dom Experience," *Economic Journal* 86 (Mar. 1974a), 32–55.
———, "Inequality Measures, Prices and Household Consumption," *Review of Economic Studies* 41 (Oct. 1974b), 493–504.
Nordhaus, William D., and John B. Shoven, "A Technique for Analyzing and Decomposing Inflation," paper presented at the National Bureau of Economic Research Conference on Price Behavior, Bethesda, Maryland, November 21–23, 1974.
Phlips, Louis, "A Dynamic Version of the Linear Expenditure Model," this REVIEW 54 (Nov. 1972), 450–458.
Polenske, Karen R., *State Estimates of the Gross National Product 1947, 1958, 1963* (Lexington, MA: Lexington Books, 1972).
Prais, S. J., and Hendrik S. Houthakker, *The Analysis of Family Budgets* (Cambridge: Cambridge University Press, 1955).
Ramsey, James B., R. Rasche, and B. Allen, "An Analysis of the Private and Commercial Demand for Gasoline," this REVIEW 57 (Nov. 1975), 502–507.
Stern, Irving, "Industry Effects of Government Expenditures: An Input-Output Analysis," *Survey of Current Business* 55 (May 1975), 9–23.
Stigler, George S., and Gary S. Becker, "De Gaustibus Non Est Disputandum," *American Economic Review* 67 (Mar. 1977), 76–90.
Sydenstricker, E., and W. I. King, "The Measurement of the Relative Economic Status of Families," *Journal of the American Statistical Association* 17 (1921), 842–857.
U.S. Bureau of Economic Analysis, "The Input-Output Structure of the U.S. Economy: 1967," *Survey of Current Business* 54 (Feb. 1974), 24–56.
———, *Definitions and Conventions of the 1967 Input-Output Study* (Washington, D.C.: U.S. Government Printing Office, 1974).
U.S. Bureau of Labor Statistics, *Survey of Consumer Expenditures, 1960–61* (Washington, D.C.: U.S. Government Printing Office, 1966).
———, *Consumer Expenditures and Income: Survey Guidelines*, Bulletin No. 1686 (1971).
U.S. Department of Commerce, *Definitions and Conventions of the 1967 Input-Output Study*, Bureau of Economic Analysis, Interindustry Economics Division (BE-51).
Vaccara, Beatrice N., "Input-Output Coefficients for the United States," in A. P. Carter and A. Brody (eds.), *Applications of Input-Output Analysis* (Amsterdam: North-Holland Publishing Company, 1970), 238–260.

THE ROBUSTNESS OF THE DURBIN-WATSON TEST

Robert Bartels and John Goodhew*

I. Introduction

Over the past decade a number of articles have appeared in this REVIEW comparing the power of the Durbin-Watson (D-W) test for autocorrelation in regression residuals with the power of the Geary test (Habibagahi and Pratschke, 1972; Harrison, 1975; and Guilkey and Schmidt, 1975). The D-W test is generally assumed to be the most powerful simple test against first order autocorrelation since the considerations of Anderson (1948) were taken into account in its development. The motivation behind the work comparing the D-W test with the Geary test therefore appears to be a concern about the robustness of the D-W test with respect to departures from the classical assumptions for the regression model, particularly the assumption of normality for the disturbance terms. The Geary test, which relies only on the signs of changes in the regression residuals and not on their magnitudes, can be expected to be a more robust test. This would compensate for its inferior power under classical assumptions.

The robustness of the D-W test has, however, never been investigated. On the other hand, there is some indirect evidence which suggests that the D-W test may be sensitive to certain kinds of departure from normality. Hannan and Kanter (1977) and Bartels (1977a) have shown that for data drawn from the class of stable distributions, which have much thicker tails than the normal distribution, the distribution of the serial correlation coefficient is much more concentrated than normal theory predicts. On the basis of these findings it could be surmised that thick-tailed disturbance distributions might seriously affect the size of the D-W test, so as to bias it in favour of the null hypothesis. As a result, the presence of autocorrelation in the disturbance terms may frequently go undetected with the consequence that the estimates of the regression parameters will be inefficient (see Nakamura and Nakamura, 1978). However, the link between the serial correlation coefficient for original observations and the D-W statistic for regression residuals is rather tenuous and not too much weight can be attached to this evidence.

In this paper we use Monte Carlo methods to obtain more direct evidence about the sensitivity of the D-W test to departures from normality. In particular, we estimate the actual rejection rates of the D-W test, for three different sets of data and a variety of non-normal disturbance distributions. In brief, the results show that the D-W test is quite robust, particularly at the 5% level of significance. Furthermore, such departures from normal theory results as were observed in the experiments always pointed towards a higher rate of rejection of the null hypothesis. It is argued that this has no serious practical consequences and hence that

Received for publication March 25, 1980. Accepted for publication April 28, 1980.

* University of Sydney.

the replacement of the D-W test by the Geary test has little justification.

II. The Monte Carlo Experiment

In matrix form the simple regression model can be written as

$$y = X\beta + \epsilon$$

where X is the $T \times k$ matrix of observations on the regression variables, β is the parameter vector, y the vector of observations on the dependent variable and ϵ is the vector of true disturbance terms. After fitting this model by least squares, the Durbin-Watson statistic is calculated from the observed residuals $e' = (e_1, \ldots, e_T)$ by $d = \sum_{t=1}^{T-1} (e_t - e_{t+1})^2 \,/\, \sum_{t=1}^{T} e_t^2$. In Monte Carlo experiments the true disturbances ϵ are known, however, and this allows us to bypass the fitting of the regression equation since it is easily shown that the regression residuals can be obtained directly as $e = M\epsilon$ where $M = [I - X(X'X)^{-1}X']$. This relationship also demonstrates that the distribution of d depends both on the distribution of ϵ and on the data matrix X, but not on the parameter values β nor on the observations y. In order to investigate the robustness of the D-W test in a variety of regression models it is therefore necessary to choose a number of data sets, X, as well as a number of different distributional forms for the disturbances.

In our Monte Carlo experiments three different data sets were chosen as representative of models occurring in practice, namely,

TEXTILES: the textiles data of Theil and Nagar (1961) consisting of 17 observations on 3 regressors,

CONSUMPTION: Klein's (1950) consumption data consisting of 21 observations on 3 regressors, and

TREND: 25 observations on the simple trend model $y_t = \alpha + \beta t + \epsilon_t$.

If we assume normality then it is possible to obtain the exact critical values of the D-W test for any given X-matrix using the formulae developed by Imhof (1961) and Pan (1968). For the textiles and consumption data these critical values have been calculated by Ward (1973); for the trend case we used the D-W upper bounds, which in this case agree with the exact critical values to at least 2 decimal places (see Guilkey and Schmidt, 1975). Using these exact critical values for the D-W test, the actual rejection rate of the test was estimated from 10,000 trials for six different distributions for the disturbances. These distributions were (a) the normal (b) the stable distribution with characteristic exponent 1.1 and skewness parameter 1 (see Fama, 1963, for a description of the stable distributions) (c) the Cauchy distribution (d) the convolution of the normal with a symmetric stable distribution having exponent 1.5 (e) a U-shaped distribution with density $f(x) = 4\tfrac{1}{2}x^{10}$ for $|x| < 1$ and (f) the centered χ^2_1 distribution.

The simulations with normal disturbances provide a benchmark for the other simulations. Distributions (b), (c) and (d) are examples of thick-tailed distributions, (c) and (d) being symmetric while (b) is extremely skewed. The relevance of such thick-tailed distributions to economics has been argued effectively by Mandelbrot (1963). In addition, Bartels (1977b) concludes, on theoretical grounds, that convolutions like (d) are appropriate models for the disturbances in regression models. Distributions (b), (c) and (d) have been included to pick up any lack of robustness the D-W test may have with respect to thick-tailed distributions. Distributions (e) and (f) have been included to determine the sensitivity of the D-W test to other departures from normality. In general, the distributions chosen represent reasonably extreme departures from normality, the aim being to present a "worst scenario" assessment of the D-W test.

The calculations in the Monte Carlo experiments were carried out in the following steps. Given one of the 3 data sets, the appropriate number of disturbance terms $\epsilon_1, \epsilon_2, \ldots, \epsilon_T$ were generated from each of the 6 disturbance distributions. For the normal distribution the random variates were generated by the Box-Muller method, for the stable distributions we used the method developed by Chambers et al. (1976), while for the remaining distributions the random variates were generated by simple transformations of uniform random numbers. For each of these six realisations of ϵ the value of the D-W statistic was then calculated. This procedure was repeated 10,000 times in order to determine the characteristics of the distribution of the D-W statistic under the different models.

III. Results

In the context of hypothesis testing the aspects of the test statistic's null distribution which are of most concern are the critical values; that is, the values of the test statistic which delineate the acceptance region from the rejection region. In our discussion of the robustness of the D-W test we therefore concentrate on the sensitivity of the significance levels associated with these critical values to departures from normality. More specifically, we examine the size of the actual type I error of the test when the nominal significance level is set at $\alpha = 0.01$, 0.025 and 0.05. Only the one-tailed critical values for positive autocorrelation are considered since, in practice, tests for negative correlation are called for relatively infrequently. Table 1 gives the estimates of the actual significance levels for the 3 different data sets and the 6 different distributions for the disturbances. One of these 6 distributions is the normal distribution, which is included both for the sake of comparison and as a check on the mechanics of the simulations. As an additional check, the exact mean and variance of the D-W statistic were calculated for each of the three data sets, and these were compared with the estimated means and variances. No irregularities were detected.

In interpreting the results in table 1 it must be remembered that the disturbance distributions used in these Monte Carlo experiments represent extreme departures from normality. The stable, Cauchy, and stable + normal distributions, for example, all have infinite variance; while the U-shaped distribution used is very U-shaped and χ^2_1 is an extreme case of a skewed distribution. When viewed in this light it can be seen that the D-W test is quite robust. For a nominated significance level of $\alpha = .05$, for instance, the estimated actual significance levels for the non-normal disturbance distributions range from 0.0487 to 0.0645. While the latter value is obviously significantly different from 0.05, it nevertheless falls well within the range 0.03 to 0.07, which Pearson and Please (1975) use as a pragmatic criterion for judging robustness.

However, for lower levels of significance the results are not nearly as reassuring. Some of the estimated significance levels are more than double the nominated level. These lower significance levels appear to be particularly sensitive to thick-tailed disturbance distributions. Surprisingly, however, the direction of the distortion is opposite to that observed by Hannan and Kanter (1977) and Bartels (1977a) for the serial correlation coefficient. Whereas in the case of the serial correlation coefficient the null hypothesis is favoured, here it is the alternative hypothesis which is accepted too frequently.

When it comes to assessing the practical implications of our simulation results it is important to note that in all cases except two the estimated significance levels in table 1 for non-normal distributions exceed the nominal significance levels. This is true for the thick-tailed as well as the U-shaped and χ^2 distributions. It appears therefore that, in general, the effect of non-normality on the D-W test is an increase in the probability of rejecting the null hypothesis. Hence there is a quite unexpected agreement between our work and the work of Nakamura and Nakamura (1978). Their investigation of the properties of regression estimates with and without a correction for autocorrelation, and assuming normality, led them to conclude that it is better erroneously to accept autocorrelation in the disturbance terms than erroneously to reject autocorrelation. The reasons for this are that firstly, easy and efficient procedures, such as the Cochrane-Orcutt procedure, exist for dealing with autocorrelation and secondly, that the failure to detect autocorrelation can lead to very inefficient estimates for the regression coefficients. Our results indicate that non-normality leads to an increased acceptance of the alternative hypothesis of autocorrelation which, according to Nakamura and Nakamura, is erring on the side of safety.

In summary, it can therefore be stated that non-normality of the disturbance terms is not a serious problem for the Durbin-Watson test. The 5% significance level is quite robust with respect to the various non-normal distributions chosen for our Monte Carlo experiments. The nominal 2½% and 1% significance levels, on the other hand, although not quite so robust, invariably result in higher rejection rates. The only practical consequence of this is that a correction for autocorrelation is made more frequently than is strictly necessary; a course of action which has already been recommended in the literature for different reasons. As a final comment it may be added that in the light of the above results and discussion there appears to be little benefit in switching from the D-W test to the Geary test, which is less powerful under normality, and whose properties for non-normal distributions are not known.

TABLE 1.—OBSERVED REJECTION RATES FOR THE DURBIN-WATSON TEST AT GIVEN LEVELS OF SIGNIFICANCE

Significance Level	Distribution of Disturbances[a] (a)	(b)	(c)	(d)	(e)	(f)	Standard Error[b]
TEXTILES							
$\alpha = .01$	.0094	.0196	.0197	.0121	.0131	.0148	.0010
.025	.0261	.0385	.0332	.0270	.0298	.0323	.0016
.05	.0523	.0619	.0587	.0524	.0546	.0544	.0022
CONSUMPTION							
$\alpha = .01$	.0103	.0151	.0095	.0107	.0125	.0144	.0010
.025	.0244	.0393	.0375	.0262	.0276	.0306	.0016
.05	.0489	.0642	.0629	.0487	.0525	.0567	.0022
TREND							
$\alpha = .01$	.0091	.0209	.0232	.0141	.0132	.0167	.0010
.025	.0221	.0411	.0399	.0301	.0306	.0354	.0016
.05	.0469	.0645	.0571	.0552	.0529	.0594	.0022

[a](a) Normal; (b) Stable with characteristic exponent 1.1 and skewness parameter 1; (c) Cauchy; (d) Convolution of normal and stable with exponent 1.5; (e) U-shaped; (f) centered χ^2_1.
[b] The standard errors are given by $\sqrt{\alpha(1 - \alpha)/10{,}000}$.

REFERENCES

Anderson, T. W., "On the Theory of Testing Serial Correlation," *Skandinavisk Aktuarietidskrift* 31 (1948), 88–116.

Bartels, Robert, "Estimation in a First Order Autoregressive Scheme with Non-normal Stable Disturbances," *Journal of Statistical Computation and Simulation* 6 (1) (1977a), 35–48.

———, "On the Use of Limit Theorem Arguments in Economic Statistics," *The American Statistician* 31 (May 1977b), 85–87.

Chambers, J. M. et al., "A Method for Simulating Stable Random Variables," *Journal of the American Statistical Association* 71 (June 1976), 340–344.

Fama, Eugene F., "Mandelbrot and the Stable Paretian Hypothesis," *Journal of Busines* 36 (Oct. 1963), 420–429.

Guilkey, David K., and Peter Schmidt, "Some Further Evidence on the Power of the Durbin-Watson and Geary Tests," this REVIEW 57 (Aug. 1975), 379–382.

Habibagahi, Hamid, and John L. Pratschke, "A Comparison of the Power of the von Neumann Ratio, Durbin-Watson and Geary Tests," this REVIEW 54 (May 1972), 179–185.

Hannan, E. J., and Marek Kanter, "Autoregressive Processes with Infinite Variance," *Journal of Applied Probability* 14 (June 1977), 411–415.

Harrison, M. J., "The Power of the Durbin-Watson and Geary Tests: Comment and Further Evidence," this REVIEW 57 (Aug. 1975), 377–379.

Imhof, J. P., "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika* 48 (3 and 4) (1961), 419–426.

Klein, Lawrence R., *Economic Fluctuations in the United States 1921–1941* (New York: John Wiley and Sons, 1950).

Mandelbrot, Benoit, "New Methods in Statistical Economics," *Journal of Political Economy* 71 (Oct. 1963), 421–440.

Nakamura, Alice, and Masao Nakamura, "On the Impact of the Tests for Serial Correlation upon the Test of Significance for the Regression Coefficient," *Journal of Econometrics* 7 (2) (1978), 199–210.

Pan, Jie-Jian, "Distributions of the Noncircular Serial Correlation Coefficients," *Translations in Mathematical Statistics and Probability* 7 (1968), 281–292.

Pearson, E. S., and N. W. Please, "Relation between the Shape of Population Distribution and the Robustness of Four Simple Test Statistics," *Biometrika* 62 (2) (1975), 223–241.

Theil, Henri, and A. L. Nagar, "Testing the Independence of Regression Disturbances," *Journal of the American Statistical Association* 56 (Dec. 1961), 793–806.

Ward, Laurel L., "Is Uncorrelating the Residuals Worth It?" MSc thesis, Department of Mathematics, McGill University, Montréal, 1973.

ENVIRONMENTAL REPERCUSSIONS AND THE ECONOMIC STRUCTURE: SOME FURTHER COMMENTS

Stuart A. Moore*

I. Introduction

In a recent contribution to this REVIEW, Steenge (1978) suggests an important modification to the input-output model which had previously been extended by Leontief to include environmental effects. Leontief (1970) added rows to show the output of pollutants by industries and, in addition, he defined a pollution abatement "industry" with a specific technology for the elimination of each pollutant. In Leontief's model, final demand sectors are assumed to tolerate a fixed level of each of the pollutants and this, together with specified levels of final demand for commodities, can be used to generate a gross output vector and quantities of eliminated pollution. In addition, an advantage of the input-output approach, as Leontief points out, is that various price relationships can be made quite explicit.

As Flick (1974) showed, and Steenge elaborated, the idea of a tolerated level of pollution is fine provided the total amount of each pollutant generated (for a specified level of final demand) is greater than, or equal to, the tolerated level. If not, the model predicts that the pollution abatement industry will produce pollution up to the tolerated level; a result which does not accord with sensible behaviour. Although, as Steenge shows, such a result is not likely to be observed in practice, he goes on to suggest a modification that has considerable intuitive attraction. Steenge's modification is to express tolerated pollution (or, equivalently, eliminated pollution) as a constant proportion of total pollution. Steenge then shows how this modified model can be further adjusted in order to investigate the implications of various pricing policies.

We will not be concerned with pricing, which is fully covered in Steenge's paper; rather, the purpose of this note will be to examine in further detail the output structure of his model. Steenge interprets the additional output requirements for pollution abatement as changes in the technological structure of industries. The alternative, but mathematically equivalent, interpretation given in this note is to define a total pollution multiplier matrix and to show that the additional output can be identified as the results of a multiplier process. A numerical example illustrates the calculation of the multiplier for different levels of pollution abatement.

Received for publication April 5, 1979. Accepted for publication January 20, 1980.

* University of Manchester.