A Comparison of Statistical Tests Distinguishing Error Dynamics and Systematic Dynamics in Distributed Lag Models

Technical Bulletin T-167
June 1990

Agricultural Experiment Station
Division of Agriculture
Oklahoma State University

A COMPARISON OF STATISTICAL TESTS DISTINGUISHING ERROR DYNAMICS AND SYSTEMATIC DYNAMICS IN DISTRIBUTED LAG MODELS

by
David A. Pyles, Eric Kocher
and Daniel S. Tilley

A COMPARISON OF STATISTICAL TESTS DISTINGUISHING ERROR DYNAMICS AND SYSTEMATIC DYNAMICS IN DISTRIBUTED LAG MODELS

by
David A. Pyles, Eric Kocher
and Daniel S. Tilley*

1. Introduction

Proper specification of econometric models frequently requires the incorporation of dynamic adjustment mechanisms to accommodate expectation generation, noninstantaneous equilibration, and other dynamic processes. The incorporation of these dynamic mechanisms often leads to distributed lag models in which lags on the dependent and independent variables are included among the regressors. However, similar lag structures also arise from static models possessing autoregressive disturbances. Consequently, if a distributed lag model were found to reasonably represent the data, the analyst might then be confronted with the question of whether the lag structure of the model should be attributed to systematic dynamic phenomena, or to auto-regressive behavior in the disturbances, or to some combination thereof.

To illustrate the problem, consider the distributed lag model:

$$(1 - \beta_1 L)y_t = \beta_0 + (\beta_2 - \beta_3 L)x_{1t} + (\beta_4 - \beta_5 L)x_{2t} + \varepsilon_t$$

where L denotes the lag operator. The ε_t are assumed to be independently and identically distributed (iid) with zero mean and variance σ^2. The lag structure of the model gives the initial impression that the impacts of shifts in x_{1t} and x_{2t} will be systematically distributed over time. However, suppose that we can formulate the model as:

$$y_t = \theta_0 + \theta_1 x_{1t} + \theta_2 x_{2t} + u_t$$
$$(1 - \rho L)u_t = \varepsilon_t$$

---

*Assistant Professor, Research Assistant and Professor, Department of Agricultural Economics, Oklahoma State University, Stillwater, Oklahoma, respectively. Research under Hatch 2037 of the Oklahoma Agricultural Experiment Station. Comments of Oral Capps, Jr. and Joseph Havlicek, Jr., are appreciated.

which implies:

$$(1 - \rho L)y_t = (1 - \rho)\theta_0 + (1 - \rho L)\theta_1 x_{1t} + (1 - \rho L)\theta_2 x_{2t} + \varepsilon_t$$

Though this formulation renders the same general lag structure, it is now clear that the dynamics of the model are to be attributed entirely to dynamic behavior in the disturbances. Since there are no systematic dynamics in the model, the full impacts of shifts in x_{1t} and x_{2t} are immediately reflected in y_t. Therefore, what initially appeared to be a dynamic model is in fact a static model with autoregressive disturbances.

For a general formulation of the problem, write the general distributed lag model as:

$$\beta(L)'\mathbf{z}_t = \varepsilon_t \tag{1}$$

where: $\mathbf{z}_t$ is a k + 1 dimensional vector of variables, the first of which is endogenous. $\beta(L)$ is a vector of polynomials in the lag operator L, and the ε_t for t = 1,,,n are iid disturbances with zero mean and variance σ^2. Distributed lag models will be denoted as $DL(m_0,m_1,,,m_k)$, where m_i denotes the order of the (i + 1)th polynomial in $\beta(L)$. Now, as a special case of (1), we have models of the form:

$$\theta(L)'\mathbf{z}_t = u_t \tag{2}$$
$$\rho(L)u_t = \varepsilon_t$$

This model is simply a distributed lag model with disturbances following an rth order autoregressive process described by the polynomial $\rho(L)$. Substitute the first of these equations into the second to obtain:

$$\rho(L)\theta(L)'\mathbf{z}_t = \varepsilon_t$$

from which it appears that (2) is the special case of (1) where:

$$\beta(L) = \rho(L)\theta(L) \tag{3}$$

The important distinction between (1) and (2) is that the latter explicitly divides the dynamic structure of the system into the systematic dynamics component $\theta(L)$, and the error dynamics component $\rho(L)$. (3) indicates that the divisibility of the dynamic structure of (1) into such components is manifested by the presence of r common factors to the polynomials in $\beta(L)$. From this perspective, models having autoregressive errors are seen as special cases of more general dynamic models.

The possibility that $\beta(L)$ may be decomposed as in (3) has at least two important implications. First, the validity of (3) can dramatically change the interpretation of

the dynamic behavior of the model. This is illustrated in the example above. Second, if the true model has the form of (2), then estimators failing to recognize the autoregressive error process will be relatively inefficient to those that do.

Unfortunately, analysts commonly estimate models having the form of (2) with little or no consideration given to the possibility that the true model follows an unreducible form of (1). The popularity of (2) has derived from the frequent discovery of empirical autocorrelation in the residuals of static models estimated with time series data. If this empirical autocorrelation is due to autoregressive disturbances, then (2) becomes the appropriate model. However, high empirical autocorrelation can also derive from misspecification of the functional form and from the exclusion of relevant regressors. Consequently, if the true model is of the form of (1), but the estimated model has relevant lags excluded from the regressors, then empirical autocorrelation is apt to be observed in the residuals because of the inability of the estimated form to capture the dynamic structure of the system. An attempt to correct for this autocorrelation by adopting (2) will merely lead to the replacement of one misspecified model with another. A better course in these cases is to first estimate a model of the form of (1). If lags are indeed found to be significant, the analyst can then test the possibility that these lags derive from autoregressive disturbances by testing the hypothesis that the parameters will admit the factorization in (3).

In this paper, simulated comparisons are made between statistical procedures designed to test the validity of the factorization in (3) for the case of first-order autoregressive disturbances. Hence, our experiments suppose a situation where the analyst is resolved, either upon a-priori or empirical grounds, as to the orders of the polynomials in $\beta(L)$, but is unresolved as to whether these polynomials are consistent with the presence of a first-order autoregressive process in the disturbances. Monte Carlo experiments are conducted in which (3) is tested using four different criteria, including the popular Wald, likelihood ratio, and Lagrangian multiplier criteria. Tests using these criteria are equivalent asymptotically; however, the tests may differ in small samples to the extent that one test might be preferred over the others. The primary purpose of the simulation experiments is to determine whether any of the criteria are indeed superior in small sample tests of (3).

Since we restrict attention to tests for first-order autoregressive disturbances, our experiments suppose a situation where the analyst is resolved that the order of autoregressive process in the disturbances is either zero or one. The assumed situation is not as restricted as it might initially appear. A model having the form of (1) can algebraically admit an autoregressive process of order no greater than the least degree of the polynomials in $\beta(L)$. Consequently, the presence of one regressor having only a single lag necessarily implies that an autoregressive process, if it exists, can be of order no greater than one. We restrict attention to the case of first-order autoregressive processes for two reasons. First, this case occurs with sufficient frequency in practice to warrant its exclusive consideration. Second, numerous difficulties are introduced when higher orders must be considered. These difficulties are discussed in Sargan and Sargan and Mehta, wherein it is shown that the traditional test criteria have severe problems when applied in such cases. Sargan and Mehta

propose an alternative criterion and a procedure wherein it may be applied to cases involving high-order autoregressive processes.

Mizon and Hendry have conducted simulation experiments similar to those presented here; however, the experiments reported in this paper differ from theirs in several important regards. These differences derive largely from the fact that Mizon and Hendry were primarily concerned with the estimation of curves approximating small sample power functions, whereas the present study is primarily concerned with the comparison of test criteria. Differences between the studies include: 1) Sample sizes used in this study are much larger than those used in Mizon and Hendry. 2) Different parameterizations are considered. In particular, we consider the both positive and negative autocorrelation, whereas the other authors consider only positive autocorrelation. 3) We consider four different test criteria, whereas only the Wald and likelihood ratio criteria were considered in the previous study. 4) We consider three different lag structures, whereas the previous study considered only two.

2. The Traditional Test Criteria

Since the constraints implying the factorization in (3) are nonlinear, it is difficult to find a reasonable test criterion having a tractable density function in finite samples. However, there are numerous criteria whose asymptotic distributions may be inferred under general assumptions. These criteria may then be used in tests wherein the asymptotic distributions are used as approximation to the finite-sample distributions. The most popular of these are the Wald, likelihood ratio, and Lagrangian multiplier criteria. These criteria are discussed here.

Let $\mathbf{y} = (y_1,.,y_n)$ be a random sample drawn from a distribution parameterized by the p-dimensional vector θ_o. Let the log-likelihood function of θ_o be $L(\theta_o;\mathbf{y})$. Henceforth, we shall omit $\mathbf{y}$ from the arguments of L to avoid notational clutter. Let $\mathbf{H}$ denote the Hessian matrix of L, and let:

$$\mathbf{I}_n^*(\theta) = -E[\mathbf{H}(\theta)]$$

Hence, $\mathbf{I}_n^*(\theta)$ denotes the information marix evaluated at θ. Moreover, let:

$$\mathbf{I}^*(\theta) = \lim_{n\to\infty} \mathbf{I}_n^*(\theta)/n$$

Denote the maximum likelihood estimator (MLE) of θ_o by $\hat{\theta}$. Let $\mathbf{h}$ be a vector of q constraints, where $q < p$. We assume that $\mathbf{h}$ is continuously differentiable, and that $\nabla\mathbf{h}(\theta_o)$ is of full row rank. Under general regularity conditions on L, $\hat{\theta}$ will be asymptotically normally distributed. Specifically:

$$n^{1/2}(\hat{\theta} - \theta_o) \xrightarrow{d} N[\mathbf{0}, \mathbf{I}^*(\theta_o)^{-1}] \tag{4}$$

Where $\xrightarrow{d}$ denotes convergence in distribution. We henceforth assume that the regularity conditions ensuring the latter hold. For a more thorough discussion of these conditions, the reader is referred to Amimiya or Dhrymes. Also, these regularity conditions ensure:

$$\mathbf{I}^*(\theta_o) = -\text{plim } \mathbf{H}(\bar{\theta})/n \tag{5}$$

where $\bar{\theta}$ is any consistent estimator of θ_o. Now, consider hypotheses of the form:

$$\begin{aligned} &H_o: \ \mathbf{h}(\theta_o) = \mathbf{0} \\ &H_a: \ \mathbf{h}(\theta_o) \neq \mathbf{0} \end{aligned} \tag{6}$$

Let $\tilde{\theta}$ maximize $L(\theta)$ subject to $\mathbf{h}(\theta) = \mathbf{0}$. If the null hypothesis is true, then under the stated assumptions, all of the following statistics converge in distribution to a chi-squared variate with q degrees of freedom:

$$\lambda_W = \mathbf{h}(\hat{\theta})'[\nabla\mathbf{h}(\hat{\theta})\mathbf{I}_n^*(\hat{\theta})^{-1}\nabla\mathbf{h}(\hat{\theta})']^{-1}\mathbf{h}(\hat{\theta})$$

$$\lambda_{LR} = -2[L(\tilde{\theta}) - L(\hat{\theta})]$$

$$\lambda_{LM} = \nabla L(\tilde{\theta})\, \mathbf{I}_n^*(\tilde{\theta})^{-1}\nabla L(\tilde{\theta})'$$

These statistics are respectively the Wald, log-likelihood ratio, and Lagrangian multiplier statistics. The log-likelihood ratio statistic is generally credited to Wilks. The Lagrangian multiplier statistic was first introduced by Rao, and is sometimes called "Rao's score statistic." A general discussion of these statistics and their applications in econometrics may be found in Engle.

The above results suggest the following decision rule for testing the hypotheses in (6):

$$\text{Reject } H_o \text{ if } \lambda > \chi^2_{q,a} \tag{7}$$

where λ is any one of the Wald, log-likelihood ratio, or Lagrangian multiplier statistics, and where $\chi^2_{q,a}$ is the $100(1 - \alpha)$th quantile of the chi-squared distribution with q degrees of freedom. The three criteria render asymptotically equivalent tests when incorporated into the later decision rule; consequently, asymptotic properties provide no basis for choice among them. However, the distributions of the statistics

may differ considerably in small samples, and in such a way that one may be preferred as a test criterion over the others.

When there is no information regarding the small sample properties of these three statistics, choice among them is often based upon computational considerations. The Wald statistic will generally be the easiest to calculate since it requires only the unrestricted MLE. Calculation of the Lagrangian multiplier statistic requires the restricted MLE, which is generally more difficult to compute than the unrestricted estimator. The log-likelihood ratio requires both the restricted and unrestricted MLEs, and will therefore be the most difficult to compute.

We now show that the Wald statistic is indeed asymptotically distributed as a chi-squared variable when the null hypothesis is true; moreover, we derive an approximation to the power function for the test in (7) when the Wald statistic is used as the test criterion. The analogous derivations for the likelihood ratio and Lagrangian multiplier statistics may be found in their original sources.

The foregoing analysis makes use of the noncentral chi-squared distribution. The density function for this distribution is:

$$f(x;d,\mu) = \sum_{i=0}^{\infty} \exp(-\mu / 2) \, [\, (\mu / 2)^i / i \, ! \,] \, f(x;d + 2i)$$

where d is the degrees of freedom; $\mu > 0$ is called the "noncentrality parameter," and f(x,d) denotes the ordinary chi-squared distribution with d degrees of freedom; hence:

$$f(x;d) = [2^{1/2} \Gamma(d/2)]^{-1} x^{d/2-1} \exp(-x/2)$$

where Γ denotes the gamma function. Observe that f(x;d,0) = f(x;d); hence, the chi-squared distribution may be regarded as the special case of the noncentral chi-squared where the noncentrality parameter is equal to zero.

The noncentral chi-squared derives its primary motivation from the normal. In particular, if $\mathbf{x}$ is a d-dimensional vector of independent normal variates having unit variances and mean vector equal to μ, then $\mathbf{x}'\mathbf{x}$ is distributed as noncentral chi-squared with d degrees of freedom and noncentrality parameter $\mu'\mu$. A useful generalization of this relation is supplied by the following theorem:

1. Theorem: Let $\mathbf{x}$ be a d-dimensional vector distributed as $N(\mu,\Sigma)$, where Σ is assumed to be positive definite, then $\mathbf{x}'\Sigma^{-1}\mathbf{x}$ is distributed as noncentral chi-squared variate with d degrees of freedom and noncentrality parameter $\mu'\Sigma^{-1}\mu$.

Proof: Since Σ is positive definite, we can find a nonsingular matrix $\mathbf{Q}$ such that $\mathbf{Q}\Sigma\mathbf{Q}' = \mathbf{I}$, and consequently, $\Sigma^{-1} = \mathbf{Q}'\mathbf{Q}$. It is easy to verify that $\mathbf{Qx} \sim N(\mathbf{Q}\mu,\mathbf{I})$; moreover:

$$\mathbf{x}'\Sigma^{-1}\mathbf{x} = (\mathbf{Q}\mathbf{x})'(\mathbf{Q}\mathbf{x})$$

The expression to the right is known to be distributed as noncentral chi-squared with d degrees of freedom and noncentrality parameter $(\mathbf{Q}\mu)'(\mathbf{Q}\mu) = \mu'\Sigma^{-1}\mu$, and the proof is complete.

If the null hypothesis is true, then under regularity conditions ensuring (4), $n^{1/2}\,\mathbf{h}(\hat{\theta})$ will be asmptotically normal. The asymptotic normality of $n^{1/2}\,\mathbf{h}(\hat{\theta})$ then implies that the Wald statistic is asymptotically distributed as chi-squared. These results may be demonstrated as follows: Since $\mathbf{h}$ is differentiable, the mean value theorem allows us to write:

$$\mathbf{h}(\hat{\theta}) = \mathbf{h}(\theta_o) + \nabla\mathbf{h}(\bar{\theta})(\hat{\theta} - \theta_o)$$

where $\bar{\theta}$ is on the line segment between $\hat{\theta}$ and θ_o. $\mathbf{h}(\theta_o) = \mathbf{0}$ by assumption; consequently, upon multiplying both sides of the latter by $n^{1/2}$, one obtains:

$$n^{1/2}\,\mathbf{h}(\hat{\theta}) = \nabla\mathbf{h}(\bar{\theta})n^{1/2}\,(\hat{\theta} - \theta_o)$$

Take the probability limit of both sides of this equation to conclude:

$$n^{1/2}\,\mathbf{h}(\hat{\theta}) \xrightarrow{d} \nabla\mathbf{h}(\theta_o)\mathbf{v} \qquad (8)$$

where $\mathbf{v}$ is known from (4) to be distributed as a multivariate normal vector with zero mean vector and covariance matrix $\mathbf{I}^*(\theta_o)^{-1}$. Therefore, $n^{1/2}\,\mathbf{h}(\hat{\theta})$ converges in distribution to a multivariate normal vector with zero mean vector and covariance marix:

$$\Omega(\theta_o) = \nabla\mathbf{h}(\theta_o)\mathbf{I}^*(\theta_o)^{-1}\nabla\mathbf{h}(\theta_o)' \qquad (9)$$

Now, we can write the Wald statistic as:

$$\lambda_W = n^{1/2}\,\mathbf{h}(\hat{\theta})'\{\nabla\mathbf{h}(\hat{\theta})[\mathbf{I}_n^*(\hat{\theta})/n]^{-1}\nabla\mathbf{h}(\hat{\theta})'\}^{-1}n^{1/2}\,\mathbf{h}(\hat{\theta})$$

Substitute (8) and (9) into the probability limit of the latter to obtain:

$$\lambda_W \xrightarrow{d} \mathbf{v}'\nabla\mathbf{h}(\theta_o)'\,\Omega(\theta_o)^{-1}\nabla\mathbf{h}(\theta_o)\mathbf{v}$$

Since $\nabla \mathbf{h}(\theta_o)\mathbf{v}$ is distributed as $N[\mathbf{O},\Omega(\theta_o)]$, the right-hand side of the latter is known by Theorem one to be distributed as chi-squared with q degrees of freedom. It will be observed that the assumed nonsingularity of $\Omega(\theta_o)$ necessitates our previous assumption that $\nabla \mathbf{h}(\theta_o)$ be of full row rank.

When the null hypothesis is false, all three of the traditional test criteria tend toward positive infinity as the sample size increases. This implies that the test in (7) will, under each criterion, have asymptotic power equal to one on the alternative hypothesis. Hence, each criterion has the desirable property of rendering a consistent test; however, this property makes the asymptotic power function an ineffective measure for test comparisons. A better measure of relative test performance is obtained by evaluating the limit of the power function on a sequence of parameters falling under the alternative hypothesis, but which approaches a parameterization satisfying the null. We thereby obtain a measure of the ability of the test to reject the null hypothesis at local alternatives. Set $\gamma = n^{1/2}\mathbf{h}(\theta_o)$, and let $\{\theta_k\}$ be a sequence whose terms satisfy:

$$\mathbf{h}(\theta_k) = \gamma / k^{1/2}$$
$$\lim_{k\to\infty} \theta_k = \theta^*$$
$$\theta_k = \theta_o \text{ for } k = n$$

Let π_k denote the power function of the test for samples of size k, and let:

$$\pi(\theta_o, n, \theta^*) = \lim_{k\to\infty} \pi_k(\theta_k)$$

π may then be used as a measure for test comparisons. The Wald, likelihood ratio, and Lagrangian multiplier criteria render asymptotically equivalent tests in the sense that under each criterion:

$$\pi = P[\chi^2(q,\mu) > \chi^2_{q,a}] \tag{10}$$
where: $\mu = \gamma'[\nabla \mathbf{h}(\theta^*)\mathbf{I}^*(\theta^*)^{-1}\nabla \mathbf{h}(\theta^*)']^{-1}\gamma$

and where $\chi^2(q,\mu)$ denotes a noncentral chi-squared variate with q degrees of freedom and noncentrality parameter μ. Hence, the three tests are equally effective at rejecting local alternatives in large samples.

The validity of (10) for the case of the Wald statistic may be shown as follows: Let $\hat{\theta}_k$ denote the maximum likelihood estimator of θ_k. Using the mean value theorem, we may write:

$$\mathbf{h}(\hat{\theta}_k) = \gamma / k^{1/2} + \nabla \mathbf{h}(\bar{\theta}_k)(\hat{\theta}_k - \theta_k)$$

where $\bar{\theta}_k$ lies on the line segment between $\hat{\theta}_k$ and θ_k. From the latter, we conclude:

$$k^{1/2}\, \mathbf{h}(\hat{\theta}_k) \xrightarrow{d} \gamma + \nabla \mathbf{h}(\theta^*)\, \mathbf{v}$$

where $\mathbf{v} \sim N[\mathbf{0}, \mathbf{I}^*(\theta^*)^{-1}]$. Hence, $k^{1/2}\, \mathbf{h}(\hat{\theta}_k)$ is asymptotically distributed as $N[\gamma, \Omega(\theta^*)]$. Now, the corresponding sequence of Wald statistics is:

$$\lambda_W = k^{1/2}\, \mathbf{h}(\theta_k)' \{\nabla \mathbf{h}(\hat{\theta}_k)[\mathbf{I}_k^*(\hat{\theta})/k]^{-1} \nabla \mathbf{h}(\hat{\theta}_k)'\}^{-1} k^{1/2} \mathbf{h}(\hat{\theta}_k)$$

Take the probability limit of this expression to obtain:

$$\text{plim}\ \lambda_W = [\gamma + \nabla \mathbf{h}(\theta^*)\mathbf{v}]' \Omega(\theta^*)^{-1} [\gamma + \nabla \mathbf{h}(\theta^*)\mathbf{v}]$$

Since $\nabla \mathbf{h}(\theta^*)\mathbf{v} \sim N[\mathbf{0}, \Omega(\theta^*)]$, then the right-hand side of the latter is known from Theorem one to be distributed as a noncentral chi-squared variate with q degrees of freedom and noncentrality parameter given in (10). Similar derivations for the log-likelihood ratio and Lagrangian multiplier statistics may be found in Gallant and Holly.

Since $\pi_n(\theta_o)$ is the nth term in a sequence converging upon $\pi(\theta_o, n, \theta^*)$, we may use π as an approximation of π_n. It will frequently prove to be very accurate, as is demonstrated in the forthcoming simulations. Moreover, π has at least three desirable properties that should be expected of a reasonable approximation to π_n. First, if θ_o satisfies the null hypothesis, then $\gamma = \mathbf{0}$, and consequently, $\mu = \mathbf{0}$; hence:

$$\pi(\theta_o, n, \theta^*) = P[\chi^2(q, 0) > \chi^2_{q,\alpha}] = \alpha$$

Therefore, π renders the appropriate asymptotic test size. Second, we should expect π to approach unity as either n approaches infinity or as $|\mathbf{h}(\theta_o)|$ approaches infinity. In either event, the noncentrality parameter μ approaches infinity also. Since the cumulative distribution function for the noncentral chi-squared distribution is a strictly increasing function of the noncentrality parameter, this then implies that π indeed approaches unity.

3. Testing Parameters of Linear Models with Normal Errors

In this section, we consider the application of the test criteria presented in the previous section to tests of nonlinear restrictions upon the coefficients of linear models having normal errors. This discussion is necessitated by the fact that the simulated tests for common factors fall within this general class of tests.

Subsequently, suppose we have a linear model whose marix representation is:

$$\mathbf{y} = \mathbf{Z}\beta_o + \varepsilon \tag{11}$$

where: $\mathbf{y}$ is an nx1 vector of observations on the dependent variable. $\mathbf{Z}$ is an nxp matrix of regressors possibly containing lags on the dependent variable, and ε is an nx1 vector of disturbances distributed as $N(\mathbf{0},\sigma_o^2\mathbf{I})$. We wish to test the null hypothesis that $\mathbf{h}(\beta_o) = \mathbf{0}$ against the alternative that $\mathbf{h}(\beta_o) \neq \mathbf{0}$, where $\mathbf{h}$ is a q-dimensional vector function for $q < p$. It is shown that under the assumption of normal disturbances, the log-likelihood ratio and Lagrangian multiplier statistics may be found as simple functions of the restricted and unrestricted MLEs of σ_o^2. We also introduce a fourth test criterion by showing that the restricted MLE of β_o may be found as the solution to an unconstrained minimization problem whose indirect objective function has the same asymptotic distribution as the log-likelihood ratio and Lagrangian multiplier statistics.

The log-likelihood function corresponding to the model in (11) is:

$$L(\beta,\sigma^2) = -(n/2)\log(2\pi\sigma^2) - (\mathbf{y} - \mathbf{Z}\beta)'(\mathbf{y} - \mathbf{Z}\beta)/(2\sigma^2)$$

Since L is differentiable, the unconstrained MLEs for β_o and σ_o^2 must satisfy:

$$\partial L/\partial\beta = (\mathbf{y} - \mathbf{Z}\beta)'\mathbf{Z}/\sigma^2 = \mathbf{0}$$
$$\partial L/\partial\sigma^2 = -n/(2\sigma^2) + (\mathbf{y} - \mathbf{Z}\beta)'(\mathbf{y} - \mathbf{Z}\beta)/2\sigma^4 = 0$$

Let $\hat{\beta}$ and $\hat{\sigma}^2$ denote the unconstrained MLEs. From the first condition, it is clear that:

$$\hat{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

Moreover, the second condition implies:

$$\hat{\sigma}^2 = (\mathbf{y} - \mathbf{Z}\hat{\beta})'(\mathbf{y} - \mathbf{Z}\hat{\beta})/n \tag{12}$$

Now, it will be observed that since $\mathbf{h}(\beta)$ does not involve σ^2, the second condition must also be satisfied by the solution which maximizes $L(\beta,\sigma^2)$ subject to $\mathbf{h}(\beta) = \mathbf{0}$. Therefore if the constrained MLEs are denoted by $\tilde{\beta}$ and $\tilde{\sigma}^2$, then we must have:

$$\tilde{\sigma}^2 = (\mathbf{y} - \mathbf{Z}\tilde{\beta})'(\mathbf{y} - \mathbf{Z}\tilde{\beta})/n \tag{13}$$

Upon substituting (12) and (13) into the log-likelihood ratio statistic, it will be found that:

$$\lambda_{LR} \equiv -2[L(\tilde{\beta}, \tilde{\sigma}^2) - L(\hat{\beta}, \hat{\sigma}^2)] = n\log(\tilde{\sigma}^2 / \hat{\sigma}^2) \tag{14}$$

Hence λ_{LR} reduces to a simple function of the unrestricted and restricted MLEs of σ_o^2.

The Lagrangian multiplier statistic for the hypothesis $\mathbf{h}(\beta_o) = \mathbf{0}$ is defined as:

$$\lambda_{LM} \equiv \nabla L(\tilde{\beta}, \tilde{\sigma}^2) \mathbf{I}_n^*(\tilde{\beta}, \tilde{\sigma}^2)^{-1} \nabla L(\tilde{\beta}, \tilde{\sigma}^2)'$$

where $\mathbf{I}_n^*$ denotes the information marix. For the present problem:

$$\mathbf{I}_n^*(\beta, \sigma^2) = \begin{bmatrix} \mathbf{Z'Z}/\sigma^2 & \mathbf{0} \\ \mathbf{0} & n/(2\sigma^4) \end{bmatrix}$$

Since the last component of $\nabla L(\tilde{\beta}, \tilde{\sigma}^2)$ is known to be equal to zero by the first-order conditions, λ_{LM} reduces to:

$$\lambda_{LM} = (\mathbf{y} - \mathbf{Z}\tilde{\beta})' \mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z}' (\mathbf{y} - \mathbf{Z}\tilde{\beta}) / \tilde{\sigma}^2$$

Using the fact that $\hat{\beta} = (\mathbf{Z'Z})^{-1}\mathbf{Z'y}$, the last equation can be reduced to:

$$\lambda_{LM} = (\hat{\beta} - \tilde{\beta})' (\mathbf{Z'Z}) (\hat{\beta} - \tilde{\beta}) / \tilde{\sigma}^2 \tag{15}$$

It happens that the numerator of this expression is equal to the difference between the restricted and unrestricted sums of squared errors, as may be seen by observing:

$$\begin{aligned} n\tilde{\sigma}^2 &= (\mathbf{y} - \mathbf{Z}\tilde{\beta})' (\mathbf{y} - \mathbf{Z}\tilde{\beta}) = [\hat{\varepsilon} - \mathbf{Z}(\tilde{\beta} - \hat{\beta})]' [\hat{\varepsilon} - \mathbf{Z}(\tilde{\beta} - \hat{\beta})] \\ &= n\hat{\sigma}^2 + (\tilde{\beta} - \hat{\beta})' (\mathbf{Z'Z})(\tilde{\beta} - \hat{\beta}) \end{aligned}$$

where $\hat{\varepsilon}$ denotes the vector of residuals for the unrestricted MLEs. Upon substituting the last relation into (15), one obtains:

$$\lambda_{LM} = n(\tilde{\sigma}^2 - \hat{\sigma}^2)/\tilde{\sigma}^2 \tag{16}$$

Hence, λ_{LM} also reduces to a simple function of the unrestricted and restricted MLEs of σ_o^2.

Using (14) and (16), λ_{LR} may be written in terms of λ_{LM} with:

$$\lambda_{LR} = n\log[n/(n - \lambda_{LM})] = -n\log(1 - \lambda_{LM}/n)$$

Using the fact that $\log(1 + x) < x$ for all $x \in (-1,\infty)$, it is clear from the later that $\lambda_{LR} > \lambda_{LM}$ for all finite n. However, it may also be seen that as n approaches infinity, λ_{LR} converges to λ_{LM}; hence, the asymptotic equivalence of the two statistics is confirmed. It has been shown by Berndt and Savin that when testing linear constraints on the coefficients of linear models, the Wald statistic will be greater than λ_{LR}; however, this inequality does not necessarily hold in the case of nonlinear constraints, as is illustrated in section six.

A difficulty with the above formulation is that calculation of $\tilde{\beta}$ requires maximization of $L(\beta,\sigma^2)$ subject to the constraint $\mathbf{h}(\beta) = \mathbf{0}$. Since constrained optimization problems are generally much more difficult to analyze and solve than unconstrained problems, calculations can generally be simplified if a formulation can be found in which only unconstrained optima are required. We now develop such a formulation, and in so doing, we arrive at yet another test criterion.

Suppose that there exists a vector function $\mathbf{g}$ mapping from E^{p-q} to E^p such that the condition $\mathbf{h}(\beta) = \mathbf{0}$ is true if and only if there exists $\theta \in E^{p-q}$ such that $\beta = \mathbf{g}(\theta)$. It will frequently be the case that parameter constraints occur naturally in this form. This happens to be the case for the common factor problem. For example, consider the DL(1,1) model:

$$y_t = \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} + \varepsilon_t$$

where the ε_t are iid disturbances. If a common factor exists in the lag polynomials of the latter model, then it must be true that the model may be alternatively written as:

$$y_t = \theta_1 x_t + u_t$$
$$u_t = \theta_2 u_{t-1} + \varepsilon_t$$

or equivalently:

$$y_t = \theta_2 y_{t-1} + \theta_1 x_t - \theta_1\theta_2 x_{t-1} + \varepsilon_t$$

Hence, the following correspondence exists between the parameters of the unrestricted and restricted models:

$$\beta_1 = g_1(\theta_1,\theta_2) = \theta_2$$
$$\beta_2 = g_2(\theta_1,\theta_2) = \theta_1$$
$$\beta_3 = g_3(\theta_1,\theta_2) = -\theta_1\theta_2$$

where the g_i would be the component functions of **g** considered above. To derive the relevant constraint for the coefficients of the unrestricted model, one must eliminate θ_1 and θ_2 from the above equations. The constraint will be found to be:

$$h(\beta_1,\beta_2) = \beta_1\beta_2 + \beta_3 = 0$$

In this case, the constraint is easily derived; however, derivation of the constraint becomes extremely difficult when additional lags are introduced. Our present point is that for the common factor problem, the constraints are initially given in terms of the vector function **g**. The derivation of **h** requires an additional step, and one which may be extremely difficult to perform.

With the existence of such **g**, we may replace the hypothesis that $\mathbf{h}(\beta_o) = \mathbf{0}$ with the hypothesis that there exists θ_o such that $\beta_o = \mathbf{g}(\theta_o)$. Moreover, we may find the maximum of $L(\beta,\sigma^2)$ subject to $\mathbf{h}(\beta) = \mathbf{0}$ simply by finding the unconstrained maximum of $L[\mathbf{g}(\theta),\sigma^2]$. If we let $\hat{\theta}$ denote the solution to this problem, then $\tilde{\beta} = \mathbf{g}(\hat{\theta})$. Now, for case of the linear model with normal errors, the log-likelihood function for θ and σ^2 is:

$$L[\mathbf{g}(\theta),\sigma^2] = -(n/2)\log(2\pi\sigma^2) - [\mathbf{y} - \mathbf{Z}\mathbf{g}(\theta)]' [\mathbf{y} - \mathbf{Z}\mathbf{g}(\theta)]/(2\sigma^2)$$

From the latter, it may be seen that the MLE of θ_o is the solution to:

$$\text{minimize } (\theta): \; S(\theta) = [\mathbf{y} - \mathbf{Z}\mathbf{g}(\theta)]' [\mathbf{y} - \mathbf{Z}\mathbf{g}(\theta)]$$

After simplifying the objective function, it will be found that:

$$S(\theta) = \hat{\sigma}^2\{[\mathbf{g}(\theta) - \hat{\beta}]' (\mathbf{Z}'\mathbf{Z}/\hat{\sigma}^2)[\mathbf{g}(\theta) - \hat{\beta}] + n\} \tag{17}$$

Hence, the minimim of $S(\theta)$ is equal to the solution of:

$$\text{minimize } (\theta): \; [\mathbf{g}(\theta) - \hat{\beta}]' (\mathbf{Z}'\mathbf{Z}/\hat{\sigma}^2)[\mathbf{g}(\theta) - \hat{\beta}] \tag{18}$$

The advantage of this formulation becomes clear in the next section where it is shown that problems of this form can generally be solved using the Gauss-Newton algorithm.

Let the optimized value of the objective function to the above problem be denoted as λ_D. Since the restricted MLE of $\hat{\sigma}^2$ is $\tilde{\sigma}^2 = S(\hat{\theta}) / n$, then from (17) it may be concluded:

$$\lambda_D = n(\tilde{\sigma}^2 - \hat{\sigma}^2) / \hat{\sigma}^2$$

Using (14) and 16) in conjunction with the latter, it is easy to confirm that the log-likelihood ratio and Lagrangian multiplier statistics may be expressed in terms of λ_D with:

$$\lambda_{LR} = n\log(1 + \lambda_D/n)$$
$$\lambda_{LM} = n\lambda_D / (\lambda_D + n)$$

Both λ_{LR} and λ_{LM} converge to λ_D as n approaches infinity; hence, the three statistics are asymptotically equivalent. Using the fact that $\log(1 + x) < x$ for $x \in (0, \infty)$, it may be concluded from the above that $\lambda_D > \lambda_{LR}$. We have already established the fact that $\lambda_{LR} > \lambda_{LM}$; hence, $\lambda_D > \lambda_{LR} > \lambda_{LM}$.

4. The Gauss-Newton Algorithm

Suppose we have a linear model of the form:

$$\mathbf{y} = \mathbf{Z}\beta_o + \varepsilon$$

where $\varepsilon \sim N(\mathbf{0}, \sigma_o^2\mathbf{I})$. Let the likelihood function for β_o and σ_o^2 be denoted by $L(\beta, \sigma^2)$. In this section, we consider calculation of the maximum of $L(\beta, \sigma^2)$ subject to the restriction $\mathbf{h}(\beta) = \mathbf{0}$. Let the unrestricted MLEs of β_o and σ_o^2 be denoted with $\hat{\beta}$ and $\hat{\sigma}^2$, and let the restricted MLEs be denoted with $\tilde{\beta}$ and $\tilde{\sigma}^2$. Following the previous section, we assume that $\mathbf{h}(\beta_o) = \mathbf{0}$ if and only if there exists θ_o such that $\beta_o = \mathbf{g}(\theta_o)$, in which case, $\tilde{\beta} = \mathbf{g}(\hat{\theta})$, where $\hat{\theta}$ is the solution to:

$$\text{minimize } (\theta): [\mathbf{g}(\theta) - \hat{\beta}]' (\mathbf{Z}'\mathbf{Z}/\hat{\sigma}^2)[\mathbf{g}(\theta) - \hat{\beta}] \qquad (19)$$

The Gauss-Newton algorithm is an iterative routine specifically designed to solve problems having the general form:

minimize $(\mathbf{x})$: $e(\mathbf{x}) = \mathbf{f}(\mathbf{x})' \mathbf{A}\mathbf{f}(\mathbf{x})$

where $\mathbf{f}$ is a differentiable m-dimensional vector function; $\mathbf{x}$ is a vector of dimension not greater than m, and $\mathbf{A}$ is an mxm positive definite matrix. The problem in (19) has precisely this form. Now, let $\bar{\mathbf{x}}$ be any realization of $\mathbf{x}$. The algorithm proceeds from $\bar{\mathbf{x}}$ by replacing $\mathbf{f}(\mathbf{x})$ in the objective function with a first-order Taylor approximation about $\bar{\mathbf{x}}$. The resulting function is:

$$\bar{e}(\mathbf{x}) = [\mathbf{f}(\bar{\mathbf{x}}) + \nabla \mathbf{f}(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})]' \mathbf{A} [\mathbf{f}(\bar{\mathbf{x}}) + \nabla \mathbf{f}(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})]$$

$\bar{e}(\mathbf{x})$ is minimized at $\mathbf{x}^*$, where:

$$\mathbf{x}^* = \bar{\mathbf{x}} - [\nabla \mathbf{f}(\bar{\mathbf{x}})' \mathbf{A} \nabla \mathbf{f}(\bar{\mathbf{x}})]^{-1} \nabla \mathbf{f}(\bar{\mathbf{x}})' \mathbf{A}\mathbf{f}(\bar{\mathbf{x}})$$

The algorithm repeatedly replaces $\bar{\mathbf{x}}$ with $\mathbf{x}^*$ until $\nabla e(\mathbf{x}^*)$ is sufficiently near the zero vector. The algorithm may be summarized as follows:

Initialization Step: Choose a starting point $\mathbf{x}_1$, and a distinguishability constant $\delta > 0$. Set $k = 1$ and go to main step.

Main Step:

1) Set $\mathbf{x}^* = \mathbf{x}_k - [\nabla \mathbf{f}(\mathbf{x}_k)' \mathbf{A} \nabla \mathbf{f}(\mathbf{x}_k)]^{-1} \nabla \mathbf{f}(\mathbf{x}_k)' \mathbf{A}\mathbf{f}(\mathbf{x}_k)$.

2) If $|\nabla e(\mathbf{x}^*)| < \delta$ then stop; otherwise, replace k with k + 1; set $\mathbf{x}_k = \mathbf{x}^*$, and repeat step one.

The algorithm breaks down upon arriving at points where $\nabla \mathbf{f}$ is not of full column rank since the matrix inversion in the main step will not be possible in such cases. However, it can be shown that if $\nabla \mathbf{f}$ is of full column rank at the optimum, then the algorithm always converges to the optimum when initiated in a neighborhood of it. It is possible for the algorithm to endlessly iterate without reducing the value of the objective function. This problem may be avoided by redefining $\mathbf{x}^*$ as:

$$\mathbf{x}^* = \mathbf{x}_k - \lambda[\nabla \mathbf{f}(\mathbf{x}_k)' \mathbf{A} \nabla \mathbf{f}(\mathbf{x}_k)]^{-1} \nabla \mathbf{f}(\mathbf{x}_k)' \mathbf{A}\mathbf{f}(\mathbf{x}_k)$$

It can be shown that if λ is sufficiently small, then $\mathbf{x}^*$ will render a lower value in the objective function than $\mathbf{x}_k$, provided that $\mathbf{x}_k$ is not a stationary point. Therefore,

with appropriate choice of λ, we can cause the algorithm to render successively lower values of e. A possible approach would be to begin with $\lambda = 1$, then successively divide this value in half until obtaining an $\mathbf{x}^*$ rendering a lower value in the objective function.

The algorithm terminates where $\nabla e(\mathbf{x}^*)$ is sufficiently near the zero vector. Since this condition will hold at all local minima of e, it is possible for the algorithm to converge at any of such points. For this reason, location of the global minimum may require initiation of the algorithm from several starting points.

5. Generation of the Data

In this section, we describe the procedures that were used in the generation of data for the simulation experiments. All experiments necessitated the generation of data for models having the form:

$$y_t = \beta_1 y_{t-1} + \ldots + \beta_q y_{t-q} + \alpha_0 x_t + \ldots + \alpha_r x_{t-r} + \varepsilon_t \tag{20}$$

$$x_t = \lambda x_{t-1} + \nu_t \tag{21}$$

Hence, all models involved only one exogenous variable; however, varying numbers of lags of this exogenous variable were included among the regressors. The exogenous variable is always generated by the first-order autoregressive process described by the second of the above equations.

Generation of the Disturbances

In all models, the ε_t and ν_t are generated as normal random variables. Most programming languages include random number generators only for the uniform distribution over the interval from zero to one (U(0,1)). However, the following theorem may be used to transform U(0,1) random variates to random variates following any desired distribution function:

2. Theorem: Let F be the cumulative distribution function corresponding to any continuous random variate, and let u be a U(0,1) variate, then $x = F^{-1}(u)$ is a random variable distributed with cumulative distribution function equal to F.

Proof: For any constant a:

$$P[x < a] = P[F^{-1}(u) < a] = P[u < F(a)] = F(a)$$

which shows the cumulative distribution function of x to be F(x), and the proof is complete.

Unfortunately, closed-form representations do not exist for the normal cumulative distribution function or for its inverse. Consequently, direct application of the theorem would necessitate the usage of approximations to these functions. However, it happens that the usage of approximations can be avoided by generating the random variates in pairs. This is illustrated in the following procedure due to Box and Muller: Let x_1 and x_2 denote the independent standard normal variates to be generated. The joint probability density for these two variates is:

$$f_{x_1, x_2}(x_1, x_2) = (1/2\pi)\exp[-(x_1^2 + x_2^2)/2]$$

Suppose we convert to polar coordinates using the transforms:

$$x_1 = r\cos(\theta)$$
$$x_2 = r\sin(\theta)$$

The joint density of r and θ is:

$$f_{r,\theta}(r,\theta) = (r/2\pi)\exp[-r^2/2]; \quad r \in [0,\infty), \ \theta \in [0,2\pi]$$

It is apparent from the nature of the problem that r and θ are independent. Of course, this may be formally verified by calculating the conditional densities. The marginal density for r is.

$$f_r(r) = r\exp[-r^2/2]; \quad r \in [0,\infty)$$

Hence, the cumulative density for r is:

$$F_r(r) = 1 - \exp[-r^2/2]; \quad r \in [0,\infty)$$

Using the latter, we can apply Theorem two to randomly generate r. In particular, if u_1 denotes a U(0,1) variate, then r can be generated with:

$$r = [-2\log(u_1)]^{1/2}$$

The maginal density for θ is simply a U(0,2π); consequently, θ can be generated with:

$$\theta = 2\pi u_2$$

where u_2 is an U(0,1) variate that is independent of u_1. x_1 and x_2 are then calculated using:

$$x_1 = r\cos(\theta) = [-2\log(u_1)]^{1/2} \cos(2\pi u_2) \quad (22)$$

$$x_2 = r\sin(\theta) = [-2\log(u_1)]^{1/2} \sin(2\pi u_2) \quad (23)$$

Marsaglia and Bray propose a method similar to that of Box and Muller, but which avoids the calculation of the sine and cosine functions. Let v_1 and v_2 be independent U(-1,1) variates conditional on $w = v_1^2 + v_2^2 \leq 1$. Hence, (v_1, v_2) may be viewed as a pair of coordinates uniformly distributed on a unit circle centered at the origin. Such coordinate pairs may be generated by generating both v_1 and v_2 from the U(-1,1) and rejecting all pairs for which $w > 1$. With use of polar coordinates, it is not difficult to confirm that w is distributed as U(0,1), and that $v_1/w^{1/2}$ and $v_2/w^{1/2}$ are distributed as the cosine and sine of a random variate distributed as U(0,2π). Moreover, both $v_1/w^{1/2}$ and $v_2/w^{1/2}$ are independent of w. Consequently, the following substitutions can be made in (22) and (23): u_1 can be replaced with w, $\cos(2\pi u_2)$ can be replaced with $v_1/w^{1/2}$, and $\sin(2\pi u_2)$ can be replaced with $v_2/w^{1/2}$. After making these substitutions, one obtains:

$$x_1 = v_1[-2\log(w)/w]^{1/2}$$

$$x_2 = v_2[-2\log(w)/w]^{1/2}$$

The latter formuli were used in the simulation experiments for the generation of standard normal variates.

Generation of Initial Values

The data series can be generated using (20) and (21) once initial values are provided; however, appropriate generation of the initial values can be a rather difficult task. If the first period corresponds to t = 1, then initiation of the sequences will require values for $(y_0,,,y_{1-q},x_1,,,x_{1-r})$. The usual approach is to generate the initial values such that their means, covariances, and autocovariances are equal to those of the sequences which they initiate. Of course, this assumes that the sequences are stationary, so that their means, covariances, and autocovariances do in fact converge upon constants with the progression of the sequences. The process generating y_t will be stationary provided that the roots to $1 - \beta_1 L - \ldots - \beta_q L^q = 0$ all lie outside the unit circle. Moreover, the process generating x_t will be stationary if $|\lambda| < 1$. All parameterizations used in the simulations are chosen such that these

stationarity conditions are satisfied. Now, since both (20) and (21) are without constant terms, both y_t and x_t have expectation equal to zero for all t. Suppose that the covariance marix of $(y_{t-1},,,y_{t-q}, x_t,,,x_{t-r})$ is Σ. Let $\mathbf{T}$ be a triangular matrix satisfying $\mathbf{T}'\mathbf{T} = \Sigma$, and let $\mathbf{w}$ be a conformable vector of independent standard normal variates. Initial values having zero mean and covariance matrix Σ can be generated with $\mathbf{Tw}$. Hence, the problem of generating initial values reduces to the problem of determining $\mathbf{T}$. In the following, we treat the construction of Σ. The problem of determining $\mathbf{T}$ from Σ is addressed in Appendix II.

The components of Σ are most easily found by formulating (20) and (21) as a vector autoregression. We may then find the components of Σ from among the components of the autocovariance matrices of the vector autoregression. Subsequently, consider the general vector autoregression of order s:

$$\mathbf{z}_t = \mathbf{A}_1\mathbf{z}_{t-1} + \ldots + \mathbf{A}_s\mathbf{z}_{t-s} + \varepsilon_t \tag{24}$$

where the $\mathbf{A}_i$ are mxm matrices of constants, and where the ε_t are iid random vectors having zero mean vector and covariance matrix Ω. Both (20) and (21) can indeed be formulated as such a model. Analysis of the general vector autoregression can be greatly expedited by reformulating it in terms of a vector autoregression of order one. This may be done as follows: Let $\mathbf{v}_i(t) = \mathbf{z}_{t-i}$ for i = 0,1, , ,s - 1; moreover, let $\mathbf{v}(t) = [\mathbf{v}_0'(t), , ,\mathbf{v}_{s-1}'(t)]'$, then (24) may be expressed as:

$$\mathbf{v}(t) = \mathbf{B}\mathbf{v}(t-1) + \xi_t \tag{25}$$

where:

$$\mathbf{B} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \ldots & \mathbf{A}_{s-1} & \mathbf{A}_s \\ \mathbf{I} & \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & & \cdot & \cdot \\ \cdot & \mathbf{0} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I} & \mathbf{0} \end{bmatrix} \qquad \xi_t = \begin{bmatrix} \varepsilon_t \\ \mathbf{0} \\ \mathbf{0} \\ \cdot \\ \cdot \\ \mathbf{0} \end{bmatrix}$$

Let the covariance matrix of $\mathbf{v}_t$ be denoted as Γ_0. To find Γ_0, postmultiply (25) by $\mathbf{v}(t)'$ and take the expectation of both sides to obtain:

$$\Gamma_0 = \mathbf{B}\Gamma_1' + \Psi \tag{26}$$

where $\Gamma_1 = E[\mathbf{v}(t)\mathbf{v}(t-1)']$, and:

$$\Psi = \begin{bmatrix} \Sigma & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

Observe that Γ_0 is symmetric, but this is not necessarily the case for Γ_1. Next, postmultiply (25) by $\mathbf{v}(t-1)'$ and take the expectation of both sides to obtain:

$$\Gamma_1 = \mathbf{B}\Gamma_0 \tag{27}$$

Upon substituting (27) into (26), one concludes:

$$\Gamma_0 = \mathbf{B}\Gamma_0\mathbf{B}' + \Psi$$

which implies:

$$\text{vec}(\Gamma_0) = \mathbf{B}\oplus\mathbf{B}\text{vec}(\Gamma_0) + \text{vec}(\Psi)$$

Hence:

$$\text{vec}(\Gamma_0) = (\mathbf{I} - \mathbf{B}\oplus\mathbf{B})^{-1}\text{vec}(\Psi)$$

where vec(.) is the vectorization operator, and $\oplus$ denotes the Kroenecker product operator. Both of these operators are discussed in APPENDIX I. The components of Σ may be found among the components of Γ_0 and Γ_1.

6. Design of the Experiments and the Results

Simulation experiments were conducted upon three distributed lag models. For each model, the presence of a common factor in the lag polynomials was tested using each of the four test criteria previously considered. The criteria were compared to determine which, if any, proved to be the best for small sample tests. In this section, we discuss the design of these experiments and present the results.

The simulated models were DL(1,1), DL(1,2), and DL(2,1). Hence, all models had the general form:

$$y_t = \beta_1 y_{t-1} + \dots + \beta_q y_{t-q} + \alpha_0 x_t + \dots + \alpha_r x_{t-r} + \varepsilon_t \tag{28}$$

with varying numbers of lags being used on the endogenous and exogenous variable. In all models, the ε_t were generated as independent standard normal variates. Also, x_t was generated in all models using:

$$x_t = .9x_{t-1} + v_t$$

where the v_t were generated as independent standard normal variates. The above process was chosen for the generation of the independent variable because it approximately characterizes a large number of economic variables, particularly when measured on an annual basis.

For the DL(1,1) model, the presence of a common factor in the lag polynomial implies that the model may be written as:

$$y_t = \psi_0 x_t + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t$$

or equivalently:

$$y_t = \rho y_{t-1} + \psi_0 x_t - \rho\psi_0 x_{t-1} + \varepsilon_t$$

Therefore, we have the following correspondence between the coefficients of the unrestricted and restricted models:

$$\beta_1 = \rho, \quad \alpha_0 = \psi_0, \quad \alpha_1 = -\rho\psi_0$$

Hence, two parameters are mapped into three coefficients. The existence of ρ and ψ_0 satisfying the above equations implies and is implied by the constraint:

$$h = \beta_1\alpha_0 + \alpha_1 = 0 \qquad (29)$$

The coefficients to (28) were chosen following procedures used in the Mizon and Hendry experiments. In particular, all coefficients in (28) were chosen in accordance with the presence of a common factor except the coefficient on the furthest lag of x_t. This coefficient was chosen so as to violate (29) by specified amounts. Specifically, we set:

$$\beta_1 = \rho, \alpha_0 = \psi_0, \quad \alpha_1 = \gamma/n^{1/2} - \rho\psi_0$$

for $\psi_0 = 1$, $\rho = (-.8,-.2,.2,.8)$, $\gamma = (0,.5,1,1.5,2)$, and $n = (20,40,60)$. Using the anlaysis in section two, it may be confirmed that as the sample size n approaches

infinity, the coefficients converge upon the null hypothesis in such a way as to cause the test criteria to converge in distribution to a noncentral chi-squared with one degree of freedom and noncentrality parameter:

$$\mu = \gamma' (\nabla h \Sigma^{-1} \nabla h)^{-1} \gamma$$

where ∇h is evaluated at the limits of the coefficients, and where Σ is the covariance of the regressors at the limits of the coefficients. The test size was always set at five percent; consequently, as sample size approaches infinity, the powers of the simulated tests should converge upon:

$$\pi = P[\chi^2(1,\mu) > \chi^2_{1,.05}] \tag{30}$$

By comparing the sample powers of the various tests with the asymptotic power above, one can determine the relative rates of convergence in the various criteria.

Values for ρ were chosen so that the relative performance of the tests could be evaluated on both high and low degrees of both positive and negative autocorrelation. 1,000 replications were simulated on all parameterizations except those corresponding to the null hypothesis ($\gamma = 0$) where 2,000 replications were used instead. The larger number of replications on the null hypothesis was used to achieve greater accuracy in the measurement of test size. Since three of the four criteria are algebraically related, it was expected that test size would become an important determinant in the selection of the best test. Suppose that the true size for a particular test is p, and that the sample power is $\hat{p}$, then the variance of $\hat{p}$ is $p(1 - p)/n$. Therefore, an approximate 99% confidence interval about some hypothesized test size, say p_o, would be $p_o \pm 2\,[p_o(1 - p_o)/n]^{1/2}$. This approximation is based upon the asymptotic normality of $\hat{p}$. Since nominal test sizes were set at five percent, 2000 replications would produce a confidence interval of radius equal to .9747 percentage points, or less than one percent.

The results for the DL(1,1) model are reported in Table I. In this and all other tables, the following notation is used: λ_W, λ_{LR}, and λ_{LM} denote the Wald, log-likelihood ratio, and Lagrangian multiplier criteria, respectively. λ_D denotes the optimal solution to the problem in (18), and π denotes the noncentral chi-squared approximation to the power function defined in (30).

Several important conclusions can be drawn from Table I. First, the sample powers are closely approximated by π, particularly for samples sizes greater than or equal to 40. Second, the experiments demonstrate that the Wald statistic is not necessarily greater than the log-likelihood ratio for tests of nonlinear constraints, since the power of the likelihood ratio test is greater than that of the Wald test under numerous parameterizations. Mizon (1977a) has also produced examples where the Wald statistic is less than the likelihood ratio in tests of nonlinear constraints.

Third, no one test appreciably outperforms the others; moreover, where one test has greater power on the alternative hypothesis than the others, it is generally found that the test also has larger size. Of course, to make fair comparison among the tests, the criteria should be adjusted so as to render equal size; however, no effort was made toward such adjustments since the unadjusted powers were felt to convincingly suggest approximate equivalence between the tests. Fourth, tests using λ_W, λ_D, and λ_{LR} consistently had estimated test size greater than the nominal size. λ_{LM} produced estimated test sizes greater than the nominal size in all cases but one. Since, $\lambda_D > \lambda_{LR} > \lambda_{LM}$, it naturally follows that among these three statistics, λ_{LM} produced tests having estimated sizes nearest to the nominal size, and tests using λ_D had estimated sizes being the furthest away. However, in all but one case, the estimated sizes for the λ_{LM} tests also proved nearer to the nominal size than those for the λ_W tests. Fifth, the powers of all tests on the alternative hypothesis were highly affected by the parameter ρ. The powers consistently increased with increases in ρ. The results indicate that the probability of type II error could be quite large if the true model is approximated by a model possessing negatively autocorrelated disturbances.

For the DL(1,2) model, a common factor in the lag polynomials implies that the model may be written as:

$$y_t = \rho y_{t-1} + \psi_0 x_t + (\psi_1 - \rho\psi_0)x_{t-1} - \rho\psi_1 x_{t-2} + \varepsilon_t$$

The coefficients of the unrestricted model in (28) will satisfy the above factorizations if and only if:

$$h = \beta_1^2\alpha_0 + \beta_1\alpha_1 + \alpha_2 = 0$$

The following parameterizations were used in the construction of the simulated models:

$$\beta_1 = \rho, \quad \alpha_0 = \psi_0, \quad \alpha_1 = \psi_1 - \rho\psi_0, \quad \alpha_2 = \gamma/n^{1/2} - \rho\psi_1$$

for $\psi_0 = 1$, $\rho = (-.8,-.2,.2,.8)$, $\psi_1 = (-.5,.5)$, and $n = (20,40,60)$. All other aspects of the experiment were the same as that for the DL(1,1) model.

The results for the DL(1,2) model are reported in Table II. All of the conclusions drawn from the previous simulation are confirmed here. π proves to be an accurate approximation of the estimated powers, particularly with 40 or more observations. As before, test powers on the alternative hypothesis are approximately equivalent, with differences usually being largely explained by differences in size. For all tests, the estimated sizes were greater than the nominal size without exception. In all but three cases, the λ_{LM} tests had estimated sizes nearer to the nominal size than did the

λ_W tests. As before, the powers of the tests increase as ρ is increased, and the risk of type II error is indicated to be quite large when the true model is approximated by a model having negatively autocorrelated disturbances. The powers appear to be invariant with respect to the parameter ψ_1, and the asymptotic powers are invariant. Indeed, the inclusion of the additional lag on the exogenous variable appears to have had little effect on the power function since both the estimated powers and asymptotic powers for the DL(1,1) and DL(1,2) models are approximately the same for equal values of γ and ρ.

For the DL(2,1) model, a common factor in the lag polynomials implies a model of the form:

$$y_t = (\rho + \theta_1)y_{t-1} - \rho\theta_1 y_{t-2} + \psi_0 x_t - \rho\psi_0 x_{t-1} + \varepsilon_t$$

The coefficients of the unrestricted model in (28) will satisfy the above factorizations if and only if:

$$h = (\alpha_1/\alpha_0)^2 + (\alpha_1/\alpha_0)\beta_1 - \beta_2 = 0$$

Coefficients for the simulated models were calculated with the following:

$$\beta_1 = \rho + \theta_1, \quad \beta_2 = -\rho\theta_1, \quad \alpha_0 = \psi_0$$
$$\alpha_1 = \psi_0\{-(\rho + \theta_1) \pm [(\rho - \theta_1)^2 + 4\gamma/n^{1/2}]^{1/2}\}/2$$

The $\pm$ term in the last relation is a consequence of the fact that for given ψ_0, r, and θ_1, there will be two values of α_1 producing $h = \gamma/n^{1/2}$. Choice between the solutions was made such that $\gamma = 0$ rendered $\alpha_1 = -\rho\psi_0$. Hence, the negative sign was chosen if $\rho > \theta_1$, but the positive sign was chosen otherwise. Parameter values were $\psi_0 = 1$, θ_1 = (-.5,.5), ρ = (-.8,-.2,.2,.8), and n = (20,40,60). All other aspects of the experiment were the same as for the previous experiments.

The results for the DL(2,1) model are reported in Table III. The most important conclusion to be drawn from these results is that Wald statistic clearly becomes inferior to the other criteria under certain parameterizations. The powers under the Wald statistic are in some instances very low; moreover, in some cases, the powers actually decreased with increasing γ. Such perverse behavior indicates that the rate of convergence in λ_W towards its asymptotic distribution is extremely slow. The other criteria still perform well; however, the rates of convergence in these are also slower than in the previous experiments, as is evidenced by the fact that π is a much poorer approximation to the estimated powers than before. Certain of the previous results are also reflected here. In particular, the estimated test size for λ_D, λ_{LR}, and λ_{LM}

still consistently exceed the nominal size; hence, λ_{LM} again performs the best insofar as approximation to the nominal size is concerned. As before, the power function appears to be decreasing in ρ, with values of ρ near -1 resulting in high probabilities of type II error. Test powers are clearly affected by θ_1, but the direction of the effect is determined partly by an interaction between θ_1 and ρ that is too complex to generalize in simplistic terms.

A further simulation of one of the DL(2,1) parameterizations was conducted using larger sample sizes with the intent of obtaining a better measure of the rate of convergence in the test criteria. The simulated parameterization corresponded to $\theta_1 = .5$, $\rho = .2$, and $\gamma = (0,.5,1,1.5,2)$. Sample sizes up to 1000 were simulated. As in the previous simulations, 1000 replications were used for $\gamma > 0$, but 2000 replications were used for $\gamma = 0$. The results of this experiment are reported in Table IV. The power estimates for sample sizes 20, 40, and 60 were lifted from Table III. The estimated powers reported in the Table IV further confirm that all criteria converge slowly under this parameterization; however, convergence in the Wald statistic is clearly the slowest. The power function for the Wald statistic continues to have power declining with γ in sample sizes as large as 125.

7. Conclusions

The most important conclusion to be drawn from the simulation experiments is that the Wald criterion can perform very poorly under certain parameterizations of the DL(2,1) model. The other criteria also exhibit slow convergence under these parameterizations; however, their performance clearly excels that of the Wald statistic. Under the DL(1,1) and DL(1,2) models, all criteria produced tests having powers very close to the asymptotic powers, particularly for sample sizes of 40 or more. These results indicate a possible generalization to be that the rate of convergence in all criteria is adversely affected by the number of lags on the dependent variable included among the regressors, and in the case of the Wald statistic the effect can be extreme.

The experiments indicate a clear tendency on the part of all criteria to produce tests with sizes exceeding the nominal size. Because of this and the algebraic relation, $\lambda_{LM} < \lambda_{LR} < \lambda_D$, the Lagrangian multiplier statistic produces tests with sizes closer to the nominal sizes than λ_{LR} and λ_D. Though the Lagrangian multiplier bears no such algebraic relationship with the Wald statistic, the simulations strongly indicate that it also excels the Wald statistic in this regard. Based upon this observation and the observations of the previous paragraph, the Lagrangian multiplier statistic appears to be the best of the four criteria for the simulated models.

The ability of all tests to reject parameterizations in the neighborhood of the null hypothesis appears highly contingent upon the value of the autocorrelation coefficient under the null. Both estimated and asymptotic powers consistently increased with increases in the signed value of the autocorrelation coefficient. The probability of type II error appears quite large for autocorrelation coefficients near -1.

TABLE I

SIMULATION RESULTS FOR DL(1,1) MODELS

Criteria	Parameters		γ				
	n	r	0	.5	1	1.5	2.0
λ_W	20	−.8	9.35	9.70	11.80	18.80	24.50
λ_D			9.60	10.30	12.30	19.50	24.80
λ_{LR}			7.75	8.60	10.40	17.20	22.70
λ_{LM}			6.20	7.00	8.80	14.60	20.60
λ_W	40	−.8	10.10	10.70	11.60	14.00	21.40
λ_D			10.50	11.30	12.60	14.50	22.20
λ_{LR}			8.70	10.20	10.90	12.20	19.60
λ_{LM}			7.15	7.90	8.70	10.50	17.20
λ_W	60	−.8	6.45	7.40	10.10	15.10	22.20
λ_D			6.50	7.30	10.20	15.20	22.80
λ_{LR}			5.80	7.10	9.80	14.60	21.60
λ_{LM}			5.45	6.90	9.30	14.10	20.80
π			5.00	5.93	8.79	13.70	20.70
λ_W	20	−.2	7.40	9.40	18.20	27.30	43.10
λ_D			8.75	11.20	19.40	29.00	44.30
λ_{LR}			7.50	9.50	16.90	25.80	41.60
λ_{LM}			6.10	7.60	14.80	23.00	37.90
λ_W	40	−.2	7.55	10.20	16.00	25.30	36.60
λ_D			8.90	11.60	18.80	28.60	40.60
λ_{LR}			7.40	10.10	16.60	25.60	36.80
λ_{LM}			5.90	8.70	13.30	22.70	33.30
λ_W	60	−.2	5.80	8.30	13.20	24.10	41.40
λ_D			6.25	8.80	14.80	25.50	42.80
λ_{LR}			6.15	8.40	13.70	24.80	41.80
λ_{LM}			5.60	7.90	12.50	23.20	40.60
π			5.00	7.07	13.50	24.50	39.38

TABLE I (Continued)

λ_W	20	.2	6.35	11.90	20.20	41.50	62.40
λ_D			8.70	14.20	23.80	45.80	68.10
λ_{LR}			7.15	12.70	21.60	42.80	64.80
λ_{LM}			6.15	10.40	18.40	39.10	61.50
λ_W	40	.2	7.00	11.00	21.70	38.10	59.70
λ_D			9.10	14.80	26.60	44.60	64.50
λ_{LR}			7.50	11.90	23.60	41.30	61.50
λ_{LM}			5.70	10.30	19.70	37.80	58.30
λ_W	60	.2	5.30	9.40	24.80	42.50	64.30
λ_D			6.25	10.40	27.70	44.00	67.20
λ_{LR}			5.90	9.70	25.70	43.60	65.40
λ_{LM}			5.35	9.10	24.70	42.40	63.90
π			5.00	9.31	22.82	44.40	67.92
λ_W	20	.8	9.80	23.00	47.60	73.80	87.90
λ_D			11.80	26.70	50.60	73.50	87.50
λ_{LR}			9.85	23.30	46.50	70.60	85.50
λ_{LM}			8.30	19.60	43.00	67.50	82.90
λ_W	40	.8	9.95	20.90	50.80	72.90	85.60
λ_D			12.25	25.30	52.10	73.70	84.30
λ_{LR}			10.25	23.10	49.60	70.20	82.90
λ_{LM}			8.40	20.30	45.60	66.40	80.30
λ_W	60	.8	7.00	19.00	52.40	82.40	95.70
λ_D			8.10	20.09	54.20	82.80	95.40
λ_{LR}			7.50	19.80	53.20	82.10	95.10
λ_{LM}			6.75	18.50	51.80	81.30	94.60
π			5.00	20.10	60.88	91.84	99.40

TABLE II

SIMULATION RESULTS FOR DL(1,2) MODELS

Criteria	Parameters			γ				
	n	Ψ_1	ρ	0	.5	1	1.5	2.0
λ_W	20	−.5	−.8	10.10	10.70	11.60	14.00	21.40
λ_D				10.50	11.30	12.60	14.50	22.20
λ_{LR}				8.70	10.20	10.90	12.20	19.60
λ_{LM}				7.15	7.90	8.70	10.50	17.20
λ_W	40	−.5	−.8	6.80	8.50	10.50	12.90	16.40
λ_D				6.85	8.40	10.70	12.50	16.30
λ_{LR}				6.45	7.60	10.20	12.00	15.20
λ_{LM}				5.70	7.00	9.00	10.90	14.70
λ_W	60	−.5	−.8	6.85	7.30	8.90	12.00	17.50
λ_D				6.90	7.50	9.10	12.40	17.50
λ_{LR}				6.35	7.10	8.60	11.50	17.00
λ_{LM}				6.05	6.50	7.90	10.90	16.50
π				5.00	5.58	7.35	10.37	14.69
λ_W	20	−.5	−.2	7.55	10.20	16.00	25.30	36.60
λ_D				8.90	11.60	18.80	28.60	40.60
λ_{LR}				7.40	10.10	16.60	25.60	36.80
λ_{LM}				5.90	8.70	13.30	22.70	33.30
λ_W	40	−.5	−.2	5.45	8.30	15.30	23.80	39.40
λ_D				5.80	9.10	16.40	25.90	40.60
λ_{LR}				5.50	8.20	15.20	24.80	39.30
λ_{LM}				5.20	7.40	14.30	23.50	37.80
λ_W	60	−.5	−.2	5.55	7.60	13.90	24.70	39.40
λ_D				6.10	7.90	14.80	26.00	40.70
λ_{LR}				5.65	7.70	14.10	24.60	39.30
λ_{LM}				5.10	7.50	13.50	23.70	38.10
π				5.00	7.00	13.21	23.84	38.29

TABLE II (Continued)

λ_W	20	−.5	.2	7.00	11.00	21.70	38.10	59.70
λ_D				9.10	14.80	26.60	44.60	64.50
λ_{LR}				7.50	11.90	23.60	41.30	61.50
λ_{LM}				5.70	10.30	19.70	37.80	58.30
λ_W	40	−.5	.2	5.45	9.70	21.60	41.10	63.30
λ_D				6.40	11.90	24.10	45.20	67.50
λ_{LR}				5.95	10.70	23.10	43.10	65.70
λ_{LM}				5.25	9.50	21.60	41.20	63.60
λ_W	60	−.5	.2	5.55	8.80	22.20	44.30	64.80
λ_D				6.25	10.20	23.90	46.90	67.30
λ_{LR}				5.75	9.30	23.30	45.90	65.60
λ_{LM}				5.65	8.60	22.20	44.20	64.40
π				5.00	9.18	22.31	43.39	66.68
λ_W	20	−.5	.8	9.95	20.90	50.80	72.90	85.60
λ_D				12.15	25.30	52.10	73.70	84.30
λ_{LR}				10.25	23.10	49.60	70.20	82.90
λ_{LM}				8.40	20.30	45.60	66.40	80.30
λ_W	40	−.5	.8	7.35	20.50	50.10	77.80	92.40
λ_D				8.15	23.90	52.80	78.70	92.10
λ_{LR}				7.75	22.30	50.70	77.10	91.60
λ_{LM}				6.90	21.20	48.70	75.70	90.90
λ_W	60	−.5	.8	6.85	21.60	53.40	83.00	92.80
λ_D				8.00	23.80	55.40	83.80	92.70
λ_{LR}				7.85	22.70	53.90	83.30	92.10
λ_{LM}				6.90	21.60	53.40	81.80	91.90
π				5.00	19.62	59.53	91.02	99.27

TABLE II (Continued)

λ_W	20	.5	−.8	9.95	10.70	11.00	15.90	20.90
λ_D				10.25	11.40	11.20	16.70	21.30
λ_{LR}				8.75	9.90	9.00	14.80	19.30
λ_{LM}				7.65	7.90	7.40	12.20	16.60
λ_W	40	.5	−.8	7.70	8.50	10.50	13.80	18.10
λ_D				7.80	8.50	10.50	13.50	18.60
λ_{LR}				7.20	7.70	9.40	13.00	17.50
λ_{LM}				6.60	7.00	8.90	12.40	16.50
λ_W	60	.5	−.8	6.35	7.00	8.50	12.30	17.50
λ_D				6.45	6.90	8.80	12.60	18.00
λ_{LR}				6.00	6.50	8.10	12.00	17.20
λ_{LM}				5.60	6.30	7.40	11.20	16.80
π				5.00	5.58	7.35	10.37	14.69
λ_W	20	.5	−.2	8.15	8.80	15.50	25.30	40.00
λ_D				9.80	10.50	18.70	27.70	43.30
λ_{LR}				8.40	9.10	16.20	25.80	40.20
λ_{LM}				6.70	7.20	14.60	22.70	37.00
λ_W	40	.5	−.2	5.80	8.80	12.60	21.90	37.30
λ_D				6.05	9.10	13.50	23.40	39.80
λ_{LR}				5.90	8.50	12.60	21.70	37.30
λ_{LM}				5.20	8.30	12.00	20.50	35.30
λ_W	60	.5	−.2	5.60	8.60	9.20	22.70	37.30
λ_D				5.90	9.00	9.10	23.50	38.00
λ_{LR}				5.45	8.70	8.80	22.90	37.50
λ_{LM}				5.30	8.50	8.30	22.20	36.30
π				5.00	7.00	13.21	23.84	38.29

TABLE II (Continued)

λ_W	20	.5	.2	6.60	12.90	20.90	38.10	61.10
λ_D				9.05	15.70	26.60	44.50	65.90
λ_{LR}				7.20	14.10	24.20	40.70	63.50
λ_{LM}				5.75	12.40	20.20	37.40	60.00
λ_W	40	.5	.2	5.85	9.80	23.80	39.30	62.80
λ_D				6.70	11.00	26.90	44.40	66.80
λ_{LR}				6.00	10.50	25.40	41.40	65.00
λ_{LM}				5.35	9.30	23.90	39.30	63.20
λ_W	60	.5	.2	5.60	10.80	23.40	45.50	60.30
λ_D				6.20	11.50	25.10	47.50	62.40
λ_{LR}				5.80	11.10	24.50	46.70	61.80
λ_{LM}				5.50	10.90	23.70	45.60	60.60
π				5.00	9.18	22.31	43.39	66.68
λ_W	20	.5	.8	10.35	20.90	47.60	71.30	84.70
λ_D				12.35	23.70	51.60	71.80	83.30
λ_{LR}				10.90	22.00	47.20	69.20	81.40
λ_{LM}				9.10	18.60	43.10	66.90	78.20
λ_W	40	.5	.8	7.30	19.50	52.20	78.10	91.80
λ_D				8.50	21.90	54.50	78.60	91.70
λ_{LR}				7.35	20.20	52.60	77.00	91.20
λ_{LM}				6.55	19.00	50.20	75.70	90.10
λ_W	60	.5	.8	6.00	16.70	53.70	82.10	94.10
λ_D				6.90	18.60	55.10	82.30	93.70
λ_{LR}				6.50	17.80	53.50	81.70	93.20
λ_{LM}				6.05	16.70	52.70	80.80	92.70
π				5.00	19.62	59.53	91.02	99.27

TABLE III

SIMULATION RESULTS FOR DL(2,1) MODELS

Criteria	Parameters			γ				
	n	θ_1	ρ	0	.5	1	1.5	2.0
λ_W	20	−.5	−.8	1.85	2.00	1.30	1.40	.90
λ_D				12.55	14.80	20.60	24.90	27.50
λ_{LR}				10.30	13.00	18.00	22.00	25.70
λ_{LM}				8.50	11.80	15.50	18.90	22.90
λ_W	40	−.5	−.8	1.70	1.30	.90	.60	.80
λ_D				7.75	11.70	16.00	21.90	28.60
λ_{LR}				7.05	10.80	14.50	20.90	27.20
λ_{LM}				6.45	9.60	13.40	19.60	25.60
λ_W	60	−.5	−.8	1.60	1.80	.70	.80	.60
λ_D				6.65	11.20	15.60	22.20	29.30
λ_{LR}				6.45	10.50	14.40	21.10	28.10
λ_{LM}				6.05	10.20	13.60	21.10	27.40
π				5.00	9.88	25.18	48.97	73.19
λ_W	20	−.5	−.2	7.55	17.80	29.40	44.00	54.70
λ_D				12.05	20.70	31.50	44.50	55.80
λ_{LR}				10.25	18.30	28.80	41.50	51.60
λ_{LM}				8.25	15.30	25.10	38.10	48.60
λ_W	40	−.5	−.2	5.40	15.10	27.80	43.10	56.60
λ_D				6.75	16.40	29.60	45.20	58.00
λ_{LR}				6.30	15.20	28.00	43.60	56.40
λ_{LM}				5.60	13.90	25.30	41.50	54.40
λ_W	60	−.5	−.2	5.60	14.50	28.70	44.40	60.00
λ_D				6.70	16.30	29.80	46.40	61.20
λ_{LR}				6.20	15.40	28.90	45.20	60.10
λ_{LM}				6.10	14.40	27.90	44.20	58.90
π				5.00	12.11	34.10	64.23	87.27

TABLE III (Continued)

λ_W	20	−.5	.2	9.20	17.60	33.10	43.80	58.90
λ_D				10.10	17.50	32.60	44.10	60.10
λ_{LR}				8.05	15.50	29.30	39.80	56.60
λ_{LM}				6.85	13.30	27.30	35.30	51.80
λ_W	40	−.5	.2	6.55	15.90	30.60	49.00	65.10
λ_D				6.45	15.90	29.00	47.40	63.30
λ_{LR}				5.70	14.70	27.80	45.90	61.80
λ_{LM}				5.30	13.00	26.60	43.80	60.00
λ_W	60	−.5	.2	6.15	12.90	30.40	49.30	67.40
λ_D				5.75	12.60	25.50	48.10	65.30
λ_{LR}				5.30	11.90	28.50	47.30	64.40
λ_{LM}				4.90	11.40	27.00	46.10	63.40
π				5.00	10.69	28.48	54.99	79.40
λ_W	20	−.5	.8	9.00	15.00	28.70	38.70	46.00
λ_D				10.65	18.30	37.00	50.40	64.40
λ_{LR}				9.25	16.10	34.20	47.40	59.60
λ_{LM}				7.75	14.00	29.40	42.20	55.90
λ_W	40	−.5	.8	8.30	17.00	33.10	48.50	63.90
λ_D				8.35	19.60	37.50	56.90	71.50
λ_{LR}				7.60	18.50	35.80	55.10	69.70
λ_{LM}				6.65	17.30	34.00	53.00	68.20
λ_W	60	−.5	.8	6.55	15.00	34.10	52.10	72.30
λ_D				6.45	16.30	36.50	57.70	75.70
λ_{LR}				6.10	15.50	34.60	55.70	74.70
λ_{LM}				5.65	14.80	33.70	54.20	74.00
π				5.00	13.11	37.93	69.78	91.06

TABLE III (Continued)

λ_W	20	−.5	−.8	12.75	10.60	8.00	5.40	5.20
λ_D				9.95	11.10	11.80	15.00	19.10
λ_{LR}				8.90	9.30	10.30	13.30	17.30
λ_{LM}				7.40	7.50	8.80	11.00	15.20
λ_W	40	−.5	−.8	10.20	5.70	6.10	3.10	2.70
λ_D				7.05	8.80	9.50	12.70	14.10
λ_{LR}				6.75	7.60	8.90	11.60	13.30
λ_{LM}				5.90	7.10	7.90	10.30	12.40
λ_W	60	−.5	−.8	9.05	7.90	3.80	3.40	3.20
λ_D				6.60	8.00	9.50	10.70	15.00
λ_{LR}				6.10	7.40	9.00	10.40	14.30
λ_{LM}				5.65	7.40	8.30	9.80	13.70
π				5.00	5.04	6.62	8.69	11.64
λ_W	20	.5	−.2	5.70	2.50	2.10	1.80	1.70
λ_D				10.70	14.40	18.00	26.40	29.30
λ_{LR}				9.20	12.60	16.10	23.70	27.40
λ_{LM}				7.25	10.10	14.00	20.90	24.10
λ_W	40	.5	−.2	5.65	3.30	1.00	1.50	.70
λ_D				6.30	10.60	13.70	23.50	27.90
λ_{LR}				5.50	10.00	13.10	22.50	25.50
λ_{LM}				4.75	9.20	11.80	21.00	24.20
λ_W	60	.5	−.2	5.15	3.00	1.70	.50	1.10
λ_D				5.95	9.50	13.80	21.60	27.90
λ_{LR}				5.30	8.90	13.30	20.50	26.40
λ_{LM}				5.10	8.50	13.00	19.70	25.60
π				5.00	6.90	12.80	22.92	36.77

TABLE III (Continued)

λ_W	20	.5	.2	3.30	1.90	2.20	1.80	2.20
λ_D				10.85	18.80	27.50	36.20	46.10
λ_{LR}				9.35	16.40	24.90	33.20	43.60
λ_{LM}				7.60	14.70	21.90	29.50	39.60
λ_W	40	.5	.2	3.45	1.70	1.20	.70	2.10
λ_D				7.10	14.80	26.30	36.70	48.40
λ_{LR}				6.50	13.50	25.10	34.50	47.70
λ_{LM}				5.75	12.10	23.80	32.40	45.50
λ_W	60	.5	.2	3.15	1.10	1.20	1.80	4.20
λ_D				6.85	13.60	23.20	34.40	53.50
λ_{LR}				6.20	13.10	22.40	33.30	52.00
λ_{LM}				5.95	12.30	21.60	32.10	50.90
π				5.00	11.91	33.31	63.00	86.34
λ_W	20	.5	.8	4.90	29.50	36.40	40.60	44.70
λ_D				10.70	45.90	67.40	79.40	85.20
λ_{LR}				8.95	42.40	64.70	77.40	83.70
λ_{LM}				7.40	38.70	61.90	74.60	81.60
λ_W	40	.5	.8	5.05	35.60	56.00	73.50	81.70
λ_D				7.85	44.60	74.20	91.20	95.50
λ_{LR}				7.25	43.50	72.70	90.40	95.30
λ_{LM}				6.35	41.80	71.00	89.70	94.60
λ_W	60	.5	.8	4.60	43.20	75.20	87.80	93.60
λ_D				5.90	49.50	83.10	94.30	97.90
λ_{LR}				5.60	48.00	82.60	94.10	97.80
λ_{LM}				5.05	46.90	81.70	93.80	97.60
π				5.00	60.92	99.40	99.99	99.99

TABLE IV

SIMULATION RESULTS FOR DL(2,1) MODELS

Criteria	Parameters			γ				
	n	θ_1	ρ	0	.5	1	1.5	2.0
λ_W	20	.5	.2	3.30	1.90	2.20	1.80	2.20
λ_D				10.85	18.80	27.50	36.20	46.10
λ_{LR}				9.35	16.40	24.90	33.20	43.60
λ_{LM}				7.60	14.70	21.90	29.50	39.60
λ_W	40	.5	.2	3.45	1.70	1.20	.70	2.10
λ_D				7.10	14.80	26.30	36.70	48.40
λ_{LR}				6.50	13.50	25.10	34.50	47.70
λ_{LM}				5.75	12.10	23.80	32.40	45.50
λ_W	60	.5	.2	3.15	1.10	1.20	1.80	4.20
λ_D				6.85	13.60	23.20	34.40	53.50
λ_{LR}				6.20	13.10	22.40	33.30	52.00
λ_{LM}				5.95	12.30	21.60	32.10	50.90
λ_W	125	.5	.2	4.00	2.50	5.80	9.40	19.10
λ_D				6.40	11.80	25.70	36.60	55.30
λ_{LR}				6.30	11.30	25.30	36.00	54.90
λ_{LM}				6.20	10.90	24.90	35.50	54.30
λ_W	250	.5	.2	4.00	4.60	10.60	23.00	39.50
λ_D				5.20	13.70	27.70	45.00	62.50
λ_{LR}				5.20	13.60	27.30	44.20	62.10
λ_{LM}				5.10	13.40	27.20	43.60	61.70
λ_W	500	.5	.2	5.30	6.50	16.70	33.70	49.90
λ_D				6.40	11.70	25.40	46.90	63.80
λ_{LR}				6.30	11.60	25.30	46.30	63.70
λ_{LM}				6.10	11.40	25.20	46.20	63.70
λ_W	1000	.5	.2	5.40	7.60	20.40	42.00	61.10
λ_D				5.90	11.30	27.60	49.10	69.50
λ_{LR}				5.80	11.20	27.60	49.10	69.50
λ_{LM}				5.70	11.10	27.60	49.00	69.50
π				5.00	11.91	33.31	63.00	86.34

References

Amimya, Takeshi. Advanced Econometrics, Cambridge Massachusetts: Harvard University Press, 1985.

Berndt, Ernst R. and N.E. Savin. "Conflict Among the Criteria for Testing Hypotheses in the Multivariate Linear Regression Model." Econometrica, 45 (1977), 1263-1278.

Box, G.E.P. and M.E. Muller. "A Note on the Generation of Random Normal Deviates." Annals of Mathematical Statistics, 29 (1958), 610-611.

Breusch, T.S. "Conflict Among Criteria for Testing Hypotheses: Extensions and Comments." Econometrica, 47 (1979), 203-207.

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

Dhrymes, Phoebus J. Econometrics, New York: Springer Verlag, 1970.

Durbin, J. "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables." Econometrica, 38 (1970), 410-421.

Engel, Robert F. "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics." Handbook of Econometrics, Vol 2. Z. Griliches and M.D. Intriligator (eds.), New York: North Holland, 1984.

Evans, G.B.A. and N.E.Savin,. "Conflict Among the Criteria Revisited: The W, LR and LM Tests." Econometrica, 50 (1982), 737-748.

Gallant, Robert A. and A. Holly. "Statistical Inference in an Implicit, Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation." Econometrica, 48 (1980), 697-720.

Gregory, Allan W. and M.R. Veall. "Formulating Wald Tests of Nonlinear Restrictions." Econometrica, 53 (1985), 1465-1468.

Hendry, David F. "Stochastic Specification in an Aggregate Demand Model of the United Kingdom." Econometrica,42 (1974), 559-578.

Hendry, David F. and J. Richard. "On the Formulation of Empirical Models in Dynamic Econometrics." Journal of Econometrics, 20 (1982), 3-33.

Kenward, Loyd R. "Autocorrelation and Dynamic Methodology with an Application to Wage Determination Models." Journal of Econometrics, 3 (1975), 179-187.

Marsaglia, G. and T.A. Bray. "A Convenient Method for Generating Random Normal Variables." Society for Industrial and Applied Mathematics Review, 7 (1964), 4-10.

Mizon, G.E. "Inferential Procedures in Nonlinear Models: An Application in a U.K. Industrial Cross Section Study of Factor Substitution and Returns to Scale." Econometrica, 45 (1977), 1221-1242.

Mizon, G.E. "Model Selection Procedures." Studies in Modern Economic Analysis, M.J. Artis and A.R. Nobay (eds.), Oxford: Basil Blackwell, 1977.

Mizon, G.E. and D.F. Hendry. "An Empirical Application and Monte Carlo Analysis of Tests of Dynamic Specification." Review of Economic Studies, 47 (1980), 21-45.

Rao, C.R. "Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation." Proceedings of the Cambridge Philosophical Society, 44 (1947), 50-57.

Rothenberg, Thomas J. "Hypothesis Testing in Linear Models When the Error Covariance Matrix is Nonscalar." Econometrica, 52 (1984), 827-842.

Sargan, J.D. "Some Tests of Dynamic Specification for a Single Equation." Econometrica, 48 (1980), 879-897.

Sargan, J.D. and F. Mehta. "A Generalization of the Durbin Significance Test and Its Applications to Dynamic Specification." Econometrica, 51 (1983), 1551-1567.

Wald, A. "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large." Transactions of the American Mathematical Society, 54 (1943), 426-482.

Wilks, S.S. "The Large Sample Distribution of the LIkelihood Ratio for Testing Composite Hypotheses," Annals of Mathematical Statistics, 9 (1938), 60-62.

Appendix I

Kronecker Products and Matrix Vectorization

Kronecker Products

Let **A** be an mxn and let **B** be pxq, then the "Kronecker product" of **A** and **B**, denoted as $\mathbf{A}\oplus\mathbf{B}$, is defined as the mpxnq marix:

$$\mathbf{A}\approx\mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \cdot & & \cdot & \\ \cdot & & \cdot & \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$$

1 Theorem: $(\mathbf{A}\oplus\mathbf{B})\,(\mathbf{C}\oplus\mathbf{D}) = (\mathbf{AC}\oplus\mathbf{BD})$ if **A** is conformable with **C** and if **B** is conformable with **D**.

2 Theorem: $(\mathbf{A}\oplus\mathbf{B})' = \mathbf{A}'\oplus\mathbf{B}'$.

3 Theorem: If **A** and **B** are invertible marices of orders m and n, then $(\mathbf{A}\oplus\mathbf{B})^{-1} = \mathbf{A}^{-1}\oplus\mathbf{B}^{-1}$.

4 Theorem: Let **A** and **B** be square matrices of orders m and n. Let the eigenvalues of **A** be denoted by λ_i for i = 1,2,,,m, and denote the eigenvalues of **B** by μ_j for j = 1,2,,,n, then the eigenvalues of $\mathbf{A}\oplus\mathbf{B}$ are:

$\lambda_i\mu_j$; i = 1,2,,,m; j = 1,2,,,n

4.1 Corollary: $\det(\mathbf{A}\oplus\mathbf{B}) = \det(\mathbf{A})^n\det(\mathbf{B})^m$.

4.2 Corollary: $\mathrm{tr}(\mathbf{A}\oplus\mathbf{B}) = \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})$.

5 Theorem: Let **A** and **B** be square marices. Let $\mathbf{a}_i$ be an eigenvector of **A** corresponding to the eigenvalue λ_i and let $\mathbf{b}_j$ be an eigenvector of **B** with eigenvalue μ_j, then $\mathbf{a}_i\oplus\mathbf{b}_j$ is an eigenvector of $\mathbf{A}\oplus\mathbf{B}$ with corresponding eigenvalue $\lambda_i\mu_j$.

Matrix Vectorization

Let **A** be an mxn marix, and let the jth column of **A** be denoted as $\mathbf{a}_j$. The "vectorized" form of **A**, denoted as vec(**A**), is defined as:

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ . \\ . \\ \mathbf{a}_n \end{bmatrix}$$

6 Theorem: Let **A** be mxn, and let **B** be nxp, then:

$$\text{vec}(\mathbf{AB}) = (\mathbf{I}_p \oplus \mathbf{A})\text{vec}(\mathbf{B}) = (\mathbf{B}' \oplus \mathbf{I}_m)\text{vec}(\mathbf{A})$$

where $\mathbf{I}_p$ denotes the identity matrix of order p.

7 Theorem: Let **A** be conformable with **B**, and let **B** be conformable with **C**, then vec (**ABC**) = (**C**′⊕**A**)vec(**B**).

8 Theorem: Let **A** be nxm and let **B** be mxn, then:

$$\text{vec}(\mathbf{B}')'\text{vec}(\mathbf{A}) = \text{tr}(\mathbf{AB}) = \text{vec}(\mathbf{A}')'\text{vec}(\mathbf{B})$$

Appendix II

Matrix Triangularization

It can be shown that if **A** is a symmetric positive definite matrix, then there exists a triangular matrix **T** such that **A = T'T.** In this appendix, we show how to construct **T.**

Let $t_{ij} = 0$ for $i < j$. Let a_{ij} denote the (i,j) element of **A**, since **A = T'T**, then:

(1)

$$a_{ij} = \Sigma_{k=1}^{i} t_{ki} t_{kj}$$

Hence:

$$a_{ii} = \Sigma_{k=1}^{i} t_{ki}^2$$

which implies:

$$t_{ii} = (a_{ii} - \Sigma_{k<i} t_{ki}^2)^{1/2}$$

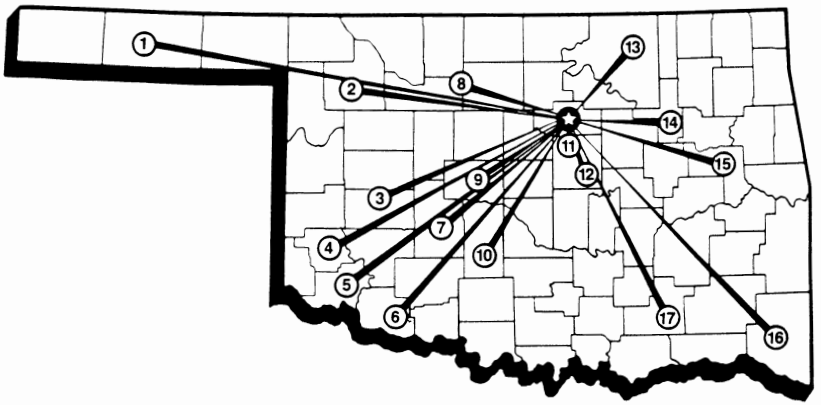
Moreover, (1) implies:

$$t_{ji} = (a_{ij} - \Sigma_{k<i} t_{ki} t_{kj}) / t_{ii}; \quad i \leq j$$

One can solve the last two equations recursively, beginning with (i,j) = (1,1), then proceeding row by row through the remainder of the marix.

Reports of Oklahoma Agricultural Experiment Station serve people of all ages, socio-economic levels, race, color, sex religion, and national origin. This publication is printed and issued by Oklahoma State University as authorized by the Dean of the Division of Agriculture and has been prepared and distributed at a cost of $625.10 for 394 copies. AI-9471 0790 TD

THE OKLAHOMA
AGRICULTURAL EXPERIMENT STATION
System Covers the State



✪ Main Station — *Stillwater and Lake Carl Blackwell*

1. Panhandle Research Station — *Goodwell*
2. Southern Great Plains Field Station — *Woodward*
3. Marvin Klemme Range Research Station — *Bessie*
4. Sandyland Research Station — *Mangum*
5. Irrigation Research Station — *Altus*
6. Southwest Agronomy Research Station — *Tipton*
7. Caddo Research Station — *Ft. Cobb*
8. North Central Research Station — *Lahoma*
9. Forage and Livestock Research Laboratory — *El Reno*
10. South Central Research Station — *Chickasha*
11. Agronomy Research Station — *Perkins*
    Fruit Research Station — *Perkins*
12. Pecan Research Station — *Sparks*
13. Pawhuska Research Station — *Pawhuska*
14. Vegetable Research Station — *Bixby*
15. Eastern Research Station — *Haskell*
16. Kiamichi Forestry Research Station — *Idabel*
17. Wes Watkins Agricultural Research and Extension Center — *Lane*

OKLAHOMA STATE UNIVERSITY
CENTENNIAL
1890 • 1990