COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 383

Trend Estimators and Serial Correlation

by

S. J. Prais* and C. B. Winsten**

February 24, 1954

1. Trend Estimation and Regression Analysis

The problem of analyzing time-series when the errors are serially correlated has been treated by a number of writers in recent years, all of whom appear to be agreed that the problem is so complex that a treatment of the problem in its full generality does not lead to results that are of interest. Thus, Cochrane and Orcutt [1] attempted to simplify the problem by conducting artificial sampling experiments on series which they thought were typical of those encountered in econometric work, and, more recently. Gurland [2] considered the regression on one determining variable in order to examine the effects of the non-stationarity of the error term. The simplification that we adopt in this paper is to consider only the special case of estimating a trend:*** this allows us to answer with some precision a number of problems concerned with the serial correlation of errors as applied to this case. At the same time, by considering the nature of the circumstances of

* I am grateful to the Department of Economics for granting me a Fellowship and to the Cowles Commission for extending their facilities during my stay in Chicago. - S.J.P.

** Cowles Commission and Oxford University Institute of Statistics.

*** In adopting this process of simplification, and in other obvious ways, we are indebted to the writings of the authors mentioned.

this case, we hope to suggest the nature of the correct answer in the general case.

In order to introduce the notation we begin by considering a familiar problem which has a beguiling though misleading simplicity[1/]. Consider the time-series

(1) $$y_t = \beta t + \eta_t$$

where t is time and runs from $-N$ to $+N$ (there being $2N + 1$ observations in all). β is the trend coefficient, y_t is the observed value of a variable at t, and η_t is an additive disturbance. If η_t is generated by the system

(2) $$\eta_t = \eta_{t-1} + \epsilon_t$$

where ϵ_t is distributed independently with zero mean and variance σ^2, then an estimator of β is given by

(3) $$b_1 = (y_N - y_{-N}) / 2N.$$

That is, the estimator is based on only the first and the last observations in the series, and all of the intermediate values are neglected.

The use of this estimator may be rationalized as follows. Consider the differences between successive values of y_t, which, on using (1) and (2), are

$$y_t - y_{t-1} = \beta t - \beta(t-1) + \eta_t - \eta_{t-1}$$

$$= \beta + \epsilon_t,$$

so that the estimation of the trend, β, can thus be reduced to the simple problem of estimating an average based on $2N$ observations subject to independent errors. On summing the $2N$ differences $y_t - y_{t-1}$ all the inter-

1/ The paradoxical nature of this example was pointed out to one of us in oral discussion by Richard Stone who considers it in his forthcoming work [3] on consumers' behavior (Vol. I, p. 307).

mediate values cancel out, and there is left the estimator given by (3). This has a variance $\sigma^2/2N$ by the usual formula for the variance of a mean.

The paradox which arises is that the above estimator seems perfectly satisfactory even though it is only based on two out of the available $2N + 1$ observations. It will subsequently appear that the considerations involved in this example are rather more complicated than has commonly been assumed in the past, and in particular, it will be shown that while (3) may be the most practical estimator to use it is not always an efficient estimator.

In this note we consider the properties of four trend estimators, one of which is the estimator just described, under various assumptions about the distribution of the error term η_t, including distributions generated by (2). One of the main objects of the exercise is to compare the estimator (3) with the classical (unweighted) least-squares estimator under these various assumptions.

In the first portion of the paper the assumption is made that the errors are drawn from a distribution which is stationary in time. This term is here used with the following meaning:

The distribution of the errors is such that

(a) the variances of η_t, denoted by $V\left\{\eta_t\right\}$, are independent of t; and

(b) the covariances between errors τ units apart in time, that is $E\left\{\eta_t\, \eta_{\tau+r}\right\}$, are independent of t, though they may of course depend on τ.

These conditions are often taken to define "stationarity in the weak sense." Further we will only be concerned with errors generated by the system

$$(4) \qquad \eta_{t+1} = p\, \eta_t + \epsilon_{t+1}$$

where p is a constant. Such a system can only generate a stationary distribution of errors if $|p| < 1$. If, in fact, the distribution is stationary, then on squaring both sides of (4) and taking expected values, there results

$$V(\eta) = p^2 \; V(\eta) + \sigma^2$$

from which it follows that

$$V(\eta) = \frac{\sigma^2}{1-p^2} .$$

Also, by repeated substitution in (4) it is possible to express $\eta_{t+\tau}$ in terms of an earlier value η_t and the intervening ϵ's as follows:

(5) $$\eta_{t+\tau} = \epsilon_{t+\tau} + p\epsilon_{t+\tau-1} + \dots + p^{\tau-1} \epsilon_{t+1} + p^{\tau} \eta_t .$$

Hence the covariance of any two residuals in the series which are τ time units apart is given by

(6) $$E\left\{ \eta_t \, \eta_{t+\tau} \right\} = p^{\tau} E\left\{ \eta_t^2 \right\} = \frac{\sigma^2 p^{\tau}}{1-p^2}$$

These results and the assumptions on which they are based will be used throughout the next four sections of the paper in which we discuss in turn the first difference estimator, a closely related estimator which we term the weighted difference estimator, the minimum variance estimator, and the classical (unweighted) least-squares estimator. In the sixth section we compare the efficiencies of these estimators under the assumption of stationarity and present the results in a number of graphs. In the seventh section a brief discussion is given of the case when the assumption of stationarity is dropped and it is shown that the efficiency of the estimator (3) depends on whether the constant term in the equation was to be estimated from the date or is known in advance.

2. The First-Difference Estimator

The estimator (3) is termed the first-difference estimator. That it is unbiased for any value of p may be seen by substituting from (1) into (3) to give

$$b_1 = [(N\beta + \eta_N) - (-N\beta + \eta_{-N})] \,/\, 2N$$

$$(7) \qquad = \beta + (\eta_N - \eta_{-N}) / \, 2N$$

and the expected value of the error term on the right-hand side is zero.

Its variance is easily calculated in taking into account the consequences of stationarity given in (6). Thus

$$V\,(b_1) = E(b_1 - \beta)^2 = E\left\{ \frac{\eta_N - \eta_{-N}}{2N} \right\}$$

$$= \frac{1}{4N^2} \left[E\,(\eta_N)^2 - 2\,E\,(\eta_N \eta_{-N}) + E\,(\eta_{-N})^2 \right]$$

$$= \frac{1}{4N^2} \left[\frac{\sigma^2}{1-p} - 2\,\frac{p^{2N}\,\sigma^2}{1-p^2} + \frac{\sigma^2}{1-p^2} \right]$$

$$(8) \qquad = \frac{\sigma^2}{2N^2} \cdot \frac{1-p^{2N}}{1-p^2}$$

This expression shows that when $p = 0$, so that we have the classical case in which the least-squares estimator is appropriate, the variance of the estimator based on only the two extreme observations is $\sigma^2/2N^2$. This is, of course, larger than need be, and as will be seen below it is considerably larger. This estimator may nevertheless be perfectly adequate in given circumstances as, for example, when the error variance is small, and the cost of obtaining the intermediate values with sufficient exactness is very high.

3. The Weighted Difference Estimator

In order to compute the efficiency of an estimator we require the variance of the minimum variance estimator on the same assumptions. It might be thought that the correct way to proceed in this case is by analogy with the justification given at the beginning of this note for the case $p = 1$. Accordingly we would compute differences of the form

$$y_t - py_{t-1} = \beta t - p\beta(t-1) + \eta_t - p\eta_{t-1}$$

(9) $$= \beta[t(1-p) + p] + \epsilon_t$$

As before, if there were $2N + 1$ original observations, there are now only $2N$ differences. Taking differences has the effect of transforming the variables so that the error terms are serially independent and it is accordingly possible to apply the classical estimating procedure. An estimator derived in this manner we term a weighted difference estimator and denote it by $\bar{b}_p$; it is given in our case by

(10) $$\bar{b}_p = \frac{\sum_t (y_t - py_{t-1})\,[t(1-p) + p]}{\sum_t [t(1-p) + p]^2}$$

On substituting for y from (1) it is seen that this estimator is unbiased and that its variance is

$$V\left\{\bar{b}_p\right\} = E\left\{\sum_{-N+1}^{N} [t\,(1-p) + p]\,\epsilon_t\right\}^2 \Big/ \left\{\sum_{-N+1}^{N} [t(1-p) + p]^2\right\}^2$$

$$= \sigma^2 \Big/ \sum_{-N+1}^{N} [t(1-p) + p]^2$$

The denominator is evaluated most simply by expanding in powers of t, thus

$$D = (1-p)^2 \; \Sigma t^2 + 2p(1-p)\,\Sigma t + \Sigma p^2$$

Now $$\sum_{-N+1}^{N} t^2 = \frac{1}{3} N(N+1)(2N+1) - N^2 = \frac{1}{3} N(2N^2 + 1),$$

$$\sum_{-N+1}^{N} t = N, \quad \text{and} \quad \sum_{-N+1}^{N} p^2 = 2Np^2$$

Hence,

$$(11) \qquad V\left\{\tilde{b}_p\right\} = \sigma^2 \Big/ \left[\tfrac{1}{3} N(2N^2+1)(1-p)^2 + 2N_p\right.$$

4. <u>The Minimum Variance Estimator</u>

The weighted difference estimator just considered is not however of minimum variance. This can most easily be seen by considering its form as p tends to zero. When $p=0$ it is apparent that a whole observation has been neglected since the estimator is then based on only $2N$ observations; for intermediate values it may therefore be presumed that it does not obtain the correct weight in the computation.

A correct procedure leading to the minimum variance estimator requires that $2N+1$ independent transformed observations be found and that these be weighted inversely in proportion to their variance. The first $2N$ of these are given by the $2N$ differences used in the weighted difference estimator each of which has a variance of σ^2; the only other available observation that is independent of these is the first observation which has a variance of $\sigma^2 / (1-p^2)$. Hence, by simply amending the formula for the weighted difference estimator in this way, we obtain the minimum variance estimator

$$(12) \qquad \hat{b}_p = \frac{\sum_{t=-N+1}^{N} (y_t - py_{t-1})\,[t(1-p)+p] - (1-p^2)\,Ny_{-N}}{\sum_{t=-N+1}^{N} [t(1-p)+p]^2 + N^2(1-p^2)}$$

The correctness of this procedure can be demonstrated more rigorously by considering the variance matrix of the errors generated by (4) on the assumption of stationarity. The result given in (6) may then be written as

$$(13) \qquad E\left\{\eta\,\eta'\right\} = \sigma^2\,\Omega = \frac{\sigma^2}{1-p^2}\begin{pmatrix} 1 & p & p^2 & \cdots & p^{2N} \\ p & 1 & p & & p^{2N-1} \\ \vdots & & & & \vdots \\ p^{2N} & p^{2N-1} & & \cdots & 1 \end{pmatrix}$$

where η now represents a column vector of the values of η_t, and η' is its transpose (analogous and obvious adjustments will be made in the subsequent matrix notation without specific comment).

By the theorem of Gauss the best linear unbiased estimator of β is

$$\hat{b}_p = (t' \Omega^{-1} t)^{-1} t' \Omega^{-1} y.$$

and this can be reduced to the classical case by working with the transformed variables

$$\tilde{t} = Gt, \quad \tilde{y} = Gy$$

where G is chosen so that

(14) $$G'G = \Omega^{-1} .$$

It may be verified that the following matrix satisfies the last condition:

(15) $$G = \begin{pmatrix} 1 & -p & 0 & \cdots & & & 0 \\ 0 & 1 & -p & \cdots & & & 0 \\ 0 & 0 & 1 & \cdots & & & 0 \\ & & & \ddots & & & \\ & & & & 1 & -p & 0 \\ & & & & 0 & 1 & -p \\ 0 & 0 & 0 & \cdots 0 & & 0 & \sqrt{1-p^2} \end{pmatrix} .$$

and this transformation yields the estimator $\bar{b}_p$ given by (10) except for the last term.

Proceeding as in the previous section it will be found that

(16) $$V \left\{ \hat{b}_p \right\} = \sigma^2 \Big/ \left[\tfrac{1}{3} N (2N^2 + 1)(1-p)^2 + N^2 (1-p^2) + 2Np \right]$$

the only difference between this expression and (11) being the middle term of the denominator which is due to the extra weight given to the first term of the series. The relationship between the variances of minimum variance and weighted difference estimators receives further comments in the Appendix.

5. The Classical Estimator

The final estimator that we wish to consider before proceeding to a discussion of efficiency is the simple (unweighted) least-squares estimator of the trend given by

$$(17) \qquad b_o = \sum_{-N}^{N} y_t \Big/ \sum_{-N}^{N} t^2 .$$

For want of a better short title, it is here termed the "classical estimator" by which we mean to indicate not that any classical statistician would have used this estimator in inappropriate circumstances, but rather that it is the estimator to use in the classical conditions of independent and homoscedastic errors. Under these conditions the estimator is of minimum variance, and on setting $p = 0$ in (16) the well-known result may be derived that

$$(18) \qquad V\left\{\hat{b}_o\right\} = \sigma^2 \Big/ \Sigma t^2 = \sigma^2 \Big/ \tfrac{1}{3} N(N+1)(2N+1).$$

Proceeding as with the estimators considered above, it is next necessary to calculate the variance of b_o under the assumption that p has any given value. The algebra is here slightly more complicated but the principle is the same. Thus, on substituting (1) into (17), there results

$$(19) \qquad b_o = \beta + \Sigma \eta t \,/\, \Sigma t^2$$

showing that the estimator is always unbiased. Further,

$$V\left\{b_o\right\} = E\left\{\Sigma \eta t \Big/ \Sigma t^2\right\}^2$$

$$(20) \qquad = E\left\{N\eta_N + (N-1)\eta_{N-1} + \ldots (-N)\,\eta_{-N}\right\}^2 \Big/ \Sigma t^2$$

On using the stationarity assumption (6) it is possible to expand (20) by collecting together terms that are equal distances apart to give a series in ascending powers of p, thus,

$$(21) \qquad V\left\{b_o\right\} = s^2\left[p^o \sum_{-N}^{N} t^2 + 2p \sum_{-N+1}^{N} t(t-1) + 2p^2 \sum_{-N+2}^{N} t(t-2) + \dots + 2p^{2N}(-N)N\right] \Big/ \left[\sum t^2\right]^2$$

$$= K\left(\sum_{-N}^{N} t^2 + 2\sum_{1}^{2N} p^r C_r\right)$$

say, where

$$(22) \qquad C_r = \sum_{-N+r}^{N} t(t-r) = \sum_{-N+r}^{N} t^2 - r \sum_{-N+r}^{r} t$$

The simplest way of summing the squares is to divide the sum into two parts and calculate

$$\sum_{-N+r}^{N} t^2 = \sum_{o}^{N} t^2 + \sum_{o}^{N-r} t^2$$

This is obvious for the values $r = 1,2,\dots,N$; but it is also valid for $r = N,\dots,2N$ since, as can easily be verified, the partition

$$\sum_{-N+r}^{N} t^2 = \sum_{o}^{N} t^2 - \sum_{0}^{r-N-1} t^2$$

which is the obvious one to choose for these values of r, leads to the same result.

It is found that

$$C_r = \frac{1}{6}\left[N(N+1)\ (2N+1) + (N-r)\ (N-r+1)\ (2N-2r+1)\right] - \frac{r^2}{2}\ (2N-r+1)$$

which simplifies to

(23) $$C_r = \frac{2}{3} N^3 + N^2 + \frac{N}{3} - r(N^2 + N) + \frac{r}{6} (r^2 - 1).$$

There does not appear to be any convenient way of summing the series in (21) with such involved expressions for C_r, and the simplest method of proceeding is therefore to use arithmetic and particular values of N and p. For small values of p it is sufficient to consider only the first few terms in the series.

6. Results for the Stationary Case

With the help of the expressions worked out in the earlier sections of this paper we can now compare the efficiencies of the various estimators. First, consider the simplest case where p in fact is zero, but the statistician, mistakenly or deliberately, uses the first-difference estimator (3) based on only two observations. How much information does he sacrifice?

The answer is given by comparing $V\{b_o\}$ given by (18) with $V\{b_o\}$ given by (8), setting $p = 0$. In an obvious notation

(24) $$\text{Eff}\left\{ b_1 \mid p = 0 \right\} = \frac{V\{b_o\}}{V\{b_1 \mid p=0\}} = \frac{2N^2}{\frac{1}{3} N(N+1)(2N+1)}$$

and for large values of N we have

(25) $$\text{Eff}\left\{ b_1 \mid p = 0 \right\} = 3/N,$$

so that the efficiency declines as the number of observations increases.

The interpretation of this result is fairly obvious. In a classical world the estimator should have been based on all the observations but in fact only two observations were chosen. Hence the efficiency of the estimator varies inversely with the number of observations that were originally available. On the other hand, if we consider as an example the case when there are only five observations available originally, so that $N = 2$, it is seen on substituting this value into (24) that the efficiency is 4/5. When there are only three observations available originally, it appears from (24) that the efficiency is unity, but a little consideration will show that this is because the first-difference estimator and the least-squares estimator have the same form under the circumstances posited.

Proceeding to a more general problem, we consider next the efficiency of the weighted difference estimator, assuming that p is known correctly. This is given by the ratio of the expressions (11) and (16); thus

$$(26) \qquad \text{Eff}\left\{\bar{b}_p\right\} = \frac{\frac{1}{3}N\,(2N^2+1)(1-p)^2 + 2Np}{\frac{1}{3}N\,(2N^2+1)\,(1-p)^2 + N^2(1-p^2) + 2Np}$$

$$= 1 - \frac{N^2\,(1-p^2)}{\frac{1}{3}N\,(2N^2+1)\,(1-p)^2 + N^2(1-p^2) + 2Np}$$

For small p and large N the loss of efficiency is approximately equal to $3/2N$, but it increases fairly rapidly as p increases. Thus for $p = 1/2$ and $N = 10$ (that is for 21 observations) the loss of efficiency is about a third.

A loss of efficiency of this magnitude on omitting only one observation is at first sight somewhat surprising; it is rather important to understand its origin in order to know to what extent the argument of the special case of estimating a trend carries over to the case of a general linear regression. The following "intuitive" explanation may therefore be in order. The variance of a regression coefficient varies inversely with the sum of squares of the determining variable: the weighted difference estimator differs from the minimum variance estimator in omitting the one extreme observation and hence the sum of squares of the determining variable is reduced. Taking into account the weight of $(1-p^2)$ to be attached to the first observation on account of the different variance of its error term, the reduction in the sum of squares due to its omission is $N^2(1-p^2)$.

On the other hand, if any of the other transformed observations had been omitted, the reduction in the sum squares would be $[t(1-p) + p]^2$, which on the average is approximately equal to $(1-p)^2$ times the average of the squares of the first N numbers: for the values of N and p considered above, the loss is about 9, whereas the omission of the first observation reduces the sum of squares in this case by 75.

It will be apparent from this argument that this loss of efficiency is particularly great for a trend, but that in a general case in which the first observation had about the same value for the determining variable as that variable had on the average, then the loss of efficiency would be considerably smaller.

The efficiency of this estimator for values of $N = 2,10$ and 50, corresponding to 5, 21 and 101 observations respectively, is shown in diagram 1. It will be seen from this that the tendency toward a decline in efficiency as p increases, is at some point offset by the fact that as p increases the relative variance of the initial observation also increases and its omission is therefore less important. This "offsetting tendency" is stronger the smaller is the value of N: indeed, for the case of 101 observations, illustrated on the graph, the efficiency curve does not begin to rise again till $p > 0{\cdot}95$.

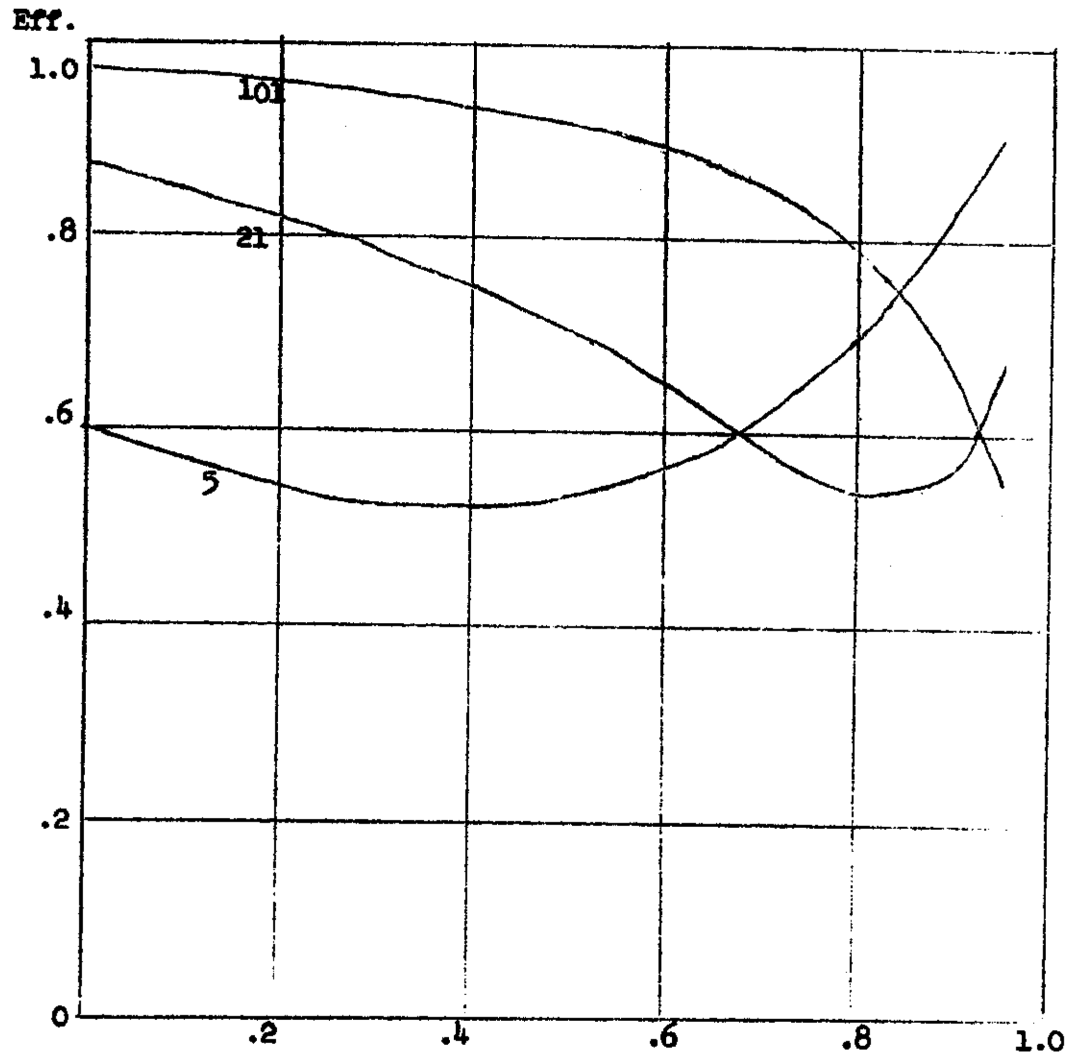

Diagram 1

The Efficiency of the Weighted Difference
Estimator for a Trend

The efficiencies of the remaining two estimators can now be more briefly examined. The values of the efficiency of the first difference estimator is shown in diagram 2 for the same three values of N; and in diagram 3 the efficiency of the classical estimator is shown for the first two of these values.

It is seen that for the typical case encountered in econometric work of twenty observations, the loss of efficiency due to using the first-difference estimator is less than a tenth as long as p is above 0.8. It follows that as long as we are fairly certain that p is in this region then it is not worth paying too much attention to estimating its value with any accuracy.

On the other hand, with 101 observations more attention has to be paid to the method of estimation if the information is to be used efficiently. Of course, it may be that with this many observations efficiency is no longer as important; but if that many observations are required in order to obtain a sufficiently accurate estimate, then a mistake in the specification of p will be more serious than if the analysis were based on fewer observations.

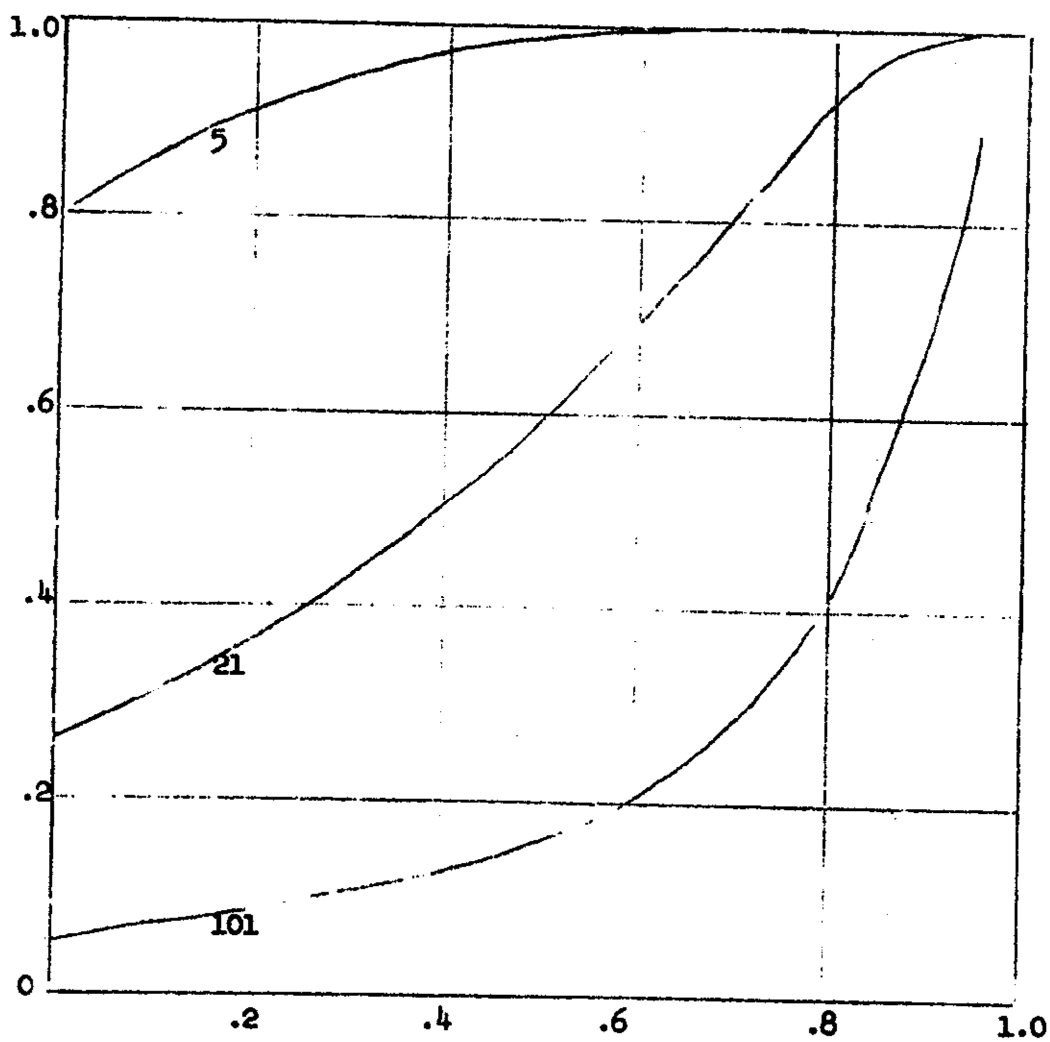

Diagram 2

Efficiency of the First-Difference Estimator

The efficiency curves for the classical estimator, however, show that for the values illustrated, the efficiency never falls below 0.8. Accordingly, if nothing at all is known about the true value of p a better estimate for the trend will be obtained in the long run if this estimator is used rather than the first-difference estimator.

The above comparisons have throughout excluded the case where the true

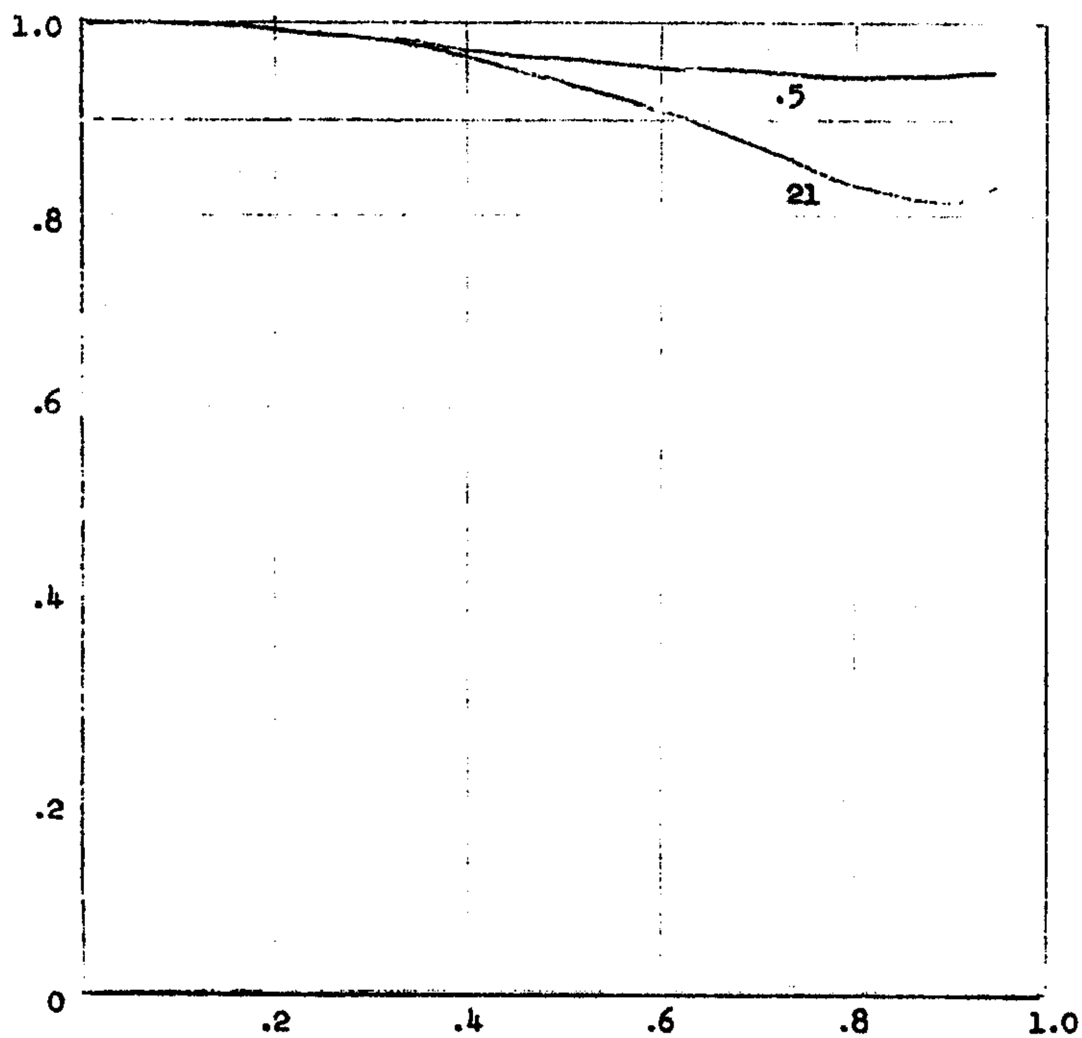

Diagram 3

Efficiency of the Classical Estimator

value of p is unity. In theory we could allow p to tend as close as we liked to unity, provided that it was finitely different from it, and the above arguments would still hold. This is however not a fruitful way of approaching the problem for, as well be seen in the next section, the case of $p = 1$ is a fundamentally different model from that so far considered.

7. Regression Analysis with Non-Stationary Errors

It has been customary in carrying out regression analyses of economic behavior based on time-series data to assume that the error terms are generated by a process that is stationary in the sense, already indicated above, that the variance of the error of each observation, and the covariances between the errors of observations separated by a given number of time periods, are independent of the origin of measurement of time. This assumption has seemed quite natural since there has been no reason to think that on the whole the degree of haphazardness of peoples' behavior varied substantially over the length of time for which data are generally available. The assumption of stationarity is, however, by no means essential for the purposes of regression analysis; indeed, the error scheme (2) of the problem mentioned at the head of this paper is non-stationary. We shall, accordingly, in this section examine some of the consequences of non-stationarity, present some reflections on the realism of the assumption and on its origins in the econometric context, and finally work out the implications for our problem of trend estimation.

We maintain the assumption that the error term is generated by the system (4) and shall now work out the covariance matrix of the errors without assuming stationarity. Reverting to the expansion (5), by taking the expected values of the squares of both sides, it is possible to express the variance of any error $\eta_{t+\tau}$ in terms of the variances of some "initial" value y_t and the intervening ϵ's. Thus

(27) $$V\left\{\eta_{t+\tau}\right\} = \sigma^2(1 + p^2 + p \quad + \dots + p^{2\tau-1}) + p^{2\tau} V\left\{\eta_t\right\} .$$

If we now express the variance of η_t as some multiple of the variance of ϵ, say

(28) $$V\left\{\eta_t\right\} = v_t \sigma^2$$

then, provided $p \neq 1$, the expression (27) may be reduced to the more concise form

(29) $$V\left\{\eta_{t+\tau}\right\} = \sigma^2\left[\frac{1-p^{2\tau}}{1-p^2} + p^{2\tau} v_t\right] .$$

The expression (27), however, is valid for all values of p including $p = 1$.

Similarly, it will be seen that for all values of p the covariance of two terms τ periods apart in time is

(30) $$\zeta\left\{\eta_{t+\tau}\, \eta_t\right\} = p^{\tau} V\left\{\eta_t\right\} = \sigma^2 p^{\tau} v_t .$$

The variances and covariances can therefore always be expressed in terms of p, σ^2 and some "initial" parameter v_t . Note that the results for the stationary case considered at the beginning of this paper emerge as a special case of these results when

(31) $$v_t = \frac{1}{1-p^2}$$

and v_t is independent of t.

We may next consider the origin and realism of the assumption of non-stationarity in econometric work and this is conveniently done by adopting, for a moment, a historical point of view. The classical, and simplest, assumption in regression analysis is that the errors in each observation are independent and of constant variance. But in applying regression analysis to time-series evidence has frequently been found that the errors are serially correlated; the question therefore arose of how to represent these intercorrelations in some simple way. The solution that has become popular is that in which the errors are generated by the first order Markoff scheme (4) with $p < 1$ and assuming stationarity.

This solution has the important advantage that the covariance matrix depends on only two independent parameters, σ^2 and p. The assumption of equal variance of errors about the regression line is maintained, and the correlation between two errors declines geometrically with this distance apart in time.

Now it may be that as a next stage in complicating the regression model it is necessary to drop the assumption that the error variance is independent of time. Thus, it may be found that when, for example, a regression analysis is carried out separately on the earlier and later halves of a set of time-series, the average variance of the residuals in the two halves differs significantly. In that case it may be appropriate to adopt a regression model which by assuming non-stationarity allows the introduction of a more complicated variance matrix of the errors which depends on the three parameters σ^2, p, and ν.

The use of non-stationarity schemes in econometric work does not, however, seem to have arisen from the kind of empirical approach just outlined, but has its origins in two rather less rigorous considerations. The first of these is that it has often been found that if regression analysis is carried out on the first differences of time series the residuals are rather more random than if the analysis is carried out on the original series. This led Cochrane and Orcutt to propose the error scheme (2). Shortly afterwards Wold [4] pointed out that this scheme was non-stationary and, hence, generally unrealistic in econometric applications.

The second line of thought is that exemplified in a recent paper by Gurland [2]. It is there supposed that the econometrician knows at which point in time the process (4) "started up" so that the value of ν could be calculated. For example, if it began at $t = 1$ with $\eta_o = 0$, then $V\left\{\eta_t\right\} = t\,\sigma^2$, and it would be known that $\nu_t = t$. It does not seem reasonable to suppose that the econometrician will have this kind of information available, at any rate in the kind of problem that he has been concerned with so far.

While the relevance of the assumption of non-stationarity is thus not at all well established it is however not inconceivable that problems may arise in which it is of importance. For this reason some results of trend estimation under this assumption are now given. Instead of proceding by the tedious method of reworking the results of the earlier sections of this paper, substituting the assumptions (27) and (30) for those given by (6), we shall examine three special cases which will bring out rather more clearly some of the interesting features of this assumption.

A. Suppose, first, that $p = 1$ and that we wish to obtain the minimum variance estimator of a trend line passing through a known origin. This is the simple problem mentioned at the beginning of this paper. Arguing, rigorously, along the lines of the argument given in section 4, we have to find $2N + 1$ transformed observations, the errors in which are independent of each other. The first $2N$ of these are the $2N$ differences between the observations, the errors of which have a variance of σ^2, and the remaining observation is the first, with a variance of $\nu\,\sigma^2$.

Assuming ν to be known we apply the classical estimating procedure to these transformed variables and find on taking the ratio of the sum of cross-products to the sum of squares

$$\hat{b}_1 = \frac{\sum_{-N+1}^{N} (y_t - y_{t-1}) \quad - Ny_{-N}/v}{2N + N^2/v}$$

$$= \frac{y_N - y_{-N} (1 + N/v)}{N(2 + N/v)}$$

This may be written as

$$\text{(32)} \qquad \hat{b}_1 = \frac{yN}{2N} \, \frac{2}{2 + N/v} \quad - \frac{y_{-N}}{2N} \, \frac{2(1 + N/v)}{2 + N/v}$$

for ease of comparison with the first-difference estimator

$$b_1 = \frac{y_N}{2N} - \frac{y_{-N}}{2N}$$

It will be seen that as $v \to \infty$ the two estimators become similar for given values of N. But for small values of N and v the two estimators are rather different in appearance. Thus, for three observations, that is $N = 1$, the first difference estimator is

$$b_1 = \tfrac{1}{2} y_1 - \tfrac{1}{2} y_{-1} ,$$

while the minimum variance estimator is:

$$\text{for } v = 1, \quad \hat{b}_1 = \tfrac{1}{3} y_1 - \tfrac{2}{3} y_{-1} ,$$

$$\text{and for } v = 2, \quad \hat{b}_1 = \tfrac{2}{5} y_1 - \tfrac{3}{5} y_{-N} .$$

B. The variance of the efficient estimator in the non-stationary case may be calculated as for the stationary case. It will be found that for a general p,

$$\text{(33)} \qquad V \left\{\hat{b}_p\right\} = \sigma^2 \Big/ \left[\tfrac{1}{3} N (2N^2 + 1) (1 - p)^2 + N^2/v + 2Np\right]$$

and this differs from the stationary case only in the middle term of the denominator which is here N^2/v and is $N^2(1 - p^2)$ in the stationary case.

C. Throughout this paper we have been concerned with trend estimation when the line passes through a known origin. One of the consequences of this, as has been seen, is that the first-difference estimator is not efficient even for $p = 1$, though it becomes efficient as the variance of the initial observation tends to infinity.

If, however, the origin is not known so that it were necessary to estimate both the parameters in the equation

$$(34) \qquad y_t = \alpha + \beta t + \eta_t$$

both the mathematics and the argument would become more complicated. When the economist is interested in estimating the slope of a line rather than its level, as is often the case, a comparison of the efficiencies of the estimates of β may be made much in the same way as when α is know to be zero.

One result must be noted in connection with the problem at the head of this paper in which $p = 1$. The minimum variance estimator of α and β may be found as in case A. It will then be seen that the 2N differences contain information only on β since the α's cancel out, and the first observation is therefore the only one that can be used to provide an estimator of α. Hence the efficient procedure is to obtain an estimate of β from the differences only; this will be

$$b = \frac{y_N - y_{-N}}{2N}$$

and will have a variance of $\sigma^2/2N$. The efficient estimator of α is then given by

$$a = y_{-N} + bN$$

the variance of which is

$$V(a) = N^2 V(b) + V(\eta_{-N})$$
$$= \sigma^2 (N/2 + \nu).$$

Hence, if a constant has to estimated as well as the slope of the line, the first-difference estimator provides an efficient estimate of the slope, but the constant term can only be estimated on the basis of more general considerations. On the other hand, it has been seen that if the constant is known then the first difference estimator is not efficient. Similar considerations apply for more general regression problems than the trend and the same result holds.

8. Some Conclusions

In this paper we have considered a number of problems that arise in analysing time-series by means of least-squares regression and have shown the precise implications for the special case of estimating a trend. The following are some of the conclusions that may be drawn from the discussion.

a) For the case of twenty observations that are typical in econometric investigations it does not appear that the estimators are highly sensitive to small changes in the assumptions.

b) This appears to be true even for the estimation of a trend by means of the first-difference formula which was seen to lead to extreme results. It was shown that even if the value of ρ was 0.8 (instead of the value 1.0 which is assumed by the first-difference estimator) the loss of efficiency is less than a tenth.

c) The weighted-difference estimator has in the past often been thought to be efficient. Our analysis has provided some insight into its efficiency, the main implication being that if the values of the determining variables for the first observation are rather different from their average values, then it is desirable to amend the computations and use the minimum variance estimator instead.

d) It was argued that the use of non-stationary error schemes is unrealistic in econometric contexts and it therefore appears that the method of applying regression analysis to the first differences of variables requires fresh examination.

APPENDIX

On a Property of Matrices Used in Weighted Difference Estimators

This Appendix gives a mathematical result which it is thought may be of interest to others working on this kind of problem.

Consider the general least-squares regression model which in a matrix notation may be written as

$$(1) \qquad y = X\beta + \epsilon.$$

If the covariance matrix of the ϵ's is $\sigma^2\Omega$ then the best linear unbiased estimator of β is

$$(2) \qquad b = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y$$

and this has a variance

$$(3) \qquad V(\hat{b}) = \sigma^2 (X'\Omega^{-1}X)^{-1}.$$

Suppose however that, perhaps because Ω is not known, a matrix $\Lambda \neq \Omega^{-1}$ (note that Λ may be singular) is used in the formula for the estimator so that instead of (2), the estimator is

$$(4) \qquad b^* = (X'\Lambda X)^{-1} X'\Lambda y.$$

The variance of b^* is then not given by (3) but by the rather more complicated expression

$$(5) \qquad V(b^*) = \sigma^2(X'\Lambda X)^{-1} X'\Lambda\Omega\Lambda X (X'\Lambda X)^{-1}$$

Now suppose that Ω is in fact given by the expression (13) in the main part of the paper and that the estimating process used depends on the taking of differences so as to transform the errors; that is, the estimator used is of the form (10). Thus, instead of using the matrix G to transform the variables, another matrix, H say, is used which differs from G only in that the last element of the leading diagonal is zero instead of $\sqrt{1 - p^2}$.

The estimating matrix Λ, which is now singular, is then given by

$$(6) \qquad \Lambda = H'H = \begin{pmatrix} 1 & -p & 0 & \cdots & 0 \\ -p & 1+p^2 & -p & & \\ 0 & -p & 1+p^2 & & \\ \vdots & & & 1+p^2 & -p \\ 0 & 0 & 0 & -p & p^2 \end{pmatrix}$$

and this differs from Ω^{-1} only in having p^2 as the least element instead of unity.

A little arduous multiplication will then show that the following remarkable property holds for this matrix:

$$(7) \qquad \Lambda \, \Omega \, \Lambda = \Lambda \, .$$

Accordingly, on substituting this result into (5) it is seen that

$$(8) \qquad V(b^*) = \sigma^2 (X'\Lambda X)^{-1}$$

that is, the expression for the variance has the same simple form (3) when the matrix Λ is used in the estimation process as when the correct matrix Ω is used. For this reason, as is otherwise obvious, the application of the classical regression formulae to the transformed variables leads to no difficulties, even though the estimator is not efficient.

The same result holds for the case when Ω is the variance matrix generated by a non-stationary process.

REFERENCES

[1] Cochrane, D., and Orcutt, G. H., "Application of Least Squares Regression to Relationship Containing Autocorrelated Error Terms", J. A. S. A., 44, 36.

[2] Gurland, J., "Autocorrelation of the Disturbances in Linear Regression", circulated as CCDP Statistics No. 368.

[3] Stone, R., et al., The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom, 1920-38, Cambridge, (forthcoming) 1954.

[4] Wold, H. "On Least Square Regression with Autocorrelated Variables and Residuals", Trans. I. S. I., Berne, 1949.