MONASH

WP 6/95

ISSN 1032-3813
ISBN 0 7326 0767 1

MONASH UNIVERSITY

ANCORA IMPARO

AUSTRALIA

THE APPLICATION OF THE DURBIN-WATSON TEST TO THE DYNAMIC REGRESSION MODEL UNDER NORMAL AND NON-NORMAL ERRORS

Maxwell L. King and David C. Harris

Working Paper 6/95
August 1995

DEPARTMENT OF ECONOMETRICS

THE APPLICATION OF THE DURBIN-WATSON TEST TO THE DYNAMIC REGRESSION MODEL UNDER NORMAL AND NON-NORMAL ERRORS

By Maxwell L. King and David C. Harris

Department of Econometrics, Monash University
Clayton, Victoria 3168, Australia

Key Words: Durbin's tests; Durbin-Watson test; modified point optimal test; Monte Carlo methods; small-disturbance asymptotics.

JEL classification nos: C12, C32.

ABSTRACT

Until recently, a difficulty with applying the Durbin-Watson (DW) test to the dynamic linear regression model has been the lack of appropriate critical values. Inder (1986) used a modified small-disturbance distribution (SDD) to find approximate critical values. King and Wu (1991) showed that the exact SDD of the DW statistic is equivalent to the distribution of the DW statistic from the regression with the lagged dependent variables replaced by their means. Unfortunately, these means are unknown although they could be estimated by the actual variable values. This provides a justification for using the exact critical values of the DW statistic from the regression with the lagged dependent variables treated as non-stochastic regressors. Extensive Monte Carlo experiments are reported in this paper. They show that this approach leads to reasonably accurate critical values, particularly when two lags of the dependent variable are present. Robustness to non-normality is also investigated.

1. Introduction

The dynamic linear regression model plays an important role in econometric modelling. The popular ordinary least squares (OLS) estimator is inconsistent if the model's disturbances are autocorrelated. It is therefore important to be able to test for autocorrelation in the disturbances of the dynamic linear regression model. Much has been written on this testing problem. The literature up until 1987 is reviewed by King (1987) and recent contributions include Dezhbakhsh (1990), Inder (1990) and King and Wu (1991).

Whether the Durbin-Watson (DW) statistic should be used to test for autocorrelation in this context has been an unusually controversial issue. As Durbin (1970) observed, the difficulty is in finding appropriate critical values. He suggested an adjustment to the DW statistic, which results, asymptotically, in a standard normal null distribution and which has become known as Durbin's h test. The adjustment involves the square root of a variance estimate and consequently breaks down when this estimate is negative. Durbin also proposed an alternative artificial regression test known as Durbin's t test. Based on Monte Carlo studies of the size and power properties of the DW, h and t tests, Kenkel (1974, 1975, 1976) recommended the use of the DW upper bound as a critical value for the DW statistic. This suggestion was reviewed and rejected by Park (1976) who conducted his own Monte Carlo comparison (Park, 1975) of the three tests. More recently, Dezhbakhsh (1990) warned against the use of the DW test in the dynamic model with the dependent variable lagged once and twice as regressors. His Monte Carlo results seem to indicate that the DW test can lack power in this situation.

In contrast, Inder (1984, 1985) used Monte Carlo methods to show that for a single lag and if appropriate critical values can be found, then the DW test is typically more powerful than Durbin's h and t tests. He (Inder, 1985, 1986) suggested the use of small-disturbance asymptotics to find appropriate DW critical values. This approach has considerable appeal because in the static model, the small-disturbance distribution (SDD) of the DW statistic is identical to its true small-sample distribution because of the statistic's invariance to the disturbance variance. Inder showed that an approximate

SDD critical value for the DW statistic is the true critical value of the statistic for the corresponding regression with the lagged dependent variables omitted. Using Monte Carlo methods, he found that these critical values generally yield sizes closer to the nominal size than do Durbin's h and t tests.

More recently, King and Wu (1991) observed that the true SDD of the DW statistic is identical to the exact distribution of the DW statistic for the corresponding regression with the lagged dependent variables replaced by their expected values. This provides a justification for the use of the familiar tables of bounds when the DW test is applied to a dynamic regression model. It also mirrors an identical result reported by Nankervis and Savin (1987) for testing linear coefficient restrictions in the dynamic regression model. They found that the SDD of the F statistic is identical to the true distribution of the F statistic from the regression with the lagged dependent variables replaced by their expected values.

A difficulty with the King and Wu finding is that the expected values of the lagged dependent variables are functions of the unknown regression coefficients. They discussed how bounds for the SDD critical value could be calculated. An alternative approach would be to estimate the expected values of the variables. One possibility is to use the lagged dependent variables as estimates of their means. This then involves calculating exact DW critical values with the lagged dependent variables treated as though they are nonstochastic. The aim of this paper is to investigate this suggestion.

The plan of the paper is as follows. The next section outlines the models and the class of tests our suggestion can be applied to. It also observes the true small-disturbance nature of the new procedure. Section 3 outlines an extensive Monte Carlo experiment designed to compare the new procedure with four existing tests in a range of circumstances including normal and nonnormal error processes. Some concluding remarks are made in the final section.

2. Theory

Consider the general linear dynamic regression model

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + x_t'\beta + u_t, \qquad (1)$$

$t = 1, \ldots, n$, where y_t is the dependent variable, x_t is a $k \times 1$ vector of exogenous variables, $\alpha = (\alpha_1, \ldots, \alpha_p)'$ and β are $p \times 1$ and $k \times 1$ parameter vectors, respectively, and u_t is a disturbance term. If there are n observations available on each variable, the parameters are estimated using the last $n - p$ observations. The model for these observations can be written as

$$y = Y_{-1}\alpha + X\beta + u, \tag{2}$$

where y and u are $(n-p) \times 1$ vectors and Y_{-1} and X are $(n-p) \times p$ and $(n-p) \times k$ matrices, respectively.

Suppose $u \sim N(0, \sigma^2\Omega(\theta))$, where σ^2 is an unknown scalar, $\Omega(\theta)$ is a positive definite symmetric matrix such that $\Omega(0) = I_{n-p}$ and θ is a $q \times 1$ unknown parameter vector. Our interest is in testing $H_0 : \theta = 0$. As observed by King and Wu (1991), this parameterization covers a number of important testing problems such as testing for, either separately or jointly, various forms of autocorrelation, heteroscedasticity and stochastic coefficients on the exogenous variables. In the context of the static model,

$$y = X\beta + u, \tag{3}$$

a locally most mean powerful invariant (LMMPI) test of H_0 against H_a : $\theta_i \geq 0$, $i = 1, \ldots, q, \theta \neq 0$, is to reject H_0 for small values of

$$s = z'Az/z'z \tag{4}$$

where z is the OLS residual vector from (3) and

$$A = -\sum_{i=1}^{q} \partial\Omega(\theta)/\partial\theta_i \bigg|_{\theta=0}$$

(see King and Wu, 1990). When $q = 1$, this reduces to King and Hillier's (1985) locally best invariant (LBI) test. Also (4) with A as the tridiagonal matrix whose main diagonal is $(1, 2, 2, \ldots, 2, 1)$ and whose leading off-diagonal elements are -1, is the DW statistic.

With respect to testing $H_0 : \theta = 0$ in (2), as King and Wu (1991) note, an obvious approach is to use s as the test statistic where z is now the OLS residual vector from (2). They showed that the SDD of s is equivalent to the exact distribution of s applied to the regression

$$y = M_{-1}\gamma + X\delta + u, \tag{5}$$

where $M_{-1} = E(Y_{-1})$. The exact form of M_{-1} depends on how the process (2) starts up. King and Wu (1991) allowed for two different sets of starting-up assumptions. Their proof also applies for any starting-up conditions that imply

$$Y_{-1} = M_{-1} + 0(\sigma). \tag{6}$$

An obvious feature of (5) is that not all the regressors are known. The elements of M_{-1} are functions of the unknown parameters α and β through the recursive formula

$$m_t = \alpha_1 m_{t-1} + \alpha_2 m_{t-2} + \ldots + \alpha_p m_{t-p} + x_t'\beta, \quad t = p+1, \ldots, n, \tag{7}$$

where $m_t = E(y_t)$ and the start-up conditions of (2) determine the values of m_i, $i = 1, \ldots, p$, required to start up (7) (see King and Wu, 1991). We also see that the SDD of s may depend on the nuisance parameters α and β through M_{-1}. Thus, typically, a test based on (4) applied to (2) can be expected to be nonsimilar. If we knew M_{-1}, this potential nonsimilarity can be removed by the calculation of critical values conditional on M_{-1}, at least in the neighbourhood of $\sigma = 0$. Note that this nonsimilarity problem is not solved by Inder's suggestion of ignoring M_{-1} or King and Wu's suggestion of calculating bounds for the SDD critical values of s to take account of the unknown regressors that make up M_{-1}.

Our proposal is to use Y_{-1} as an estimate of M_{-1} and thus calculate exact critical values of s conditional on the observed Y_{-1} values. We are replacing the expected value of a random matrix by its observed value; a simplification that is often used in econometrics, particularly in estimating information matrices. There is also a small-disturbance justification of this approximation which can be argued as follows.

For any given realization of u and hence Y_{-1}, the distribution of s treating Y_{-1} as nonstochastic but u as $N(0, \sigma^2 I_n)$ is well defined. However, this distribution changes with each realization of u because Y_{-1} changes. The exact proof used by King and Wu (1991, pp. 148-149) can be used to prove that each one of these distributions converges to the distribution of s applied to (5) as σ tends to zero, *ceteris paribus.* The fact that Y_{-1} is now treated as non-stochastic makes no difference. The key relationship is (6). In other words, the distribution we suggest should be used to obtain critical values or p-values converges to the SDD of s applied to (2) as σ tends to zero. Thus our approach has a small-disturbance justification. Furthermore, it results in an approximately similar test for σ close to zero.

In summary, therefore, our suggested procedure involves ignoring the fact that Y_{-1} in (2) is stochastic and calculating either an exact critical value or a p-value as we would if Y_{-1} were a non-stochastic matrix of regressors. Of course, there is the obvious question of how well this procedure works in practice, particularly when σ is relatively large. The remainder of this paper addresses this issue.

3. The Monte Carlo Experiment

A Monte Carlo experiment was conducted to compare the accuracy of the above testing procedure in the context of testing

$$H_0 : \rho = 0$$

against

$$H_a : \rho > 0$$

in (2) when the disturbances follow the stationary first-order autocorrelation process,

$$u_t = \rho u_{t-1} + \epsilon_t, \qquad |\rho| < 1 \tag{8}$$

where the innovations, ϵ_t, are mutually independent with mean zero and variance σ^2. The experiment involved both one lag ($p = 1$) and two lags ($p = 2$) of the dependent variable in (2). Both normal and non-normal innovations of (8) were used.

3.1 The Tests

Two versions of the DW test were used in the study. The first, denoted DWE, involved the procedure outlined in section 2. Observe that the critical value used in this procedure is a function of the realized values of y_{t-1} and hence of y_t. It therefore has to be recalculated for each new sample; in other words a separate critical value is needed for each iteration. An equivalent method of applying the test is to calculate, using Imhof's (1961) algorithm, an exact p-value for the DW statistic while treating the observed y_{t-1} as nonstochastic. In fact at each iteration, the calculated p-value can be viewed as the test statistic and the desired significance level as the critical value. In the study, we implemented the test in this manner by calculating the exact p-value at each iteration using a modified version of Koerts and Abrahamse's (1969) FORTRAN version of Imhof's algorithm.

The second version of the DW test, denoted DWI, is the application of Inder's (1985, 1986) suggestion of using the exact DW critical value from the regression with the lagged dependent variables omitted. It may appear on the surface that the DWE and DWI tests are the same test with different critical values and therefore different sizes. This is not the case because the critical values of the DWE test are calculated as a function of the y_t values while those of the DWI test are not and therefore remain constant from iteration to iteration. This implies that the two tests do not share the same set of critical regions. The DWI test is expected to be more nonsimilar than the DWE test.

The third test is Inder's (1990) modified point optimal test denoted IMPO. This involves first estimating (2) using OLS in order to obtain $\hat{\alpha}_i$, $i = 1, \ldots, p$, the OLS estimates of the coefficients of the lagged dependent variables. Then King's (1985) point optimal invariant (POI) test is applied to the static regression

$$y^* = X\beta + u \tag{9}$$

where

$$y_t^* = y_t - \hat{\alpha}_1 y_{t-1} - \ldots - \hat{\alpha}_p y_{t-p}, \qquad t = p+1, \ldots, \text{n}.$$

Define

$$\begin{aligned}
\tilde{y}_{p+1} &= (1-\rho_1^2)^{1/2} y_{p+1}^*, & \\
\tilde{y}_t &= y_t^* - \rho_1 y_{t-1}^*, & t &= p+2, \ldots, n, \\
\tilde{x}_{p+1} &= (1-\rho_1^2)^{1/2} x_{p+1}, & \\
\tilde{x}_t &= x_t - \rho_1 x_{t-1}, & t &= p+2, \ldots, \text{n}.
\end{aligned}$$

Then the test statistic is

$$\tilde{z}(\rho_1)'\tilde{z}(\rho_1)/\hat{z}'\hat{z},$$

where $\tilde{z}(\rho_1)$ is the OLS residual vector from the regression of $\tilde{y}_t$ on $\tilde{x}_t, t = p+1, \ldots, n, \hat{z}$ is the OLS residual vector from (9), and ρ_1 is a constant such that $0 < \rho_1 < 1$. The test is made operational by a choice of ρ_1 value. We followed Inder's suggestion of $\rho_1 = 0.5$. Inder also showed that an approximate SDD critical value for this test is the exact critical value of King's (1985) POI test applied to (9). These critical values, calculated using the methodology in King (1985), were used in our experiment.

The fourth test is Durbin's (1970) t test which is conducted as a test of the significance of z_{t-1} in the OLS regression of z_t on $z_{t-1}, y_{t-1}, \ldots, y_{t-p}, x_t$, $t = p+2, \ldots, n$, where z_t are the OLS residuals from (2). The final test is Durbin's (1970) h test. Its test statistic is

$$h = (1 - d/2)[(n-p)/\{1-(n-p)\hat{V}(\hat{\alpha}_1)\}]^{1/2}$$

where d is the DW statistic calculated for (2), and $\hat{V}(\hat{\alpha}_1)$ is the usual estimate of the variance of the OLS estimator of α_1 from (2). Asymptotically under H_0, h has a $N(0,1)$ distribution. Unfortunately, in small samples h sometimes cannot be calculated because $\{1-(n-p)\hat{V}(\hat{\alpha}_1)\}$ is negative. In this study, we only use those replications where h can be calculated to estimate the probability of the test rejecting H_0.

3.2 Experimental Design

The Monte Carlo experiment involved applying the five tests to data generated according to (2) and (8) with $p = 1$ and $p = 2$. In order to generate start-up values for (2), we followed Inder (1985, 1986, 1990) and assumed that $y_1, \ldots, y_p$ have constant mean equal to $E(y_{p+1})$ and that deviations from this mean follow the stationary $AR(p)$ process

$$v_t = \alpha_1 v_{t-1} + \alpha_2 v_{t-2} + \ldots + \alpha_p v_{t-p} + u_t$$

in which u_t is generated as (8).

The following data sets and parameter values were used:

$X1$: The exogenous regressors are those from Durbin and Watson's (1951) annual U.K. consumption of spirits example including the constant dummy. $k = 3; \sigma = 0.1, 500$ for $n = 30$ and $\sigma = 0.25, 2000$ for $n = 69$.

$X2$: The exogenous regressors are a constant dummy, quarterly Australian private capital movements commencing 1968(1), Australian Government capital movements and these latter two variables lagged one quarter. $k = 5; \sigma = 500, 1000$ for $n = 30$ and $\sigma = 1500, 5000$ for $n = 60$.

$X3$: The exogenous regressors are a constant dummy, real US GNP commencing 1947(1) (see Maddala and Rao, 1973). $k = 2; \sigma = 50, 500$ for both $n = 30$ and $n = 70$.

$X4$: The exogenous regressors are the eigenvectors corresponding to the five smallest eigenvalues of the DW A matrix. Note that the constant dummy is the eigenvector corresponding to the smallest (zero) root. $k = 5; \sigma = 2, 50$ for both $n = 30$ and $n = 70$.

$X5$: The exogenous regressors are the eigenvectors corresponding to the zero and four largest eigenvalues of the DW A matrix. $k = 5; \sigma = 0.5, 500$ for $n = 30$ and $\sigma = 0.5, 5$ for $n = 70$.

$X6$: The exogenous regressors are a constant dummy and four independent $AR(1)$ regressors generated artificially via

$$x_{it} = 0.95 x_{it-1} + \eta_{it}$$

where

$$\eta_{it} \sim N(0, 1), \qquad t = 0, 1, \ldots, n$$

and $x_{i0} \sim N(0, 1.333)$. $k = 5; \sigma = 2.5, 10$ for both $n = 30$ and $n = 70$.

All β values were set to one and when $p = 1, \alpha$ was set to 0.25, 0.5, 0.75, 0.99. We also set $\alpha = 1.0$ in the case of $p = 1$ with y_0 having a starting value of zero. For $p = 2$, the α values used were $(\alpha_1, \alpha_2) = (0.25, 0.25), (0.5, 0.25)$, $(0.9, 0.09), (1.25, -0.5)$. The σ values were chosen so that the range of average R^2 values span 0.5 and 0.9.

The above data sets were chosen because they exhibit a range of behaviour. $X1$ and $X3$ have been used in a number of earlier studies. $X2$ was chosen because its regressors exhibit large fluctuations and a strong degree of seasonality. The regressors of $X4$ and $X5$ correspond to the upper and lower bounding distributions of the DW statistic in the static model. They were included in the hope that they would show extremes in behaviour of the various tests. All tests were conducted at nominal significance levels of 0.10, 0.05, 0.025 and 0.01. Sizes and powers were calculated at $\rho = 0, 0.3, 0.6, 0.9$.

In all cases, the random innovations, ϵ_t, of (8) were generated as $N(0, \sigma^2)$ variates by applying the Box-Muller transformation to [0,1] uniform random variables as outlined by King and Giles (1984). 2000 replications were used throughout.

For $p = 1$, the experiments were repeated with non-normal pseudo-random innovations generated by the algorithm proposed by Ramberg and Schmeiser (1972, 1974). This algorithm is

$$r(p) = \theta_1 + (p^{\theta_3} - (1-p)^{\theta_4})/\theta_2, \qquad 0 \leq p \leq 1,$$

where $r(p)$ is the generated pseudo-random innovation, p is a uniform pseudo-random variate, θ_1 is a location parameter, θ_2 is a scale parameter and θ_3 and θ_4 are shape parameters. Tables of θ values that allow for a wide variety of distributions are provided by Ramberg, Tadikamalla, Dudewicz and Mykytka (1979). Following Lee (1992) and Brooks and King (1994), the non-normal distributions we used are:

(i) the distribution with zero skewness and kurtosis of six, $\theta = (0, -0.1686, -0.0802, -0.0802)'$,

(ii) the distribution with zero skewness and kurtosis of nine, $\theta = (0, -0.3203, -0.1359, -0.1359)'$ and

(iii) the distribution with skewness of one and kurtosis of six, $\theta = (-0.379, -0.0562, -0.0187, -0.0388)'$.

3.3 Results

Because the complete experiment involved the calculation of 9600 test sizes and 28,800 test powers in the case of stationary α values and a further 1440 sizes and 4320 powers in the case of $\alpha = 1, p = 1$, we have attempted to focus on the main patterns, placing particular emphasis on the size results. We began by calculating the absolute differences between the estimated sizes and the nominal sizes. Average values of these absolute differences, the maximum absolute difference plus a decomposition of this average into deviations above and below the nominal level for stationary α values are presented in Tables I-III. The average differences above (below) the nominal level were calculated by setting all deviations below (above) the nominal level to zero and then recomputing the average. In this way, the average deviations above and below the nominal level add up (allowing for rounding errors) to the average absolute differences from the nominal level. Each table contains the average values for each data set as well as overall values for small (30) and moderate $(60-70)$ sample sizes. Because the results for each of the non-normal error distributions are very similar, only those for symmetric disturbances with a kurtosis of 9 are presented. The greatest variability of sizes when the error distributions vary, occur for the h test and for design matrix $X6$. An important issue is whether the accuracy of the DWE procedure decreases as σ increases. Table IV presents the overall averages decomposed by small and large σ values. Results for the nonstationary case of $p = 1$ and $\alpha = 1$ are presented in Table V. Ideally, all numbers in Tables I to V should be zero (no difference from the nominal level) so the closer they are to zero the better.

We shall first consider the size results for one lag $(p = 1), |\alpha| < 1$ and normal disturbances given in Table I. An obvious feature is that the average accuracy of each of the three SDD based tests (DWE, DWI and IMPO) is almost always better than that of either of the large sample asymptotic tests (Durbin's h and t tests). The DWI and IMPO tests have very similar average accuracies with the DWE test not far behind. Durbin's h test is always the least accurate test. For $X4$ and $X5$, the two data sets chosen to show extreme results, the gap in accuracy between the two classes is greatest, particularly for $X5$. Typically, the average accuracy of all tests improves as the sample size increases, although $X2$ provides a clear exception. As might be expected, the improvement in accuracy is greatest for the large sample asymptotic tests. We also see from Table IV that average accuracies of the five tests decrease as the variance σ^2 increases. A somewhat unexpected finding is that when σ^2 increases, the actual increase in average absolute differences of estimated and nominal sizes is greatest for Durbin's h and t tests.

TABLE I

Average absolute differences of estimated and nominal sizes, their decomposition to contributions of estimated sizes above and below the nominal size and maximum absolute differences for $p = 1, |\alpha| < 1$ and normal disturbances.

	$n = 30$						$n = 60 - 70$				
	DWE	DWI	IMPO	t	h		DWE	DWI	IMPO	t	h
						$X1$					
Average	.036	.027	.028	.046	.111		.034	.030	.030	.051	.105
Above	.030	.017	.016	.042	.107		.020	.015	.015	.047	.102
Below	.006	.010	.012	.004	.004		.014	.015	.015	.004	.003
Maximum	.200	.134	.134	.146	.257		.148	.108	.107	.131	.254
						$X2$					
Average	.016	.015	.015	.018	.059		.025	.020	.020	.027	.062
Above	.008	.002	.002	.015	.057		.014	.008	.007	.026	.062
Below	.008	.013	.013	.003	.002		.011	.012	.013	.001	.000
Maximum	.051	.066	.067	.056	.177		.083	.084	.088	.088	.185
						$X3$					
Average	.032	.026	.028	.062	.135		.030	.027	.028	.035	.084
Above	.021	.011	.012	.059	.132		.014	.010	.010	.032	.082
Below	.011	.015	.016	.003	.003		.016	.017	.018	.003	.002
Maximum	.154	.094	.100	.160	.275		.111	.096	.098	.090	.227
						$X4$					
Average	.178	.136	.137	.194	.196		.110	.092	.092	.138	.208
Above	.178	.135	.136	.194	.196		.101	.082	.081	.138	.208
Below	.000	.001	.001	.000	.000		.009	.010	.010	.000	.000
Maximum	.544	.464	.469	.404	.358		.429	.384	.380	.283	.331
						$X5$					
Average	.026	.029	.030	.236	.308		.031	.030	.030	.117	.182
Above	.009	.006	.007	.236	.308		.012	.008	.008	.117	.182
Below	.017	.023	.023	.000	.000		.019	.022	.022	.000	.000
Maximum	.095	.096	.098	.411	.533		.098	.099	.100	.222	.360
						$X6$					
Average	.033	.020	.019	.026	.052		.020	.015	.014	.028	.051
Above	.032	.017	.015	.011	.039		.014	.007	.006	.014	.039
Below	.001	.003	.004	.015	.012		.006	.008	.008	.014	.012
Maximum	.168	.119	.110	.084	.147		.142	.102	.096	.070	.180
						All X					
Average	.054	.042	.043	.097	.143		.042	.036	.036	.066	.116
Above	.046	.031	.031	.093	.140		.029	.022	.021	.062	.113
Below	.008	.011	.012	.004	.003		.012	.014	.014	.004	.003

TABLE II

Average absolute differences of estimated and nominal sizes, their decomposition to contributions of estimated sizes above and below the nominal size and maximum absolute differences for $p = 2$ and normal disturbances.

	$n = 30$						$n = 60 - 70$				
	DWE	DWI	IMPO	t	h		DWE	DWI	IMPO	t	h
						$X1$					
Average	.024	.031	.032	.024	.057		.037	.040	.041	.022	.073
Above	.006	.004	.005	.007	.055		.006	.006	.006	.010	.072
Below	.018	.027	.027	.017	.002		.031	.034	.035	.012	.001
Maximum	.060	.072	.077	.074	.158		.094	.097	.098	.070	.232
						$X2$					
Average	.023	.033	.034	.014	.071		.034	.038	.039	.017	.062
Above	.004	.005	.005	.007	.071		.005	.006	.006	.006	.062
Below	.019	.028	.029	.007	.000		.029	.032	.033	.011	.000
Maximum	.072	.081	.085	.047	.166		.094	.096	.096	.062	.208
						$X3$					
Average	.028	.036	.037	.025	.079		.038	.041	.042	.015	.078
Above	.005	.005	.006	.013	.078		.006	.006	.006	.009	.078
Below	.023	.031	.031	.012	.001		.032	.035	.036	.006	.000
Maximum	.076	.085	.089	.060	.217		.096	.096	.100	.059	.198
						$X4$					
Average	.045	.029	.028	.028	.036		.028	.030	.032	.056	.134
Above	.041	.017	.015	.013	.026		.008	.006	.006	.045	.129
Below	.004	.012	.013	.014	.009		.020	.024	.026	.011	.005
Maximum	.260	.152	.147	.089	.090		.068	.073	.090	.164	.284
						$X5$					
Average	.036	.039	.040	.143	.233		.043	.044	.045	.033	.105
Above	.005	.005	.006	.143	.233		.006	.006	.006	.032	.105
Below	.031	.034	.034	.000	.000		.037	.038	.038	.001	.000
Maximum	.098	.098	.098	.416	.450		.100	.100	.100	.108	.216
						$X6$					
Average	.014	.024	.024	.036	.028		.025	.031	.032	.024	.029
Above	.004	.003	.003	.005	.011		.004	.004	.005	.003	.020
Below	.010	.021	.021	.031	.017		.021	.027	.027	.021	.009
Maximum	.034	.054	.058	.087	.080		.072	.080	.081	.066	.111
						All X					
Average	.028	.032	.033	.045	.084		.034	.038	.038	.028	.080
Above	.011	.007	.007	.032	.079		.006	.006	.006	.018	.078
Below	.017	.025	.026	.013	.005		.028	.032	.032	.010	.002

TABLE III

Average absolute differences of estimated and nominal sizes, their decomposition to contributions of estimated sizes above and below the nominal size and maximum absolute differences for $p = 1, |\alpha| < 1$ and symmetric disturbances with kurtosis = 9.

	$n = 30$						$n = 60 - 70$				
	DWE	DWI	IMPO	t	h		DWE	DWI	IMPO	t	h
						$X1$					
Average	.035	.024	.025	.043	.102		.031	.028	.029	.037	.087
Above	.031	.017	.016	.038	.098		.018	.012	.012	.033	.084
Below	.004	.007	.009	.005	.004		.013	.016	.017	.004	.003
Maximum	.187	.128	.125	.124	.213		.140	.105	.104	.103	.219
						$X2$					
Average	.016	.017	.021	.025	.071		.023	.019	.020	.026	.066
Above	.009	.003	.003	.021	.069		.013	.008	.007	.024	.066
Below	.007	.014	.017	.004	.002		.010	.011	.013	.002	.000
Maximum	.074	.066	.079	.075	.193		.089	.078	.086	.088	.214
						$X3$					
Average	.030	.022	.025	.058	.122		.027	.026	.027	.029	.075
Above	.022	.011	.010	.055	.119		.012	.008	.008	.026	.073
Below	.009	.011	.014	.003	.003		.015	.018	.019	.003	.002
Maximum	.148	.086	.082	.138	.240		.092	.092	.096	.076	.201
						$X4$					
Average	.167	.131	.130	.176	.180		.098	.080	.079	.123	.194
Above	.167	.131	.130	.176	.180		.092	.072	.069	.123	.194
Below	.000	.000	.000	.000	.000		.006	.008	.010	.000	.000
Maximum	.530	.468	.470	.381	.346		.405	.351	.350	.254	.310
						$X5$					
Average	.024	.028	.030	.216	.278		.032	.030	.032	.109	.173
Above	.008	.005	.005	.216	.278		.012	.008	.008	.109	.173
Below	.016	.023	.025	.000	.000		.020	.022	.024	.000	.000
Maximum	.092	.094	.098	.388	.511		.098	.098	.099	.198	.328
						$X6$					
Average	.047	.028	.025	.019	.067		.023	.019	.021	.030	.076
Above	.046	.026	.023	.016	.067		.015	.008	.007	.022	.069
Below	.001	.002	.002	.003	.000		.008	.011	.014	.008	.007
Maximum	.144	.095	.085	.038	.113		.102	.076	.068	.068	.165
						All X					
Average	.053	.042	.042	.089	.137		.039	.034	.034	.059	.112
Above	.047	.032	.031	.087	.135		.027	.019	.018	.056	.110
Below	.006	.010	.011	.002	.002		.012	.015	.016	.003	.002

TABLE IV

Average absolute differences of estimated and nominal sizes and their decomposition to contributions of estimated sizes above and below the nominal size classified by error variance and sample size.

	$n = 30$					$n = 60 - 70$				
	DWE	DWI	IMPO	t	h	DWE	DWI	IMPO	t	h
					$p = 1$, Normal Errors, $\lvert\alpha\rvert < 1$					
					Small σ					
Average	.043	.036	.036	.076	.106	.033	.029	.029	.054	.091
Above	.036	.024	.024	.068	.099	.022	.016	.016	.048	.086
Below	.007	.012	.012	.008	.007	.011	.013	.013	.006	.005
					Large σ					
Average	.065	.049	.049	.118	.181	.050	.042	.042	.078	.140
Above	.057	.039	.038	.118	.181	.036	.027	.027	.076	.139
Below	.008	.010	.011	.000	.000	.014	.015	.015	.001	.001
					$p = 2$, Normal Errors					
					Small σ					
Average	.028	.032	.033	.053	.062	.034	.037	.038	.027	.055
Above	.010	.006	.006	.036	.054	.006	.006	.006	.014	.051
Below	.018	.026	.027	.017	.008	.028	.031	.032	.013	.004
					Large σ					
Average	.028	.032	.032	.037	.106	.035	.038	.039	.028	.106
Above	.012	.007	.007	.027	.104	.006	.006	.006	.021	.105
Below	.016	.025	.025	.010	.002	.029	.032	.033	.007	.001
					$p = 1$, Symmetric Errors, kurtosis $= 9$, $\lvert\alpha\rvert < 1$					
					Small σ					
Average	.044	.037	.038	.066	.102	.034	.030	.031	.051	.096
Above	.039	.027	.026	.062	.100	.022	.015	.015	.046	.092
Below	.005	.010	.012	.004	.002	.012	.015	.016	.005	.003
					Large σ					
Average	.062	.046	.047	.112	.171	.045	.037	.038	.067	.128
Above	.055	.037	.036	.112	.171	.032	.023	.022	.066	.127
Below	.007	.009	.011	.000	.000	.013	.014	.016	.001	.001

TABLE V

Average absolute differences of estimated and nominal sizes, their decomposition to contributions of estimated sizes above and below the nominal size and maximum absolute differences for $p = 1, \alpha = 1$ and normal disturbances.

	$n = 30$						$n = 60 - 70$				
	DWE	DWI	IMPO	t	h		DWE	DWI	IMPO	t	h
						$X1$					
Average	.051	.039	.040	.048	.073		.038	.031	.032	.036	.042
Above	.049	.031	.031	.030	.056		.038	.027	.028	.021	.029
Below	.002	.008	.009	.018	.017		.000	.004	.004	.015	.013
Maximum	.198	.131	.136	.092	.164		.143	.106	.108	.066	.098
						$X2$					
Average	.005	.016	.014	.021	.020		.020	.010	.010	.012	.014
Above	.005	.000	.000	.000	.000		.020	.009	.008	.004	.008
Below	.000	.016	.014	.021	.020		.000	.001	.002	.008	.006
Maximum	.019	.034	.029	.051	.044		.061	.034	.030	.036	.028
						$X3$					
Average	.014	.010	.010	.022	.016		.014	.006	.007	.014	.009
Above	.013	.003	.003	.000	.001		.014	.004	.005	.000	.001
Below	.001	.007	.007	.022	.015		.000	.002	.002	.014	.008
Maximum	.052	.029	.029	.069	.060		.043	.020	.020	.047	.036
						$X4$					
Average	.215	.159	.158	.208	.159		.143	.114	.113	.094	.116
Above	.215	.159	.158	.208	.159		.143	.114	.113	.089	.110
Below	.000	.000	.000	.000	.000		.000	.000	.000	.005	.006
Maximum	.546	.460	.464	.391	.354		.427	.376	.372	.278	.327
						$X5$					
Average	.018	.013	.014	.153	.179		.015	.008	.007	.058	.073
Above	.018	.005	.006	.153	.179		.015	.006	.005	.058	.073
Below	.000	.008	.008	.000	.000		.000	.002	.002	.000	.000
Maximum	.074	.032	.035	.131	.154		.041	.016	.016	.131	.154
						$X6$					
Average	.018	.013	.016	.032	.025		.006	.006	.006	.032	.030
Above	.018	.010	.010	.000	.000		.006	.000	.000	.000	.000
Below	.000	.003	.006	.032	.025		.000	.006	.006	.032	.030
Maximum	.064	.039	.042	.082	.082		.011	.020	.015	.068	.060
						All X					
Average	.054	.042	.042	.081	.078		.039	.030	.029	.041	.047
Above	.053	.035	.035	.065	.066		.039	.027	.026	.029	.037
Below	.001	.007	.007	.016	.012		.000	.003	.003	.012	.010

When critical values are required to control the size of a nonsimilar test, the conventional approach is to control the maximum probability of a Type I error. Therefore, many would only regard true sizes above the nominal level as bad. If we focus only on that part of the average caused by estimated sizes above the nominal size, we see that the relative performance of the DWE, DWI and IMPO tests is even better than that of Durbin's two tests. The only exception to this observation occurs for the artificial data set $X6$. There is also a slight decline in the relative average accuracy of the DWE test compared with the DWI and IMPO tests. These conclusions should be tempered slightly by the maximum differences between estimated and nominal sizes. They suggest that there can be circumstances in which the DWE test is the least accurate in terms of actual size. On the other hand, we see that for $X2$ and $X5$ it is the most robust test because it has the smallest maximum differences for these design matrices.

For two lags and normal disturbances (Table II), there is a noticeable and somewhat unexpected improvement in accuracy of both of Durbin's tests. In particular, the DWE and Durbin's t tests have reasonably similar levels of accuracy with the DWE test having an advantage for $n = 30$ while Durbin's t test is almost always the most accurate test for larger sample sizes. Table IV suggests that increasing σ^2 typically does not affect the average accuracy of the DWE, DWI, IMPO and Durbin's t tests but does result in a significant drop in accuracy for Durbin's h test.

Of the SDD based tests, the DWE test is typically the more accurate while the DWI and IMPO tests have almost identical levels of accuracy. The relative improvement of the DWE test going from $p = 1$ to $p = 2$ is not unexpected because the DWE approach attempts to take account of the effects of all regressors on the test statistic's null distribution. If we focus only on that part of the average caused by estimated sizes above the nominal size, the relative performance of the SDD based tests looks much better. Also the relative performance of the DWI and IMPO tests compared to the DWE test shows a slight improvement.

The results for $p = 1, |\alpha| < 1$ and the three non-normal disturbances are almost identical and therefore only one set is given in Table III. Overall the ranking and patterns are similar to those in Table I, indicating that the sizes of all five tests are rather robust to the particular nonnormal distributions considered in this study.

Table V summarizes the size results for the unit root case, $\alpha = 1$, when $p = 1$ under normal disturbances. In general there is little difference between the $\alpha = 0.99$ and $\alpha = 1$ results. Overall, the average accuracy of each of the three SDD based tests is better than that of the large sample asymptotic tests although the differences are much smaller than for $|\alpha| < 1$. Also average accuracy continues to increase as n increases. The true size of the DWE test is nearly always higher than the nominal size. This is true to a lesser degree for the DWI and IMPO tests while for large n, Durbin's h and t tests have true sizes below the nominal size on a greater proportion of occasions. As one might expect, relative average accuracy of the three SDD based tests is greatest for small σ values. In fact for large n and large σ, there is very little difference between the average accuracies of all five tests.

With respect to comparing the powers of the five tests, the task is complicated by the fact that different tests have different sizes which vary with values of the coefficients of the lagged dependent variable. The best we can do is to look for tests that have lower sizes and higher powers than other tests. Our main aim is to check whether the new DWE procedure results in a loss of power compared to the other tests. Tables VI-IX provide selected sizes and powers for $X1$ at the 5% nominal level for $p = 1$ and $p = 2$.

For $p = 1$ and normal disturbances, we could not find a case where the DWI test has lower size and higher power than the DWE test. In all circumstances, DWE never has a lower probability of rejecting the null than the DWI test. Therefore, we could find no evidence to suggest our test procedure results in a reduction in power. A comparison of the DWE and IMPO tests reveals a similar picture for $X1 - X5$. Only for $X6$ could we find evidence that the IMPO test is more powerful than the DWE test. The differences appear to be small but significant. The comparison of the powers of Durbin's tests and the DWE test is made very difficult in some cases because of the rather high sizes of Durbin's tests. Nevertheless, we were able to find evidence that the DWE test has a power advantage over both Durbin's

t and h tests, particularly when $\alpha = 0.75, 0.99$ or 1.0. The evidence was more conclusive in the case of the h test. These conclusions apply equally for both small and large values of σ.

For the case of $p = 2$ and normal disturbances, the DWE test is always more powerful than the DWI test. An obvious explanation is that the DWE test has higher size. However for $X5$, we did notice some evidence suggesting the DWE procedure has a steeper sloping power curve than DWI when $\alpha = 0.25, 0.5$ and σ is small. As for $p = 1$, we found the DWE test to always have higher power than IMPO, except for $X6$. In the situations for $X6$ where IMPO is more powerful than DWE, the differences are generally smaller than those found for $p = 1$. Again the comparison of the DWE test with Durbin's tests, particularly the h test, is complicated by the inflated sizes of these latter tests. However, we found evidence of greater power for DWE in many cases, especially for large ρ. In only one situation, for $X6$ and small σ, did DWE lose power compared to t.

4. Concluding Remarks

This paper provides a small-disturbance justification for applying the DW test to the dynamic linear regression model and using critical values calculated by treating all regressors as exogenous. Our Monte Carlo results support this suggestion. They show that the new procedure provides reasonably accurate critical values although there can be situations in which its accuracy is questionable. While slightly less accurate than the DWI and IMPO critical values for one lag of the dependent variable, our study suggests the new procedure results in slightly more accurate critical values when two lags are present. The sizes of all five tests for autocorrelation appear to be relatively robust to nonnormality in the disturbances.

We also found that the new procedure does not result in a loss of power. In the case of the dynamic linear regression model with two lags of the dependent variable, our results again support the use of the DW test. This contradicts the findings of Dezhbakhsh (1990) who warned against the use of the DW test in this situation.

TABLE VI

Estimated sizes and powers of the five tests for $X1, p = 1, n = 30, \sigma = 0.1$, normal disturbances and a nominal level of 5% .

	Test						Frequency of	
α_1	ρ	DWE	DWI	IMPO	t	h	no h test	Average R^2
0.25	0	0.020	0.014	0.010	0.070	0.168	379	0.752
	0.3	0.110	0.086	0.082	0.222	0.305	223	0.756
	0.6	0.352	0.271	0.281	0.366	0.428	81	0.767
	0.9	0.542	0.467	0.466	0.438	0.484	16	0.810
0.50	0	0.060	0.043	0.041	0.074	0.153	177	0.868
	0.3	0.294	0.237	0.243	0.303	0.371	77	0.867
	0.6	0.663	0.594	0.600	0.581	0.625	15	0.865
	0.9	0.818	0.774	0.779	0.717	0.739	1	0.882
0.75	0	0.095	0.077	0.072	0.051	0.080	29	0.958
	0.3	0.461	0.392	0.400	0.336	0.371	13	0.956
	0.6	0.826	0.785	0.788	0.716	0.742	3	0.946
	0.9	0.940	0.924	0.925	0.878	0.883	0	0.943
0.99	0	0.050	0.039	0.039	0.012	0.014	0	0.996
	0.3	0.392	0.323	0.330	0.176	0.191	0	0.996
	0.6	0.827	0.776	0.779	0.632	0.656	0	0.994
	0.9	0.964	0.948	0.948	0.892	0.900	0	0.986
1.00	0	0.044	0.033	0.032	0.009	0.012	0	1.000
	0.3	0.365	0.306	0.310	0.147	0.160	0	1.000
	0.6	0.795	0.756	0.760	0.592	0.613	0	1.000
	0.9	0.957	0.942	0.942	0.869	0.887	0	1.000

TABLE VII

Estimated sizes and powers of the five tests for $X1, p = 1, n = 69, \sigma = 2000$, normal disturbances and a nominal level of 5% .

α_1	Test						Frequency of no h test	Average R^2
	ρ	DWE	DWI	IMPO	t	h		
0.25	0	0.002	0.001	0.002	0.129	0.272	346	0.094
	0.3	0.066	0.053	0.055	0.290	0.381	19	0.261
	0.6	0.420	0.371	0.376	0.507	0.561	0	0.513
	0.9	0.738	0.704	0.705	0.668	0.699	0	0.817
0.50	0	0.022	0.017	0.016	0.141	0.227	36	0.251
	0.3	0.374	0.328	0.337	0.515	0.564	0	0.458
	0.6	0.886	0.853	0.859	0.873	0.887	0	0.677
	0.9	0.988	0.985	0.986	0.977	0.981	0	0.890
0.75	0	0.090	0.076	0.078	0.135	0.175	3	0.516
	0.3	0.736	0.694	0.696	0.722	0.748	0	0.688
	0.6	0.992	0.990	0.989	0.985	0.989	0	0.831
	0.9	1.000	1.000	1.000	0.999	1.000	0	0.949
0.99	0	0.135	0.112	0.116	0.102	0.119	0	0.856
	0.3	0.851	0.817	0.819	0.763	0.787	0	0.917
	0.6	0.999	0.997	0.998	0.995	0.996	0	0.960
	0.9	1.000	1.000	1.000	1.000	1.000	0	0.991
1.00	0	0.135	0.112	0.117	0.101	0.117	0	0.870
	0.3	0.843	0.805	0.810	0.754	0.784	0	0.925
	0.6	0.999	0.997	0.998	0.993	0.996	0	0.964
	0.9	1.000	1.000	1.000	1.000	1.000	0	0.992

TABLE VIII

Estimated sizes and powers of the five tests for $X1, p = 2, n = 30, \sigma = 0.1$, normal disturbances and a nominal level of 5%.

(α_1, α_2)	ρ	Test DWE	DWI	IMPO	t	h	Frequency of no h test	Average R^2
(0.25,0.25)	0	0.013	0.010	0.009	0.037	0.106	360	0.866
	0.3	0.014	0.009	0.008	0.049	0.147	505	0.867
	0.6	0.023	0.012	0.008	0.052	0.184	513	0.865
	0.9	0.340	0.285	0.276	0.103	0.451	264	0.871
(0.5,0.25)	0	0.010	0.007	0.006	0.020	0.073	321	0.955
	0.3	0.015	0.009	0.007	0.031	0.134	442	0.955
	0.6	0.108	0.064	0.051	0.032	0.216	345	0.950
	0.9	0.560	0.498	0.487	0.144	0.566	106	0.937
(0.9,0.09)	0	0.024	0.013	0.013	0.026	0.036	117	0.995
	0.3	0.211	0.144	0.143	0.053	0.168	113	0.995
	0.6	0.609	0.517	0.521	0.081	0.471	62	0.994
	0.9	0.895	0.857	0.859	0.406	0.797	6	0.986
(1.25,-0.5)	0	0.033	0.013	0.013	0.019	0.048	109	0.967
	0.3	0.161	0.083	0.081	0.069	0.115	27	0.969
	0.6	0.442	0.291	0.288	0.181	0.280	7	0.971
	0.9	0.760	0.629	0.622	0.430	0.552	5	0.975

TABLE IX

Estimated sizes and powers of the five tests for $X1, p = 2, n = 69, \sigma = 2000$, normal disturbances and a nominal level of 5%.

(α_1, α_2)	ρ	Test DWE	DWI	IMPO	t	h	Frequency of no h test	Average R^2
(0.25,0.25)	0	0.000	0.000	0.000	0.091	0.228	440	0.184
	0.3	0.000	0.000	0.000	0.146	0.304	612	0.379
	0.6	0.001	0.001	0.000	0.090	0.215	550	0.645
	0.9	0.280	0.253	0.216	0.090	0.483	146	0.886
(0.5,0.25)	0	0.000	0.000	0.000	0.084	0.227	436	0.424
	0.3	0.001	0.001	0.000	0.087	0.238	621	0.628
	0.6	0.038	0.024	0.017	0.012	0.207	213	0.819
	0.9	0.492	0.472	0.447	0.106	0.560	16	0.949
(0.9,0.09)	0	0.026	0.017	0.016	0.037	0.147	215	0.827
	0.3	0.320	0.280	0.273	0.033	0.456	64	0.903
	0.6	0.717	0.694	0.682	0.062	0.754	2	0.956
	0.9	0.935	0.929	0.925	0.447	0.935	0	0.990
(1.25,-0.5)	0	0.009	0.004	0.004	0.017	0.044	5	0.760
	0.3	0.205	0.136	0.136	0.187	0.246	0	0.860
	0.6	0.719	0.641	0.638	0.591	0.678	0	0.928
	0.9	0.981	0.967	0.967	0.946	0.968	0	0.973

Acknowledgements

This research was supported by an ARC grant. We are grateful to Fong Lai, Vladimir Rouderfer and Alan Morgan for research assistance and to Aman Ullah and two referees for constructive suggestions.

References

Brooks, R.D. and M.L. King, (1994), Testing Hildreth-Houck against return to normalcy random regression coefficients, *Journal of Quantitative Economics* 10, 33-52.

Dezhbakhsh, H., (1990), The inappropriate use of serial correlation tests in dynamic linear models, *Review of Economics and Statistics* 72, 126-132.

Durbin, J., (1970), Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, *Econometrica* 38, 410-421.

Durbin, J. and G.S. Watson, (1951), Testing for serial correlation in least squares regression II, *Biometrika* 38, 159-178.

Imhof, P.J., (1961), Computing the distribution of quadratic forms in normal variables, *Biometrika* 48, 419-426.

Inder, B.A., (1984), Finite-sample power of tests for autocorrelation in models containing lagged dependent variables, *Economics Letters* 14, 179-185 and 16, 401-402.

Inder, B.A., (1985), Testing for first order autoregressive disturbances in the dynamic linear regression model, Unpublished Ph.D. Thesis (Monash University, Clayton).

Inder, B.A., (1986), An approximation to the null distribution of the Durbin-Watson statistic in models containing lagged dependent variables, *Econometric Theory* 2, 413-428.

Inder, B.A., (1990), A new test for autocorrelation in the disturbances of the dynamic linear regression model, *International Economic Review* 31, 341-354.

Kenkel, J.L., (1974), Some small-sample properties of Durbin's tests for serial correlation in regression models containing lagged dependent variables, *Econometrica* 42, 763-769.

Kenkel, J.L., (1975), Small-sample tests for serial correlation in models containing lagged dependent variables, *Review of Economics and Statistics* 57, 383-386.

Kenkel, J.L., (1976), Comment on the small-sample power of Durbin's h-test, *Journal of the American Statistical Association* 71, 96-97.

King, M.L., (1985), A point optimal test for autoregressive disturbances, *Journal of Econometrics* 27, 21-37.

King, M.L., (1987), Testing for autocorrelation in linear regression models: A survey, in M.L. King and D.E.A. Giles, eds., *Specification analysis in the linear model* (Routledge and Kegan Paul, London) 19-73.

King, M.L. and D.E.A. Giles, (1984), Autocorrelation pre-testing in the linear model: Estimation, testing and prediction, *Journal of Econometrics* 25, 35-48.

King, M.L. and G.H. Hillier, (1985), Locally best invariant tests of the error covariance matrix of the linear regression model, *Journal of the Royal Statistical Society* B47, 98-102.

King, M.L. and P.X. Wu, (1990), Locally optimal one-sided tests for multiparameter hypotheses, paper presented at the 6th World Congress of the Econometric Society, Barcelona.

King, M.L. and P.X. Wu, (1991), Small-disturbance asymptotics and the Durbin-Watson and related tests in the dynamic regression model, *Journal of Econometrics* 47, 145-152.

Koerts, J. and A.P.J. Abrahamse, (1969), *On the theory and application of the general linear model*, (Rotterdam University Press, Rotterdam).

Lee, J.H.H., (1992), Robust Lagrange multiplier and locally-most-mean-powerful based score tests for ARCH and GARCH regression disturbances, *mimeo*, Monash University.

Maddala, G.S. and A.S. Rao, (1973), Tests for serial correlation in regression models with lagged dependent variables and serially correlated errors, *Econometrica* 41, 761-774.

Nankervis, J.C. and N.E. Savin, (1987), Finite sample distributions of t and F statistics in an $AR(1)$ model with an exogenous variable, *Econometric Theory* 3, 387-408.

Park, S., (1975), On the small-sample power of Durbin's h test, *Journal of the American Statistical Association* 70, 60-63.

Park, S., (1976), Rejoinder to 'Comments on the small-sample power of Durbin's h-test', *Journal of the American Statistical Association* 71, 97-98.

Ramberg, J.S. and B.W. Schmeiser, (1972), An approximate method for generating symmetric random variables, *Communications of the Association for Computing Machinery* 15, 987-990.

Ramberg, J.S. and B.W. Schmeiser, (1974), An approximate method for generating asymmetric random variables, *Communications of the Association for Computing Machinery* 17, 78-87.

Ramberg, J.S. P.R. Tadikamalla, E.J. Dudewicz and E.F. Mykytka, (1979), A probability distribution and its uses in fitting data, *Technometrics* 21, 201-215.