Title
Approximate Power Functions for Some Robust Tests of Regression Coefficients

Author
Rothenberg, Thomas J.

Publication Date
1984

Peer reviewed

APPROXIMATE POWER FUNCTIONS FOR SOME ROBUST TESTS

OF REGRESSION COEFFICIENTS

Thomas J. Rothenberg

Department of Economics
University of California, Berkeley

Research Papers in Economics No. 84-1

Summary

Edgeworth expansions are developed for the distribution functions of some test statistics in the normal linear regression model where the error covariance matrix is unknown. Tests based on generalized least squares estimates and also on ordinary least squares estimates are considered. In both cases, adjustments to the asymptotic critical values are found and approximate local power calculated. The approximations are applied to a number of examples, including heteroscedasticity and autocorrelation.

APPROXIMATE POWER FUNCTIONS FOR SOME ROBUST TESTS OF REGRESSION COEFFICIENTS

Thomas J. Rothenberg*

1. INTRODUCTION

In the linear statistical model, tests on individual regression coefficients can be contructed using the familiar t-statistic. If the errors are normal, independent, and homoscedastic, these tests have the correct size and possess optimal power properties. In practice, however, we are rarely sure about the proper specification of the errors. Since the t-test is not robust to misspecification of the error distribution, alternative tests based on a more richly parameterized model are often considered. In the present paper we examine the properties of a family of tests in the normal regression model when the error covariance matrix is unknown. The evaluations are based on second-order asymptotic approximations to the sampling distributions for local alternatives.

Let X be a known $n \times K$ nonrandom design matrix having full column rank. The random vector $y = (y_1,\ldots,y_n)'$ is normal with mean $X\beta$ and covariance matrix Σ. The K-dimensional vector of regression coefficients β and the $n \times n$ matrix Σ are both unknown. The problem is to use the observed value of y to test the null hypothesis H_0: $c'\beta = c'\beta_0$ against the one-sided alternative H_A: $c'\beta > c'\beta_0$, for given K-dimensional vectors c and β_0.

If a reasonable estimate $\hat{\beta}$ is available and one can estimate its variance, it is natural to construct a Wald-type test statistic of the form

$$T = \frac{c'(\hat{\beta} - \beta_0)}{(\text{est var } c'\hat{\beta})^{1/2}} \tag{1.1}$$

and to reject if T takes on large values. An example is the least squares t-ratio

$$(1.2) \qquad T_0 = \frac{c'(X'X)^{-1}X'y - c'\beta_0}{[s^2c'(X'X)^{-1}c]^{1/2}}$$

where s^2 is the residual sum of squares divided by $n - K$. If Σ is a scalar matrix, the appropriate critical value for T_0 can be calculated from the Student t distribution. If Σ is not scalar, the variance estimate used in T_0 is biased and the t-test is invalid. To protect against this possibility, the econometrician might postulate that Σ lies in some family of covariance matrices $\mathcal{S}$ and then construct a test having (approximately) the right size for any Σ in $\mathcal{S}$. In particular, let θ be a p-dimensional parameter vector taking values in Θ, some open set in p-dimensional Euclidean space. Let $\Sigma(\theta)$ be a smooth function such that $\mathcal{S}$ is the range of $\Sigma(\theta)$ as θ varies over Θ. The econometrician is assumed to construct a test statistic of the form (1.1) using the parametric model

$$(1.3) \qquad y = X\beta + u, \qquad u \sim N[0,\Sigma(\theta)], \qquad \theta \in \Theta \quad .$$

It is useful to distinguish two different cases depending on how large p is:

<u>Case I</u>. The parameter vector θ has low dimension (p much smaller than the sample size n); if the parametric model is correct, θ can be well estimated by $\hat{\theta}$ (for example, by maximum likelihood or by a regression on least squares residuals). The seemingly unrelated regressions model and models where the errors follow a low-order autoregression are typical examples of this case. An obvious test statistic is the "studentized" generalized least squares (GLS) estimate

$$T_1 = \frac{c'(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y - c'\beta_0}{[c'(X'\hat{\Sigma}^{-1}X)^{-1}c]^{1/2}} \tag{1.4}$$

where $\hat{\Sigma} \equiv \Sigma(\hat{\theta})$. (Note that the usual t-statistic T_0 is a special case where $p = 1$ and $\Sigma = \theta I$.)

Case II. The parameter vector θ has high dimension (p approximately equal to n); all of the elements of θ cannot generally be well estimated. The heteroscedastic model where the errors are independent but nothing is known about the variances is one example. The autocorrelation model where the u_i are a stationary stochastic process but nothing is known about the autocorrelation function is another example. With such a large family $\mathcal{S}$, it might be unwise to attempt generalized least squares. However, the ordinary least squares (OLS) estimate is normal with mean β and covariance matrix $(X'X)^{-1}X'\Sigma X(X'X)^{-1}$. Even though θ cannot be well estimated, there often exists a sample matrix $\tilde{\Sigma}$ such that $X'\tilde{\Sigma}X/n$ is a good estimate of $X'\Sigma X/n$. If so, a possible test statistic is

$$T_2 = \frac{c'(X'X)^{-1}X'y - c'\beta_0}{[c'(X'X)^{-1}X'\tilde{\Sigma}X(X'X)^{-1}c]^{1/2}} \; . \tag{1.5}$$

This idea of constructing an OLS based test whose (asymptotic) size is robust under quite general unknown covariance structure is explored in White (1980).

These two cases are not exhaustive. In particular, a third case where p is large but still considerably less than n is of some importance in econometrics. For example, in the autocorrelation model where the error spectrum is known to be smooth, Hannan (1963) and Amemiya (1973) present asymptotically efficient GLS estimates of β under the assumption that both p and n/p tend to infinity. Tests based on such estimates are worth exploring. Unfortunately, the method of asymptotic expansions employed in this paper does not apply to these tests. Hence, we shall confine our attention to the two cases described above.

The choice of an appropriate model for Σ is difficult since in practice little is known about the error process. If one chooses a low-dimensional parametric family $\Sigma(\theta)$ and the true error covariance matrix is not contained in it, the resulting GLS test will have the wrong size, even when the sample is large. If one chooses a high-dimensional family as in Case II, one may guarantee the correct size (at least for large samples) but at the cost of reduced power. Optimal choice necessarily depends on the strength of one's convictions about the error structure. But quantifying the size error and power loss in a number of typical situations can certainly be an aid.

In most situations the exact probability distributions of test statistics like T_1 and T_2 are difficult to derive. Critical regions are usually justified by asymptotic arguments and calculated from limiting normal distributions. When the sample is large, the use of an estimated Σ matrix in (1.4) and (1.5) instead of the true value typically produces negligible error. It is not so obvious, however, that this error can safely be ignored when samples are small. If the normal critical value is used to form a rejection region, the actual size of the test may differ substantially from the desired level Furthermore, the actual power of the test may be considerably lower than that predicted by ignoring the randomness of $\hat{\Sigma}$ and $\tilde{\Sigma}$. In the following sections we sketch a method for obtaining higher-order asymptotic approximations to the distributions of test statistics like T_1 and T_2. We calculate adjusted critical values that yield tests having the correct size to order n^{-1}. In addition, we compute approximate local power functions for the size-adjusted tests. The approximations are derived in Sections 2-4 for quite general covariance structures. The resulting formulae are then evaluated for some simple cases in Sections 5 and 6. For moderate sample sizes, the second-order approximations are found to differ considerably from the first-order asymptotic approximations. In particular, the size of a robust test based on T_1 or T_2 is often much larger than its nominal value. The loss in power resulting from using an estimated Σ instead of its true value depends very much on the particular problem. For example, the power loss is typically large for the robust OLS test which protects against error dependence; the loss is surprisingly low for the OLS test which protects against heteroscedasticity.

The use of higher-order asymptotic expansions under local alternatives to evaluate testing procedures is discussed in the survey papers by Bickel (1974), Pfanzagl (1980), and Rothenberg (1982, 1984a). The approach has been applied to the problem of testing regression coefficients in models with unknown error covariance matrix in Rothenberg (1984b), where quite general multiparameter Wald, likelihood ratio, and Lagrange multiplier tests are compared. Unfortunately, due to the complexity of the asymptotic expansions in terms of noncentral chi square distributions, the resulting approximate power functions are difficult to interpret. In the present paper, we concentrate on one-dimensional hypotheses and develop relatively simple approximations based on the normal distribution. Our approach is similar to that employed by Albers (1978) and Durbin (1983) in investigating the effects of autocorrelation. Indeed, our results can be viewed as extensions of theirs to a broader class of models.

2. DISTRIBUTION THEORY: CASE I

The GLS test statistic using the true error covariance matrix,

$$(2.1) \qquad \bar{T}_1 = \frac{c'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y - c'\beta_0}{[c'(X'\Sigma^{-1}X)^{-1}c]^{1/2}} ,$$

is normal with unit variance and mean

$$(2.2) \qquad \delta_1 = \frac{c'(\beta - \beta_0)}{[c'(X'\Sigma^{-1}X)^{-1}c]^{1/2}} .$$

Denoting the standard normal distribution function by $\Phi(\cdot)$ and its upper α percentage point t_α, we can write the power function for the size-α critical region $\bar{T}_1 > t_\alpha$ as

$$(2.3) \qquad \Pr[\bar{T}_1 > t_\alpha] = \Phi(\delta_1 - t_\alpha).$$

We are interested in the distribution of the feasible test statistic T_1 where Σ has been estimated from the data on the basis of the low-dimensional parametric model (1.3). We shall assume that the true covariance matrix lies in $\mathcal{S}_1$, the range of $\Sigma(\cdot)$ over the parameter space $\Theta \subset R^p$. Suppose that the model is identified so that θ is the unique vector in Θ yielding the true covariance matrix Σ. Then, if $\hat{\theta}$ is a good estimate of θ, we would expect T_1 to behave rather like $\bar{T}_1$ and to have a similar distribution function.

To make this idea more precise and to develop better approximations to the distribution of T_1, it is useful to embed our problem in a sequence indexed by the sample size and to examine the limiting behavior of our sample statistics as n tends to infinity. We let the parameter space Θ and the parameter vectors δ_1 and θ be fixed and consider a sequence of estimates $\hat{\theta}_n$, $n \times K$ design matrices X_n, and $n \times n$ covariance matrices Σ_n such that $n^{-1}X_n'\Sigma_n^{-1}X_n$ and $n^{-1}X_n'\hat{\Sigma}_n^{-1}X_n$ stay positive definite and (stochastically) bounded. For this sequence, we develop an order n^{-1} Edgeworth approximation to the distribution of T_{1n}.

The assumption that δ_1 stays fixed in our sequence of problems requires some justification. For any given β in H_A, δ_1 is a function of $X_n'\Sigma_n^{-1}X_n$ and will tend to infinity as the sample size grows. Our interest, however, is in comparing tests in situations where power is moderate--say, near one-half--which implies a value of δ_1 near t_α. Thus, it seems reasonable to hold δ_1 fixed even when letting n tend to infinity. This is equivalent to treating β as a function of n and considering a sequence of local alternatives $c'\beta_n$ tending to $c'\beta_0$. Local asymptotic analysis should give reasonable approximations for that portion of thepower curve we are most interested in.

To verify that the approximations we develop actually have errors $o(n^{-1})$, numerous regularity conditions must be imposed on the sequence of estimates $\hat{\theta}_n$, design matrices X_n, and covariance matrices Σ_n. These conditions are discussed in Rothenberg (1984a, 1984b) and will not be repeated here. Since our main purpose is to obtain useful results for applications, we shall not attempt to establish rigorously the order of the approximation error, but simply sketch the formal development of the Edgeworth expansion. To avoid cumbersome notation, we shall drop the subscript n in the following derivations.

By definition, the test statistic T_1 can be written as

$$(2.4) \qquad T_1 = \frac{\bar{T}_1 + n^{-1/2}Z}{(1 + n^{-1/2}S)^{1/2}}$$

where S and Z are the standardized random variables

$$S = \sqrt{n}\,\frac{c'(X'\hat{\Sigma}^{-1}X)^{-1}c - c'(X'\Sigma^{-1}X)^{-1}c}{c'(X'\Sigma^{-1}X)^{-1}c}$$

(2.5)

$$Z = \sqrt{n}\,\frac{c'(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y - c'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y}{[c'(X'\Sigma^{-1}X)^{-1}c]^{1/2}} .$$

The distribution of T_1 depends on the particular estimate used for $\hat{\theta}$. In practice, all of the commonly used estimates (including the MLE and those obtained by a regression on least-squares residuals) have the property that they are even functions of u and are invariant to changes in β. (See, for example, Breusch (1980).) As shown in Rothenberg (1984a), this invariance implies that the pair of random variables (S,Z) is distributed independently of $\bar{T}_1$. When p is small and the mapping $\Sigma(\cdot)$ defining $\mathcal{S}_1$ is smooth, the standardized estimator $\sqrt{n}(\hat{\theta} - \theta)$ typically has a well-behaved limiting distribution as do S and Z. In fact, under reasonable regularity conditions, the joint distribution of (S,Z) can be approximated with error $o(n^{-1})$ by a distribution possessing bounded moments of the form[1/]

$$E(S) = \frac{m(\theta)}{\sqrt{n}} + O(n^{-1}), \qquad E(Z) = 0, \qquad Cov(S,Z) = O(n^{-1})$$

(2.6)

$$Var(S) = V_S(\theta) + O(n^{-1}), \qquad Var(Z) = V_Z(\theta) + O(n^{-1})$$

where V_S, V_Z, and m are smooth functions of θ. Then the algorithm given in Appendix A can be employed to obtain a formal order n^{-1} Edgeworth approximation to the distribution of T_1. For any Σ in $\mathcal{S}_1$ and any nonrandom scalar t

$$\Pr[T_1 \le t] \simeq \Phi\left[t\left(1 - \frac{A_1(t,\theta)}{2n}\right) - \delta_1\left(1 - \frac{B_1(t,\theta)}{2n}\right)\right] \tag{2.7}$$

where

$$\begin{aligned} A_1(t,\theta) &= \frac{1}{4}(1 + t^2)V_S(\theta) + V_Z(\theta) - m(\theta) \\ B_1(t,\theta) &= \frac{1}{4}\, t^2 V_S(\theta) + V_Z(\theta) \ . \end{aligned} \tag{2.8}$$

The approximate distribution (2.7) depends on the asymptotic variances of S and Z and on the $n^{-1/2}$ term in the expansion of the mean of S. Explicit expressions for these moments are easily obtained. Denote the derivative matrices $\Sigma_i \equiv \partial\Sigma/\partial\theta_i$ and $\Sigma_{ij} \equiv \partial^2\Sigma/\partial\theta_i\partial\theta_j$ for $i,j = 1,\ldots,p$; define

$$d = \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}c \ , \qquad D = \Sigma^{-1} - \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} \ . \tag{2.9}$$

Then, by Taylor expansion around the true value θ, S and Z can be approximated by

$$\begin{aligned} S &\simeq \sum_i \frac{d'\Sigma_i d}{d'\Sigma d}\sqrt{n}(\hat\theta_i - \theta_i) + \frac{1}{2\sqrt{n}}\sum_i\sum_j \frac{d'(\Sigma_{ij} - 2\Sigma_i D\Sigma_j)d}{d'\Sigma d}\, n(\hat\theta_i - \theta_i)(\hat\theta_j - \theta_j) \\ Z &\simeq -\sum_i \frac{d'\Sigma_i Du}{(d'\Sigma d)^{1/2}}\sqrt{n}(\hat\theta_i - \theta_i) \ . \end{aligned} \tag{2.10}$$

Suppose $\sqrt{n}(\hat\theta - \theta)$ is asymptotically normal with $p \times p$ covariance matrix $\Lambda(\theta) = [\lambda_{ij}(\theta)]$ and that the approximate distribution of $\hat\theta_i$ has mean

$$E(\hat\theta_i) = \theta_i + \frac{\mu_i(\theta)}{n} + O(n^{-1}) \qquad i = 1,\ldots,p \ . \tag{2.11}$$

The required moments are then given by[2/]

$$v_S(\theta) = \sum_i \sum_j \lambda_{ij} \frac{d'\Sigma_i d}{d'\Sigma d} \frac{d'\Sigma_j d}{d'\Sigma d}$$

$$v_Z(\theta) = \sum_i \sum_j \lambda_{ij} \frac{d'\Sigma_i D \Sigma_j d}{d'\Sigma d} \tag{2.12}$$

$$m(\theta) = \sum_i \mu_i \frac{d'\Sigma_i d}{d'\Sigma d} + \frac{1}{2} \sum_i \sum_j \lambda_{ij} \frac{d'(\Sigma_{ij} - 2\Sigma_i D \Sigma_j) d}{d'\Sigma d} \quad .$$

Under appropriate smoothness conditions these moments can be well estimated for large n by replacing θ by $\hat{\theta}$.

3. DISTRIBUTION THEORY: CASE II

The OLS test statistic using the true error covariance matrix,

$$(3.1) \qquad \bar{T}_2 = \frac{c'(X'X)^{-1}X'y - c'\beta_0}{[c'(X'X)^{-1}X'\Sigma X(X'X)^{-1}c]^{1/2}} ,$$

is normal with unit variance and mean

$$(3.2) \qquad \delta_2 = \frac{c'(\beta - \beta_0)}{[c'(X'X)^{-1}X'\Sigma X(X'X)^{-1}c]^{1/2}} .$$

The power function for the size-α critical region $\bar{T}_2 > t_\alpha$ is

$$(3.3) \qquad \Pr[\bar{T}_2 > t_\alpha] = \Phi(\delta_2 - t_\alpha) .$$

If the column space of X is spanned by K characteristic vectors of Σ, the test statistics $\bar{T}_1$ and $\bar{T}_2$ are identical; otherwise, $\delta_1 > \delta_2$ and the GLS test is uniformly more powerful than the OLS test.

Again, we are interested in the distribution of the feasible test statistic T_2 where now Σ has been estimated from the data using a high-dimensional parametric model. By definition

$$(3.4) \qquad T_2 = \frac{\bar{T}_2}{\left(1 + \frac{W}{\sqrt{n}}\right)^{1/2}}$$

where W is the standardized variance estimate

$$(3.5) \qquad W = \sqrt{n}\, \frac{c'(X'X)^{-1}X'(\tilde{\Sigma} - \Sigma)X(X'X)^{-1}c}{c'(X'X)^{-1}X'\Sigma X(X'X)^{-1}c} .$$

Let $\mathscr{S}_2$ be the set of n × n covariance matrices Σ such that $X'\tilde{\Sigma}X/n$ is a well-behaved estimate of $X'\Sigma X/n$. To make this notion precise, we embed our problem in a sequence indexed by the sample size and examine the limiting behavior of our statistics. Again, we let β depend on n so that δ_2 is fixed and consider a sequence of estimates $\tilde{\Sigma}_n$, true covariance matrices Σ_n, and design matrices X_n such that $X_n'\tilde{\Sigma}_n X_n/n$ is stochastically bounded. We assume that, for all Σ_n in the set $\mathcal{S}_2$, the joint probability distribution of $(\bar{T}_{2n}, W_n)$ can be approximated with error $o(n^{-1})$ by a distribution possessing bounded moments. Typically, the conditional moments of this approximate distribution take the form

$$\sqrt{n}E(W|\bar{T}_2) = a[(\bar{T}_2 - \delta_2)^2 - 1] + b + O(n^{-1/2})$$

(3.6)

$$\mathrm{Var}(W|\bar{T}_2) = V_W + O(n^{-1})$$

where a, b, and V_W depend on X and Σ, but not on $\bar{T}_2$ or δ_2. (Again, we drop the subscript n to simplify notation.) Then, applying the algorithm of Appendix A, we find the formal order n^{-1} Edgeworth approximation

$$(3.7) \qquad \Pr[T_2 < t] \simeq \Phi\left[t\left(1 - \frac{A_2(t,\Sigma)}{2n}\right) - \delta_2\left(1 - \frac{B_2(t,\delta_2,\Sigma)}{2n}\right)\right]$$

for arbitrary nonrandom t, where

$$A_2(t,\Sigma) = \frac{1}{4}(1 + t^2)V_W - a(t^2 - 1) - b$$

(3.8)

$$B_2(t,\delta_2,\Sigma) = \frac{1}{4}t^2V_W + at(\delta_2 - 2t) \quad .$$

When $\tilde{\Sigma}$ is a function of the ordinary least-squares regression residuals, the parameters a, b, and V_W depend only on a few basic quantities. Define the $n \times n$ matrices

$$(3.9) \qquad M = I - X(X'X)^{-1}X' , \qquad Q = n(M\Sigma M - \Sigma)$$

and the n dimensional vectors

$$(3.10) \qquad x = nX(X'X)^{-1}c , \qquad z = \frac{M\Sigma x}{\sqrt{x'\Sigma x/n}} .$$

The least-squares residual vector $\hat{u}$ can be decomposed into the sum of two terms, one proportional to $\bar{T}_2 - \delta_2 \equiv Y$ and the other independent of $\bar{T}_2$:

$$\begin{aligned} \hat{u} &= y - X(X'X)^{-1}X'y = Mu \\ (3.11) \qquad &= \frac{M\Sigma x}{x'\Sigma x} x'u + M\left(I - \frac{\Sigma x x'}{x'\Sigma x}\right)u \\ &\equiv z\frac{Y}{\sqrt{n}} + e . \end{aligned}$$

The vector e is normal and uncorrelated with the N(0,1) variable Y. A similar decomposition can often be found for the standardized variance estimate W. We develop explicit formulae for two important special cases.

(a) Heteroscedasticity Suppose the errors are known to be independent, but the variances are totally unknown. Then, Σ is a diagonal matrix with typical element $\sigma_i^2 = E(u_i^2)$ and a natural choice for $\tilde{\Sigma}$ is the diagonal matrix with typical element $\hat{u}_i^2$. We find

$$\begin{aligned} W &= \sqrt{n}\,\frac{\sum x_i^2(\hat{u}_i^2 - \sigma_i^2)}{\sum x_i^2\sigma_i^2} \\ (3.12) \qquad &= \sqrt{n}\,\frac{\sum x_i^2(e_i^2 - \sigma_i^2)}{\sum x_i^2\sigma_i^2} + \frac{\sum x_i^2 z_i^2}{\sum x_i^2\sigma_i^2}\frac{Y^2}{\sqrt{n}} + 2Y\frac{\sum x_i^2 z_i e_i}{\sum x_i^2\sigma_i^2} . \end{aligned}$$

The conditional moments of W given $\bar{T}_2$ take the form (3.6) with

$$(3.13) \qquad a = \frac{\sum x_i^2 z_i^2}{\sum x_i^2 \sigma_i^2}\,, \qquad b = \frac{\sum x_i^2 q_{ii}}{\sum x_i^2 \sigma_i^2}\,, \qquad V_W = 2n \frac{\sum x_i^4 \sigma_i^4}{(\sum x_i^2 \sigma_i^2)^2}\,.$$

Under reasonable regularity conditions on the sequences of x_i, z_i, and σ_i, the coefficients a, b, and V_W are $O(n^0)$ and can be well estimated (when n is large) by replacing σ_i^2 by $\hat{u}_i^2$ and by replacing σ_i^4 by $\frac{1}{3}\hat{u}_i^4$ in (3.9), (3.10), and (3.13).

<u>(b) Autocorrelation</u> Suppose the errors are known to be stationary with unknown autocovariances $\gamma_k = E(u_t u_{t+k})$. The error covariance matrix Σ has elements $\sigma_{ij} = \gamma_{|i-j|}$ and a natural choice for $\tilde{\Sigma}$ has elements $\tilde{\sigma}_{ij} = \hat{\gamma}_{|i-j|}$ for some set of estimated autocovariances $\hat{\gamma}_k$. A reasonable estimate is

$$(3.14) \qquad \begin{aligned} \hat{\gamma}_k &= \frac{1}{n-k} \sum_{t=1}^{n-k} \hat{u}_t \hat{u}_{t+k} \\ &= \frac{1}{n-k} \sum_{t=1}^{n-k} \left[e_t e_{t+k} + z_t z_{t+k} \frac{Y^2}{n} + (z_t e_{t+k} + z_{t+k} e_t) \frac{Y}{\sqrt{n}} \right] \end{aligned}$$

for $k = 0,1,\ldots,n-1$. Define $r_k = \frac{1}{n} \sum_t x_t x_{t+k}$, $\bar{r}_k = \frac{1}{n-k} \sum_t z_t z_{t+k}$, and $\tilde{\gamma}_k = \frac{1}{n-k} \sum_t e_t e_{t+k}$ for $k = 0,1,\ldots,n-1$. It will be convenient to define these autocorrelations for negative k by setting $\hat{\gamma}_k = \hat{\gamma}_{-k}$, $r_k = r_{-k}$, etc. Then, using the convention that these autocorrelations are zero for $|k| \geq n$, we can write

$$(3.15) \qquad \begin{aligned} W &= \sqrt{n} \frac{\Sigma r_k (\hat{\gamma}_k - \gamma_k)}{\Sigma r_k \gamma_k} \\ &= \sqrt{n} \frac{\Sigma r_k (\tilde{\gamma}_k - \gamma_k)}{\Sigma r_k \gamma_k} + \frac{\Sigma r_k \bar{r}_k}{\Sigma r_k \gamma_k} \frac{Y^2}{\sqrt{n}} + \frac{\varepsilon Y}{\sqrt{n}} \end{aligned}$$

where the summation runs from $-\infty$ to ∞. The $\tilde{\gamma}_k$ and ε are, by construction, independent of Y. Hence, $\sqrt{n}\, E(W|\, Y)$ has the form (3.6) with

$$(3.16) \qquad a = \frac{\Sigma r_k \bar{r}_k}{\Sigma r_k \gamma_k}\ , \qquad\qquad b = \sqrt{n}\, E(W) = \frac{\Sigma r_k \bar{q}_{kk}}{\Sigma r_k \gamma_k}$$

where $\bar{q}_{kk} = \mathrm{tr}(X'X)^{-1}X'X_{-k}(X'X)^{-1}X'\Sigma X - 2\ \mathrm{tr}(X'X)^{-1}X'\Sigma X_{-1}$ is the (approximate) bias of $n\hat{\gamma}_k$. (The lagged crossproduct matrices $X'X_{-k}$ and $X'\Sigma X_{-k}$ are formed in the usual manner by summing over the $n - |k|$ common observations.)

It is not at all obvious that W has a limiting distribution. Indeed, in the simplest special case where X is a single column of ones and the u_t are white noise, one can verify that W blows up with increasing sample size. If, however, both the r_k and the γ_k die off rapidly as k increases, it can be demonstrated that W is stochastically bounded and that the coefficients a and b are $O(n^0)$. Suppose the r_k and γ_k are square summable as n tends to infinity. Then the (limiting) convolution of the two autocovariance series

$$(3.17) \qquad c_k = \sum_{j=-\infty}^{\infty} r_j \gamma_{j+k}$$

is also square summable. From the approximate moments given by Fuller (1976, p. 239), we have

$$\mathrm{Var}[\sum_i r_i \sqrt{n}(\tilde{\gamma}_i - \gamma_i)] = 2 \sum_i \sum_j \sum_k r_i r_j \gamma_{i+k} \gamma_{j+k} + o(n^0).$$

Rearranging the terms in the sum and letting n tend to infinity, we find

$$(3.18) \qquad V_W = \frac{2\sum c_k^2}{(\sum r_k \gamma_k)^2}\ .$$

The asymptotic moments of W have somewhat simpler representation in the frequency domain. Let $f(\lambda)$ be the spectral density function for the error process u; that is, f is the Fourier transform of the autocovariance sequence γ_k. Suppose, when n is large, the Fourier transform of the autocovariance sequence r_k behaves like the (continuous) spectral density function $g(\lambda)$. Likewise, suppose the Fourier transform of the $\bar{r}_k$ behaves like the continuous spectral density $\bar{g}(\lambda)$. Then,

$$a = \frac{\int_{-\pi}^{\pi} \bar{g}(\lambda) g(\lambda)\, d\lambda}{\int_{-\pi}^{\pi} f(\lambda) g(\lambda)\, d\lambda}, \qquad V_W = \frac{2\int_{-\pi}^{\pi} f^2(\lambda) g^2(\lambda)\, d\lambda}{\left[\int_{-\pi}^{\pi} f(\lambda) g(\lambda)\, d\lambda\right]^2}.$$

A similar representation is available for b. Indeed, our procedure for forming T_2 in the autocorrelated case is essentially equivalent to transforming the model $y = X\beta + u$ to the frequency domain and treating it as a heteroscedastic model. The parameters a, b, and V_W are given by (3.13), where the variables are reinterpreted as discrete Fourier transforms of the original variables and Σ is reinterpreted as the diagonal matrix of spectral elements.[3/]

4. SIZE-CORRECTED TESTS AND THEIR DEFICIENCIES

The approximate distribution functions (2.7) and (3.7) can be employed to construct critical regions having the correct size to order n^{-1}. Consider the critical value

$$(4.1) \qquad t_1 = t_\alpha\left(1 + \frac{\hat{A}_1}{2n}\right) \equiv t_\alpha\left(1 + \frac{\frac{1}{4}(1 + t_\alpha^2)V_S(\hat{\theta}) + V_Z(\hat{\theta}) - m(\hat{\theta})}{2n}\right)$$

where t_α is the normal critical value satisfying $\Phi(t_\alpha) = 1 - \alpha$. As long as the true error covariance matrix lies in $\mathcal{S}_1$ so that the expansion (2.7) is valid, the critical region $T_1 > t_1$ has size $\alpha + o(n^{-1})$. Since $\text{plim}(\hat{A}_1 - A_1) = 0$ under our assumptions, the use of an estimated A_1 in forming t_1 produces an error of smaller order than n^{-1}. The functions $V_S(\theta)$ and $V_Z(\theta)$ are readily calculated from (2.12); they depend only on the derivatives of $\Sigma(\theta)$ and on the asymptotic variance of $\hat{\theta}$. The function $m(\theta)$ is more tedious to calculate since it depends on the n^{-1} term in the Nagar expansion of the bias of $\hat{\theta}$. Some examples are given in Sections 5 and 6.

A similar argument can be used for tests based on T_2. Consider the critical value

$$(4.2) \qquad t_2 = t_\alpha\left(1 + \frac{\hat{A}_2}{2n}\right) \equiv t_\alpha\left(1 + \frac{\frac{1}{4}(1 + t_\alpha^2)\tilde{V}_W - \tilde{a}(t_\alpha^2 - 1) - \tilde{b}}{2n}\right)$$

where $\tilde{V}_W$, $\tilde{a}$, and $\tilde{b}$ are estimates of V_W, a, and b. As long as the true Σ lies in $\mathcal{S}_2$ so that the expansion (3.7) is valid and $\text{plim}(\tilde{A}_2 - A_2) = 0$, the critical region $T_2 > t_2$ has size $\alpha + o(n^{-1})$. Again, V_W is typically easy to calculate since it depends only on the first-order asymptotic properties of W. The parameters a and b are more tedious to calculate since they depend on the higher-order asymptotic properties.

It is useful to investigate the sign and magnitude of the terms A_1 and A_2 which measure the size error resulting from using the asymptotic critical value t_α instead of the corrected values. The simplest example, useful for comparison purposes, is the Student t-statistic (1.2). If Σ is in fact scalar with variance σ^2, T_0 has the representation (2.4) with $S = \sqrt{n}(s^2 - \sigma^2)/\sigma^2$ and $Z = 0$. Since $\mathrm{Var}(S) = 2n/(n - K)$, the second-order approximate critical value is

$$t_0 = t_\alpha\left(1 + \frac{1 + t_\alpha^2}{4(n - K)}\right) \tag{4.3}$$

which may be recognized as the first two terms of the Cornish-Fisher expansion of the percentage point of the Student distribution. For a test of size $\alpha = .05$, t_α^2 is approximately three and the corrected critical value is about $100/(n - K)$ percent greater than the asymptotic value. In evaluating the magnitude of the correction term in more complicated cases, it will be convenient to set t_α^2 at 3 and n at 50. Then the Student value of about two percent can serve as a simple benchmark.[4/]

The terms A_1 and A_2 depend on the specific family of covariance matrices $\mathcal{A}$ postulated and on the specific estimate $\Sigma(\hat{\theta})$ or $\tilde{\Sigma}$. Some idea of the range of possible values can be seen from the examples presented in Sections 5 and 6 below. The following general propositions, however, can be deduced from the formulae given in Sections 2 and 3:

1. If $\Sigma(\hat{\theta})$ is an unbiased estimate of Σ, then A_1 is necessarily positive. (This follows from the fact that $c'(X'\Sigma^{-1}X)^{-1}c$ is a concave function of Σ; hence, by Jensen's inequality, m is nonpositive.)

2. If Σ is in fact scalar, a is zero. Hence, if $\tilde{\Sigma}$ is an unbiased estimate of a Σ matrix that is close to being scalar, A_2 will be positive.

From these two facts it appears that there is a general tendency for unbiased estimates of Σ to yield test statistics with positive A values. Such test statistics, when compared to the asymptotic critical value t_α, will reject the null hypothesis too often. Monte Carlo results by MacKinnon and White (1982) are consistent with this observation.

An approximate power function for the size-adjusted test based on T_1 is

$$Pr[T_1 > t_1] \simeq \Phi\left[\delta_1\left(1 - \frac{B_1}{2n}\right) - t_\alpha\right]$$

where B_1 is given in (2.8). An approximate power function for the size-adjusted test based on T_2 is

$$Pr[T_2 > t_2] \simeq \Phi\left[\delta_2\left(1 - \frac{B_2}{2n}\right) - t_\alpha\right]$$

where B_2 is given in (3.8). Comparing these second-order approximate power functions with the exact power functions for the tests based on $\bar{T}_1$ and $\bar{T}_2$ given in (2.3) and (3.3), we see that the B's measure the loss of power due to using an estimated Σ. In fact, they are the asymptotic deficiencies as defined by Hodges and Lehmann (1970). The term B_1 is (to a first-order of approximation) the number of additional observations needed when using the critical region $T_1 > t_1$ to attain the same power as the critical region $\bar{T}_1 > t_\alpha$. Similarly, B_2 is the number of additional observations needed when using the region $T_2 > t_2$ to attain the same power as the region $\bar{T}_2 > t_\alpha$.

Since V_S and V_Z are nonnegative, the deficiency for the test based on T_1 is also nonnegative. It is an increasing function of t_α^2 but is independent of the alternative δ_1. The deficiency for the test based on T_2, however, depends on the specific alternative δ_2 and may even take on negative values for sufficiently large $|a|$. Because an inefficient estimate of β is used in the numerator of the test statistic T_2, noise in the denominator may help! Again, the magnitude of B_1 and B_2 is illustrated for some typical examples in Sections 5 and 6.

In choosing among parametric models for Σ, there is a trade-off between guaranteeing the correct size and obtaining highest power. Generally, the higher the dimensionality of the set $\mathcal{S}$ of permitted covariance matrices, the lower is the power of the test based on that model. Traditional first-order asymptotic theory, however, suggests one striking exception: tests based on GLS statistics like T_1 and tests based on robust OLS statistics like T_2 always dominate the simple t-test based on T_0. If Σ is scalar, all three tests are asymptotically equivalent; if Σ is nonscalar, the t-test has the wrong size, whereas the other tests often have the correct (asymptotic) size. If the desire is to maximize power subject to the constraint that the size is correct, one would never perform a t-test unless absolutely sure of the error specification. The cost of insuring correct size is zero, asymptotically.

When second-order terms are taken into account, this conclusion must be modified. The size-adjusted tests based on T_0, T_1, and T_2 do not have identical second-order power functions in the scalar case. When higher-order approximations are employed, the cost of insuring correct size is found to be nonzero. In the following sections, we calculate this cost for a number of simple examples.

5. A HETEROSCEDASTIC EXAMPLE

Consider the single-regressor model

$$(5.1) \qquad y_i = \beta x_i + u_i \qquad i = 1,\dots,n$$

where the u_i are independent normal errors with zero means and variances $\sigma_i = E(u_i^2)$. If the error variances are known to be equal, the usual t-test would be optimal for the null hypothesis $\beta = \beta_0$ against $\beta > \beta_0$. Its power function is, approximately,

$$(5.2) \qquad \Phi\left[\delta_0\left(1 - \frac{t_\alpha^2}{4n}\right) - t_\alpha\right]$$

where $\delta_0 = (\beta - \beta_0)(\Sigma x_i^2)^{1/2}/\sigma$ and σ^2 is the common unknown variance. The deficiency of this t-test compared to the test with known σ^2 is $t_\alpha^2/2$.

If one were not sure that the error variances were constant, one might postulate a low-dimensional parametric model for Σ and form a GLS test statistic. Suppose, for example, that the n observations are classified into p subgroups where the error variances are assumed constant within a group but may vary across groups. Let $R = [r_{ij}]$ be a $n \times p$ selection matrix such that $r_{ij} = 1$ if observation i lies in group j; $r_{ij} = 0$ if observation i does not lie in group j. Then, the natural GLS test statistic is

$$(5.3) \qquad T_1 = \frac{\sum_{j=1}^{p} \sum_{i=1}^{n} (y_i - x_i\beta_0)x_i r_{ij}/\hat{\sigma}_j^2}{\left[\sum_{j=1}^{p} \sum_{i=1}^{n} x_i^2 r_{ij}/\hat{\sigma}_j^2\right]^{1/2}}$$

where σ_j^2 is some reasonable estimate of the error variance of subgroup j. The simplest unbiased estimate is

$$(5.4) \qquad s_j^2 = \frac{1}{n_j - 1} RSS_j \qquad j = 1,\ldots,p$$

where RSS_j is the residual sum of squares from the OLS regression using only the n_j data points in group j. An alternative estimate could be formed from the residuals from a pooled regression on all the data. Since these two estimates are asymptotically equivalent, the choice affects only the value of m in (2.12). The size adjustment term A_1 depends on the particular way of estimating the σ_j^2, but the deficiency term B_1 does not.

If p is small and all the n_j moderately large (say, 15 or more), the distribution of T_1 should be well approximated by its formal order n^{-1} Edgeworth expansion. Under the assumption that the error variances are indeed constant within groups, the approximate distribution of T_1 is given by (2.7) with

$$(5.5) \qquad V_S = 2 \sum_{j=1}^{p} \frac{\eta_j^2}{n_j}, \qquad V_Z = 2 \sum_{j=1}^{p} \frac{\eta_j(1 - \eta_j)}{n_j},$$

$$\delta_1 = (\beta - \beta_0)\left[\sum_i \sum_j x_i^2 r_{ij} \sigma_j^{-2}\right]^{1/2}$$

where

$$(5.6) \qquad \eta_j = \frac{\sum_i x_i^2 r_{ij} \sigma_j^{-2}}{\sum_i \sum_k x_i^2 r_{ik} \sigma_k^{-2}} \qquad j = 1,\ldots,p$$

is the fraction of the sample information coming from subgroup j. If the unbiased estimates (5.4) are used in forming T_1, we find that $m = -V_Z$.

To get an idea of the magnitude of the size correction term A_1 and the deficiency term B_1, it is convenient to examine the special case where each subgroup has the same number of observations. Then $n_j = n/p$ and the correction terms become

(5.7)
$$A_1 = 4p + \frac{1}{2}p(t_\alpha^2 - 7)\sum \eta_j^2$$
$$B_1 = 2p + \frac{1}{2}p(t_\alpha^2 - 4)\sum \eta_j^2 \quad .$$

For typical values of α, deficiency is slightly less than 2p. That is, one needs approximately 2p additional observations to compensate for not knowing the p error variances. Depending on the distribution of the η_j, the size-correction term A_1 can range from 2p to 4p - 2 for $t_\alpha^2 = 3$. In practice, one might expect large x_i^2 values to be associated with large error variances, implying an A_1 at the upper end of that range. In the canonical case where n = 50, $t_\alpha^2 = 3$, and the η_j are all equal, the corrected critical value t_1 is 4p - 2 percent greater than the asymptotic value t_α.

The use of test statistics like T_0 and T_1 presupposes considerable knowledge about the error variances. An alternative test statistic, requiring essentially no information, is

$$T_2 = \frac{\sum (y_i - \beta_0 x_i) x_i}{\left(\sum x_i^2 \hat{u}_i^2\right)^{1/2}} \tag{5.8}$$

where $\hat{u}_i$ is the OLS residual for observation i. This statistic is asymptotically N(0,1) under H_0 for very general patterns of heteroscedasticity. Furthermore, as long as the x_i and σ_i are well behaved, the distribution of T_2 possesses a valid order n^{-1} Edgeworth approximation for local alternatives. The approximate distribution has the form (3.7) with

$$V_W = 2n \frac{\sum x_i^4\sigma_i^2}{\left(\sum x_i^2\sigma_i^2\right)^2} , \qquad \delta_2 = \frac{(\beta - \beta_0)\sum x_i^2}{\left(\sum x_i^2\sigma_i^2\right)^{1/2}}$$

(5.9)

$$a = n \sum_i \left(\frac{x_i^2\sigma_i^2}{\sum x_j^2\sigma_j^2} - \frac{x_i^2}{\sum x_j^2}\right)^2 , \qquad b = a - \frac{1}{2} V_W \ .$$

The size-correction term A_2 and the deficiency B_2 for the test based on (5.8) depend on the joint distribution of the x_i^2 and σ_i^2. If the σ_i^2 are all equal and the x_i are distributed as a zero mean normal variate, $a = 0$ and $V_W = -2b = 6$. Then, deficiency is $3t_\alpha^2/2$; when $t_\alpha^2 = 3$ and $n = 50$, the corrected critical value is approximately nine percent greater than the asymptotic value. To get a more general picture of the range of values the size correction can take, suppose that the pairs (x_i, σ_i) are distributed as zero mean normal variates with correlation coefficient r. When $t_\alpha^2 = 3$, we find after a bit of calculation

$$A_2 = 9\left(1 + \frac{4r^2}{1 + 2r^2}\right) .$$

Hence, when $n = 50$ and the x_i^2 are highly correlated with the σ^2, the corrected critical value t_2 is about twenty percent greater than the asymptotic value. The deficiency B_2 depends on the alternative and typically is negative for small values of δ_2.

It is interesting to investigate the relative merits of the tests based on T_0, T_1, and T_2. In general, however, the tests are incomparable. If the σ_i^2 are not all equal, the test based on T_0 has the wrong size. Likewise, the test based on T_1 will have the wrong size if the error variance are not constant within groups. Only in the special case of homoscedastic errors will all three tests be (approximately) valid. Thus we shall conduct our comparison in this special case.

When the σ_i^2 are all equal, the three asymptotic noncentrality parameters δ_0, δ_1, and δ_2 are also equal. Hence, the three power functions are, to second order, determined by the deficiency parameters:

$$B_0 = \frac{1}{2} t_\alpha^2$$

$$B_1 = \frac{1}{2} t_\alpha^2 \sum \frac{\eta_j^2}{n_j} n + 2 \sum \frac{\eta_j(1 - \eta_j)}{n_j} n \tag{5.10}$$

$$B_2 = \frac{1}{2} t_\alpha^2 n \frac{\sum x_i^4}{\left(\sum x_i^2\right)^2} .$$

If the η_j/n_j are approximately constant and the x_i normal,

$$B_0 = \frac{1}{2} t_\alpha^2 , \qquad B_1 = \frac{1}{2} t_\alpha^2 + 2(p - 1) , \qquad B_2 = \frac{3}{2} t_\alpha^2 .$$

It appears that there is little power loss from using the robust test statistic T_2 when it was not needed. However, if p is large, the power loss from using the GLS statistic T_1 can be substantial. When there is in fact no problem of heteroscedasticity, attempting to do generalized least squares can introduce considerable noise.

6. AN AUTOCORRELATION EXAMPLE

Consider the single-regressor time series model:

$$(6.1) \qquad y_t = x_t\beta + u_t \qquad\qquad t = 1,\ldots,n$$

where the u_t are n successive realizations of a mean zero stationary normal stochastic process. The usual t-test for the hypothesis $\beta = \beta_0$ is based on the assumption that the errors are white noise. An alternative is to fit a low-dimensional parametric model for the errors and to construct a test based on generalized least squares. For example, one might postulate the stationary AR(1) model:

$$(6.2) \qquad u_t = \rho u_{t-1} + \varepsilon_t \qquad\qquad \varepsilon_t \sim N(0,\sigma^2)$$

and form the test statistic

$$(6.3) \qquad T_1 = \frac{\sum (y_t - \hat{\rho}y_{t-1} - \beta_0 x_t + \beta_0 x_{t-1}\hat{\rho})(x_t - \hat{\rho}x_{t-1})}{\hat{\sigma}\left[\sum (x_t - \hat{\rho}x_{t-1})^2\right]^{1/2}}$$

where $\hat{\rho}$ and $\hat{\sigma}$ are efficient estimates under the assumed model.[5/] Another alternative is to calculate all the autocovariances of the OLS residual series $\hat{u}$:

$$\hat{\gamma}_k = \frac{1}{n-k}\sum_{t=1}^{n-k} \hat{u}_t\hat{u}_{t+k} \qquad\qquad k = 0,1,\ldots,n-1$$

and form the "robust" OLS statistic

$$(6.4) \qquad T_2 = \frac{\sum_t (y_t - \beta_0 x_t)x_t}{\left[\sum_t \sum_s x_t x_s \hat{\gamma}_{|t-s|}\right]^{1/2}} .$$

If the x_t and u_t are well behaved, T_1 and T_2 are asymptotically normal and their distributions can be approximated by Edgeworth expansions as in Sections 2 and 3.

Suppose the errors really are AR(1) so that the postulated model (6.2) is correct. Then the Edgeworth approximation for T_1 takes the form (2.7) with asymptotic noncentrality parameter

$$\delta_1 = \frac{(\beta - \beta_0)\left[\sum (x_t - \rho x_{t-1})^2\right]^{1/2}}{\sigma} .$$

Explicit expressions for the variables S and Z are given in Appendix B; their asymptotic variances are

$$V_S = 2 + 4(1 - \rho^2)\left[\frac{\sum x_{t-1}(x_t - \rho x_{t-1})}{\sum (x_t - \rho x_{t-1})^2}\right]^2$$

(6.5)

$$V_Z = 2 - V_S + \frac{\sum\sum (x_{t+1} - 2\rho x_t + x_{t-1})(x_{s+1} - 2\rho x_s + x_{s-1})\rho^{|t-s|}}{\sum (x_t - \rho x_{t-1})^2} .$$

The asymptotic mean of S depends on the particular estimates used for ρ and σ in (6.3). It is convenient to use the modified maximum likelihood estimates:

$$\hat{\rho} = \frac{n}{n - 2} \frac{\sum \tilde{u}_t \tilde{u}_{t-1}}{\sum \tilde{u}_{t-1}^2}$$

(6.6)

$$\hat{\sigma}^2 = \frac{1}{n - 3} \sum (\tilde{u}_t - \hat{\rho}\tilde{u}_{t-1})^2$$

where $\tilde{u}_t = y_t - x_t\tilde{\beta}$ and $\tilde{\beta}$ is the MLE for β;[6/] the summation ranges from $t = 2$ to $t = n$. Then, as shown in Appendix B, $\sqrt{n}E(S)$ is given by

$$(6.7) \qquad m = \frac{1}{2}(V_S - 2) - \frac{(1 - \rho^2)\sum x_t^2}{\sum (x_t - \rho x_{t-1})^2} .$$

These formulae simplify in some interesting special cases. For example, if $\rho = 0$,

$$(6.8) \qquad \begin{aligned} V_S &= 2(1 + 2r_1^2) \\ V_Z &= 2(1 + r_2 - 2r_1^2) \\ m &= 2r_1^2 - 1 \end{aligned}$$

where $r_i = \sum x_t x_{t-i} \Big/ \sum x_t^2$. The formulae also simplify when the x_t behave like an AR(1) process. That is, if r_i is approximately r^i for constant r less than one in absolute value, then

$$(6.9) \qquad \begin{aligned} V_S &= 2 + 4(1 - \rho^2)\left(\frac{r - \rho}{1 + \rho^2 - 2r\rho}\right)^2 \\ V_Z &= 2\,\frac{(1 - r^2)(1 - \rho^2)^2}{(1 - r\rho)(1 + \rho^2 - 2r\rho)^2} \\ m &= \frac{1 - \rho^2}{1 + \rho^2 - 2r\rho}\left[\frac{2(r - \rho)^2}{1 + \rho^2 - 2r\rho} - 1\right] . \end{aligned}$$

The approximate distribution of T_2 has the form (3.7) with parameters a, b, and V_W given in (3.16) and (3.18). When both u_t and x_t behave like AR(1) processes, the formulae simplify. After considerable algebra, we find

$$\delta_2 = \sigma^{-1}(\beta - \beta_0)\left[(1 - \rho^2)\frac{1 - r\rho}{1 + r\rho} \sum x_t^2\right]^{1/2}$$

(6.10)

$$a = \frac{2\rho^2}{1 - \rho^2}\,\frac{1 - r^2}{1 - r^2\rho^2}$$

$$b = -\left[\frac{1 + r^2}{1 - r^2} + \frac{8r\rho}{1 - r^2\rho^2}\right]$$

$$V_W = 2\,\frac{1 - r^2\rho^2}{(1 - r^2)(1 - \rho^2)}\left[1 + \left(\frac{r + \rho}{1 + r\rho}\right)^2\right] + \frac{4r\rho}{1 - r^2\rho^2} \quad .$$

The formulae in (6.9) and (6.10) can be used to get an idea of the magnitude of the size adjustment and deficiency factors defined in Section 4. Some typical values are given in Table 1 for the case $t_\alpha^2 = 3$; the deficiencies for T_2 are calculated at $\delta_2 = t_\alpha$ where asymptotic power is one-half. Note that the values for A are the percent error (for a sample of 50) in using the asymptotic critical value instead of the corrected value.

Again, it is of interest to compare the relative merits of the tests based on T_0, T_1, and T_2 in the special case where $\rho = 0$ and all three tests are valid. The three asymptotic noncentrality parameters δ_0, δ_1, and δ_2 are equal and the second-order power functions are determined by the deficiency terms

$$B_0 = \frac{1}{2}\, t_\alpha^2$$

(6.11)

$$B_1 = \frac{1}{2}(1 + 2r_1^2)t_\alpha^2 + 2(1 + r_2 - 2r_1^2)$$

$$B_2 = \frac{1}{2}\, t_\alpha^2 \sum_{k=-\infty}^{\infty} r_k^2 \quad .$$

Since only two covariance parameters are estimated in forming the GLS test statistic T_1, its deficiency is small (less than five typically). The test based on T_2, on the other hand, has a large deficiency if the x_t are highly autocorrelated.

Table 1

Size-adjustment factors and asymptotic deficiencies for GLS and robust OLS tests in the AR(1) model ($t_\alpha^2 = 3$)

		GLS		OLS	
ρ	r	A_1	B_1	A_2	B_2
	0	5.0	3.5	3.0	1.5
	0.5	5.0	3.8	5.0	2.5
0	0.7	5.0	4.0	8.8	4.4
	0.9	5.0	4.3	28.6	14.3
	1.0	5.0	4.5	-	-
	0	3.6	2.6	3.0	0.5
	0.5	5.0	3.5	9.3	3.3
0.5	0.7	6.5	4.7	15.1	6.2
	0.9	9.3	7.6	37.6	17.4
	1.0	11.0	10.5	-	-
	0	2.8	2.1	3.0	-1.4
	0.5	3.6	2.6	11.4	2.4
0.7	0.7	5.0	3.5	19.1	6.3
	0.9	10.0	7.7	45.3	19.6
	1.0	19.0	18.5	-	-

7. CONCLUSIONS

The examples developed in Sections 5 and 6 are meant to be illustrative and certainly cannot be viewed as realistic descriptions of actual econometric application. Nevertheless, they do suggest that second-order size correction terms and deficiencies can be large even in very simple models where asymptotic theory might be expected to work well. Since the magnitudes of the order n^{-1} adjustments seem to be rather sensitive to the design matrix X and error covariance matrix Σ, it is difficult to make generalizations. It would appear, however, that the null rejection probabilities of robust regression tests are often considerably greater than their nominal level.

Although the approximate distributions for T_1 and T_2 were derived under the assumption of normality, it would not be difficult to dispense with that assumption. The resulting approximations will depend on the third and fourth cumulants of the error distribution. As long as these cumulants are estimable, second-order size-adjusted critical values for tests based on T_1 and T_2 can be constructed and their power functions approximated. However, once normality has been dropped, it is no longer reasonable to restrict attention to procedures based on least squares. Andrews (1982), for example, constructs robust point estimates of the M-type for regression coefficients in models with possible error auto-correlation and gross contamination. Tests based on such estimates are worth considering but are beyond the scope of the present analysis.

The approximations developed in this paper, although probably considerably better than the usual asymptotic approximations, are not necessarily very accurate for small samples. The method of Edgeworth expansions produces relatively simple approximation formulae and is applicable to a wide variety of econometric models. For any particular test statistic, however, alternative methods tailored to the problem at hand and yielding more accurate results are often available. Nevertheless, the second-order size and power corrections derived here should provide a reasonable indication of the magnitudes involved. Moreover, they provide some insight into the key features of the design matrix which determine the sampling distributions of commonly used test statistics.

Appendix A: EDGEWORTH APPROXIMATIONS

The approximate distributions for the test statistics are derived using an algorithm developed in Cavanagh (1983). Consider the statistic

$$\text{(A.1)} \qquad T_n = X_n + \frac{P_n}{\sqrt{n}} + \frac{Q_n}{n}$$

where X_n has distribution function F_n and density function f_n. Suppose the random variables P_n and Q_n possess bounded moments as n tends to infinity and the conditional moments

$$p_n(x) \equiv E(P_n | X_n = x), \qquad q_n \equiv E(Q_n | X_n = x), \qquad v_n(X) \equiv \mathrm{Var}(P_n | X_n = x)$$

are smooth functions of x. Define the derivatives

$$p_n'(x) \equiv \frac{dp_n}{dx}, \qquad v_n'(x) \equiv \frac{dv_n}{dx}, \qquad c_n(x) \equiv \frac{d \log f_n(x)}{dx}.$$

The formal Edgeworth expansion to the distribution function of T_n is the inverse Fourier transform of the power series expansion of its characteristic function, arranged by ascending powers of $n^{-1/2}$. Dropping the subscript n for notational convenience, we expand the characteristic function of T_n to order n^{-1} as follows:

$$\begin{aligned}
\Psi_n(t) &\equiv Ee^{itT} \simeq Ee^{itX}\left[1 + \frac{itP}{\sqrt{n}} + \frac{itQ}{n} + \frac{(itP)^2}{2n}\right] \\
&= E_X e^{itX}\left[1 + \frac{itp(X)}{\sqrt{n}} + \frac{itq(X)}{n} + \frac{(it)^2[v(X) + p^2(X)]}{2n}\right] \\
&\simeq E_X \exp\left\{it\left(X + \frac{p(X)}{\sqrt{n}} + \frac{q(X)}{n}\right)\right\} + \frac{(it)^2}{2n} E_X e^{itX} v(X) .
\end{aligned}$$

The final term can be rewritten using integration by parts. If the density f(x) is everywhere differentiable, we have

$$-\frac{(it)^2}{2n}\int e^{itx}v(x)f(x)dx = \frac{it}{2n}\int e^{itx}[v'(x) + v(x)c(x)]f(x)dx \;.\;.$$

Hence, $\Psi_n(t)$ is approximately $Ee^{ith(X)}$, where

$$h(X) = X + \frac{p(X)}{\sqrt{n}} + \frac{2q(X) - v'(X) - v(X)c(X)}{2n} \quad . \tag{A.2}$$

Since T and h(X) have the same characteristic function to order n^{-1}, they have the same formal Edgeworth expansion to that order. Thus,

$$\begin{aligned} Pr[T \le x] &\simeq Pr[h(X) \le x] \simeq Pr[X \le h^{-1}(x)] \\ &= F[h^{-1}(x)] \;. \end{aligned}$$

Reversion of the series (A.2) and reintroduction of the subscripts yields the final result:

$$Pr[T_n \le x] \approx F_n\left[x - \frac{p_n(x)}{\sqrt{n}} + \frac{2p_n(x)p_n'(x) - 2q_n(x) + v_n'(x) + c_n(x)v_n(x)}{2n}\right] . \tag{A.3}$$

Under mild regularity conditions on the conditional moments, it can be verified that (A.3) is a valid approximation with error $o(n^{-1})$ as n tends to infinity.

The test statistics T_1 and T_2 given in (2.4) and (3.4) can be put into the form (A.1) by power series expansion:

$$T_1 = \bar{T}_1 + \frac{2Z - \bar{T}_1 S}{2\sqrt{n}} + \frac{3\bar{T}_1 S^2 - 4SZ}{8n} + O_p(n^{-3/2})$$

$$T_2 = \bar{T}_2 - \frac{\bar{T}_2 W}{2\sqrt{n}} + \frac{3\bar{T}_2 W^2}{8n} + O_p(n^{-3/2}) .$$

Since $\bar{T}_1$ and $\bar{T}_2$ are normal, $F(x) = \Phi(x - \delta)$ and $c(x) = (\delta - x)$. Furthermore, in both cases, $p(x)$ is $O(n^{-1/2})$ so the term $p(x)p'(x)/n$ is $O(n^{-2})$ and can be ignored. Eqs. (2.7) and (3.7) then follow from the general approximation formula (A.3).

Appendix B: CALCULATIONS FOR AUTOCORRELATION EXAMPLE

Suppose $u_1, \ldots, u_n$ are normal random variables with zero means and co-variances $E(u_i u_j) = \sigma_u^2 \rho^{|i-j|}$. Then the log likelihood function for the model (6.1) is

$$-\frac{n}{2} \log \sigma^2 + \frac{1}{2} \log (1 - \rho^2) - \frac{1}{2} \sigma^{-2} \left[\sum_{t=2}^{n} (u_t - \rho u_{t-1})^2 + (1 - \rho^2) u_1^2 \right]$$

where $\sigma^2 = (1 - \rho^2)\sigma_u^2$ and $u_t = y_t - \beta x_t$. The maximum likelihood estimates are

$$\tilde{\rho} = \frac{n - 1}{n} \sum_{t=2}^{n} \tilde{u}_t \tilde{u}_{t-1} \Big/ \sum_{t=3}^{n} \tilde{u}_{t-1}^2 + o_p(n^{-1})$$

(B.1)

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{t=2}^{n} (\tilde{u}_t - \tilde{\rho} \tilde{u}_{t-1})^2 + \frac{1 - \tilde{\rho}^2}{n} \tilde{u}_1^2$$

where $\tilde{u}_t = y_t - \tilde{\beta} x_t$ and $\tilde{\beta}$ is the MLE for β. Defining $z_t = x_t - \rho x_{t-1}$ and $\varepsilon_t = u_t - \rho u_{t-1}$, we can write

$$\tilde{\beta} - \beta = \sum_{t=2}^{n} z_t \varepsilon_t \Big/ \sum_{t=2}^{n} z_t^2 + O_p(n^{-1}).$$

Our modified maximum likelihood estimates (6.6) can be written as

$$\hat{\rho} = \frac{n}{n - 2} \sum_{t=2}^{n} u_t u_{t-1} \Big/ \sum_{t=2}^{n} u_{t-1}^2 + \frac{1}{n} \sigma_u^{-2} (\tilde{\beta} - \beta)^2 \sum_{t=2}^{n} z_t x_{t-1}$$

(B.2)

$$+ \frac{1}{n} \sigma_u^{-2} (\tilde{\beta} - \beta) \sum_{t=2}^{n} (2\rho x_t - x_{t+1} - x_{t-1}) u_t + o_p(n^{-1})$$

$$\hat{\sigma}^2 = \frac{1}{n - 3} \sum_{t=2}^{n} [\varepsilon_t - (\tilde{\beta} - \beta) z_t - (\tilde{\rho} - \rho) u_{t-1}]^2 + o_p(n^{-1}) .$$

Let $\Delta_\sigma \equiv \sqrt{n}(\hat{\sigma}^2 - \sigma^2)/\sigma^2$ and $\Delta_\rho = \sqrt{n}(\hat{\rho} - \rho)$. To order $n^{-1/2}$, their moments are

$$\text{(B.3)} \qquad \begin{aligned} &\operatorname{Var} \Delta_\sigma = 2, \qquad \operatorname{Var} \Delta_\rho = 1 - \rho^2, \qquad \operatorname{Cov}(\Delta_\sigma, \Delta_\rho) = 0 \\ &E\Delta_\sigma = 0, \qquad E\Delta_\rho = -\frac{1 - \rho^2}{\sqrt{n}} \frac{\sum z_t x_{t-1}}{\sum z_t^2} . \end{aligned}$$

Details can be found in Sheehan (1982).

Let $\hat{z}_t = x_t - \hat{\rho} x_{t-1}$ and $\hat{\varepsilon}_t = u_t - \hat{\rho} u_{t-1}$. Then, the test statistic (6.3) has the representation (2.4) with

$$\text{(B.4)} \qquad \begin{aligned} S &= \sqrt{n}\left(\frac{\hat{\sigma}^2}{\sigma^2} \frac{\sum z_t^2}{\sum \hat{z}_t^2} - 1\right) \\ &\simeq \Delta_\sigma + \frac{2\Delta_\rho \sum z_t x_{t-1}}{\sum z_t^2}\left(1 + \frac{\Delta_\sigma}{\sqrt{n}} + \frac{2\Delta_\rho}{\sqrt{n}} \frac{\sum z_t x_{t-1}}{\sum z_t^2}\right) - \frac{\Delta_\rho^2}{\sqrt{n}} \frac{\sum x_t^2}{\sum z_t^2} \end{aligned}$$

and

$$\text{(B.5)} \qquad \begin{aligned} Z &= \frac{\sqrt{n}}{\sigma} \left(\frac{\sum \hat{z}_t \hat{\varepsilon}_t}{\sum \hat{z}_t^2} - \frac{\sum z_t \varepsilon_t}{\sum z_t^2}\right)\left(\sum z_t^2\right)^{1/2} \\ &\simeq \Delta_\rho (\sigma^2 \sum z_t^2)^{-1/2} \sum_t \left[(2\rho x_t - x_{t-1} - x_{t+1}) u_t - \frac{2 \sum z_s x_{s-1}}{\sum z_s^2} z_t \varepsilon_t\right] . \end{aligned}$$

The asymptotic moments for S and Z can be calculated from these stochastic expansions.

FOOTNOTES

*Research for this paper was supported by the National Science Foundation under grant SES8112805 and by the Institute of Business and Economic Research at the University of California, Berkeley. The paper was completed while the author was an academic visitor at the London School of Economics. I am grateful to Christopher Cavanagh, James Durbin, Gene Savin, and James Stock for their helpful comments and suggestions.

1. It is not necessary to assume that S and Z possess moments. We require only that their approximate distribution has moments.
2. For many problems, Σ^{-1} has a simpler form than Σ. In these cases one can make use of the fact that $\Sigma_i = -\Sigma\Omega_i\Sigma$ and $\Sigma_{ij} = 2\Sigma\Omega_i\Sigma\Omega_j\Sigma - \Sigma\Omega_{ij}\Sigma$, where Ω_i and Ω_{ij} are derivatives of $\Omega \equiv \Sigma^{-1}$. It is often convenient, as in Rothenberg (1984a, 1984b) to conduct the entire analysis in terms of Ω rather than Σ.
3. The discrete Fourier transforms of the u_i have variances that are asymptotically (but not exactly) equal to the power spectrum. Our spectral estimates, the Fourier transforms of the $\hat{\gamma}_k$, adjust for this bias and hence differ slightly from the usual periodogram values.
4. For $\alpha = 0.05$, a one percent error in the critical value t corresponds to approximately a three percent error in the size of the test. Thus, when $n = 50$, an A value of five implies a fifteen percent error in the significance level.
5. The sums run from $t = 2$ to $t = n$. This statistic is not quite of the form (1.4) since ρ-differing is only approximately the same as generalized least squares. Except for a modification in the definition of δ_1, the analysis is unaffected by this approximation.

6. To the order of approximation employed here, the distribution of T_1 is unaffected if $\tilde{\beta}$ is replaced by some other asymptotically efficient estimate. Hence, $\tilde{u}_t$ may be calculated as residuals from the second round of an interative scheme. If OLS residuals are used, the mean of S is somewhat more complicated.

REFERENCES

Albers, W., 1978, Testing the mean of a normal population under dependence, Annals of Statistics 6, 1337-1344.

Amemiya, T., 1973, Generalized least squares with an estimated auto-covariance matrix, Econometrica 41, 723-732.

Andrews, D., 1982, A model for robustness against distributional shape and dependence over time, Ph.D. thesis (University of California, Berkeley).

Bickel, P., 1974, Edgeworth expansions in nonparametric statistics, Annals of Statistics 2, 1-20.

Breusch, T., 1980, Useful invariance results for generalized regression models, Journal of Econometrics 13, 327-340.

Cavanagh, C., 1983, Hypothesis testing in models with discrete dependent variables, Ph.D. thesis (University of California, Berkeley).

Durbin, J., 1983, The price of ignorance of the autocorrelation structure of the errors of a regression model; in: S. Karlin, T. Amemiya, and L. Goodman (eds.), Studies in econometrics, time series, and multivariate statistics (Academic Press, New York), 243-253.

Fuller, Wayne, 1976, Introduction to statistical time series (Wiley, New York).

Hannan, E. J., 1963, Regression for time series; in: M. Rosenblatt (ed.), Time series analysis (Wiley, New York).

Hodges, J. L. and E. L. Lehmann, 1970, Deficiency, Annals of Mathematical Statistics 41, 783-801.

MacKinnon, J. and H. White, 1982, Inference in the presence of heteroskedasticity of unknown form: Evidence on finite sample properties, Discussion paper 82-18 (Department of Economics, University of California, San Diego).

Pfanzagl, J., 1980, Asymptotic expansions in parametric statistical theory; in: P. R. Kirshnaish (ed.), Developments in statistics, Vol. 3 (Academic Press, New York).

Rothenberg, T. J., 1982, Comparing alternative asymptotically equivalent tests; in: W. Hildenbrand (ed.), Advances in econometrics (Cambridge University Press, Cambridge), 255-262.

Rothenberg, T. J., 1984a, Approximate normality of generalized least squares estimates, Econometrica 52, in press.

Rothenberg, T. J., 1984b, Hypothesis testing in linear models when the error covariance matrix is nonscalar, Econometrica 52, in press.

Rothenberg, T. J., 1984c, Approximating the distributions of econometric estimators and test statistics; in: Z. Griliches and M. Intriligator (eds.), Handbook of econometrics, Vol. 2 (North-Holland, Amsterdam), in press.

Sheehan, D., 1981, Approximating the distributions of some time-series estimators and test statistics, Ph.D. thesis (University of California, Berkeley).

White, H., 1980, A heteroskedasticity-consistent covariance estimator and a direct test for heteroskedasticity, Econometrica 48, 817-838.