69

On Testing for Uniformity of Fit in Regression:

An Econometric Case Study *

John L. Pratschke

It is frequently necessary to test regression results for uniformity of fit - i.e. to test the randomness of the distribution of the error terms. Such a test is necessary because an indication of a systematic trend in the pattern of residuals would suggest that the regression equation fitted to the data did not properly reflect the curvature of the true relationship between the variables regressed. If, for instance, the residuals after regression fell into three groups of positive residuals, followed by a group of negative residuals, followed again by positive residuals, then the implication would clearly be that the estimated regression function misrepresented the true relationship. The problem is, perhaps, most familiar in the analysis of serial correlation of errors in time-series analysis. The problem also arises in regressions based on cross-section data, provided that the regression estimates of the residuals after regression are first ordered in ascending order of the independent variable. In a two-variable case, the analogy with time-series studies is then complete, because time is there the independent variable and the residuals are automatically ordered in ascending order of the time variable. In cross-section studies, the independent variable (if there is but one) is the logical, and indeed only, possible choice with which to order the regression residuals. Where there is more than one independent variable,

* This note should be regarded as a revised version of Memorandum Nos. 60 (revised) and 64 combined.

the more important of the two appears the logical choice, though the possibility of ordering the residuals by any other independent variable does exist. (Again, the analogy with time-series results exists in the case of multiple regression studies: time is generally assumed to be the relevant independent variable when the Durbin-Watson d-test is applied [2,3]). When applying such tests to cross-section data, the term "uniformity of fit" is used instead of "serial correlation of errors".

It is the purpose of this study to describe and compare a number of possible tests for uniformity of fit. It will be obvious from the foregoing that the best known test is a modified version of the Durbin-Watson d-test (op. cit.), where the residuals after regression are reordered in ascending order of the major independent variable, and the d-test then applied in the ordinary way as

$$d = (e_j - e_{j-1})^2 / e_j^2$$

where e_j is the jth residual

Griliches et alia [6] have reported that, in their view, the d-test may be unduly influenced by even one aberrant observation. For this reason, they recommend that alternative non-parametric tests be developed, and the results compared with those yielded from the application of the d-test.

Among the first to propose such a non-parametric test were Stevens [9] , Wald and Wolfowitz [11] and Swed and Eisenhart [10] . Their test, when applied to the examination of regression residuals - as has recently been discussed by Draper and Smith [1] - examines the possibility of sequences of residuals of similar sign (+ or -) being drawn at random from a normal population. The test, which may be referred to as the "Runs Test" is, essentially, a development of the well-known statistical problem of whether or not two independent samples are derived from one continuous distribution. It may be shown (see Mood [7]) that the probability distribution of u, where u is the number of runs, or sequences of one or more residuals of the same sign, tends to normality, with mean and variance defined as

$$E(u) = 2n_1n_2/(n_1 + n_2) + 1$$

$$\text{and } \sigma_u = 2n_1n_2/(2n_1n_2 - n_1n_2)/(n_1 + n_2)^2(n_1 + n_2 - 1)$$

where

n_1 is the number of residuals with positive signs, n_2 is the number of residuals with negative signs, and N (= $n_1 + n_2$) is the total number of residuals. This tendency toward normality is confirmed by Geary [5] when N = 40.

Indeed, Geary (op. cit.) has recently proposed a simpler test, in which account is taken only of the number of changes in sign (+ or -) between successive residuals. Clearly, the number of sign-changes τ is one less than the number of runs, i.e. $\tau = u - 1$. The probability distribution of τ is defined by the point binomial distribution.

Another non-parametric test suggested by Griliches et alia (op. cit) compares the sign of the i th residual with that of the i + 1 th and the frequencies of the observed combination of signs of successive residuals are arranged in a 2 x 2 contingency table of the form

		Sign of the i th Residual		
		+	-	Total
Sign of the i + 1 th Residual	+	a_{11}	a_{12}	$a_{11} + a_{12}$
	-	a_{21}	a_{22}	$a_{21} + a_{22}$
	Total	$a_{11} + a_{21}$	$a_{12} + a_{22}$	$N - 1 = \Sigma_i \Sigma_j a_{ij}$

and chi-squared (x^2) is defined as

$$x^2 = \frac{(N-1)(/a_{11}a_{22} - a_{21}a_{22}/ -(N - 1)/)^2}{(a_{11}+a_{12})(a_{21}+a_{22})(a_{12}+a_{22})(a_{11}+a_{21})}$$

when Yates' correction is used.

If there is positive first order autocorrelation of residuals, one would expect a_{11} and a_{22} to be significantly larger than a_{21} and a_{12}. The null hypothesis of no autocorrelation (i. e. of uniformity of fit) is tested by applying the ordinary chi-squared test (with one degree of freedom) to the table.

These four tests, - the modified Durbin-Watson d-test, the Runs Test, the Geary sign-change test, and the Chi-squared test were applied to the residuals obtained when estimating eighteen different algebraic forms of the Engel function for each of five major expenditure groups. The function forms used are set out in Table A1 of the Appendix. Sixteen observations only were available from which to estimate the regressions. The critical values of the Durbin-Watson d are given in Table 1 following:

Table 1. Critical Values of Durbin-Watson d

No. of Independent Variables	Probability Level	d_L	d_U
1	0.05	1.10	1.37
2		0.98	1.54
3		0.86	1.73
1	0.01	0.84	1.09
2		0.74	1.25
3		0.63	1.44

Source: Durbin and Watson (op. cit)

Table 2: Significance Levels of u

When $\epsilon <$ 0.50, u_ϵ is the largest integer u' for which

$$P\{u \leqslant u'\} \leqslant \epsilon$$

When $\epsilon >$ 0.50, u_ϵ is the smallest integer u' for which

$$P\{u \leqslant u'\} \leqslant \epsilon$$

N = 16		u_ϵ			
n_1	n_2	0.01	0.05	0.95	0.99
1	15	-	-	-	-
2	14	-	2	5	5
3	13	2	3	7	7
4	12	3	4	9	9
5	11	3	4	11	11
6	10	3	5	11	13
7	9	4	5	12	13
8	8	4	5	12	13
9	7	4	5	12	13
10	6	3	5	11	13
11	5	3	4	11	11
12	4	3	4	9	9
13	3	2	3	7	7
14	2	-	2	5	5
15	1	-	-	-	-

Source: Swed and Eisenhart (op. cit).

The application of the Runs Test is complicated by the fact that N = 16; instead of using the normal approximation, the exact probability distribution of u, as tabulated by Swed and Eisenhart (op. cit) was used. Their table is easily manipulated to give Table 2, which shows, for $1 < n_1 < 16$, given N = 16, the values of u_ϵ, where (i) u_ϵ is the largest integer u' for which $P(u \leq u') \geq \epsilon$ when $\epsilon < 0.50$; and (ii) u_ϵ is the smallest integer u' for which $P(u \leq u') \geq \epsilon$ when $\epsilon > 0.50$. The significance levels chosen are the conventional 0.05, 0.01, 0.95 and 0.99 per cent levels.

(Table 2)

The cumulative point binomial distribution, for N = 16, is given in Table 3. It may be seen that the probability of having three or less sign changes is approximately 2 per cent. Similarly, the probability of having ten or more sign changes is approximately 6 per cent - both on the null hypothesis that the residuals are randomly distributed. The

Table 3: Cumulative Point Binomial Distribution

τ	P	τ	P
0	0.0000	8	.6964
1	0.0005	9	.8491
2	0.0037	10	.9408
3	0.0176	11	.9824
4	0.0592	12	.9963
5	0.1509	13	.9995
6	0.3036	14	1.0000
7	0.5000	15	1.0000

discontinuity of the distribution makes it impossible to establish precise 1 per cent or 5 per cent probability levels.

(Table 3)

The convention adopted here is to take $P(\tau \leqslant 2) \sim 1$ per cent and $P(\tau \leqslant 4) \sim 5$ per cent, since these are the values of τ whose probabilities most nearly approximate to the 1 and 5 per cent levels. Thus, especially at the 1 per cent level we are discriminating somewhat against τ - what we have styled the 1 per cent level is in reality the 0.004 per cent **level.**. On the other hand, we are being a little "kind" to τ at the 5 per cent level by using P = 0.06.

The small numbers of observations also raises difficulties in the utilization of the 2 x 2 table. With $N = 16$ (i.e. $N - 1 = \Sigma_i \Sigma_j a_{ij}$) the individual cell entries are generally small, and the correctness of using chi-squared in such cases has, of course, been the subject of some controversy. The Fisher Exact Probability Test was used instead, utilizing the significance levels as tabulated by Finney [4]

The results from the four tests on the residuals are set out in Table 4 following

(Table 4)

The concordance between the Durbin-Watson d-test, the Runs Test and the Geary sign-change test is quite good: the surprising result is the poor showing of the Fisher Exact test when applied to this type of data. Leaving aside the results of the Fisher Exact test, it is noteworthy that of the eight cases where d is significant at the 5 per cent level, τ is significant in five cases, of which one was at the 1 per cent level. Comparable figures for u are four significant, of which two were significant at the 1 per cent level. In general, both u and τ appear to give results broadly similar to those obtained using the d-test, as may be seen from Table 5.

(Table 5)

The interesting results for τ and u where d is inconclusive and also worth noting.

The Economic and Social Research Institute
Dublin.

13 May 1970.

Table 4: Comparison of Alternative Forms of the Engel Function using Four Tests for Uniformity of Fit.

Function Type	Function No.	Food				Clothing				Fuel and Light				Housing				Sundries			
		Significance Appraisal																			
		d	u	τ	Fisher Exact	d	u	τ	Fisher Exact	d	u	τ	Fisher Exact	d	u	τ	Fisher Exact	d	u	τ	Fisher Exact
	1.1					ϕ	*	*													
Linear	1.2		*	*		ϕ	*	*											*	*	
	1.3													*	**	*					
	2.1	ϕ				**		*						ϕ				**	**	**	*
Semi-log	2.2	ϕ		*		*		*										**	**	**	*
	2.3	*												ϕ							
	3.1																		*	*	
Double-log	3.2		*															ϕ	*	*	
	3.3	*	*	*		**				ϕ	**	*		**				**	**	*	
	4.1	ϕ				ϕ	**	**						**	*	*		**	*	*	
Log-inverse	4.2	*	*	*				*						**	*	*		**	*	*	
	4.3	*								*				ϕ							
Linear in w_i	5.1	ϕ	**	**	*	ϕ				ϕ	**	**	*					*	**	**	*
	5.2	**	**	**	*					ϕ	*	*						**	**	**	*
Semi-log in w_i	6.1																				
	6.2																		*	*	
Leser	7.1													*							
	7.2																				

Notes

** Indicates significance at the 1 per cent level. * Indicates significance at the 5 per cent level. ϕ Indicates an inconclusive d test.

Table 5: Concordance of Results Using τ , d, and u Statistics.

Significance Levels of d	Significance Levels of u	d	τ	u
0.01	.01	12	4 } 10	5 } 9
	.05		6 }	4 }
	n. s.		2	3
0.05	.01	8	1 } 5	2 } 4
	.05		4 }	2 }
	n. s.		3	4
Inconclusive	.01	17	3 } 9	4 } 8
	.05		6 }	4 }
	n. s.		8	9
Not Significant	.01	54	- } 5	- } 5
	.05		5 }	5 }
	n. s.		49	49
Total	.01	12 } 20	12 } 29	11 } 26
	.05	8 }	17 }	15 }
	Inconclusive	17	-	-
	n. s.	53	61	64

Note n. s. Indicates not significant.

Table A.1: Algebraic Forms of the Engel Function Fitted to Data of Five Major Expenditure Groups.

Function Type	Function No.	Form of Engel Function
Linear	1.1	$v_i = \alpha_i + \beta_i E + \Upsilon_i n + \epsilon_i$
	1.2	$v_i = \alpha_i + \beta_i E + \Upsilon_i \log n + \epsilon_i$
	1.3	$v_i/n = \alpha_i + \beta_i (E/n) + \epsilon_i$
Semi-log	2.1	$v_i = \alpha_i + \beta_i \log E + \Upsilon_i n + \epsilon_i$
	2.2	$v_i = \alpha_i + \beta_i \log E + \Upsilon_i \log n + \epsilon_i$
	2.3	$v_i = \alpha_i + \beta_i \log (E/n) + \epsilon_i$
Double-log	3.1	$\log v_i = \alpha_i + \beta_i \log E + \Upsilon_i n + \epsilon_i$
	8.2	$\log v_i = \alpha_i + \beta_i \log E + \Upsilon_i \log n + \epsilon_i$
	3.3	$\log (v_i/n) = \alpha_i + \beta_i \log (E/n) + \epsilon_i$
Log-inverse	4.1	$\log v_i = \alpha_i + \beta_i /E + \Upsilon_i n + \epsilon_i$
	4.2	$\log v_i = \alpha_i + \beta_i /E + \Upsilon_i \log n + \epsilon_i$
	4.3	$\log (v_i/n) = \alpha_i + \beta_i (n/E) + \epsilon_i$
Linear in w_i	5.1	$w_i = \alpha_i + \beta_i E + \Upsilon_i n + \epsilon_i$
	5.2	$w_i = \alpha_i + \beta_i E + \Upsilon_i \log n + \epsilon_i$
Semi-log in w_i	6.1	$w_i = \alpha_i + \beta_i \log E + \Upsilon_i n + \epsilon_i$
	6.2	$w_i = \alpha_i + \beta_i \log E + \Upsilon_i \log n + \epsilon_i$
Leser	7.1	$w_i = \alpha_i + \beta_i \log E + \Upsilon_i /E + \delta_i n + \epsilon_i$
	7.2	$w_i = \alpha_i + \beta_i \log E + \Upsilon_i /E + \delta_i \log n + \epsilon_i$

Notes

v_i is average weekly household expenditure on i

E is average weekly household total expenditure ($E = \Sigma_i v_i$)

n is average household size

w_i is average weekly household expenditure proportion on i (i.e. $w_i = v_i / E$)

References Cited

[1.] Draper, N.R., and Smith, H.: Applied Regression Analysis, New York: John Wiley & Sons., Inc., 1966.

[2.] Durbin, J., and Watson, G.S.: "Testing for Serial Correlation in Least Squares Regression, I", Biometrika, Vol. 37, p. 409, 1950.

[3.] - : "Testing for Serial Correlation in Least Squares Regression, II", Biometrika, Vol. 38, p. 159, 1951.

[4.] Finney, D. J.: "The Fisher-Yates Test of Significance in 2 x 2 Contingency Tables", Biometrika, Vol. 35, p. 145, 1948.

[5.] Geary, R. C.: "Relative Efficiency of Count of Sign Changes for Assessing Residual Autoregression in Least Squares Regression", Biometrika (in Press), 1970.

[6.] Griliches, Z., Maddala, G. S., Lucas, R., and Wallace, N: "Notes on Estimated Aggregate Quarterly Consumption Functions", Econometrica, Vol. 30, No. 3, 1962.

[7.] Mood, A. M.: Introduction to the Theory of Statistics, New York: McGraw-Hill Book Co., 1950.

[8.] Prais, S. J., and Houthakker, H. S.: The Analysis of Family Budgets, Cambridge U. P., 1955.

[9.] Stevens, W. L.: "Distribution of Groups in a Sequence of Alternatives", Annals of Eugenics, Vol. 9, Part I, January 1939.

[10.] Swed, F. S., and Eisenhart, C.: "Tables for Testing Randomness of Grouping in a Sequence of Alternatives", Annals of Mathematical Statistics, Vol. 14, No. 1, March 1943.

[11.] Wald, A., and Wolfowitz, J.: "On a Test Whether Two Samples are from the Same Population", Annals of Mathematical Statistics, Vol. 11, No. 2, June 1940.