Economic and Social Review Vol. 9 No. 1

Randomisation and the von Neumann Function: A Variance Formula and a Problem

R. C. GEARY*
The Economic and Social Research Institute, Dublin

IN a paper of many years ago (Geary 1952) what was termed the *contiguity ratio* was introduced, to determine whether, in probability, a statistical map has a pattern or whether the mapped statistics are distributed at random. This ratio is really a two-dimensional version of the von Neumann (1941) statistic, more familiar as that tabled for null-hypothesis normal OLS residuals by J. Durbin and G. S. Watson. Geary was also concerned with the OLS residual problem. He approached it in two ways, by randomisation and by classical OLS regression theory, his instruments being means and variances of the contiguity ratio.

A difficulty with randomisation treatment was expressed as follows:—

"The problem is to determine if there is a contiguity effect, i.e. if c (the contiguity ratio) has a significantly low value after the elimination of q independent variables by the least square method. As far as randomisation is concerned, it would appear that the test developed in this section can be applied formally, the z being the remainders after the contributions of the independent variables have been removed. To a certain extent the writer shares the misgivings of some other students about the validity of the randomization approach in its application to regression remainders. As each successive independent variable is removed, should not the degree of freedom be diminished? It does not seem so. What happens is that the variance (or range) of the remainders diminish as the effect of each independent variable is allowed for, the test becoming indeterminate when the number of independent variables (originally with mean zero) is one less than the number of observations n, i.e., when all the remainders are zero. Accordingly the formal application of the randomization procedure, without diminution of the number of degree of freedom, does not result in *obvious* inconsistency: we can conceive of cases

* The paper benefited from comments by R. N. Vaughan.

where c will be significantly low even after removal of the effect of $(n - 2)$ independent variables. Since doubts remain, however, the writer considered it desirable to examine the problem from the classical sampling aspect. In any case it will be interesting to compare the results of the two approaches. In the practical aspect the randomisation method has the advantage that it can be applied without the assumption of universal normality in the n observations,regarded as a random sample".

As far as the writer is aware the degrees of freedom problem has never been discussed in this application: the controversy in another context between K. Pearson and R. A. Fisher is part of statistical history. One of the objects of the present communication is to invite statisticians to discuss the problem.

The contiguity ratio context is too esoteric for a suitable discussion. The problem arises in the much simpler single dimension of the ratio. But the writer is unaware of any randomisation treatment of the von Neumann statistic, so he ventures to give one here without any claim to originality. One result is remarkable, as will be seen.

One is given a sample of n measures of any kind (they may be raw values, OLS residuals etc.), $x_1, x_2 \ldots, x_n$, ordered in a particular way. From a given function (e.g., the von Neumann ratio) one wants to make inferences about the character of the sample (is it probably non-normal, autoregressed etc.?) One considers the $n!$ permutations of the sample values for each of which the test function has a value. These $n!$ values are regarded as forming a frequency distribution. If the single value of the function found for the given ordered sample is near the ends of the frequency distribution (i.e., beyond the .05, .01 etc., limits) one rejects the hypothesis, exactly as in ordinary theory. A feature of the test is that no assumption is made about the frequency distribution from which the sample of n is drawn. In theory one could calculate the moments of the function—or at least the first four moments—and so estimate the frequency distribution using e.g., the Pearson curve system. Here we deal only with the first two moments, the mean and the variance, which suffice for most practical purposes.

The test function cannot usefully be symmetrical in $(x_1, x_2, \ldots, x_n)$ because then all the $n!$ values would be the same. The essence of the von Neumann ratio d is that it is not symmetrical (for $n>2$) as it assumes that the sample elements are arrayed in a particular way. In fact, assuming, without loss of generality, that:—

$$\sum_{i=1}^{n} x_i = 0 \tag{1}$$

as will always be the case with OLS residuals, d is given by—

$$d = \sum_{i=2}^{n} (x_i - x_{i-1})^2 \Big/ \sum_{i=1}^{n} x_i^2 = N/D \tag{2}$$

The numerator N is assymetrical, the denominator D symmetrical, i.e., D has the same value in all permutations. We need concern ourselves only with the numerator N. It is the fact of constant D that makes the calculation of moments of d exactly calculable. This is also the classical case when the sample is a normal one because then d is a homogeneous function of degree zero, with $nr^2 = \Sigma\, x_i^2$ in the denominator. The fact that when the sample is normal r is independent of d (Geary, 1933) makes the exact calculation of the moments of d, and hence the estimation of the frequency of d (as by Durbin-Watson) possible.

If f is any polynomial function of $(x_1, x_2 \ldots, x_n)$ ordered in a particular way the randomisation mean $M\ (f)$ of f is the sum of f for all the permutations divided by $n!$. To find the mean of d^2 given by (2) or, in effect, N^2 we have to deal with terms in x_i^4, $x_i^3\, x_{i'}$, $x_i^2\, x_{i'}\, x_{i''}$ and $x_i\, x_{i'}\, x_{i''}\, x_{i'''}$, all subscripts different. On taking means we may disregard subscripts and insert mean values of these terms, having regard only to exponents. These mean values may be written (in a notation which is obvious) (4), (31), (22), (211), (1111). Note that (31) = (13) etc.

Square (1) and take means. There are n of type x_1^2 and $n\ (n-1)/2$ of type $x_i x_{i'}$, $i' \neq i$. Hence:

$$n\ (2) + (11)\ [2n\ (n-1)]/2 = 0 \tag{3}$$

or

$$(11) = -\ (2)/(n-1) \tag{4}$$

As in (4), we can express all terms in two or more variables in single variable expressions. As an example of the method of derivation, we have

$$\Sigma x_i\ \Sigma x_i^3 = 0 \tag{5}$$

Multiplying out and taking means:

$$n\ (4) + n\ (n-1)\ (31) = 0 \tag{6}$$

or

$$(31) = -\ (4)/(n-1) \tag{7}$$

The derivation of other randomisation means we need is a little more complicated. We shall be content to give the results:

$$\begin{aligned}(22) &= [n\ (2)^2 - (4)]/(n-1)\\ (211) &= [2\ (4) - n\ (2)^2]/(n-1)\ (n-2)\\ (1111) &= 3\ [n\ (2)^2 - 2\ (4)]/(n-1)\ (n-2)\ (n-3)\end{aligned} \tag{8}$$

From (2),

$$D = n\ (2) \tag{9}$$

Expanding the numerator of (2)

$$N = (x_1^2 + x_n^2) + 2 \sum_{i=2}^{n-1} x_i^2 - 2 \sum_{i=2}^{n} x_i x_{i-1} \tag{10}$$

Hence taking means

$$M\ (N) = 2\ (2) + 2\ (n-2)\ (2) + 2\ (n-1)\ (2)/(n-1), \tag{11}$$

using (4). Hence $M\ (N) = 2n\ (2)$. Then:

$$M\ (d) = M\ (N)/D = 2, \tag{12}$$

using (9).

The algebra of the calculation of $M\ (d^2)$ or, in effect, $M\ (N^2)$ is onerous but the result is simple. We regard N, given by the right side of (10) as three terms $(A + B + C)$ with square $(A^2 + 2AB + \ldots + C^2)$ and aggregate the terms, having regard to coefficients and numbers of terms of each kind, x_i^4, $x_i^3\, x_i$, etc., which, on taking means are replaced by (4), (31) etc. Then, gathering terms we find:

$$\begin{aligned}M\ (N^2) = 2\ (2n-3)\ (4) - 8\ (2n-3)\ (31) + 2\ (2n^2 - 4n + 3)\ (22)\\ - 8\ (n-2)\ (n-3)\ (211) + 4\ (n-2)\ (n-3)\ (1111).\end{aligned} \tag{13}$$

Using (7) and (8) and collecting terms:

$$M\ (N^2) = 2n\ [(2n^2 - 3)\ (2)^2 - (4)]/(n-1). \tag{14}$$

As $M\ (d^2) = M\ (N^2)/D^2$ with $D = n\ (2)$:

$$\begin{aligned}\text{var}\ (d) &= M\ (d^2) - [M\ (d)]^2\\ &= 2\ [(2n-3) - b_2]/n\ (n-1)\end{aligned} \tag{15}$$

where $b_2 = (4)/(2)^2$ the familiar kurticity statistic in normal theory in which in fact its population value β_2 is 3.

As a check on the quite elaborate, if elementary, algebra, consider the case of $n = 2$. There is then but a single value of d given by (2), for in this case d is symmetrical in (x_1, x_2). $(4) = (x_1^4 + x_2^4)/2 = x_1^4$ since $x_1 + x_2 = 0$ and $(2) = x_1^2$. Hence $b_2 = 1$. Substituting then $n = 2$ and $b_2 = 1$ in (15) we find var $(d) = 0$, as we should.

Of course (15) is $O(n^{-1})$, which means that, with increasing n, d tends in probability towards 2 (see (12)). What, as announced above, is remarkable is that the coefficient of b_2 is $O(n^{-2})$. This implies that the variance is nearly independent of the frequency distribution from which the random sample of n (arrayed in any order) is drawn. As an example take $n = 20$—one would scarcely be interested in e.g., residual autocorrelation for fewer observations—and $b = 1$, and 6, a range probably covering most distributions. Values of standard deviation (= square root variance) of var (d) given by (15) are—

Value of b_2	s.d. of d
1	0.4353
6	0.4039

The difference is of small importance, having regard to the uses of the statistic d.

There is an interest in comparing the lower critical probability points derived from randomisation with the well-known d_L of Durbin-Watson (1951). Assuming normality, $b_2 = 3$ and, since the mean is 2, lower .05 and .01 NHP critical points would be given respectively by $(2 - 1.9600s) = c_1$ and $(2 - 2.5759s) = c_2$, s being the randomisation standard deviation, s^2 given by (15). Following are comparisons with d_L for various values of n, with k', the number of regression terms, equalling 1 and 2:—

n	s	*.05 probability*	*Durbin-Watson* d_L		*.01 probability*	*Durbin-Watson* d_L	
		c_1	$k' = 1$	$k' = 2$	c_2	$k' = 1$	$k' = 2$
20	0.4230	1.17	1.20	1.10	0.91	0.95	0.86
30	0.3523	1.31	1.35	1.28	1.09	1.13	1.07
40	0.3080	1.40	1.44	1.39	1.21	1.25	1.20
50	0.2770	1.46	1.50	1.46	1.29	1.32	1.28
60	0.2538	1.50	1.55	1.51	1.35	1.38	1.35
70	0.2356	1.54	1.58	1.55	1.40	1.43	1.40
80	0.2207	1.57	1.61	1.59	1.43	1.47	1.44
90	0.2084	1.59	1.63	1.61	1.46	1.50	1.47
100	0.1980	1.61	1.65	1.63	1.49	1.53	1.50

The regularity in relationship is marked: from $n = 30$ on in no case do the c figures differ from the DW by more than 0.04. The differences, however, though small, are systematic, not random.

Clearly the correspondence between the c and the d_L is good. As is well known, in Durbin-Watson there is a zone of indecision characterised by limits d_L and d_U. As the previous table shows, the randomisation approach strongly favours d_L. The difference between d_L, and d_U is the greater the smaller the value of n so that in the following table attention is confined to $n = 20, 30, 40$.

	.05 *probability*				.01 *probability*			
N	$k' = 1$		$k' = 2$		$k' = 1$		$k' = 2$	
	$d_L - c_1$	$d_U - c_1$	$d_L - c_1$	$d_U - c_1$	$d_L - c_2$	$d_U - c_2$	$d_L - c_2$	$d_U - c_2$
20	.03	.24	−.07	.27	.04	.24	−.05	.36
30	.04	.18	−.03	.26	.04	.17	−.02	.25
40	.04	.14	−.01	.20	.04	.13	−.01	.19

There is not a shadow of doubt that, with $k' = 1,2, d_L$ is closer to the randomisation value than is d_U, to the extent that there seems no point in referring to d_U in testing for absence of residual autoregression. As the randomisation formula is almost distribution-free (as regards the disturbances) the same may now be said of the d_L series. Obviously the randomisation limits (c_1 and c_2) themselves can confidently be used for adjudging significance if too much precision be not attached to the null-hypothesis probability (i.e. to use "near" .05, .01 etc. instead of exact .05, .01 etc.). When n is not too small the foregoing assumption that the $n!$ values are normally distributed is close enough for the purpose of the approximate probability statement.

In their original 1951 paper Durbin and Watson furnish tables only for the lower critical values of the von Neumann ratio, namely, d_L and d_U. The upper critical values should also be considered: the null hypothesis (i.e., no residual autocorrelation) should also be rejected when the actual ratio exceeds $4 - d_L$ (4 being the algebraic maximum of the ratio). The latter situation is far more rare, in actual practice, than the former. It can occur only when consecutive values of the calculated disturbances tend to oscillate in sign from *plus* to *minus*, when, in fact the change in signs test tau (Geary 1970)[1] is near n, the number of observations, instead of near $n/2$ when the disturbance values are in random order.

1. None of the several papers on the tau test questioned the validity of the randomisation test applied to OLS disturbances. All comments bore on the discriminatory power of the test, compared with the DW and other tests. None of the commentators objected that the randomisation test did not take account of degrees of freedom, or denied that, having regard to its simplicity and convenience, tau was more efficient than might have been expected.

The main point is, however, that, at the upper rejection limit, the discrepancies between the normal theory randomisation critical values and the DW values are exactly those shown in the previous tables.

Comparison between randomisation and DW null hypothesis limits has been confined to $k' = 1$ and $k' = 2$, i.e., for estimated OLS disturbances after removal of one and two indvars respectively. Comparison is not so satisfactory for regressions with more than two indvars. The nub of the problem raised in this paper and stated at the outset is that we have but one randomisation critical limit for given probability; there are k' values of d_L. With randomisation the manner by which the OLS disturbances are estimated (i.e., without regard to number of indvars) is disregarded: we simply have a time sequence of numbers (positive and negative since their sum must be zero) and the null hypothesis is that they are probably in non-random order. The fact that the estimated disturbances are related because their sum is zero is of no importance unless n is small. Basically the approach is the same in using the tau test, on the classical approach, on the contrary,with k' indvars there are $(k' + 1)$ linear relations between the disturbances so that in significance-testing the number of independent variables involved is not n but $(n - k' - 1)$, the number of degrees of freedom.

An obvious approach is as follows. If, in (15), one regards n as number of degrees of freedom established in the ordinary way, would the resulting values of c_1 and c_2 correspond to the Durbin-Watson values of d_L for $k' = 1, 2, \ldots, 5$. The answer is "No". Attention may be confined to the most testing case of number of observations equal to 20.

With 1 to 5 indvars, degrees of freedom in the disturbance range from 18 to 14 and using (15) and, always assuming normality, the c_1 values for $n = 18$ and 14 are respectively 1.13 and 1.04, range 0.09. But the Durbin-Watson range for d_L from $k' = 1$ to $k' = 5$ is 0.41. This comparison relates to probability .05. At .01 probability the contrast is also marked: for c_2 the range is 0.15, for k' 0.35. If any correction for number of degrees of freedom is necessary using randomisation, it will not be on these lines.

The most important conclusion is that for practical purposes the Durbin-Watson d_L system may be accepted without the prior assumption that OLS disturbances are normally distributed. It is true that this has been shown here only for $k' = 1$ and 2. There is little reason to doubt that it is true for all values of k'. The randomisation approach has the merit of being far simpler algebraically than Durbin-Watson and can stand on its own, especially when n is not small.

REFERENCES

DURBIN, J and G. S. WATSON, 1951. "Testing for serial correlation in least squares regression 11", *Biometrika,* Vol. 38.

GEARY, R. C., 1933. "A general expression for the moments of certain symmetrical functions of normal samples". *Biometrika,* Vol. XXV.

GEARY, R. C., 1952. "The contiguity ratio and statistical mapping", *The Incorporated Statistician,* Vol. 5, No. 3.

GEARY, R. C., 1970. "Relative efficiency of count of sign changes for assessing residual autoregression in least squares regression". *Biometrika,* Vol. 57.

VON NEUMANN, J., 1941. "Distribution of the ratio of the mean square successive difference to the variance", *The Annals of Mathematical Statistics,* XII, No. 4.