Is the Science of Econometrics any Use for Short-term Forecasting? An Enquiry based mainly on Irish Educational Statistics 1947-1967

R. C. GEARY

1. INTRODUCTION [1]

As his title indicates, the writer, wisely or unwisely, proposes to hang a very general and important question on the highly specialized inquiry with which this paper mainly deals: in fact to make a *prima facie* case for a full-scale inquiry into the measure of success, or otherwise, of econometricians the world over, attained in recent years, using orthodox or unorthodox methods. Most of the theoretical problems seem to have been solved and, thanks to the computer, a vast range of experimental results is now available. The time has come for their appraisal. Since results at the start (say twenty years ago) were, in general, poor, improvements must have been effected—otherwise it is to be presumed that these methods would not be persisted with. We want to learn about these improvements. How were they attained? Are results systematically appraised and how? From the writer's limited experience but fairly extensive reading he is inclined to take a somewhat pessimistic view: he earnestly hopes he is wrong. The following reflections mirror his gloom and are set down so that others may point out to him the error, if any, of his ways.

The writer finds that most of the works which he has read about economic models are characterized by a particular kind of approach. They start with an impressive examination on traditional non-statistical lines of the data and the putative relations between them, in qualitative terms. Then the statistical models of several equations are set up and the equations solved, i.e. the coefficients

1. This paper is an extended version of an address delivered in Paris on 10 September 1968 to the Seminar on Quantative Techniques for Educational Planning conducted by OECD. The writer desires to express his thanks to Professor Abdul G. Khan for organisation of the data for the computer and advice generally, to Mrs. Aine Hyland for data assembly and compilation and to Miss Treasa O'Donoghue for help with non-computerized calculations.

estimated (using least squares (LS) or two-stage LS), usually with their accompanying t-statistics, and the equations with their r's, Durbin-Watson statistics (DW) and residuals. Sometimes the equations, each purporting to explain a single endo-variable, have but a tenuous connection with the economic analysis which preceded their setting-up. The equations seem to own their inception more to commonsense than to economic theory, analysts allowing themselves much flexibility in changing explanatory variables, selecting the "best" equations from literally hundreds of computer runs. Empiricism is all, or nearly all.

The results are then discussed, usually with reference to the coefficients having the signs (+ or −) which theory ordains. One finds the coefficient estimates used to establish elasticities and the like, ignoring the writer's warning[2] that the coefficients have individual validity only in simple (i.e. two variable) regression. The writer holds that the only use of systems is for forecasting, either straight or hypothetical, the latter to answer the question "if we alter the control variables, what will be the effect on the economy?" Yes, the writer has felt for some time that model making in the social sciences is facing something of a crisis, if this is not too strong a word. He would not be surprised, from past experiences in other fields, to find that such an opinion is "in the air". The time for calm reappraisal (it need not be "agonizing") has come.

A few general speculations as to why things go wrong. Modern statistical science hypothecates almost laboratory conditions of control of experiment. These conditions can be satisfied literally in the laboratory (the writer recalls[3] applying Robert Boyle's original 17th century data to proving statistically the truth of his law PV=constant) in field experiments such as establishing, by reference to a probability scale, the superiority of one fertiliser over another, and in random sampling social surveys, whereby confidence limits can be established for some measurement which it would be impracticable to ascertain by complete inquiry. For this kind of work statistical science is entirely adequate, indeed essential. The errors are "under control", as the saying is, the term "errors" being understood as in the ordinary sense, mainly errors of measurement, though in field experimentation there is a further hypothetical, but plausible extension of the idea. The point is that we can organize the experiment in advance, so that errors of estimate will be within prescribed limits. For the purposes of econometrics, statistical theory had to be developed and, in general, these developments have been mathematically elegant and extensive in strictly algebraic terms. But (as happened before, long years ago, and will happen again) the Scarlet Woman (as mathematics have been termed elsewhere[4]) has largely taken over, with all her charm and seduction.

2. R. C. Geary, Some Remarks about Relations between Stochastic Variables: a Discussion Document, *Review of the International Statistical Institute,* 1963.

3. R. C. Geary, Boyle's Law as an Inherent Relation between Observations subject to Sampling Errors, Letter to *Nature,* 1943.

4. Commentary, by R. C. Geary, in *Europe's Future in Figures,* North Holland Publishing Company, 1962.

Much of the trouble lies in the error term of our economic models, i.e. sets of equations. If the analyst writes down one equation of the set as $y=f(x)+u$, y being an endo-variable and f some function of other endogenous and predetermined variables, he is making a statement about how the economic system works. To the extent that he is wrong (in this and other equations of the system) in his form of equation, the error term u and its fellows in other equations have to bear the brunt of his errors. Of course, as stated above, we have some controls such as r (the coefficient of correlation between actual and estimate), s (measuring the magnitude of error of estimate of the endo-variable) and the DW test for establishing the non-autoregression of residuals; but these are not enough. The blunt truth is that we are far from knowing in measurable terms how the social system works (i.e. our equations are descriptive and empirical, but not functional) and we are asking too much of our error terms. There is also, of course, the trouble that, even if the form of our equations is theoretically correct, the coefficients may be changing: this aspect also requires investigation.

The computer, always improving in capacity and general efficiency, can be a snare. One notices an inevitable tendency of researchers, having programmed even large systems, to run their data through the machines in all kinds of combinations of variables and equations with a view to selecting "the best", i.e. by reference to r and DW. The writer must not be too censorious, for he will do it himself, as we shall see in what follows; but he will not make the mistake of thinking that the traditional probabilistic tests apply to his "best", in particular the F-test of equation significance.

2. *A STUDY OF IRISH EDUCATIONAL DATA*

This section is intended less as a serious analysis of Irish education (on which the writer has no pretensions to expertise) than as a statistical exercise. Statistics here are the end and not, as they usually are (and normally indeed should be) a means towards the end. It is hoped, however, that the raw data displayed in Appendix Tables A1–A4, will be found useful to educationists, whatever view be taken of the 50 regression equations given in Table 1. Table A1 consists of ancillary data used as independent variables in some of the equations. Attention is directed to the notes to the tables.

The 1947–1966 *Equations:* The 50 equations purport to explain the behaviour during the years 1947–1966 of 20 variables, the plan being to provide 4 equations each for the variables S (pupils), T (teachers), P (pupil-teacher ratio), N (teachers in training), for each type of education, subscripts indicating type of education (1 primary, 2 secondary, 3 vocational, 4 university). There are 6 regressions for G (government expenditure on education and one for each (C cost per pupil).

The general approach was to set up for each dependent variable a theory of explanation and then to modify it by changing the explanatory (or independent)

TABLE 1: *Regression Equations, Ireland*, 1947-1966

Notation

Subscript to indicate type of education: 1 primary, secondary, 3 vocational, 4 university.
Superscript indicates time-lag; blank means current.
All money values are at constant (1958) prices, found by deflating current price values by the implicit GNP price index, unless otherwise available.
Student-Fisher t values in brackets under coefficient value. Coefficients deemed insignificant when $t < 1$.

Dependent variables

F = Private expenditure on education
G = Government expenditure on education
H = Total expenditure on education (government and private)
S = Enrolment number
T = Teachers number
P = S/T = Pupil/Teacher ratio
C = H/S = Cost per pupil
J = F/S = Cost to persons per pupil
N = Teachers in training number

Independent variables

E = Total Government expenditure on goods and services
B = Births number
Y = GNP
M = Industrial output index
D = Teachers' salaries total
Q = Population number
F = Private expenditure on education
J = F/S = Private expenditure per pupil
t = Time in annual units.

Number	*Equation (error term omitted)*	*r*	*s.e.*	*DW*	*Other variables included in regression but with insignificant coefficients*
1	$S_1 = 261{\cdot}73 + \underset{(11{\cdot}22)}{1{\cdot}7931t} + \underset{(1{\cdot}03)}{0{\cdot}3788B^{-6}} + \underset{(1{\cdot}46)}{0{\cdot}5184B^{-10}} + \underset{(1{\cdot}65)}{0{\cdot}5811B^{-11}} + \underset{(1{\cdot}60)}{0{\cdot}5809B^{-12}} + \underset{(2{\cdot}56)}{0{\cdot}7842B^{-13}}$	·994	2·78	1·50	B^{-7}, B^{-8}, B^{-9}

Table 1 continued

Number	Equation (error term omitted)	r	s.e.	DW	Other variables included in regression but with insignificant coefficients
2	$S_1 = 255{\cdot}54 + 1{\cdot}7759B^{-13}$ $(1{\cdot}72)$	·914	9·76	n.c.	B^{-6}, B^{-7}, B^{-8}, B^{-9}, B^{-10}, B^{-11}, B^{-12}
3	$S_1 = 267{\cdot}42 + 2{\cdot}2426t + 0{\cdot}3667 (\sum_{i=6}^{13} B^{-i})$ $(13{\cdot}37) \quad (7{\cdot}99)$	·979	3.98	0·56†	None
4	$S_1 = 169{\cdot}85 + 0{\cdot}6083 \sum_{i=6}^{13} B^{-i}$ $(4{\cdot}37)$	·718	13·11	n.c.	None
5	$S_2 = 70{\cdot}45 + 2{\cdot}4757t - 1{\cdot}7090J_2$ $(18{\cdot}04) \quad (2{\cdot}12)$	·988	1·67	0·62†	None
6*	$S_2 = -80{\cdot}83 + 6{\cdot}8507t$ $(1{\cdot}99)$	·497	8·86	n.c.	None
7	$S_2 = -3{\cdot}68 + 1{\cdot}8510t + 0{\cdot}0796Y$ $(4{\cdot}28) \quad (1{\cdot}37)$	·990	2·84	0·34†	Y^{-1}, Y^{-2}
8	$S_2 = 3{\cdot}40 + 2{\cdot}0738t + 0{\cdot}1128Y^{-1}$ $(5{\cdot}03) \quad (2{\cdot}08)$	·988	2·92	n.c.	Y^{-2}
9	$S_3 = 5{\cdot}32 + 0{\cdot}6290t + 0{\cdot}1396M$ $(3{\cdot}91) \quad (2{\cdot}07)$	·992	0·97	0·81†	M^{-1}, M_3^{-2}
10	$S_3 = 0{\cdot}82 + 0{\cdot}2771M - 0{\cdot}2045M$ $(3{\cdot}50) \quad (1{\cdot}45)$ $+ 0{\cdot}1599M^{-1}$ $(1{\cdot}99)$	·984	1·34	n.c.	None
11	$S_3 = -0{\cdot}26 + 0{\cdot}0922M + 0{\cdot}6303S_3^{-2}$ $(2{\cdot}27) \quad (1{\cdot}79)$	·994	0·81	·078	t, M^{-1}
12	$S_3 = -1{\cdot}21 + 0{\cdot}0938M + 0{\cdot}7417S_3^{-2}$ $(2{\cdot}39) \quad (5{\cdot}54)$	·994	0·79	n.c.	M^{-1}
13	$S_4 = 15{\cdot}52 + 0{\cdot}7065t - 0{\cdot}1839J_4$ $(51{\cdot}29) \quad (13{\cdot}09)$	·998	0·20	2·21	None
14*	$S_4 = 9{\cdot}39$	·047	2·94	n.c.	J_4
15	$S_4 = -5{\cdot}91 - 0{\cdot}2990t + 0{\cdot}2530S_2^{-3}$ $(5{\cdot}61) \quad (1{\cdot}34)$ $+ 0{\cdot}2276S_2^{-4}$ $(1{\cdot}67)$	·998	0·19	0·73	S_2^{-1}, S_2^{-2}
16	$S_4 = -0{\cdot}11$	·788	1·96	n.c.	S_2^{-1}, S_2^{-2}, S_2^{-3}, S_2^{-4}
17	$T_1 = 6{\cdot}08 + 0{\cdot}0460t + 0{\cdot}4772T_1^{-4}$ $(2{\cdot}23) \quad (5{\cdot}96)$	·996	0·07	2·04	N_1^{-1}, N_1^{-2}, N_1^{-3}

Table 1 continued

Number	*Equation (error term omitted)*	*r*	*s.e.*	*DW*	*Other variables included in regression but with insignificant coefficients*
18	$T_1 = 4{\cdot}35+0{\cdot}0012N_1^{-1}+0{\cdot}0011N_1^{-3}$ $(2{\cdot}31)$ $(2{\cdot}61)$ $+0{\cdot}5794T_1^{-4}$ $(7{\cdot}85)$	·994	0·08	n.c.	N^{-2}
19	$T_1 = 15{\cdot}76+0{\cdot}1272t-0{\cdot}0076S_1$ $(13{\cdot}91)$ $(2{\cdot}57)$	·988	0·11	n.c.	None
20	$T_1 = -0{\cdot}34+0{\cdot}0291S_1$ $(6{\cdot}45)$	·835	0·36	0·16†	None
21	$P_1 = 3{\cdot}26-0{\cdot}0265t+0{\cdot}9432P_1^{-1}$ $(1{\cdot}56)$ $(3{\cdot}80)$	·912	0·40	1·94	P_1^{-2}
22*	$P_1 = 35{\cdot}94-0{\cdot}0668t$ $(2{\cdot}07)$	·439	0·83	n.c.	None
23	$P_1 = 1{\cdot}59+1{\cdot}0609P_1^{-1}$ $(4{\cdot}47)$	·897	0·42	n.c.	P_1^{-2}
24	$P_1 = 0{\cdot}68+0{\cdot}9785P_1^{-1}$ $(8{\cdot}35)$	·896	0·41	n.c.	None
25	$P_2 = 7{\cdot}70+0{\cdot}1066t+0{\cdot}4428P_2^{-1}$ $(2{\cdot}38)$ $(1{\cdot}77)$	·977	0·23	2·01	P_2^{-2}
26	$P_2 = 11{\cdot}58+0{\cdot}1613t$ $(17{\cdot}47)$	·972	0·24	n.c.	None
27	$P_2 = 0{\cdot}52+0{\cdot}7682P_2^{-1}$ $(3{\cdot}24)$	·968	0·26	n.c.	P_2^{-2}
28	$P_2 = 0{\cdot}59+0{\cdot}9653P_2^{-1}$ $(16{\cdot}02)$	·967	0·26	n.c.	None
29	$P_3 = 3{\cdot}15+1{\cdot}0337P_3^{-1}-0{\cdot}2841P_3^{-2}$ $(4{\cdot}60)$ $(1{\cdot}29)$	·926	0·35	n.c.	t
30	$P_3 = 11{\cdot}59+0{\cdot}0993t$ $(3{\cdot}94)$	·681	0·65	n.c.	None
31	$P_3 = 2{\cdot}84+1{\cdot}0364P_3^{-1}-0{\cdot}2557P_3^{-2}$ $(4{\cdot}74)$ $(1{\cdot}30)$	·926	·035	2·30	None
32	$P_3 = 3{\cdot}01+0{\cdot}7702P_3^{-1}$ $(9{\cdot}83)$	·918	0·35	n.c.	None
33*	$P_4 = 7{\cdot}97+0{\cdot}5794P_4^{-1}$ $(2{\cdot}30)$	·519	0·48	1·95	P_4^{-2}, t
34*	$P_4 = 13{\cdot}01$	·154	0·52	n.c.	t
35*	$P_4 = 7{\cdot}59+0{\cdot}5887P_4^{-1}$ $(2{\cdot}42)$	·516	·046	n.c.	P_4^{-2}
36*	$P_4 = 6{\cdot}43+0{\cdot}5012P_4^{-1}$ $(2{\cdot}40)$	·493	0·46	n.c.	None

Table 1 continued

Number	*Equation (error term omitted)*	*r*	*s.e.*	*DW*	*Other variables included in regression but with insiinificant coefficients*
37	$N_1 = 561{\cdot}60+33{\cdot}58611t$ $(11{\cdot}00)$	·975	47·52	0·83†	$(D_1/T_1)^{-1}$, $(D_1/T_1)^{-2}$
38	$N_1 = 528{\cdot}22+32{\cdot}8222t$ $(11{\cdot}70)$	·974	46·79	n.c.	$(D_1T/_1)^{-1}$
39	$N_1 = -663{\cdot}84+1{\cdot}5948(D_1/T_1)^{-1}$ $(1{\cdot}57)$ $+1{\cdot}1198(D_1/T_1)^{-2}$ $(1{\cdot}23)$	·762	134·93	n.c.	None
40	$N_1 = -634{\cdot}13+2{\cdot}6363(D_1/T_1)^{-1}$ $(4{\cdot}63)$	·737	136·85	n.c.	None
41	$C_1 = -5{\cdot}82+0{\cdot}03161D_1/T_1+0{\cdot}0407Y/Q$ $(8{\cdot}52) \quad (6{\cdot}82)$	·982	·056	1·80	None
42*	$C_2 = 46{\cdot}93+0{\cdot}0550D_2/T_2$ $(2{\cdot}80)$ $-0{\cdot}0471Y/Q$ $(1{\cdot}53)$	·665	1·94	n.c.	None
43*	$C_3 = 8{\cdot}87+0{\cdot}0628D_3/T_3$ $(1{\cdot}37)$	·412	5·49	n.c.	Y/Q
44*	$C_4 = -60{\cdot}75+0{\cdot}3023Y/Q$ $(1{\cdot}41)$	·687	14·71	1·76	D_4/T_4
45	$G = 1{\cdot}80+0{\cdot}3085t+0{\cdot}1600E^{-1}$ $(4{\cdot}26) \quad (1{\cdot}77)$	·986	0·56	2·31	E, F^{-2}
46	$G = -12{\cdot}56+0{\cdot}1834E+{\cdot}01543E^{-1}$ $(2{\cdot}03) \quad (1{\cdot}18)$	·968	0·81	n.c.	E^{-2}
47	$G = -12{\cdot}12+0{\cdot}1826E+0{\cdot}2050E^{-1}$ $(2{\cdot}07) \quad (2{\cdot}18)$	·967	0·80	n.c.	None
48	$G = 2{\cdot}33+0{\cdot}3187t+0{\cdot}1388E$ $(4{\cdot}83) \quad (2{\cdot}75)$	·983	0·58	n.c.	None
49	$G = 2{\cdot}86+0{\cdot}3319t+0{\cdot}2030E^{-1}$ $(5{\cdot}19) \quad (3{\cdot}00)$ $-0{\cdot}0724E^{-2}$ $(1{\cdot}00)$	·985	0·55	2·33	None
50	$G = -10{\cdot}95+0{\cdot}3681E$ $(14{\cdot}18)$	·958	0·87	n.c.	None

*Equation not significant at ·95 probability level. †Significant of residual auto-correlation at ·95 probability level.

variables: the theory will in each case be clear from the list of notation. The number of independent variables ranged from 1 to 9. Many types of regressions were essayed, including "straight" (i.e. using current independents only), lagged independents, lagged dependents. About half the equations contain time t as a specific independent, the writer preferring this treatment to the "Δ approach" As might be expected, in general inclusion of specific t greatly improved the regression, as the r values indicate. Only the coefficients deemed "significant" (on a very liberal interpretation) are shown: variables included in the regression with coefficients with t-value less than unity are indicated in the last column. The approach is frankly experimental and empirical: classical stochastic considerations are in abeyance.

Forty of the equations are significant at the .95 probability level, most of them highly so (as will appear later), by the F-test. The exceptions are nos. 6, 14, 22, 33–36, 42–44. The DW's had to be calculated on a desk machine, and the computer to which the writer had access did not provide for this statistic. Consequently, only 20 out of 50 were calculated; and of these 7 indicated residual serial correlation. As will presently appear, however, there is a much simpler statistic available for assessing residual randomness which will serve for most purposes with the superlative advantage of being calculable in a period of seconds.

Having regard to the main purpose of this paper, namely statistical methodology, comments are meagre in the text on the showing of Table 1, implying that the writer expects the reader to draw his own inferences. The fact of the possession of the raw data (as in Tables A1–A4) is immensely more important than any analysis thereof, however sophisticated. The writer is probably in a small minority amongst statisticians in his assertion that, not only should simple analysis be included as part of econometrics, but should be regarded as the more important part. By simple analysis is meant graphics, percentages, scatter diagrams and the like. When the writer was young, statisticians generally accepted the view that "analysis is like extracting the juice from an orange; 90 per cent can be produced by squeezing with your hands; you have to use a machine (i.e. sophisticated methods) to get the last 10 per cent." Despite the great development of statistical science in the intervening years, there remains, in the writer's view, a large measure of truth in the simile. Simple methods carry conviction with decision-makers, who are rarely economists, statisticians or mathematicians. The really useful truths (i.e. useful from the welfare point of view) are usually simple truths. And it is a good discipline to try to simplify and rationalize findings if they have come to light through esoteric methods. So the writer has a growing respect for what are termed "naive" methods, amply exploited in what follows.

The Objective Showing of Table 1: Despite the apparent derogation of Table 1 in this paper, there can be no question about its objective utility, even if the showing is negative. The "derogation" is from a technical point of view, of little interest to educationists who will note, for example, that equations nos. 1–4 attempt to explain numbers of primary schoolchildren (S_1) by births 6–13 years before. Of these no. 1 is far the best, though clearly there was present a strong secular trend (indicated by large coefficient of t): more children are going to school now than formerly.

The best of the $4S_2$ (secondary school pupils) is no. 7. Unfortunately the DW is significantly low, showing that a full explanation of the vagaries of S_2 has not been found. The equations show, however, that, even apart from the rise in income (Y) there is a strong secular trend upwards. The $4S_3$ (vocational school pupils) equations are very good r-wise apart from too low DW's. It seems almost certain that increased industrialization has been a major cause of the rise in S_3.

The S_4 (university students) series (nos. 13–16) is very interesting. While trend is of overwhelming importance (no. 13), when trend is allowed for, private expenditure per pupil in real terms (J_4) becomes important negatively; the lower the cost the greater the number of students, and vice versa, apart from the powerful secular trend upwards; yet J_4 by itself has no influence (no. 14)! No. 15 shows that, especially when (negative) allowance is made for trend, numbers in the university are influenced by number of secondary students in the previous years. No. 16 is an instance of the phenomenon to which attention has recently been directed,[5] of equation significance with all coefficients insignificant.

No. 17 is the most satisfactory of the $4T_1$ (primary teachers) equations—note high r and DW near 2. It marks a consistency with teachers in training.

The best behaved of the P series are the four equations for P_2 (secondary pupil-teacher ratio). No relation was ascertainable for P_4. Apart from trend t, all the causative variables were lagged dependents.

The attempt to explain N_1 (number of student teachers, primary of course) by average real salary of primary teachers was rather unsuccessful, as nos. 37–40 show. Clearly t takes up a large part of the variation (nos. 37–38, almost identical for obvious reasons) but the DW indicates that the relation is incomplete.

Only one of the C (real cost per pupil) series is significant, namely C_1, the equation for which yielding a high r (=.98) and a DW near 2. This equation says that C_1 is explained partly by average teacher's salary and partly by average GNP. Both have highly significant coefficients.

The theory involved in the 6 equations relating to G (government expenditure on education) is that this expenditure depends largely on total government expenditure. Clearly trend t is an important element here also, indicating expenditure on education constantly increasing at a steeper gradient than general expenditure.

Statistical Appraisal: The regressions shown in Table 1, relating to the 20 years 1947–1967, were also produced for the 11 years 1947–1957, the main object being to compare the estimates of the dependent variables from the two sources during the years in common 1947–1957, inclusive. Another object was to compare the "forecasts" for 1958 and 1959, derived from the 1947–1957, with actuals for these

5. R. C. Geary and C. E. V. Leser: Significance Tests in Multiple Regression, *The American Statistician*, February 1968.

years. The appraisal that follows leans much more heavily on the latter than on the longer series, mainly because of interest in forecasting potential. In fact, the so-called "actuals" in 1967 are either provisional and subject to amendment or were not available when the paper was written. No such qualifications apply to 1958 and 1959.

Naive versus Calculated Forecasts: The naive forecasts were derived for each of the dependent variables separately by speculative extrapolation, having regard only to the trend in that variable alone during the previous 3–5 years. No calculations were involved. The "calculated" values are those derived from the two series of regression formulae. For purposes of comparison, "calculated" was accorded a score of 1 if nearer than naive in absolute value to actual, 0 if further away, and 1/2 if equidistant (to smallest digit shown). Calculated values were produced only when the equation was significant. Mainly for this reason the number of "forecasts" for each year is less than 50, the number of equations in Table 1. See Table 2. Since the comparisons rather sensationally favour the naive forecasts, the writer wishes to give an assurance that in making these "forecasts" the "actuals" were carefully concealed. The game is fairly played.

Scores are:—15 (37) for 1958, 17 (37) for 1959, 10 (30) for 1967, bracketed figures indicating number of comparisons. It will be noted that the results are consistent: in all three regression results are better than the naive in less than half the cases. As between the various entities one notices that regression scores consistently well only for P_3. There is a fair measure of scoring consistency between 1958 and 1959: on average score, 12 units, $r=.72$. There is, however, no appreciable consistency between average scores in 1959 and 1967.

Consistency of Calculated Values with The Regression Periods: We have two sets of estimates for the 11 years 1947–1957, one based on the whole regression period 1947–1966, the other on the earlier period 1947–1957. The computer sheets have the regression estimates for each year for each entity juxtaposed with the actual. The residuals, actual *minus* estimated, are also printed. We regard the two series of estimates as consistent if they are similar. How are we to measure such consistency in a succinct way?

We have, of course, the regression coefficients for each entity and some people, we think misguidedly, would like to have the two sets of coefficients systematically compared. Apart from this procedure's involving a highly complex table, difficult to interpret, we regard attention to individual coefficients in time series analysis as waste of time. When one's independents are time series, inevitably highly intercorrelated, it is a matter of chance which variable or variables take over the burden of relationship to the dependent, coefficient-wise. For instance, as the present study amply shows, omission of insignificant independents can have a substantial effect on the coefficients of the variables remaining, but a negligible effect on the dependent estimates. The latter are far more stable, in multivariable regression, than the elements of the coefficient sector. In multivariable regression we should concern ourselves with dependent variable estimation, and little else.

Probably the best way to determine the stability of the estimator would be to compare sum squares difference between the two estimates with sum squares residuals 1947–1957 given on the computer sheets. If the ratio of the former to the latter is small we infer consistency. But how small? Having regard to the fact that the residuals themselves have only (T-k) degrees of freedom, the mathematization of the test would appear a formidable task indeed. For our purpose a simple sign test of residuals will suffice, while we recognize that it is less efficient (i.e. less sensitive) than full treatment. Table 3 shows the sign score (maximum 11) of residuals for all usable equations, i.e. if the residual signs (+ or −) in a given year are the same, the score is 1, otherwise 0.

TABLE 3: *Sign Score from Comparison of Signs of Residuals, 1947–1957*

Eq. no.	*Var.*	*Score*	*Eq. no.*	*Var.*	*Score*	*Eq. no.*	*Var.*	*Score*
3	S_1	9	21	P_1	7	34	P_4	8
4	S_1	7	22	P_1	6	37	N_1	11
7	S_2	5	23	P_1	10	38	N_1	10
8	S_2	5	24	P_1	10	39	N_1	8
9	S_3	5	25	P_2	8	40	N_1	10
10	S_3	5	26	P_2	10	41	C_1	10
11	S_3	10	27	P_2	11	45	G	11
12	S_3	9	28	P_2	11	46	G	10
15	S_4	7	29	P_3	11	47	G	10
17	T_1	9	30	P_3	9	48	G	10
18	T_1	10	31	P_3	10	49	G	9
19	T_1	9	32	P_3	11	50	G	10
20	T_1	10						

By this test the two sets of equations may be regarded as consistent. Out of 37, no fewer than 6 are absolutely so (i.e. score 11), while a further 14 have only one aberrant case (i.e. score 10). The full frequency distribution is given in Table 4.

TABLE 4: *Frequency Distribution of Sign Scores*

Score	*Frequency*
5	4
6	1
7	3
8	3
9	6
10	14
11	6
	37

If one wants a test of significance one can have recourse to a coin-tossing analogue with $p=1/2$, possibility of successes being 0 to 11. The (two-sided) random probability of 0 or 1 failures/successes is $2\ (1+11)/2^{11}=12/1024=.0117$. We actually find .54 (=20/37)!

But has this kind of consistency any relation to forecasting efficiency as adjudged by calculated *v.* naive score? Table 5 is derived from Tables 2 and 3.

TABLE 5: *Relationship between Consistency and Efficiency*

	1958		1959		1967	
Items	A	B	A	B	A	B
Highly consistent (Score 10 or 11)	11	20	12½	20	5½	14
Others	4	17	4½	17	4½	16
Total	15	37	17	37	10	30

A: Score on calculated *v.* naive
B: Total cases

It is quite clear that in 1958 and 1959 the highly consistent items had proportionately the higher score on the calculated *v.* naive test (e.g. 1958, 11/20 compared with 4/17.) The 1967 comparison, which in the same direction, is less decisive. This result qualitatively is to be expected since 1958 and 1959 are within one of the regressions while 1967 is outside them.

Standard Error of Forecast: Table 6 sets the problem. The table shows the number of cases in which the actual values lay within the range $\pm$ 2 s of the calculated

TABLE 6: *Number of Actual less Calculated (in Absolute Value) less than estimated Twice Equation Standard Error*, 1958, 1959, 1967

	1958		1959		1967	
Item	*No. of eqs.*	*Act.-cal.* $< 2s$	*No. of eqs.*	*Act.-cal.* $< 2s$	*No. of eqs.*	*Act.-cal.* $< 2s$
S_1	2	1	2	1	4	1
S_2	2	2	2	1	2	2
S_3	4	0	4	2	4	3
S_4	1	0	1	0	1	0
T_1	4	1	4	1	4	4
P_1	4	3	4	3	3	3
P_2	4	4	4	4	4	2
P_3	4	4	4	4	4	4
P_4	1	1	1	1	—	—
N_1	4	3	4	3	4	4
C_1	1	1	1	1	—	—
G	6	6	6	6	—	—
Total	37	26	37	27	30	23

TABLE 2: *Comparison of Calculated and Naive Forecasts*, 1958, 1959, 1967

		Regression 1947–1959								*Regression 1947–1966*			
Eq. No.	*Vari-able*	1958				1967				1967			
		Act.	*Calcd.*	*Naive*	*Score*	*Act.*	*Calcd.*	*Naive*	*Score*	*Act.*	*Calcd.*	*Naive*	*Score*
1	S_1	490·7	—	491·0	—	492·3	—	494·0	—	496·5	489·8	496	0
2	S_1	490·7	—	491·0	—	492·3	—	494·0	—	496·5	474·2	496	0
3	S_1	490·7	492·3	491·0	0	492·3	496·8	494·0	0	496·5	492·8	496	0
4	S_1	490·7	479·0	491·0	0	492·3	476·4	494·0	0	496·5	465·7	496	0
7	S_2	66·2	66·6	65·4	1	69·6	68·4	68·4	1/2	103·6	99	104	0
8	S_2	66·2	65·3	65·4	0	69·6	69·3	68·4	1	103·6	98	104	0
9	S_3	23·8	25·0	22·9	0	24·6	24·6	23·5	1	40·5	38·5	40·0	0
10	S_3	23·8	22·1	22·9	0	24·6	21·7	23·5	0	40·5	42·0	40·0	0
11	S_3	23·8	22·5	22·9	0	24·6	23·5	23·5	1/2	40·5	40·6	40·0	1
12	S_3	23·8	21·6	22·9	0	24·6	22·6	23·5	0	40·5	39·5	40·0	0
15	S_4	9·3	8·9	9·2	0	10·0	9·1	9·4	0	18·0	17·5	17·7	0
17	T_1	13·6	14·1	13·5	0	13·8	14·3	13·6	0	14·7	14·7	14·8	1
18	T_1	13·6	13·6	13·5	0	13·8	13·8	13·6	1	14·7	14·8	14·8	1/2
19	T_1	13·6	13·2	13·5	0	13·8	13·2	13·6	0	14·7	14·7	14·8	1
20	T_1	13·6	13·4	13·5	0	13·8	13·4	13·6	0	14·7	14·1	14·8	0
21	P_1	36·2	37·6	36·4	0	35·8	37·8	36·5	0	33·8	33·5	33·5	1/2
22	P_1	36·2	36·8	36·4	0	35·8	36·9	36·5	0	33·8	—	33·5	—
23	P_1	36·2	36·6	36·4	0	35·8	36·4	36·5	1	33·8	33·4	33·5	0
24	P_1	36·2	36·4	36·4	1/2	35·8	36·2	36·5	1	33·8	33·4	33·5	0
25	P_2	13·4	13·3	13·2	1	13·8	13·4	13·3	1	14·2	14·8	14·5	0
26	P_2	13·4	13·3	13·2	1	13·8	13·5	13·3	1	14·2	15·0	14·5	0
27	P_2	13·4	13·2	13·2	1/2	13·8	13·4	13·3	1	14·2	14·6	14·5	0
28	P_2	13·4	13·2	13·2	1/2	13·8	13·4	13·3	1	14·2	14·6	14·5	0
29	P_3	13·0	13·1	12·6	1	13·1	13·3	12·7	1	13·4	12·9	12·9	1/2
30	P_3	13·0	13·5	12·6	0	13·1	13·7	12·7	0	13·4	13·7	12·9	1
31	P_3	13·0	12·8	12·6	1	13·1	13·0	12·7	1	13·4	12·8	12·9	0
32	P_3	13·0	12·8	12·6	1	13·1	13·0	12·7	1	13·4	12·9	12·9	1/2
34	P_4	12·1	12·0	11·8	1	12·9	11·9	11·8	1	—	—	—	—
37	N_1	942	916	995	1	950	959	1055	1	1190	1188	1240	1
38	N_1	942	889	995	1/2	950	943	1055	1	1190	1184	1240	1
39	N_1	942	788	995	0	950	805	1055	0	1190	1177	1240	1
40	N_1	942	804	995	0	950	759	1055	0	1190	1199	1240	1
41	C_1	21·1	21·3	22·6	1	21·7	21·6	22·6	1	..			—
45	G	16·2	16·6	17·0	1	16·5	17·4	17·0	0	..			—
46	G	16·2	14·9	17·0	0	16·5	15·1	17·0	0	..			—
47	G	16·2	15·0	17·0	0	16·5	15·2	17·0	0	..n.a.		n.c.	—
48	G	16·2	16·8	17·0	1	16·5	17·4	17·0	0	..			—
49	G	16·2	16·8	17·0	1	16·5	17·7	17·0	0	..			—
50	G	16·2	14·8	17·0	0	16·5	15·2	17·0	0	..			
Total score:					17/37				17/37				10/30

forecasts. If the models were correct, the number of observations T infinite and the residual error term normally distributed, then $s=\sigma$ and approximately 5 % of the actuals would be outside the calculated forecasts or, on the basis of total cases 37, 37 and 30, about 2, 2, 1–2 respectively. The table shows that outsiders were much more numerous, 11, 10 and 7 respectively, not inconsistent one with another.

The main reason for the aberration is that T is a small number. If our (single equation) model in matrix form is

$$Y=X\beta+u,$$

where Y and u are $T\times 1$, X $T\times k$, β $k\times 1$ and u is homoskedastic, variance σ^2, mean zero and non-autoregressed. This model is assumed valid in the forecast period (e.g. 1958 and 1959), i.e. the estimated vector b of β is based on the data matrix (e.g. for 1947–1957). Let the forecast independent vector be ξ (transpose ξ'). Then the variance of the difference $(Y-Y_c)$ between what will transpire Y and Y_c, the regression estimate, is

$$E\,(Y-Y_c)^2=\sigma^2\,\lambda^2$$

where[6]

$$\lambda^2=1+\xi'(X'X)^{-1}\xi \tag{5}$$

Then the Student-Fisher $t=(Y-Y\;)/\lambda s$, with d.f. (of s)$=T-k$, where k is the number of coefficients (including constant).

As an example, consider regressions involving time t only. In the 1947–1957 regression $T=11$. Hence d.f. is 9. It is easily calculated that, for the 1958 forecast, $\lambda^2=1.3273$, so that $\lambda=1.1521$. The .95 null-hypothesis probability t-value for 9 d.f. is 2.262. Hence the .95 confidence range for $(y-y_c)/s$ is $\pm$ 2.61 $(=1.1521\times 2.262)$ instead of 2 assumed in Table 6. The results shown are not, in general, inconsistent with theory.

There would be no theoretical difficulty about calculating λ for each equation for each of the three regression forecasts, namely those for 1958, 1959 and 1967. We consider this unnecessary. We have been concerned merely to establish the point that, even when the model applies to the periods forecasted, and the time series short as they inevitably tend to be, the error multiplier λ may be substantially greater than unity and the t range greater than for normal theory. Given the estimate of the error variance σ^2, the forecasts are correspondingly the more imprecise. The cardinal need of ensuring that s^2 is small has added force.

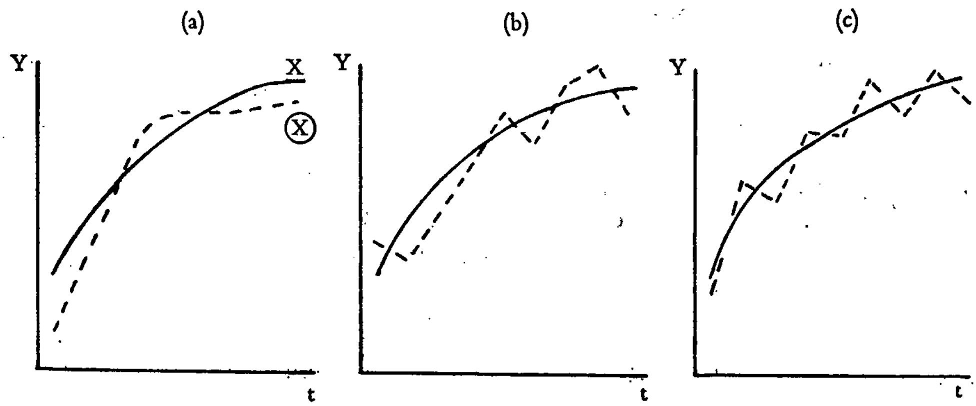
It might now be argued that the calculated *v.* naive test, dealt with earlier, falls to the ground since the calculated estimates are unfirm, in the sense that all that

6. It is worth remarking that the variance for any vector ξ included in the regression is $[1-\xi'(X'X)^{-1}\xi]\sigma^2$. In neither case is it σ^2.

can be said of them is that (corresponding to a working probability, .95, .99 etc.) they lie between rather wide confidence limits, which can be estimated. We could logically have taken any figure between the limits. We do not accept this argument. Both calculated and naive forecasts were based on the same body of data. If the statistician defends his regression estimates in the foregoing manner, the naive estimator could counter that *his* figures are also imprecise, if he cannot define his limits of error with the confidence of his rival.

Some Remarks on the Durbin-Watson Test (DW)[7]*:* All the DWs were calculated for the significant 1947–1957 equations, only some for the longer 1947–1966 series. Interest naturally centres on the low DWs. Lowest values in ascending order (figures in brackets indicating equation no.) for the 1947–1957 series are:— 0.41 (30), 05.6 (4), 0.74 (3), 0.74 (39), 0.78 (40), 0.81 (37), 0.82 (38). These are the only DWs less than unity and may be regarded as the only equations with significant residual autoregression, recalling that the time period is only 11 years. As there are 37 equations in all with only 7 aberrant DW-wise, this phenomenon is not major in the present instance. Nevertheless the question rises: what do we do about equations with low DWs and satisfactorily high s, i.e. with low standard errors?

What happens in practice is best illustrated diagrammatically:—



In each case the broken line illustrates the actual value of the dependent variable, the firm line the regression calculated value. In all three cases, of course, the sum of the deviations is zero. Situation (a) represents the case of a low DW; in (b) the deviations are randomly ordered and yield a DW of about 2; in (c), idealised in

7. Testing for Serial Correlation in Least Squares Regression, *Biometrika*, 37, 38 (1950–51).

the diagram, deviations tend to alternate in sign (i.e. the sequence is $-, +, -, +, \ldots, -$) yielding a DW above 2. (It is easy to show that DW cannot exceed 4). From the short-term forecasting point of view there is nothing to be done in situation (b): for good or ill, granted our data, the forecast must be the regression estimate, *simpliciter.*

But suppose we find ourselves in something like situation (a). The regression forecast a year ahead would be at position ×. Clearly this could not be accepted, granted the relation of the curves (which we are aware of). We would try to correct somehow to attain "forecast" marked X, i.e. by a negative addition to the original regression estimate.

Many attempts were made to evolve a technique of correction in cases of low DW. The most hopeful, *ab initio*, appeared to be to fit the three first orthopolynomials (the first being t itself) to the residuals. For example with equation 3, DW=.74, the fitted t-curve, was significant. (We tried 5 orthos here, but the 4th and 5th made insignificant contributions). The F value was 14.68, in excess of the .98 (3, 5 d.f.) of 12.06. But the extrapolated corrections for 1958 and 1959 were respectively -1.4 and -6.4.

Having acquired some respect for his own perspicacity as a "naive" forecaster, the writer tried his luck with his naive forecasts of the residuals for comparison with those obtained by orthopolynomial extrapolation. Comparing each with actuals in the 7 cases of low DW in the two years, in 6 cases the naive was superior to the calculated. The latter method according is *not* recommended. But neither is the naive applied to the residues better than the naive applied to the original data! The score (in each case out of 7) for the original naive is 4½ in 1958 and 5 in 1959.

To conclude on a more hopeful note for those who have to labour with a desk machine to calculate their DWs: A count of successive sign changes of residuals (a matter of seconds) will often do just as well, to establish randomness. Table 7 shows the emphatic relationship between DW and number of sign changes with the 1947–1957 data.

TABLE 7: *DW and Residual Sign Changes*

No. of cases	*No. of Sign changes*	*Average DW*
4	2	0·62
3	3	1·11
6	4	1·63
10	5	2·09
5	6	2·23
6	7	2·47
2	8	3·09
1	9	3·13
37	—	—

The correlation between DW and number of sign changes, r is reasonably high, in fact .86 on our data. If T be the number of residuals ($=37$ in our example) in the null hypothesis case the frequency distribution is the point-binomial with $p=\frac{1}{2}$, exponent $(T-1)$. Probably this τ test is less sensitive than the DW, as using less information: this aspect is being examined[8]. Anyway, it is always worth trying because of its simplicity.

Do Good Regressions Yield Better Forecasts?: A "good" regression is, by definition, one with an r near unity and a DW near 2. We invented an empirical measure of relative goodness, namely the ratio of actual value of F to its .95 null-hypothesis probability value: for instance, in the 1947–1966 series, equation 25, the actual F is 109.65 whereas the .95 probability critical value (with 3, 16 d.f.) is 4.08, giving a goodness ratio of 27. The lowest value is unity, since equations with lower values were eliminated as insignificant. Forecasting efficiency of the equation was adjudged, as above, by the score (1, 1/2 or zero) in the contrast calculated *versus* naive. The DW adequacy condition was met by omiting equations with DWs less than unity.

Simple regressions between goodness and forecasting efficiency are $-.094$ for 1957, $-.217$ for 1958 (both based on 1947–1957 regressions, 30 pairs), and .154 for 1967 (based on 1947-1966 regressions, 14 pairs). None is significant. There is no relationship between forecasting ability and quality of regression, as defined.

3. CRITIQUE OF TWO OTHER MODELS

Leser's Model[9]*:* Conrad Leser is one of our most ingenious model-makers. His latest, consisting of four behaviouristic equations and one identity, yields coefficients of determination (R^2) exceeding .93 for all four endogenous variables, *prima facie* a very satisfactory result as these variables are first differences (Δ). The high R^2s are helped, it is true, by two dummy variables z and z' to take care of exceptional figures, e.g. due to the incidence of import levies in 1956. The time unit is a year and the data refer to the years 1953–65 which yield 12 sets of first differences.

The object of the present note is to test the model during the years 1947–1953 by comparison of actual and estimated values of Leser's four endogenous variables in these years: a test by rearcast. Basic data are given in Appendix B Tables B1 and B2. For algebraic convenience we change Leser's notation (op. cit., page 3) as indicated in tables, using Y and X for endogenous and exogenous variables respectively. In our notation Leser's equations (*op. cit.*, page 5) are (omitting dummies and error terms):—

8. R. C. Geary, Relative Efficiency of Count of Sign Changes for Assessing Residual Auto-regression in Least Squares Regression, *Biometrika* (in press).

9. Appendix 2 of *The Irish Economy in* 1967, Economic and Social Research Institute, *Paper No.* 39, Dublin, August 1967.

$$Y_1 = 1.200 + 0.8238X_1 + 0.5574X_2$$
$$Y_2 = 1.623 + 0.5491X_1 + 0.9214X_2 + 1.7976X_3$$
$$Y_3 = -3.980 + 0.9762Y_2$$
$$Y_4 = -32.520 + 0.6469Y_3 + 508.81X_4$$

The notation makes plain the recursive character of the model. Also, each equation is identified. The reduced form is as follows:—

$$Y_1 = 1.200 + 0.8238X_1 + 0.5574X_2$$
$$Y_2 = 1.623 + 0.5491X_1 + 0.9214X_2 + 1.7967X_3$$
$$Y_3 = 2.396 + 0.5360X_1 + 0.8995X_2 + 1.7548X_3$$
$$Y_4 = -34.070 + 0.3467X_1 + .05819X_2 + 1.1352X_3 + 508.81X_4$$

The calculated values of the Y are found by substitution of the X as given in Table B2. Calculated and actual values of the Y are given in Table B3.

The comparisons are generally disappointing, even when allowance is made for the effects on imports especially of the Korean War. An obvious difficulty in this reverse forecasting is that, as Table B1 shows, most of the data, at the Δ level, fluctuate for year to year in quite fantastic degree. ΔM is a case in point: there must be a great accumulation and decumulation of import stocks going on all the time. Rather similarly with ΔY (for all its appearance of regularity at the Y level): the values are seen to range from £7m to £58m.

To eliminate partly accidental year-to-year fluctuations we compare ΣY, calculated and actual, using Table B3:

	Calcd. £m	*Act.* £m
ΣY_1	97·7	55·2
ΣY_2	175·9	169·6
ΣY_3	147·8	144·1
ΣY_4	60·7	103·6
Total	482·1	472·5

We give the total only as a curiosity: it compares very well! By this aggregate test Y_2 (GNP less government current expenditure) and Y_3 (personal disposible income) emerge very well; not so Y_1 (imports) and Y_4 (personal expenditure).

As with the education data we compare forecasts calculated from the model with those obtained using a naive model. The naive model is as follows:—

$M = 0{\cdot}39Y$; $Y = 3{\cdot}0D$ (D=money); $C = 0{\cdot}73Y$;
$Z = 1{\cdot}15C$ (Z=personal income); $G = 0{\cdot}118Y$;
$T = 0{\cdot}055C$ (T=taxation on personal income).
$Y_1 = \Delta M$; $Y_2 = \Delta Y - \Delta G$; $Y_3 = Y^d = \Delta Z - \Delta T$; $Y_4 = \Delta C$

The coefficients were based on experience in 1953–1965. The naive model uses one exogenous variable (money=currency+current accounts) instead of Leser's four; and six coefficients instead of Leser's sixteen. *Ceteris paribus* Leser's model should therefore yield much more accurate results than the naive model.

Absolute values of deviations from actual for both models are shown in Table B4. Six year average changes compare as follows (£ million):—

	Leser	*Naive*	*Actual year-to-year*
Y_1	14·2	15·6	21·6
Y_2	14·3	16·4	28·3
Y_3	6·5	8·8	24·0
Y_4	8·6	7·9	17·3

Three out of four rearcasts are more accurate, though negligibly so compared with the far larger magnitude of actuals.

This investigation was undertaken to appraise the forecasting power of Leser's model which the writer hoped to use in conjunction with money variables in a more extended model. Rearcasting may not be fair to the model. A very general impression prevails that the structure of the Irish economy changed drastically in 1958. If this be so, the coefficients have probably changed also, even if the model is functionally sound. While sympathizing with Leser's having to use annual data for as long a period as possible, one must question the validity of his straddling two distinct periods (i) 1953–1957 and (ii) 1958–1965 with his data. For all their brevity in years it might be well to estimate and compare the coefficients for the two periods 1947–1958 and 1958–1966 using Leser's model.

Goldfeld's Model: S.M. Goldfeld[10] has produced a fairly large model for U.S.A. with 32 endogenous variables and equations (including a few identities). Most of these variables pertain to banking (such as demand deposits, time deposits, borrowing, excess reserves, four interest rates etc., distinguishing "town" and "non-town" districts). Six endo-variables were non-financial macros: GNP, durable and nondurable consumption, fixed and inventory investment and disposable income. 48 quarterly sets of observations were used from III 1950–II 1962. All the equations (except those for interest rates and non-financial items) were of the form $\Delta x=\beta x_{-1}+$ linear expression in endo- and exo-variables, the latter considerable in number. Most of the equations contained at least 10 coefficients (including 3 dummies for seasonality correction). R^2 (corrected for degrees of freedom), standard error of estimate and DW are given for each of the 21 behaviouristic equations. The complete model was solved by two stage LS. There are a great number of subgroups examined for relationships. Generally a very thorough job

10. Commercial Bank Behaviour and Economic Activity *North Holland Publishing Company,* 1966.

was done, of its kind. The author might have been wise to omit the many coefficients he found insignificant by the *t*-test, and recomputing, so reducing the number of explanatory variables.

We are here interested only in the forecasting power of the model, as distinct from economic analysis, preliminary and final, which the author gives in full measure. He also gives a table (p. 171) of short term predictions, for the two quarters following those to which his equations relate, namely III and IV 1962, for 21 variables. We prefer to examine *changes* between the quarters, as a more rigorous, but more realistic, test: standard errors of estimate (also given by the author) juxtaposed with absolute predictions tend to make the latter look better than they are. As regards the first column in the following table we need not be specific in describing the entities, or their units, granted our present objective: The signs test to which appeal is often made—"the signs are right"—are here subject to the qualification that so many of the actual are + that we must suspect a general rise in the economy (or perhaps a seasonal rise), affecting endo- and exo- variables alike.

Variable Number	*Change IV '62-III' 62* Actual (A)	Predicted (P)	*Ratio* $\|P-A\|s$	*Ratio* $\|A\|S$
1	2	3	4	5
1	+113	+ 22	2·21	2·8
2	+ 4	+ 7	0·09	0·1
3	+206	+ 82	0·93	1·5
4	+ 18	+ 49	1·29	0·8
5	+ 70	+280	2·84	0·9
6	+ 85	+146	1·28	1·8
7	+ 48	− 16	1·20	0·9
8	− 33	− 13	0·59	1·0
9	+ 48	+ 30	1·03	2·8
10	+ 16	+ 16	0·00	3·6
11	+ 76	+ 79	0·71	18·0
12	+ 45	+ 26	4·06	9·6
13	+ 56	+ 92	2·35	3·7
14	+ 41	+ 31	0·19	0·8
15	+ 11	+ 12	0·81	8·9
16	+ 67	+ 48	2·42	8·5
17	+ 31	+ 29	0·28	4·3
18	+207	+155	1·20	4·8
19	+ 7	− 1	0·89	0·8
20	− 16	− 2	0·41	0·5
21	− 21	− 1	0·85	0·9

Note: Cols. 4–5: *s*=standard error of estimate (*op. cit.*, p. 171), Author's figure of 0·823 for s.e. of var. no. 6 corrected to 0·478.

The object of column 4 is to test the credibility of the forecasts as a frequency series. If s were an unbiased estimate of the standard error of the calculated Δ forecast then, with the normality assumption, we would expect about 5% of the series to lie outside the range $\pm\ 2\ s$ or 1 out of 21. We find 5, the main reason being that, as shown in the text proper, the estimated standard error is not s but λs where $\lambda > 1$: theoretically λ is always calculable, but not by us in the present instance because we lack the data: we guess that λ may be 1·2 or 1·3. Even so, the deviations are somewhat on the high side, the value of 4·06 for variable no. 12 looking very unlikely. We recognise that the author's s.e. (our s) is not ideal for our present purpose (some are for Δ's, some for straight data) but they must serve. The trouble really is the absolute magnitude in the difference between the figures in columns 2 and 3. Can we be satisfied with ultimately finding a rise of 113 in variable 1 while a rise of 22 was anticipated? While a few of the predicted changes are very accurate, the predictions generally, as measurements, are hardly satisfactory.

Column 5 makes the point more precisely. Suppose that the changes of column 2 could be regarded as typical in magnitude—it would, of course, have been preferable to use, for analysis, averages of absolute values of changes, but the author does not furnish his raw data. A really sound short-term forecasting model should have the property that the typical changes should be many times the standard error of estimate—perhaps the multiple should be 5 or 6. Only 4 (those numbered 11, 12, 15, 16) of the 21 variables satisfy this condition.

No depreciation of the work of the two scholars cited here is to be implied. On the contrary, the writer esteems these works highly. He happens to have selected them for particularly close study in the hope that they would be useful to him in the construction of a large model on which he is at present engaged.

4. *CONCLUSION*

The writer's answer to the question in his title is "Yes" but, as a result of the present investigation, his reading, and his own work in other fields, the Yes must be a qualified one, an act of faith in future work, rather than an affirmation of achievements so far. Of course scepticism has been a feature for a long time amongst econometricians (the controversy between the Dutch—econometric—and British schools of short-term macro-forecasting some 15 years ago will be recalled) but seems to have receded in recent years. The author suggests that the time for reappraisal on an international scale is overdue. Recognition of inadequacy must precede improvement in method.

The writer's intention was to produce, on traditional lines, a model of Irish educational data, for the OECD assignment. He decided to start with single equation regression: if this proved promising, but only if (he felt), should one proceed with models of many equations. The present paper completes the first stage. The results are disappointing, so much so that there does not seem to be much point in going to the second stage with existing data. Others may be more successful; it is for this reason that basic appendix Tables A1–A4 have been provided. More ingenious researchers may be able to evolve other (and better) exogenous (or independent) series.

Yet our single equations are very typical of those supplied by other investigators at the present time: they are no worse, anyway. Our methods of appraisal may have some interest: there are, of course, others. The point is that, in most cases, such appraisal is generally conspicuous by its absence, though there are honourable exceptions; the writer has in mind statistical, as distinct from general, credibility, appeal to which is rather more common. The writer wonders how many models (single—or many—equation) would stand up to the methods of appraisal used here.

Quite recently, and therefore long after the foregoing was written, a very interesting and useful address by our erstwhile colleague, C. E. V. Leser, has been published.[11] Though we must not commit him to sharing our views, he seems to do so implicitly by citing several examples of failures of models to forecast reasonably well in the short term and does not give a single case of a model successful in this respect. He does, however, remind us of the unremitting attention towards improvement given in the Netherlands to their model by H. Theil and other Dutch scholars.

For forecasting efficiency it is not enough that "the errors are under control" (in the stochastic sense). It is of paramount importance that these errors should be small in relation to that magnitude of our estimates, and our efforts in model making must be unremittingly directed towards making them so.

It must be borne in mind that the writer's criticism in the paper bears on only one aspect, namely the short term forecasting potential of models. His criticisms do *not* apply to planning (i.e. the medium and long term aspect) for which econometric models are essential and apt to their task.

To end with some speculations as to how short-term forecasting systems in the social sciences can be improved. The writer does not consider that there is a promising future in macro-models for short-term forecasting: one must disaggregate in the hope that in the disaggregated units our established econometric theory will have greater validity in practice, be these units groups of individuals, sectors of the economy, areas etc. Disaggregation means homogenization. If our units are small enough, our systems can, we hope, be the simpler and, taking a line from the great breaks-through of science, we may hope that, in the social sciences, simplicity is an aspect of truth.

Most social scientists are not econometricians and, as Larochefoucauld almost remarked, they may be philosophical (if not a little pleased) by their econometric friends' discomfiture. Did they not say so from the beginning? The writer can confront them only with a statement of faith: the way and the light will be found through measurement and combinations of measurements. We must go on trying, the critical apparatus being an essential part of our methodology.

11. "Can Economists foretell the future?" An Inaugural Lecture. *Leeds University Press*, 1969.

APPENDIX A

TABLE A1: *Some General Data, Ireland,* 1947–1967

Year	*Population*	*Births*	GNP *Total*	GNP *Per Head*	GNP *Price*	*Industrial Output Index*
	Q	*B*	*Y*	*Y/Q*		*M*
	000	000	£m.	£	(1958=1)	(1953=100)
1947	2,974	69·0	500·8	168	0·663	62·1
1948	2,985	65·9	525·6	176	0·695	70·9
1949	2,981	64·2	552·6	185	0·708	80·9
1950	2,969	63·6	557·0	188	0·715	92·4
1951	2,961	62·9	564·3	191	0·744	93·4
1952	2,949	64·6	580·1	197	0·823	93·5
1953	2,945	62·6	598·1	203	0·877	100·0
1954	2,933	62·5	603·9	206	0·874	105·0
1955	2,909	61·6	616·0	212	0·894	107·8
1956	2,898	60·7	608·3	210	0·919	104·0
1957	2,885	61·2	611·8	212	0·949	99·0
1958	2,853	59·5	601·1	211	1·000	101·9
1959	2,846	60·2	632·7	222	1·007	110·9
1960	2,832	60·7	658·5	233	1·025	118·7
1961	2,818	59·8	688·9	244	1·052	128·8
1962	2,829	61·8	707·6	250	1·102	137·9
1963	2,849	63·2	731·1	257	1·144	146·0
1964	2,862	64·1	771·7	270	1·227	157·1
1965	2,873	63·4	786·7	274	1·284	164·6
1966	2,884	62·2	800·0	277	1·326	170·0
1967	2,895	61·3	833·0	288	1·376	(185)

Notes: Values are at constant (1958) prices. Figures for 1967 in Tables 1–4 are provisional.

TABLE A2: *Number of Pupils and Teachers, Ireland, 1947–1967*

Year	*Pupils (S)*					*Teachers (T)*					*Student Primary Teachers*
	Primary	*Secondary*	*Vocational*	*University*	*Total*	*Primary*	*Secondary*	*Vocational*	*University*	*Total*	
	000	000	000	000	000	No.	No.	No.	No.	No.	No.
1947	444·1	42·9	(13·0)	7·6	(507·6)	12,772	3,584	1,261	574	18,192	541
1948	446·1	43·8	(14·0)	7·5	(511.4)	12,612	3,671	1,270	582	18,135	637
1949	445·2	45·4	(15·0)	7·8	(513·4)	12,821	3,863	1,345	555	18,584	643
1950	449·4	47·1	(17·0)	7·9	(521·4)	12,870	3,844	1,404	595	18,713	640
1951	452·1	48·6	(18·0)	7·7	(526·4)	12,792	3,929	1,380	602	18,703	633
1952	460·8	50·2	19·0	7·9	537·9	12,888	4,043	1,444	626	19,001	655
1953	468·7	52·2	20·3	8·1	549·3	13,000	4,170	1,504	626	19,300	681
1954	472·5	54·0	20·5	8·3	555·3	13,144	4,097	1,612	621	19,474	690
1955	479·5	56·4	20·9	8·5	565·3	13,231	4,417	1,659	677	19,984	755
1956	486·6	59·3	21·3	8·6	575·8	13,262	4,564	1,725	726	20,277	860
1957	488·2	62·4	22·5	8·9	582·0	13,402	4,739	1,767	745	20·653	935
1958	490·7	66·2	23·8	9·3	590·0	13,554	4,957	1,817	759	21,087	942
1959	492·3	69·6	24·6	10·0	596·5	13,753	5,032	1,881	773	21,439	950
1960	491·9	73·4	26·3	10·6	602·2	13,866	5,178	1,968	826	21,838	s46
1961	490·0	76·8	27·2	11·4	605·4	14,032	5,282	2,051	884	22,249	986
1962	484·6	80·4	28·3	12·4	605·7	14,091	5,630	2,160	971	22,852	987
1963	484·4	84·9	29·7	13·3	612·3	14,218	5,908	2,300	1,403	23,469	987
1964	487·2	89·2	32·4	14·4	623·2	14,297	6,161	2,457	1,095	24,010	1,111
1965	490·2	93·0	34·8	15·2	633·2	14,469	6,477	2,638	1,166	24,750	1,146
1966	493·2	98·7	37·5	16·6	646·0	14,614	6,795	2,912	1,251	25,572	1,191
1967	496·5	103·6	40·5	18·0	658·6	14,686	7,248	3,014	1,350	26,298	1,190

Note: Figures for 1967 are provisional.

TABLE A3: *Expenditure on Education, Ireland,* 1947–1966

£000

Year	*Personal (F)*					*Central Government (G)*					*Overall Total Expend.*
	Primary	*Secondary*	*Vocational*	*University*	*Total*	*Primary*	*Secondary*	*Vocational*	*University*	*Total*	
1947						6,722	1,083	687	451	10,157	
1948						7,702	1,238	807	574	11,628	
1949			n.a.			7,952	1,271	830	604	12,099	n.a.
1950						8,045	1,260	872	579	12,191	
1951						8,494	1,318	834	718	12,721	
1952						10,154	1,608	937	705	14,926	
1953	—	951	79	363	1,394	9,628	1,516	943	744	14,334	15,728
1954	—	1,017	82	390	1,488	10,154	1,524	1,016	1,062	15,301	16,789
1955	—	1,130	85	422	1,637	10,345	1,900	1,081	764	15,621	17,258
1956	—	1,151	107	452	1,710	10,081	1,871	1,221	688	15,404	17,114
1957	—	1,233	113	509	1,855	11,029	1,914	1,073	839	16,493	18,348
1958	—	1,343	116	520	1,979	10,377	2,318	1,096	783	16,152	18,131
1959	—	1,435	120	550	2,104	10,701	2,187	1,188	819	16,521	18,625
1960	—	1,510	124	611	2,245	11,187	2,339	1,223	1,048	17,440	19,685
1961	—	1,506	139	645	2,290	10,671	2,411	1,300	1,072	16,970	19,260
1962	—	1,606	165	682	2,453	11,077	2,486	1,369	1,100	17,546	19,999
1963	—	1,681	162	746	2,589	11,492	2,785	1,630	1,591	19,068	21,657
1964	—	1,736	154	753	2,693	11,371	2,543	1,425	1,816	1,8540	21,233
1965	—	1,939	191	779	2,921	12,555	3,017	1,481	2,122	20,700	23,621
1966	—	1,949	238	800	2,987	13,244	3,263	2,303	1,855	21,110	24,097

Note: See Notes to Table A1. Data for 1967 not available.

TABLE A4: *Pupil/Teacher Ratio, Average Salary, and Private Expenditure, Ireland,* 1947–1967

Year	*Pupil/Teacher Ratio (P=S/T)*				*Average Salary (D/T)*				*Private Expend.*	
	Primary	*Secondary*	*Vocational*	*University*	*Primary*	*Secondary*	*Vocational*	*University*	*Per Sec. Pupil*	*Per Univ. Pupil*
					£	£	£	£	£	£
1947	34·8	12·0	(10·3)	13·2	448·5	163·5	572·6	836·2	..	
1948	35·4	11·9	(11·0)	12·9	539·1	190·4	661·4	841·9	..	
1949	34·7	11·8	(11·2)	14·1	519·8	183·0	622·3	882·9	..	n.a.
1950	34·9	12·3	(12·1)	13·3	501·6	180·5	615·3	821·8	..	
1951	35·3	12·4	(13·0)	12·8	533·8	199·0	587·6	902·0	..	
1952	35·8	12·4	13·2	12·6	583·1	228·5	653·0	869·0	..	
1953	36·1	12·5	13·5	12·9	566·5	225·4	615·7	851·4	18·2	44·8
1954	35·9	13·2	12·7	13·4	605·0	259·5	671·8	906·6	18·8	47·0
1955	36·2	12·8	12·6	12·6	600·6	251·1	676·9	846·4	20·0	49·6
1956	36·7	13·0	12·3	11·8	581·1	229·8	644·6	847·1	19·4	52·6
1957	36·4	13·2	12·7	11·9	613·6	226·0	672·3	880·5	19·8	57·2
1958	36·2	13·4	13·0	12·3	581·2	312·9	638·4	866·9	20·3	55·9
1959	35·8	13·8	13·1	12·9	575·1	275·0	662·9	910·7	20·6	55·0
1960	35·5	14·2	13·4	12·8	578·7	280·0	656·5	929·8	20·6	57·6
1961	34·9	14·5	13·3	12·9	577·9	288·0	675·8	942·3	19·6	56·6
1962	34·4	14·3	13·1	12·8	575·7	283·5	640·3	914·5	20·0	55·0
1963	34·1	14·4	12·9	12·8	598·6	311·9	697·0	934·8	19·8	56·1
1964	34·1	14·5	13·2	13·2	573·4	273·7	580·4	905·9	20·0	52·3
1965	33·9	14·4	13·2	13·0	653·9	359·4	705·1	926·2	20·9	51·3
1966	33·7	14·5	12·9	13·3	695·2	353·3	696·1	928·0	19·8	48·2
1967	33·8	14·2	13·4	13·3	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

See *Notes* to Table A1.

APPENDIX B

TABLE B1: *Basic Data*

Values in £m

Years	$\Delta M(Y_1)$	ΔY	ΔT	ΔZ	$\Delta Y^d(Y_3)$	$\Delta C(Y_4)$	ΔG	$\Delta I(X_1)$	ΔX	ΔB_a	$(Y_d)_{-1}$	$(C)_{-1}$	$\Delta e(X_3)$
1952–53	9·3	46·9	0·8	34·4	33·6	28·8	5·4	—0·1	11·2	2·8	388·9	352·7	6·1
1951–52	—30·7	57·9	1·8	31·2	29·4	13·4	2·6	4·0	22·0	4·3	359·5	339·3	5·7
1950–51	47·5	21·4	2·2	27·2	25·0	25·4	8·8	13·0	16·1	—2·7	334·5	313·9	7·2
1949–50	30·9	7·0	0·6	12·7	12·1	14·4	3·4	10·3	10·4	—5·3	322·4	299·5	2·8
1948–49	—6·4	26·2	1·4	22·3	20·9	7·4	—0·9	13·3	3·5	3·1	301·5	292·1	3·1
1947–48	4·6	33·2	1·9	25·0	23·1	14·2	3·7	9·4	14·8	8·4	278·4	277·9	6·9

Notes: Principal source of data: NIE 1964, Appendix 4. Notation: Leser, *op. cit.*, page 3; T=Taxation on personal income (Z).

TABLE B2: *Exogenous Variables*

Values in £m

Years	X_1	X_2	X_3	X_4
1952–53	−0·1	14·0	6·1	·093083
1951–52	4·0	26·3	5·7	·056189
1950–51	13·0	13·4	7·2	·061584
1949–50	10·3	5·1	2·8	·071030
1948–49	13·3	6·6	3·1	·031177
1947–48	9·4	23·2	6·9	·001796

Notes: see TABLE B1. $X_2=\Delta X+\Delta B_a$,
$X_4=[(Y_d-C)/Y_d]_{-1}$

TABLE B3: *Comparison of Calculated and Actual Endogenous Variables (£m)*

	Y_1		Y_2		Y_3		Y_4	
Years	*Calc.*	*Act.*	*Calc.*	*Act.*	*Calc.*	*Act.*	*Calc.*	*Act.*
1952–53	8·9	9·3	25·4	41·5	20·8	33·6	28·3	28·8
1951–52	−19·2	−30·7	38·3	55·3	33·4	29·4	17·7	13·4
1950–51	19·4	47·5	34·0	12·6	29·3	25·0	17·7	25·4
1949–50	12·5	30·9	17·0	3·6	12·6	12·1	11·8	14·4
1948–49	−15·8	−6·4	20·6	27·1	16·1	20·9	−6·2	7·4
1947–48	21·9	4·6	40·6	29·5	35·6	23·1	−8·6	14·2

Notes: Y_1, Y_3, Y_4, actual, see TABLE B1. $Y_2=\Delta Y-\Delta G$

TABLE B4: *Absolute Values of Deviations from Actual using Leser's (L) and a Naive (N) Model (£m)*

	Y_1		Y_2		Y_3		Y_4	
Years	*L*	*N*	*L*	*N*	*L*	*N*	*L*	*N*
1952–53	0·4	0·6	16·1	19·3	12·8	13·5	0·5	10·4
1951–52	11·5	22·6	17·0	36·8	4·0	12·6	4·3	1·9
1950–51	28·1	39·4	21·4	5·9	4·3	8·1	7·7	10·0
1949–50	18·4	24·7	13·4	10·2	0·5	0·2	2·6	3·1
1948–49	9·4	5·6	6·5	0·1	4·8	3·6	13·6	15·0
1947–48	17·3	0·6	11·1	20·3	12·5	14·7	22·8	6·8