Relative Efficiency of Regression using Original Data or First Differences: the case of Autocorrelated Disturbances.[1]

PETER NEARY

In two recent articles, Geary [2] and Tillman [6] have compared the efficiency of using data in absolute form or in the form of first differences in regression analysis of economic time series. Both authors take a highly unfavourable view of the widely-used method of first differences: Geary concludes that this method "will usually be highly inefficient" and suggests instead the inclusion of a time trend along with the independent variable X in its original form. Tillman extends the discussion to the case of multiple regression, and argues that taking first differences is not an appropriate way of overcoming the difficulties associated with multicollinearity among the independent variables.

The purpose of this note is to show that the conclusions of both authors depend crucially on the restrictive assumptions they make concerning the presence or absence of serial correlation in the error terms. Tillman mentions this problem but does not discuss its implications for the relative efficiency of absolute and first difference regression. Geary, on the other hand, confines his attention to two extreme cases: that of regularity of the original error term u, and regularity of the first difference error u^*. He thus ignores the far more likely situation of the truth lying somewhere between the two extremes. It will be shown below that adopting a more flexible approach goes a considerable way towards redressing the balance in favour of first differences.

1. The author wishes to thank R. C. Geary and B. J. Whelan for helpful comments on an earlier draft of this paper.

1. *The General Model*

For a more general approach, consider the model:

$$y = X\beta + u. \tag{1}$$

We make all the usual assumptions, except that no restrictions are placed on the variance-covariance matrix of the u's:

$$E(uu') = \sigma^2 V_1 \tag{2}$$

Where V_1 is any $n \times n$ positive semi-definite matrix.

The best linear unbiased estimator of β is given by Aitken's generalised least squares:

$$\hat{\beta} = (X'V_1^{-1}X)^{-1}X'V_1^{-1}y \tag{3}$$

with covariance, cov $\hat{\beta} = \sigma^2 X'V_1^{-1}X)^{-1}$.
Whereas, applying OLS to (1) yields the estimator:

$$b = (X'X)^{-1}X'y$$

with true covariance, cov $b = \sigma^2(X'X)^{-1}X'V_1X(X'X)^{-1}$.

The next step is to take first differences of (1). This may be represented by premultiplying the model by the $(n-1) \times n$ first differencing matrix T, to give:

$$Ty = TX\beta + Tu$$

or

$$y^* = X^*\beta + u^* \tag{4}$$

The covariance matrix of the disturbances is now

$$\begin{aligned} E(u^*u^{*\prime}) &= E(Tuu'T') \\ &= \delta^2 TV_1T' \end{aligned}$$

which for convenience we may write as $\sigma^2 V_2$.
Applying generalised least squares to (4) gives the estimator:

$$\begin{aligned} \hat{\beta}^* &= (X^{*\prime}V_2^{-1}X^*)^{-1}X^{*\prime}V_2^{-1}y^* \\ &= (X'T'V_2^{-1}TX)^{-1}X'T'V_2^{-1}Ty \end{aligned}$$

with covariance, cov $\hat{\beta}^* = \sigma^2(X'T'V_2^{-1}TX)^{-1}$.
Similarly, the OLS estimator of β in (4) is:

$$b^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$$

$$= (X'T'TX)^{-1}X'T'Ty$$

with covariance, cov $b^* = \sigma^2(X'T'TX)^{-1}X'T'V_2TX(X'T'TX)^{-1}$.

We are now faced with four different estimators[2] of β though our primary interest lies in comparing cov b with cov b^*. Following Tillman, we obtain, by successive applications of the generalised Gauss-Markov theorem:

$$\text{cov}\,\hat{\beta} \leqslant \text{cov}\,b$$

and, cov $\hat{\beta} \leqslant$ cov $\hat{\beta}^* \leqslant$ cov b^*.

In interpreting these inequalities, two points should be made. First, as we would expect, $\hat{\beta}$ has unambiguously the smallest variance of all four estimators. This means that, except for the limiting cases of u regular (V_1 an identity matrix of order n) and u^* regular (V_2 an identity matrix of order $n-1$), the most efficient estimator of β is given by *neither* OLS with original data nor OLS with first differences. Instead the minimum variance estimator is given by the generalised least squares formula as in (3) above, which is identical to the transformation proposed by Kadiyala [3]. Of course, calculation of this requires a knowledge of V_1, which in practice is unknown and so must be estimated. However, if a consistent estimator of V_1 is substituted into equation (3), the resulting estimator may be shown to have the same asymptotic distribution as (3), and the available evidence suggests that it will perform nearly as well in small samples.

All this is well known. Of more immediate interest is the second point, which is that, in general, we can say nothing about the relationship between cov b and cov b^*. This situation is obviously very different from that considered by Geary and Tillman. Not only is an unambiguous statement about the relative merits of applying OLS to original data and to first differences now impossible, but their relative efficiency may be expected to vary greatly from model to model, depending on the form of the X matrix, as well as on the structure of the (true) V matrix. In order to proceed further, we must therefore turn to some particular examples.

2. It may be noted that, if we add an additional row to T, making it a square matrix, and if T^{-1}exists, then the formula for $\hat{\beta}^*$ reduces to that for β. The simplest way of doing this is to add to T a new first row: [1, 0 . . . 0]; and this is in fact identical to the procedure mentioned by Geary on p. 552 of his article, of adding an additional relation, $Y_1 = a + \beta X_1 + u_1$, to the $n-1$ first differenced relations, to make n relations in all. However, while this simplification of the algebra is attractive, the resulting T matrix is in fact not a first differencing matrix in the generally accepted sense of the term.

(2) *Some Special Cases*

For simplicity we consider the case of a single independent variable, and assume that the error term follows a first-order Markov scheme:

$$Y_t = a + \beta X_t + u_t$$

$$u_t = \rho u_{t-1} + e_t \qquad t = 1, \ldots n$$

where e_t is regular, and ρ is a parameter which can take on any value (though for stability we require $|\rho| \leqslant 1$).

It may be noted in passing that the two structures for u_t considered by Geary are both special cases of this model: obviously u_t is regular when $\rho = 0$, while, on taking first differences:

$$Y_t^* = \beta X_t^* + u_t^*$$

where

$$u_t^* = u_t - u_{t-1}$$

$$= e_t - (1 - \rho) u_{t-1}.$$

Thus u_t^* is regular when $\rho = 1$. In the light of this, we would expect the method of OLS applied to original data, which implicitly assumes that ρ is zero, to be relatively more efficient the closer ρ actually is to zero. Conversely, the method of first differences, which implicitly assumes that ρ is unity, may be expected to be more efficient as ρ tends towards unity.

To examine this further, we consider some particular forms of X matrix. In each case we evaluate, for different values of ρ, the formulae for the variances of b, the least squares estimate of β using the original data, and b^*, the least squares estimate of β using first differences. This may be done either by calculating the matrix expressions already given, or by working out their scalar equivalents (the latter are given in many econometrics textbooks; see, e.g., Kmenta [4] pp. 276 and 291). The ratio of the two variances, $F = \text{var } b / \text{var } b^*$, is then a measure of the efficiency of the first difference estimator relative to that of the original data estimator.

$$\textit{Case } 1: X_t = t (t = 1, 2, \ldots n)$$

For this, the first case considered by Geary, Table 1 gives the variance ratio for different values of n and ρ.(3,4) The first column, for $\rho = 0$, is the case considered by Geary, where $F = 6(n-1)/n(n+1)$; obviously first differences are here highly inefficient. However, as ρ increases, it is only natural to expect that the method

3. The last column of this and succeeding tables may be ignored for the present. They are discussed in section 4 below.

4. As p goes from 0 to -1, $-$the case of negative serial correlation—the efficiency of first differences decreases even further.

of first differences—which as we have said implicitly assumes a value for ρ of $+1$ —becomes relatively more efficient. That this is the case is apparent from the table: for given n, F increases monotonically with ρ for all values of ρ up to ·8. Moreover, for reasonably large values of ρ, F is greater than unity in a significant number of cases, increasingly so the smaller the sample size. However, it does appear that for large n the relative efficiency of first differences tends towards zero irrespective of ρ.

TABLE 1: *Ratio of variances of b and for different values of n and ρ*

Case 1: X={1, 2, . . . n}

n \ ρ	*Ratio of true variances for different values of ρ* 0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	*Ratio of "apparent" variances*
4	·900	·940	·970	·991	1·006	1·016	1·021	1·024	1·024	1·023	·600
6	·714	·786	·856	·921	·977	1·021	1·051	1·068	1·073	1·069	·289
8	·583	·659	·741	·826	·910	·987	1·050	1·091	1·110	1·105	·167
10	·491	·564	·645	·736	·833	·932	1·024	1·095	1·133	1·134	·109
20	·271	·322	·382	·455	·544	·655	·792	·954	1·117	1·207	·029
30	·187	·224	·269	·325	·395	·487	·610	·779	1·002	1·211	·013
40	·143	·172	·207	·252	·309	·385	·491	·645	·878	1·176	·007
50	·115	·139	·169	·206	·254	·318	·410	·547	·770	1·119	·005
70	·083	·101	·123	·150	·186	·236	·307	·417	·609	·987	·002
100	·059	·071	·087	·107	·133	·169	·222	·306	·459	·811	·001

Case 2: $X_t = -\{T^2, -(T-1)^2, - - -1, 1, - - -(T-1)^2, T^2\}$ where $2T=n$

Table 2 gives the same variance ratio for Geary's second case, designed to illustrate the situation where the values of the independent variables tend to cluster around the median. As before, the first column corresponds to Geary's equation (1.19), but again the relative efficiency of first differences increases markedly (and monotonically) with ρ. In fact, if we accept Geary's conjecture that this case is more typical than Case 1, then the argument for first differences at least in small samples is strengthened, especially if ρ is believed to equal or exceed ·8. Again, however, as n tends towards infinity the relative efficiency of first differences appears to approach zero, even more quickly than in Case 1.

Case 3: $X_t = rX_{t-1}+v_t$ (v_t a regular random variable)

Arguably a more realistic situation than either of the first two cases is that where the independent variable is itself generated by an auto-regressive scheme with parameter r. The author has been unable to derive the variance ratio in this case for small samples; however, for large n they are easily derived, and are shown by Kmenta (*loc. cit.*) to be approximately equal to:

$$\text{var } b = \frac{\sigma^2 u}{\Sigma x^2} \left\{ \frac{1+\rho r}{1-\rho r} \right\}$$

$$\text{var } b^* = \frac{\sigma^2 u(1-\rho)}{\Sigma x^2} \left\{ \frac{1}{1-r} + \frac{1-\rho}{2(1-\rho r)} \right\}$$

Therefore,

$$F = \frac{\text{var } b}{\text{var } b^*} = \frac{2(1+\rho r)(1-r)}{(1-\rho)\{4-(1+r)(1+\rho)\}}.$$

It is of some interest to investigate the properties of this asymptotic variance ratio. Firstly, it is easily established that

$$\frac{\partial \log F}{2\partial \rho} = \frac{r}{1+\rho r} + \frac{1}{1-\rho} + \frac{1+r}{4-(1+r)(1+\rho)}$$

$$>0, \text{ for } r<1.$$

Thus, for a given value of r, F increases monotonically with ρ. The partial

TABLE 2: *Ratio of variances of b and b* for different values of n and ρ*

Case 2: [illegible] $= -T^2, -(T-1)^2, \ldots -1, 1, \ldots (T-1)^2, T^2\}$, *where* $2T=n$

n \ ρ	*Ratio of true variances for different values of ρ*										*Ratio of "apparent" variances*
	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	
4	·712	·788	·857	·917	·966	1·005	1·034	1·056	1·071	1·080	·647
6	·441	·508	·588	·683	·789	·904	1·018	1·123	1·209	1·270	·367
8	·352	·404	·468	·549	·651	·778	·930	1·097	1·262	1·395	·240
10	·299	·344	·399	·468	·558	·676	·833	1·031	1·258	1·469	·170
20	·177	·208	·245	·291	·348	·425	·534	·705	·997	1·472	·053
30	·127	·151	·179	·215	·260	·319	·403	·532	·771	1·293	·025
40	·099	·118	·141	·170	·208	·257	·327	·433	·629	1·113	·015
50	·081	·097	·117	·141	·173	·216	·275	·368	·534	·965	·010
70	·059	·071	·086	·105	·130	·163	·210	·284	·416	·757	·005
100	·042	·051	·062	·076	·094	·119	·155	·212	·315	·578	·003

derivative with respect to r cannot be uniquely signed, however, so we must have recourse to calculating the value of F for different values of the two parameters.

First, for r equal to -1, a highly unrealistic case, $F+1$ irrespective of ρ; that is, the two methods are equally efficient. For r equal to $+1$, F equals zero. This case is of interest, for it is, in fact, identical with Geary's Case 1 as n increases towards infinity. Geary's findings are therefore doubly biased against the method of first differences—both by his restrictive examples and by his treatment of autocorrelation. For values of r between these two extremes, Table 3 shows that the method of first differences is frequently more efficient, and for high values of ρ will usually be considerably so. Interestingly enough, even for random X ($r=0$) the method of first differences is more efficient for values of ρ greater than ·268.

3. *Practical Implications*

The preceding examples have shown that, in these particular cases, the method of first differences will be considerably more efficient than was claimed by Geary and Tillman, as the degree of residual autocorrelation increases. Obviously the extent to which this conclusion will hold in general depends on how realistic these examples are. Fortunately some evidence is available on this point. A study by Ames and Reiter [1] of 100 annual series for the US economy found that a first order autoregressive scheme, of the kind labelled Case 3 above, with an average autocorrelation coefficient of ·84, was an excellent representation of many of the series they considered.

Furthermore, if we accept the rationale most commonly given for the inclusion of stochastic disturbance terms—namely, that they reflect the influence of variables omitted from the regression—this evidence suggests that in many time series applications we may expect u to follow a first-order Markov scheme with a fairly high value of ρ. Obviously, the findings of Ames and Reiter are only suggestive, and are highly restrictive in that they refer only to annual data on the US economy. Nevertheless, pending further research on this topic, they do imply that in a majority of time-series studies the method of first differences may be expected to be more efficient than that using original data.

Finally, it may be noted that for any given X matrix, the value of F contingent on different values of ρ may be calculated directly. This has been done for the example given by Geary, where Y is the Irish money supply, and X is GNP at current prices, 1949–1965; the results are shown in table 4.[5] An approximately unbiased estimate of ρ, calculated from the residuals of the regression of Y on X,[6] is ·535. From the table, we can therefore conclude that first differences are indeed

5. I have been unable to replicate exactly the results given by Geary. The data used here gave the following results, which are similar, but not identical to Geary's: $b=·264$ with ESE ·00577 ($r=·997$); and $b^*=·259$ with ESE ·02492 ($r=·809$).

6. The formula used was È=È+2(ɪ+È₁),n, where È₁, is the autocorrelation coefficient calculated directly from the residuals. See Wallis [7] for further details.

TABLE 3: *Asymptotic ratio of variances of b and b* for different values of ρ and r*

Case 3: $X_t = rX_{t-1} + v_t$

ρ / r	Asymptotic ratio of true variances for different values of ρ: 0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	Ratio of "apparent" variances
−0·9	·974	·988	1·004	1·024	1·050	1·086	1·138	1·224	1·393	1·895	3·8
−0·8	·947	·974	1·005	1·045	1·097	1·168	1·272	1·443	1·780	2·785	3·6
−0·7	·918	·957	1·004	1·063	1·140	1·245	1·401	1·656	2·162	3·668	3·4
−0·6	·888	·939	1·000	1·077	1·178	1·318	1·524	1·863	2·537	4·543	3·2
−0·5	·857	·918	·993	1·087	1·212	1·385	1·641	2·063	2·903	5·410	3·0
−0·4	·824	·894	·982	1·093	1·241	1·445	1·750	2·255	3·260	6·266	2·8
−0·3	·788	·868	·967	1·094	1·263	1·498	1·851	2·437	3·606	7·109	2·6
−0·2	·750	·838	·947	1·089	1·278	1·543	1·941	2·606	3·938	7·935	2·4
−0·1	·710	·804	·923	1·077	1·285	1·577	2·020	2·761	4·252	8·742	2·2
0	·667	·766	·893	1·058	1·282	1·600	2·083	2·899	4·545	9·524	2·0
0·1	·621	·724	·856	1·031	1·268	1·609	2·129	3·014	4·812	10·272	1·8
0·2	·571	·677	·813	·993	1·241	1·600	2·154	3·102	5·043	10·977	1·6
0·3	·519	·623	·760	·944	1·199	1·571	2·151	3·155	5·229	11·621	1·4
0·4	·462	·564	·698	·881	1·137	1·516	2·114	3·160	5·351	12·179	1·2
0·5	·400	·496	·625	·801	1·053	1·429	2·031	3·103	5·385	12·609	1·0
0·6	·333	·421	·538	·702	·939	1·300	1·889	2·958	5·286	12·833	·8
0·7	·261	·335	·436	·579	·790	1·117	1·664	2·685	4·979	12·701	·6
0·8	·182	·238	·315	·427	·595	·862	1·321	2·213	4·316	11·862	·4
0·9	·095	·127	·172	·237	·338	·504	·802	1·411	2·966	9·282	·2

TABLE 4. *Ratio of variances of b and b* for different values of* ρ

Example: X=Irish GNP, current prices, 1949–1965 (n=17)

Ratio of true variances for different values of ρ										Ratio of "apparent" variances
0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	
·166	·195	·231	·278	·341	·426	·549	·732	1·006	1·384	·066

inferior to the original data as a method of estimating the true β, though they are considerably less inferior than Geary suggests.

4. *True versus "apparent" variances*

The discussion so far has been exclusively concerned with the true variances of b and b^*, the OLS estimators of β, based on original data and on first differences, respectively. Of course, calculation of these requires a knowledge of the true value of ρ (or, in general, a knowledge of the true V_1 matrix). In practice, however, such knowledge is not available, and it is likely that estimates of the true variances will be calculated by mechanically applying the standard OLS formulae. However, these will yield biased estimates of, not only the variance b^*, as Tillman has pointed out, but also of the variance of b. The resultant "apparent,' variances will be:

$$\hat{\text{cov}}\, b = \sigma^2(X'X)^{-1}$$

$$\hat{\text{cov}}\, b^* = \sigma^2(X^{*\prime}X^*)^{-1} = \sigma^2(X'T'TX)^{-1}.$$

Not only are these estimates biased, but they are also completely invariant with respect to the true value of ρ. In other words, irrespective of the value of ρ, taking first differences of a particular X matrix will appear to have a specific effect on the variance of the estimate of β, but this "apparent" effect is not in any way an indication of the true gain or loss in efficiency.

As for the quantitative magnitude of this "apparent" efficiency ratio, the upper and lower bounds calculated by Tillman for the ratio of the generalised "apparent" covariances are still valid, though the ratio should now be written: $|\hat{\text{cov}}\, b^*|/|\hat{\text{cov}}\, b|$ (instead of * 1côv b.1 1/cov 1 b). Since these bounds are extremely wide, however, they give no indication as to the likely effect of taking first differences on the "apparent" variance of the estimators. Nevertheless, as with the true variance ratio, it is possible to calculate the "apparent" variance ratio for any specific X matrix, and this has been done, for the examples already considered, in the last column of Tables 1 to 4.[7] It is evident that except in Table 3, the "apparent" variance ratio seriously underestimates the gain in efficiency to be obtained from using first differences, and that in all tables the degree of underestimation increases rapidly with ρ. A comparison of the true and "apparent" ratios in these tables should be a sombre warning, if any is needed, against the mechanical application of standard formulae and library programmes.

5. *Other reasons for choosing between absolute data and first differences*

So far this note has followed these of Geary and Tillman in concentrating on only one aspect of the choice between original data and first differences, namely the question of efficiency.[8] However, there are other issues to be considered. Suits [5] for example, lists four other reasons, apart from a possible gain in efficiency, why first differences may be preferred to original data in time-series regressions:

(1) Econometric equations, as originally formulated, frequently involve stocks, for which data are not currently available. However, the first difference in stocks (ignoring depreciation, which may be expected to follow a fairly smooth pattern) is well represented by current purchases, on which data are more readily available.

(2) In forecasting applications, the present position is known, and the important question is what change from that position will result from projected changes in other factors. The use of first differences serves to focus the analysis on these changes.

(3) The use of first differences minimises the effect of slowly moving variables such as population, tastes, technical change, etc., without explicitly introducing them into the analysis.

(4) Equations estimated in first differences are less likely to be affected by data revisions, which usually alter the level at which variables are measured, rather than their year-to-year variations.

It may be noted that these reasons are of a more practical nature than the issue of relative efficiency with which this note has been mainly concerned. This is not

7. Analytically, the expressions for these "apparent" variance ratios are as follows: for Case 1, $12/n(n+1)$; for Case 2, $80(n—2n+3)/n(n+1)(3n^2+6n-4)$; and for Case 3, $2(1-r)$. It should be noted that these formulae are only approximate, since they make no allowance for the fact that the estimates of σ^2 from original data and from first differences will not in general be equal, and will moreover be biased. However, since this fact merely increases the variability of the "apparent" efficiency ratio it does not affect the conclusion in the text, namely, that this ratio gives no indication of the true gain or loss in efficiency.

8. Since all the estimators of β considered are unbiased, comparison between them on the basis of their variances is identical to comparing between them using the mean square error criterion.

in any way to underestimate their importance, however, and a practising econometrician would do well to keep them in mind in addition to the question of efficiency.

It may also be remarked that Suits does not mention any reduction in the degree of multicollinearity as an argument favouring the use of first differences. In fact, as Tillman [6] has argued, the issue of multicollinearity is quite irrelevant to the choice between absolute data and first differences. The illusion that taking first differences will overcome the problem of multicollinearity presumably derives from the fact that a transformation from X to X^* such that the inter-correlations between the variables in X^* are less than those between the variables in X, will usually produce diagonal elements of $(X^{*\prime}X^*)^{-1}$ which will be smaller than the corresponding elements of $(X'X)^{-1}$. However this does not mean that the transformation will therefore reduce the variances of the estimates of β. For these variances depend not on the matrices given here, but on the more complicated expressions for cov b and cov b^* given in Section 1 above. Which of these latter is the smaller will depend not just on X, but also on V_1, that is, on the degree of autocorrelation in the original model. To say that taking first differences overcomes the problem of multicollinearity is just another example of the error noted in Section 4. It may indeed reduce the "apparent" variance ratio, but whether the true variance of the estimate of β will be reduced depends primarily on the V_1 matrix.

6. *Summary and Conclusion*

This note should not be interpreted as advocating the indiscriminate use of first differences in regression. It merely seeks to show that the conclusions reached by Geary and Tillman concerning the efficiency of this method relative to that of original data are unduly pessimistic. By contrast with them the following points have been made:

(1) If we permit the possibility of an unspecified degree of autocorrelation in the residuals, then nothing can be said in general about the relative efficiency of absolute data and first differences in regression.

(2) For the special case of a first order Markov scheme with parameter ρ, the relative efficiency of first differences increases with ρ, at least in some special cases considered. Moreover, for realistic kinds of X matrix and realistic values of ρ, this method will be more efficient than that using original data.

(3) The "apparent" variances of b and b^*, calculated by mechanically applying the standard OLS formulae, are a hopeless indicator of the relative merits of the two methods.

(4) In choosing between OLS with original data and with first differences, there are a number of issues in addition to that of relative efficiency which should be kept in mind. However, it is never correct to say that the use of first differences

leads to more efficient estimates because it reduces the degree of multicollinearity.

Finally, two further points may be made. Firstly, this note has concentrated on comparing different estimators on the basis of their relative efficiency in the estimation of the parameter vector β. However, everything that has been said may be applied equally well to the choice between estimators from the point of view of their predictive ability. This is because, to each estimator b, with variance-covariance matrix $\sigma^2 W$, there corresponds a predictor for Y_{n+t}: $p = X'_{n+t} b$, with variance $\sigma^2 (1 + X'_{n+t} W X_{n+t})$, where X_{n+t} is the vector of observations on the independent variables in the prediction period.[9] Obviously, choosing b so as to minimise W must inevitably minimise the prediction variance also.

Secondly, it should be repeated that the choice between b and b^*, which has been the focus of this note, will in almost all situations be a "second-best" one. This is because, strictly speaking, if ρ is not exactly equal to either zero or unity, then the most efficient estimator of β is neither b nor b^*, but β, Aiken's Generalised Least Squares Estimator (formula (3) above). Two further sets of tables, similar to those given in Section 2, could be produced to illustrate the loss of efficiency involved in using original data or first differences instead of the optimal GLS estimator. Of course, it may sometimes happen that for some reason—whether ease of computation, or because we are more concerned about other defects in our model than about the presence and severity of autocorrelation—we wish to confine our attention to the two options of using either b or b^*. But it should not be forgotten that, except in two extreme cases (regularity of u and regularity of u^*) there is a loss of efficiency involved.

Trinity College, Dublin

REFERENCES

[1] Ames, E., and S. Reiter: "Distributions of Correlation Coefficients in Economic Time Series", *Journal of the American Statistical Association*, September 1961, pp. 637–656.

[2] Geary, R. C.: "Two Exercises in Simple Regression", *Economic and Social Review*, Vol. 3, No. 4, July 1972.

[2a] Goldberger, A. S.: "Best linear unbiased prediction in the generalised linear regression model", *Journal of the American Statistical Association*, June 1962, pp. 369–375.

[3] Kadiyala, K. R.: "A Transformation Used to Circumvent the Problem of Autocorrelation", *Econometrica*, Vol. 36, No. 1, January 1968.

[4] Kmenta, J.: *Elements of Econometrics*, Macmillan, 1971.

[5] Suits, D. B.: "Forecasting and Analysis with the Econometric Model", *American Economic Review*, Vol. 52, No. 1, March 1962, pp. 104–132.

[6] Tillman, J. A.: "The Efficiency of Taking First Differences in Regression Analysis: A note", *Economic and Social Review*, Vol. 4, No. 4, July 1973.

[7] Wallis, K. F.: "Lagged Dependent Variables and Serially Correlated Errors: A Reappraisal of Three-Pass Least Squares"; *Review of Economics and Statistics*, Vol. 49, pp. 555–567, 1967.

9. This assumes that we have no information concerning the covariance between the sample disturbances and the disturbance in the prediction period. If such information is available, Goldberger [2*a*] shows how it may be used to calculate a predictor with still smaller variance.