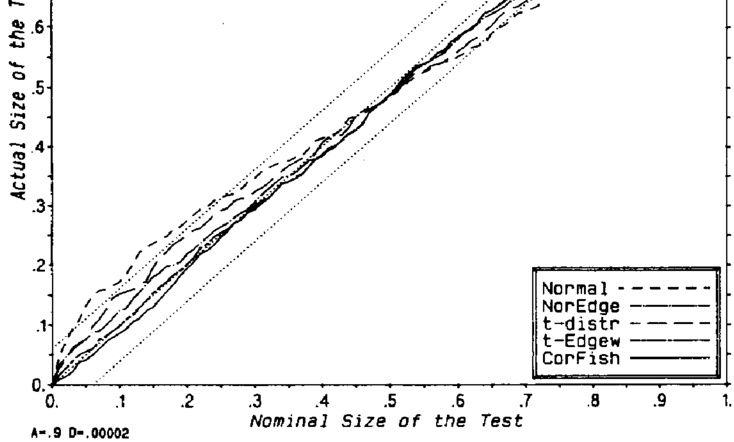

Figure 1. Size corrections of the t test for β_2, $\gamma' = (\gamma_1, 0, 0, 0)$ and $A = .9$.

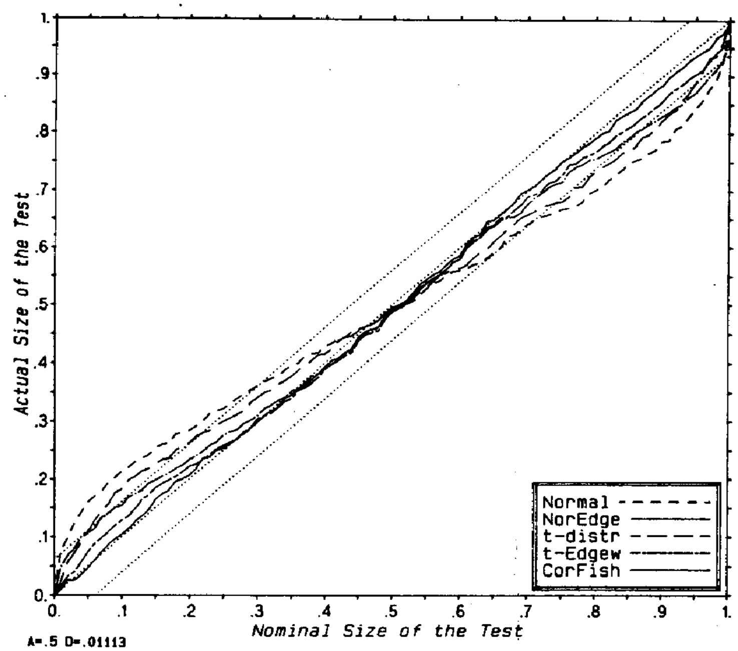

Figure 2. Size corrections of the t test for β_3, $\gamma' = (\gamma_1, 1, 1, 1)$ and $A = .5$.

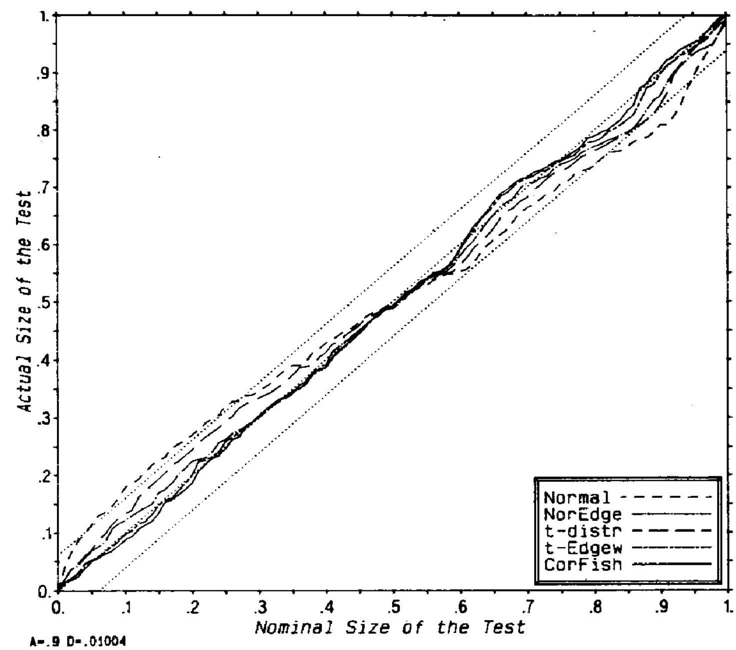

Figure 3. Size corrections of the t test for β_4, $\gamma' = (\gamma_1, 1, 0, 0)$ and $A = .9$.

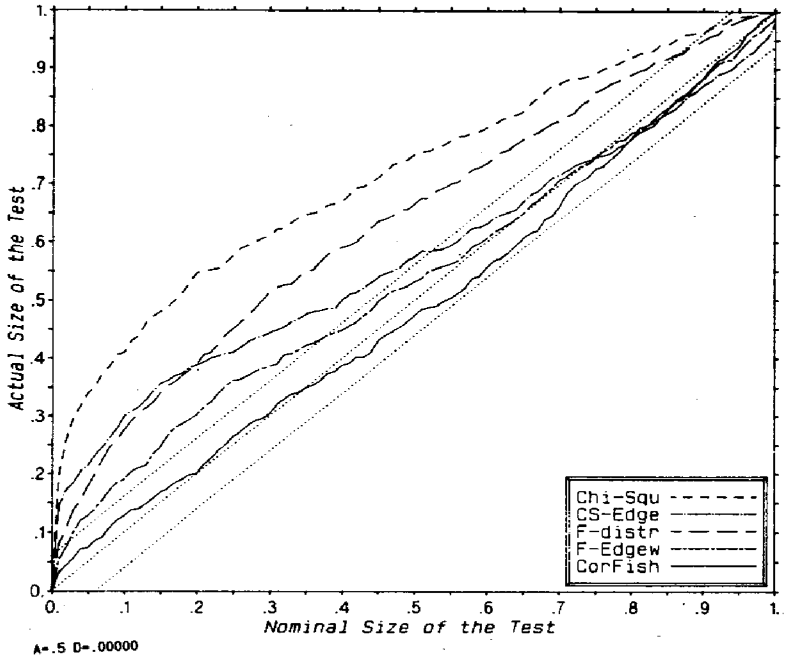

Figure 4. Size corrections of the F test, $\gamma' = (\gamma_1, 0, 0, 0)$ and $A = .5$.

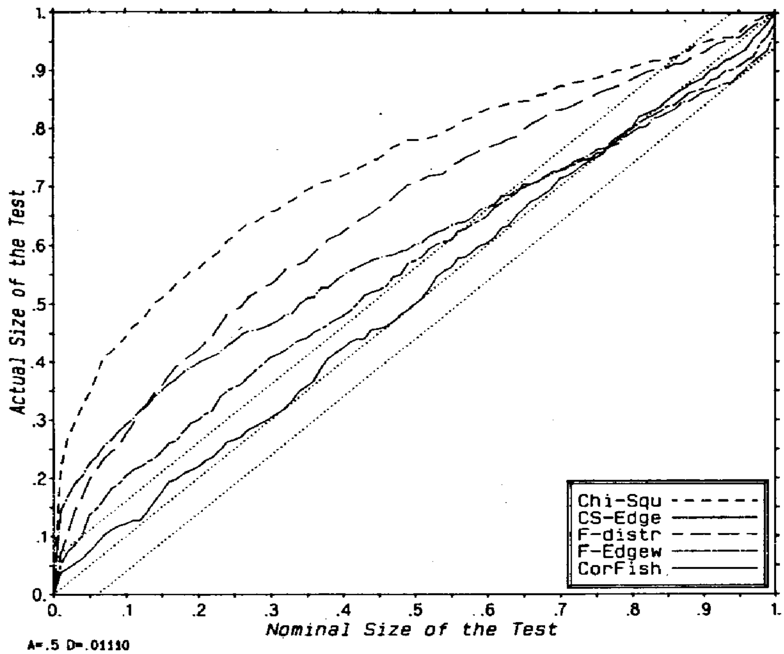

Figure 5. Size corrections of the F test, $\gamma' = (\gamma_1, 1, 1, 1)$ and $A = .5$.

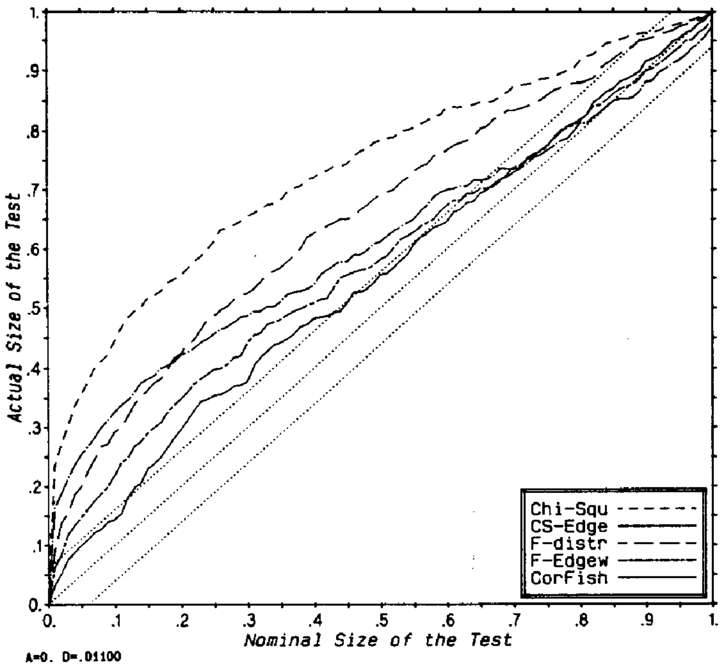

Figure 6. Size corrections of the F test, $\gamma' = (\gamma_1, 1, 1, 0)$ and $A = 0$.

In Figures 1 and 3 all the curves lie entirely inside the band with the exception of the curve corresponding to the N testing procedure, and the additional exception of a very small part of the curve of the T testing method in Figure 1. On the contrary, in Figure 2 only the curves of the CFT and TE size corrections lie inside the band. It must be noted that, according to the properties of the band discussed previously, all the curves lying entirely inside the band are, at the 5% level of significance, statistically identical so that it is meaningless to rank their performance. For the curves, however, intersecting the band a ranking is resulted in, which is consistent with the information of Table 1. It is noteworthy that for p-values greater than or equal to almost 10% the NE size correction performs better than the conventional t test.

In Figures 4 and 5 only the curve corresponding to the CFF testing procedure lies entirely inside the band, whereas in Figure 6 all the curves have large parts outside the band. Thus, in Figures 4 and 5, the difference between the actual and nominal size of the CFF test is statistically insignificant, at the 5% significance level. Figures 4 through 6 confirm the ranking of the various F testing methods resulted in Table 2 for p-values less than or equal to almost 10%. It must be noted, however, that for p-values somewhat greater than 10% the X2E size correction performs better than the conventional F test.

Figures 1 and 4 show that the performance of the Edgeworth or Cornish-Fisher size corrections is good even in cases where the population errors are homoscedastic, i.e. the error covariance matrix is a scalar matrix.

Perhaps more importantly, the tables and diagrams presented above confirm our theoretical considerations about the superiority of the locally exact (degrees of freedom adjusted) tests over the asymptotic (unadjusted) tests. The fact that often the uncorrected t or F tests are superior to the Edgeworth-corrected normal or chi-square tests was quite unexpected and it should be taken seriously into account in applied research.

7. CONCLUDING REMARKS

In this paper we derived Edgeworth approximations for the distribution function of some GLS test statistics based on the t and F distributions, and the corresponding Cornish-Fisher corrections of the test statistics. Both the approximations and the corrections are locally exact, i.e. they reduce to the exact formulae when the error covariance matrix is known up to a multiplicative factor.

We conducted a Monte-Carlo study in order to compare the performance of these corrections. Our results indicate that the locally exact Edgeworth corrections are generally better than the traditional Edgeworth corrections. Furthermore, our experiments indicate that the Cornish-Fisher corrections perform better than the locally exact Edgeworth corrections almost everywhere in the parameter space.

Finally, from (4.13) it is clear that when the sample size is small, the Cornish-Fisher corrected F statistic can assume negative values, a somehow troublesome possibility since the Cornish-Fisher corrected F statistic is assumed to be distributed as an F variable. Fortunately, negative values are encountered only when the uncorrected F statistic assumes very large positive values, i.e. in cases where the null hypothesis is rejected with very high probability by the uncorrected F test. Negative values of the corrected F statistic are extremely unlikely, when the null hypothesis is true. Therefore, in applied research a negative Cornish-Fisher corrected F statistic implies the rejection of the null hypothesis with very high probability. Nevertheless, this can be a very useful warning indicating that the sample size is too small relative to the observed correlations so that the asymptotic methods are not applicable.

Needless to say that this problem is not encountered only in Cornish-Fisher corrections. When the sample size is small, the Edgeworth corrections can produce negative critical values. In both cases the use of the F distribution instead of the chi-square distribution reduces the possibility of such over-corrections, since the F statistic (4.9) and the critical values of the F distribution are smaller in magnitude than the chi-square statistic (4.2) and the critical values of the chi-square distribution. This is the reason why we did not consider in this paper the Cornish-Fisher correction of the chi-square statistic.

The use of the Cornish-Fisher corrections in applied research is indeed very simple. Once the corrected statistics have been calculated, they can be treated exactly as the corresponding statistics in the classical linear model. Therefore, we hope that these corrections can be a useful tool in applied statistical and econometric research.

APPENDIX

For any stochastic quantity (scalar, vector, or matrix) Y_τ, we write $Y_\tau = \omega(\tau^i)$ if for every $n>0$ there exists an $\varepsilon>0$, such that

$$\Pr[\,|Y_\tau/\tau^i| > (-\log\tau)^\varepsilon] = o(\tau^n) \text{ as } \tau\to 0, \tag{A.1}$$

where $|\cdot|$ is the Euclidean norm. The use of this order is motivated by the fact that, if two stochastic quantities differ by a quantity of order $\omega(\tau^i)$, then under general conditions the distribution function of the one provides an asymptotic approximation to the distribution function of the other, with an error of order $O(\tau^i)$. Moreover, the orders $\omega(\cdot)$ and $O(\cdot)$ have similar operational properties. For details see Magdalinos (1986).

Proof of Lemma 1: It is easily proved that under the null hypothesis (3.1) the t statistic (3.2) admits a stochastic expansion of the form

$$t = t_0 + \tau t_1 + \tau^2 t_2 + \omega(\tau^3), \tag{A.2}$$

where

$$\begin{aligned} t_0 &= k'b = e'b/(e'Ge)^{1/2}, \\ t_1 &= k'b_* - k'b(\delta_0 + g_*)/2, \\ t_2 &= -k'b_*(\delta_0 + g_*)/2 + k'b\,[3\left(\delta_0^2 + g_*^2\right) + 2\,\delta_0\, g_*]/8, \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} &k = e/(e'Ge)^{1/2},\;\; b = GX'\Omega u/\sqrt{T},\;\; b \sim N(0,G),\;\; b_* = \hat{G}X'\hat{\Omega}Mu, \\ &\hat{G} = (X'\hat{\Omega}X/T)^{-1},\;\; M = I - X(X'\Omega X)^{-1}X'\Omega, \\ &g_* = k'G_*k,\;\; G_* = \sqrt{T}(\hat{G} - G). \end{aligned} \tag{A.4}$$

Let s be an imaginary number. Also, let $I(x)$ and $i(x)$ be the distribution and density functions respectively of a standard normal variable. Taking expectations in (A.2) we find that the characteristic function of the statistic (3.2) is

$$\begin{aligned} \varphi_t(s) &= E[\exp(s(t_0 + \tau t_1 + \tau^2 t_2))] + O(\tau^3) = \\ &= \varphi(s) + \frac{\tau^2}{2} s(s\pi_1 + s^3\pi_1)\,\varphi(s) + O(\tau^3), \end{aligned} \tag{A.5}$$

where

$\varphi(s) = E[\exp(st_0)]$,

$$\pi_1 = \sum_{i=1}^{k} \sum_{j=1}^{k} \lambda_{ij}\left(g_i g_j + g_{ij}^* - g_{ij}\right) + \sum_{i=1}^{k} g_i(\lambda_{i0} - \mu_i) + \lambda_0 - \mu_0,$$

$$\pi_2 = \frac{1}{4}\left[\sum_{i=1}^{k} \sum_{j=1}^{k} \lambda_{ij} g_i g_j + 2\sum_{i=1}^{k} g_i \lambda_{i0} + \lambda_0\right], \qquad \text{(A.6)}$$

$g_i = k'G_i k, \quad g_{ij} = k'G_{ij}k,$

$G_i = -GA_iG, \quad G_{ij} = G(A_iGA_j - A_{ij}/2)G.$

Dividing (A.5) by $-s$ and inverting the Fourier transform we complete the proof of Lemma 1. Magdalinos (1986) proves the validity of this approximation.

Proof of Lemma 2: Let $I(x)$ and $i(x)$ be the distribution and density functions respectively of a standard normal variable. Also, let $I_{T-n}(x)$ and $i_{T-n}(x)$ be the distribution and density functions respectively of a t variable with T-n degrees of freedom. Fisher (1925) shows that the Edgeworth expansions of the t distribution and density functions are

$$I_{T-n}(x) = I(x) - (1/4)\tau^2(1+x^2)xi(x) + O(\tau^4),$$

and (A.7)

$$i_{T-n}(x) = i(x) + O(\tau^2),$$

where $\tau = T^{-1/2}$. Using Lemma 1, (A.7) and simple algebra we find

$$\Pr(t \le x) = I_{T-n}(x) - \frac{\tau^2}{2}(p_1 + p_2x^2)xi_{T-n}(x) + O(\tau^3). \qquad \text{(A.8)}$$

As $\delta \to 0$, γ becomes a fixed known vector, so that

$\Lambda = 0, \quad \lambda = \mu = 0, \quad \lambda_0 = 2, \quad \mu_0 = 0,$

and (A.9)

$p_1 = p_2 = 0.$

Therefore, the approximation is locally exact.

Proof of Theorem 1: Under the null hypothesis (3.1), the t statistic (3.2) can be easily shown to admit a stochastic expansion of the form

$$t = t_0+\tau t_1+\tau^2 t_2+\omega(\tau^3), \tag{A.10}$$

where the first term in the expansion is

$$t_0 = e'b/(e'Ge)^{1/2}, \quad b = GX'\Omega u/\sqrt{T}. \tag{A.11}$$

Therefore, the Cornish-Fisher corrected t statistic (3.11) admits a stochastic expansion of the form

$$\hat{t} = t_0+\tau t_1+\tau^2 (t_2-t_3)+\omega(\tau^3), \tag{A.12}$$

where

$$t_3 = (p_1+p_2 t_0^2)t_0/2. \tag{A.13}$$

Let s be an imaginary number, and let $\varphi_t(s)$, $\varphi(s)$ be the characteristic functions of the t statistic (3.2) and of a standard normal variable respectively. Using (A.12) we find that the characteristic function of the Cornish-Fisher corrected statistic $\hat{t}$ is

$$\hat{\varphi}_t(s) = \varphi_t(s) - \tau^2 s\, E[\exp(st_0)t_3]+O(\tau^3) =$$

$$= \varphi_t(s) - \frac{\tau^2}{2} s[p_1 s+p_2(3s+s^3)]\, \varphi(s) + O(\tau^3). \tag{A.14}$$

Dividing by $-s$, inverting the Fourier transform and using Lemma 2 we find'

$$\Pr(\hat{t} \le x) = \Pr(t \le x) + \frac{\tau^2}{2}(p_1+p_2x^2)xi_{T-n}(x) + O(\tau^3) =$$

$$= I_{T-n}(x) - \frac{\tau^2}{2}(p_1+p_2x^2)xi_{T-n}(x) + \frac{\tau^2}{2}(p_1+p_2x^2)xi_{T-n}(x) + O(\tau^3) =$$

$$= I_{T-n}(x) + O(\tau^3). \tag{A.15}$$

Working as in the proof of Lemma 2 we find that as $\delta \to 0$, $p_1=p_2=0$, so $\hat{t} = t$ is exactly distributed as a t variable with T-n degrees of freedom.

Proof of Lemma 3: Lemma 3 follows immediately from Rothenberg's (1984b) Proposition 2. The validity of this approximation is proved in Magdalinos (1986).

Proof of Lemma 4: Let $F_r(x)$ and $f_r(x)$ be the distribution and density functions respectively of a chi-square variable with r defrees of freedom. Also, let $F^r_{T-n}(x)$ and $f^r_{T-n}(x)$ be the distribution and density functions respectively of an F variable with r and T-n degrees of freedom. Then we can easily show that the Edgeworth expansions of the F distribution and density functions are

$$F^r_{T-n}(x) = F_r(rx) + \frac{\tau^2}{2}(r-2-rx)\, rxf_r(rx) + O(\tau^4),$$

and (A.16)

$$f^r_{T-n}(x) = rf_r(rx) + O(\tau^2).$$

From Lemma 3, (A.16), the definitions of the statistics (4.2) and (4.9) and the use of simple algebra we find

$$\Pr(v \le x) = F^r_{T-n}(x) - \tau^2(q_1+q_2x)x\, f^r_{T-n}(x) + O(\tau^3), \tag{A.17}$$

where

$$q_1 = h_1/r+(r-2)/2, \quad q_2 = h_2/(r+2)-r/2. \tag{A.18}$$

As $\delta \to 0$, γ becomes a fixed known vector, so

$$\Lambda = 0, \quad \lambda = \mu = 0, \quad \lambda_0 = 2, \quad \mu_0 = 0,$$

and (A.19)

$$h_1 = -r(r-2)/2, \quad h_2 = r(r+2)/2.$$

From (A.18) and (A.19) we have that

$$q_1 = q_2 = 0, \tag{A.20}$$

which means that the approximation is locally exact.

Proof of Theorem 2: It can be easily shown that under the null hypothesis (4.1) the F statistic (4.9) admits a stochastic expansion of the form

$$v = v_0 + \tau v_1 + \tau^2 v_2 + \omega(\tau^3), \tag{A.21}$$

where the first term in the expansion is

$$v_0 = b'Qb/r, \qquad b = GX'\Omega u/\sqrt{T}. \tag{A.22}$$

Therefore, the Cornish-Fisher corrected F statistic (4.13) admits a stochastic expansion of the form

$$\hat{v} = v_0 + \tau v_1 + \tau^2 (v_2 - v_3) + \omega(\tau^3), \tag{A.23}$$

where

$$v_3 = (q_1 + q_2 v_0) v_0. \tag{A.24}$$

Let s be an imaginary number, and let $\varphi_v(s)$, $\varphi_r(s)$ be the characteristic functions of the F statistic (4.9) and of a chi-square variable with r degrees of freedom respectively. From (A.23) it is implied that the characteristic function of the Cornish-Fisher corrected statistic $\hat{v}$ is

$$\hat{\varphi}_v(s) = \varphi_v(s) - \tau^2 sE[\exp(sv_0)v_3] + O(\tau^3) =$$

$$= \varphi_v(s) - \tau^2 s[q_1 \varphi_{r+2}(s/r) + q_2 \frac{r+2}{r} \varphi_{r+4}(s/r)] + O(\tau^3). \tag{A.25}$$

Also, it can be shown that

$$(rx)f_r(rx) = rf_{r+2}(rx), \qquad (rx)^2 f_r(rx) = r(r+2)f_{r+4}(rx). \tag{A.26}$$

Dividing (A.25) by $-s$, inverting the Fourier transform and using Lemma 4, (A.16), and (A.26) we find

$$\begin{aligned} \Pr(\hat{v} \le x) &= \Pr(v \le x) + \tau^2 [q_1 r f_{r+2}(rx) + q_2 \frac{r+2}{r} r f_{r+4}(rx)] + O(\tau^3) = \\ &= \Pr(v \le x) + \tau^2 [q_1 rx f_r(rx) + q_2 rx^2 f_r(rx)] + O(\tau^3) = \\ &= \Pr(v \le x) + \tau^2 (q_1 + q_2 x) rx f_r(rx) + O(\tau^3) = \\ &= F^r_{T-n}(x) - \tau^2 (q_1 + q_2 x) x f^r_{T-n}(x) + \tau^2 (q_1 + q_2 x) x f^r_{T-n}(x) + O(\tau^3) = \\ &= F^r_{T-n}(x) + O(\tau^3). \end{aligned} \tag{A.27}$$

Working as in the proof of Lemma 4 we find that as $\delta \to 0$, $q_1 = q_2 = 0$, so $\hat{v} = v$ is exactly distributed as an F variable with r and T-n degrees of freedom.

Proof of Proposition 1: It can be easily shown that

$$A_i = -\sum_{t=1}^{T} \sigma_t^{-4} z_{ti} x_t x_t' / T, \quad A_{ij}^* = \sum_{t=1}^{T} \sigma_t^{-6} z_{ti} z_{tj} x_t x_t' / T, \quad A_{ij} = 2A_{ij}^* \tag{A.28}$$

so that (5.9) are implied from the definition (3.4). Also note that the nxn positive semidefinite matrix P has r positive eigenvalues, so that

$$P = W\mathcal{L}W' = \sum_{i=1}^{r} \lambda_i w_i w_i' = \sum_{i=1}^{r} h_i h_i', \quad h_i = \lambda_i^{1/2} w_i, \tag{A.29}$$

where $\mathcal{L}$ is the diagonal matrix of eigenvalues of P, W is the matrix with columns the standardized eigenvectors of P, and λ_i ($i=1,\dots,r$) are the r positive eigenvalues of P. In matrix notation

$$P = W\mathcal{L}W' = (W\mathcal{L}^{1/2})(W\mathcal{L}^{1/2})' = YY', \qquad Y = W\mathcal{L}^{1/2}, \tag{A.30}$$

where Y is a $n \times r$ matrix with columns the vectors h_i. Then it can be shown that

$$x_t' P x_s = x_t' YY' x_s = \sum_{i=1}^{r} x_t' h_i h_i' x_s. \tag{A.31}$$

Using definition (4.3) and (A.31) we find that the j-th element of the vector c is

$$c_j = \mathrm{tr} A_j P = -\sum_{t=1}^{T} \sigma_t^{-4} z_{tj} x_t' P x_t / T = -\sum_{i=1}^{r} \sum_{t=1}^{T} \sigma_t^{-6} \left(h_i' x_t \right)^2 z_{tj} z_t' \gamma, \tag{A.32}$$

from which the first of (5.10) is implied. The rest of the proof is similar and omitted.

Proof of Proposition 2: We define the $T \times 1$ vectors $\bar{u}$, ε and $\bar{\varepsilon}$ with elements

$$\bar{u}_t = u_t^2 - \sigma_t^2,$$

$$\varepsilon_t = 2u_t e_t - \tau e_t^2, \qquad e_t = x_t' BX'u/\sqrt{T}, \tag{A.33}$$

$$\bar{\varepsilon}_t = 2u_t \bar{e}_t - \tau \bar{e}_t^2, \qquad \bar{e}_t = x_t' GX'\Omega u/\sqrt{T}.$$

Collecting terms of the same order of magnitude in (5.2) we find that the sampling error of $\hat{\gamma}_{GQ}$ is

$$\delta_*^{GQ} = \sqrt{T}\left(\hat{\gamma}_{GQ} - \gamma\right) = d_{*1}^{GQ} - \tau d_{*2}^{GQ},$$

where (A.34)

$$d_{*1}^{GQ} = \bar{B}Z'\bar{u}/\sqrt{T}, \quad d_{*2}^{GQ} = \bar{B}Z'\varepsilon/\sqrt{T}.$$

Expanding (5.3) in a Taylor series and collecting terms of the same order of magnitude we find that the sampling error of $\hat{\gamma}_A$ is

$$\delta_*^A = \sqrt{T}\left(\hat{\gamma}_A - \gamma\right) = d_{*1}^A - \tau d_{*2}^A + \omega\left(\tau^2\right),$$

where (A.35)

$$d_{*1}^A = \bar{G}\left(Z'\Omega^2\bar{u}/\sqrt{T}\right),$$

$$d_{*2}^A = \bar{G}\left(Z'\Omega^2\varepsilon/\sqrt{T}\right) - 2\sum_{i=1}^{k}\bar{G}\left(Z'\Omega\Omega_i\bar{u}/\sqrt{T}\right)d_{1i}^{GQ} + 2\sum_{i=1}^{k}\overline{GA_i G}\left(Z'\Omega^2\bar{u}/\sqrt{T}\right)d_{1i}^{GQ},$$

and d_{1i}^{GQ} is the i-th element of the vector d_{*1}^{GQ}. Similarly, from (5.4) we find that the sampling error of $\hat{\gamma}$ for all α is

$$\delta_*^\alpha = \sqrt{T}\left(\hat{\gamma}_\alpha - \gamma\right) = d_{*1}^A - \tau d_{*2}^\alpha + \omega\left(\tau^2\right),$$

where (A.36)

$$d_{*1}^A = \bar{G}\left(Z'\Omega^2\bar{u}/\sqrt{T}\right),$$

$$d_{*2}^\alpha = \bar{G}\left(Z'\Omega^2\bar{\varepsilon}/\sqrt{T}\right) - 2\sum_{i=1}^{k}\bar{G}\left(Z'\Omega\Omega_i\bar{u}/\sqrt{T}\right)d_{1i}^A + 2\sum_{i=1}^{k}\overline{GA_i G}\left(Z'\Omega^2\bar{u}/\sqrt{T}\right)d_{1i}^A,$$

and d_{1i}^{A} is the i-th element of the vector d_{*1}^{A}. Since, as $\alpha \to \infty$ the IA estimators converge to the ML estimator, it is clear that the sampling error of the ML estimator admits a stochastic expansion that is the same with (A.36).

It is easily shown that

$$E\left(d_{*1}^{GQ}\right)=E\left(d_{*1}^{A}\right)=0,\quad E\left(d_{*2}^{GQ}\right)=\overline{B}\xi,\quad E\left(d_{*2}^{\alpha}\right)=\overline{G}\xi_2,$$

$$E\left(d_{*2}^{A}\right)=\overline{G}\xi_1+4\sum_{i=1}^{k}G\left[\overline{A}_i\overline{g}_i-\left(Z'\Omega_i\Omega^{-1}Z/T\right)\overline{b}_i\right], \tag{A.37}$$

$$E\left(d_{*1}^{GQ}d_{*1}^{GQ\,\prime}\right)=2\overline{B\Gamma B},\quad E\left(d_{*1}^{A}d_{*1}^{A\,\prime}\right)=2\overline{G},$$

where $\overline{g}_i$ is the i-th coloumn of the matrix $\overline{G}$, $\overline{b}_i$ is the i-th coloumn of the matrix $\overline{B}$, and $\overline{A}_i = (Z'\Omega\Omega_i Z/T)$. Using the expectations (A.37) we prove the formulae (5.12), (5.13), and (5.14) for Λ and μ.

From (2.3) we take that

$$\hat{\sigma}^2=\left[u'\hat{\Omega}u-b'\left(X'\hat{\Omega}X/T\right)b+\omega(\tau)\right]/(T-n), \tag{A.38}$$

where

$$b=GX'\Omega u/\sqrt{T}=\sqrt{T}\left(\hat{\beta}-\beta\right)+\omega(\tau), \tag{A.39}$$

and $\hat{\beta}$ is the feasible GLS estimator of β. We define the scalars

$$w_0=\sqrt{T}\left(u'\Omega u/T-1\right),\quad w_i=\sqrt{T}\left(u'\Omega_i u/T+a_i\right),$$

where (A.40)

$$a_i=-E\left(u'\Omega_i u/T\right)=\sum_{t=1}^{T}\sigma_t^{-2}z_{ti}/T,$$

and the $k\times 1$ vectors w, a with elements w_i, a_i respectively. Expanding (A.38) in a Taylor series and collecting terms of the same order of magnitude we find that, for any estimator of γ whose sampling error admits a stochastic expansion of the form (A.34) - (A.36), we have

$$\delta_0 = \sqrt{T}\left(\hat{\sigma}^2 - 1\right) = \sigma_0 + \tau\sigma_1 + \omega\left(\tau^2\right),$$

where (A.41)

$$\sigma_0 = w_0 - a'd_{*1}, \quad \sigma_1 = w'd_{*1} + d_{*1}'\bar{A}d_{*1} + a'd_{*2} - b'Ab + n.$$

For all the estimators of γ we have that

$$E(w_0 d_{*1}) = 2\gamma, \qquad E(w'd_{*1}) = -2k. \tag{A.42}$$

Using the expectations (A.42) we find

$$\begin{aligned} \lambda_0 &= 2 - 4a'\gamma + a'\Lambda a, \\ \lambda &= 2\gamma - \Lambda a, \\ \mu_0 &= \mathrm{tr}\left(\bar{A}\Lambda\right) - 2k - a'\mu. \end{aligned} \tag{A.43}$$

Substituting $a = \bar{A}\gamma$ and the values of Λ and μ for the different estimators of γ we complete the proof.

REFERENCES

Amemiya T. (1977), "A Note on a Heteroscedastic Model", *Journal of Econometrics*, 6, 365-370; and "Corrigenta", *Journal of Econometrics*, 8, 275.

Breusch T. S. (1980), "Useful Invariance Results for Generalized Regression Models", *Journal of Econometrics*, 13, 327-340.

Cornish E. A. and R. A. Fisher (1937), "Moments and Cumulants in the Specification of Distributions", *Revue de l' Institute International de Statistic*, 4, 1-14.

Dhrymes P. J. (1969), "Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3SLS Estimated Parameters", *The Review of Economic Studies*, 36, 213 -226.

_________(1971), "Equivalence of Iterative Aitken and Maximum Likelihood Estimators for a System of Regression Equations", *Australian Economic Papers*, 10, 20-24.

_________(1978), *Introductory Econometrics* (New York: Springer Verlang).

Durbin J. (1973), "Distribution Theory for Tests Based on the Sample Distribution Function", *Society for Industrial and Applied Mathematics*, Philadelphia, Pa, Section 2.4.

Fisher R. A. (1925), "Expansion of Student's Integral in Powers of n^{-1}", *Metron*, 5, 109-112.

Fisher R. A. and E. A. Cornish (1960), "The Percentile of Distributions Having Known Cumulants", *Technometrics*, 2, 209-225.

Goldfeld S. M. and R. E. Quandt (1972), *Nonlinear Methods in Econometrics* (Amsterdam: North Holland).

Hildreth C. and J. P. Houck (1968), "Some Estimators for a Linear Model with Random Coefficients", *Journal of the American Statistical Association*, 63, 584-595.

Hill G. W. and A. W. Davis (1968), "Generalized Asymptotic Expansions of the Cornish-Fisher Type", *Annals of Mathematical Statistics*, 39, 1268-1273.

Kunimoto N., K. Morimune and Y. Tsukuda (1983), "Asymptotic Expansions of the Distribution of the Test Statistic for Overidentifying Restrictions in a System of Simultaneous Equations", *International Economic Review*, 24, 199-215.

Magdalinos M. A. (1983), *"Applications of the Refined Asymptotic Theory in Econometrics"*, Ph. D. Thesis, University of Southampton.

_________(1985), "Improving some Instrumental Variables Test Procedures", *Econometric Theory*, 1, 240-261.

_________(1986), *Stochastic Expansions and Asymptotic Approximations*, Discussion Paper 36, Athens School of Economics.

Magee L. (1989), "An Edgeworth Test Size Correction for the Linear Model with AR (1) Errors", *Econometrica*, 57, 661-674.

McDonald G. C. and D. I. Galarneau (1975), "A Monte Carlo Evaluation of some Ridge-Type Estimators", *Journal of the American Statistical Association*, 70, 407-416.

Morimune K. and Y. Tsukuda (1984), "Testing a Subset of Coefficients in a Structural Equation", *Econometrica*, 52, 427-448.

Rothenberg T. J. (1982), "Comparing Alternative Asymptotically Equivalent Tests", in Hildenbrand W. (ed.), *Advances in Econometrics* (Cambridge: University Press).

_________(1984a), "Approximate Normality of Generalized Least Squares Estimates", *Econometrica*, 52, 811-825.

_________(1984b), "Hypothesis Testing in Linear Models when the Error Covariance Matrix is Nonscalar", *Econometrica*, 52, 827-842.