SOME APPROACHES TO THE ESTIMATION AND DISCOUNTING OF INTERPOLATED OBSERVATIONS

Roger J. BOWDEN[1]
University of Auckland

Research Memorandum No. 98
March 1976

The interpolation of economic time series has a long and consistently unfashionable history as an object of study. Nevertheless economists have from time to time been forced by the inadequacy of their data or the ambition of their constructs to consider ways of filling in the gaps where observations are missing. The problem known as "missing observations" has been fairly intensively researched in the statistical literature (for a review, see A.A. Afifi and E.M. Elashoff (1966)). The tenor of this work has been to design parameter estimators which compensate, as it were, for the missing data, rather than stressing the estimation of the missing observations themselves. The economist, however, is often interested in testing a number of alternative hypotheses, perhaps of differing functional form. To design a different estimation procedure for each particular hypothesis under test would represent a substantial and possibly unjustified use of research time. Thus it is useful to be able to estimate in a relatively simple way the missing observations themselves, and once having achieved the complete series, to know what modifications must be made to standard tests in screening a number of alternative hypotheses.

There are thus two related tasks to which the present paper addresses itself. The first is to find a relatively simple and general method of estimation of the observations to be interpolated. The use of a related series for which observations are readily available on the required basis was investigated by M. Friedman (1962). G.C. Chow and An-Loh Lin (1971) showed how to formalize and generalize this work in the context of least-squares estimators. To complement this work we need now to pay some attention to methods based on likelihood principles. The method of maximum-likelihood was applied by M. Drettakis (1973) in a missing-observations study, to generate estimates of observations missing at the start of a series for a simultaneous equation model. In the present paper we study the more usual situation where observations are missing in a regular way throughout the series. Although we investigate briefly a method based on true maximum likelihood principles, we reject at in favour of a "quasi-likelihood" procedure which perhaps has more in common with Bayesian techniques.

Given that one has obtained point estimates, with associated variances or covariances, there remains the problem of how the now-complete data series is to be deployed. Under the sort of research strategy described above, we might be interested in running regressions in order to test a set of different hypotheses. Intuition suggests that tests based on interpolated data ought not to be quite so conclusive as those employing a complete series. In other words, interpolated data should be discounted in some way, a problem which has not, to this author's knowledge, been raised in a systematic manner. As we shall see, the problem of estimation, on the one hand, and *discounting*, on the other, are intimately related. For the starting point for discounting formulas are the variances and covariances of the interpolations as yielded by the method of estimation. More than this, the approach adopted to the problem of estimation is of importance for the specification of the regression models, and hence for the meaning of the discounting process.

We consider in this paper the problem of interpolation, strictly defined. Typically this arises in connection with a stock series, such as the quantity of money at a particular point in time. The particular example we employ is the problem of interpolating observations on the net worth (wealth) of a community. A related problem (see Chow and Lin (1971)) is that of distribution, typically applicable to a flow series. Thus an available quarterly income series may have to be assigned to monthly observations, under the constraint that the sum of the monthly observations should add to the given monthly total. The limitation we impose in discussing only the pure interpolation problem is necessitated by space requirements, although we think that similar techniques, appropriately modified for the adding-up constraint, could be applied to the distribution problem.

The scheme of the paper is as follows. In section I we interpret and compare two distinct approaches based on the likelihood function. We work in terms of a linear auto-regressive process of the first order in which every second observation is to be interpolated. To infuse our discussion with some living substance we cast it as the problem of interpolating annual wealth observations. In section II we extend this basic work to cover (a) higher-order interpolation - say, every second and third observation to be interpolated (b) higher-order autoregressive formulas and (c) to the case where interpolative formula may be non-linear, based in this instance on a more realistic approach to the problem of interpolating a wealth series. We include also (d) a short discussion of some sampling problems. Finally, section III contains the ideas and techniques associated with the problem of discounting the interpolated data in carrying out regressions.

I. The Estimation Problem

In a static situation, the problem of interpolation using a related series is a straightforward application of linear regression theory (see e.g. A.S. Goldberger (1964) pp.167-171). We shall instead choose an interpolative procedure based on an autoregressive relationship, for it would seem a strange series indeed if prior and later values of the series to be interpolated were not to affect (over and above what can be explained by a presumed relationship with a related series) the observation to be interpolated. To give the discussion a little flesh and blood we shall work in terms of a concrete but simple problem; it will, however, be apparent that our treatment of this problem is quite general so far as a first-order autoregressive system is concerned, and higher-order systems will be treated later.

Suppose, then, that direct observations have been made at annual intervals of a wealth (net worth) series W for a community, but that 6-monthly figures are required for the project under investigation. We have available figures on the required basis for income y, consumption c, and a rate of interest variable (or index) r, which we believe representative of the movement of yields in the economy as a whole. We might then write the model:

$$(1) \qquad w_t = (1 + r_t)(w_{t-1} - c_t) + y_t + \varepsilon_t,$$

where the subscript t refers to 6-month periods, and ε_t is a disturbance term. As an estimate of the LHS the RHS of (1) will be subject to various errors. Thus the discrete nature of (1) may not accurately reflect the process of continuous discounting[2]. The choice of the index r_t also involves an index-number problem[3]. One would presume that the error involved here would be smaller, the more perfect the capital market involved. There may also be errors in the observation of the flow quantities c_t and y_t, for such magnitudes are usually more accurately estimated on an annual basis. We allow for such errors by adding the disturbance term ε_t. In the present section we shall simply specify that $E\varepsilon_t = 0$, $E\varepsilon_t^2 = \sigma^2$ and ε_t serially uncorrelated. A more realistic treatment of sources of error is sketched in section II(c); our purpose at the moment is expository. Finally, we suppose that observations $w_1, w_3, w_5, \ldots w_T$ are available, so that T is odd. Our object is to estimate $w_2, w_4 \ldots w_{T-1}$, and to attach estimates of error thereto.

Let us, then, consider equation (1) and suppose for the moment that the missing observations were in fact known. The Jacobian of the transformation from e to w is unity. Thus we can construct a likelihood function as

$$(2)\quad L = \frac{1}{(2\pi\sigma^2)^{\frac{T}{2}}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{t=2}^{T} (w_t - (1+r_t)w_{t-1} - x_t)^2,$$

where for the sake of convenience we have written $x_t = y_t - (1+r_t)c_t$. Now suppose, on the other hand, that every second observation is not known. Our first method of estimation proceeds by simply treating $w_2, w_4, \ldots, w_{T-1}$ as parameters and maximising (2) with respect to these unknown quantities. The resulting estimates, which are (as we shall see) not true maximum-likelihood estimates we shall call quasi-likelihood (QL).

Thus, differentiating (2) (or better, its logarithm) with respect to the parameters $w_2, w_4, \ldots$ we obtain

$$(3)(a)\quad \hat{w}_2 = \frac{1}{1+(1+r_3)^2}((1+r_2)w_1 + (1+r_3)w_3 + x_2 - (1+r_3)x_3)$$

$$(3)(b)\quad \hat{w}_4 = \frac{1}{1+(1+r_5)^2}((1+r_4)w_3 + (1+r_5)w_5 + x_4 - (1+r_5)x_5)$$

$$\ldots,$$

as the resulting estimators. These are independent of the value of $\hat{\sigma}^2$, which can be obtained in terms of the now-complete series as

$$\frac{1}{T}\sum_{t=2}^{T}(w_t-(1+r_t)w_{t-1}-x_t)^2.$$

The reader can verify that these estimators satisfy the elementary consistency requirement that if $r_t = r$, $y_t = y$, $c_t = c$ for all t and $\varepsilon_t = 0$, they correctly reproduce the values $w_2, w_4, \ldots, w_{T-1}$.

We turn now to the interpretation and justification of these QL estimators. Let us consider the following representative cell (it will be useful for later comparison to choose w_4 instead of w_2 as the observation to be interpolated):

$$\text{(4)(a)} \qquad w_4 = (1+r_4)\bar{w}_3 + x_4 + \varepsilon_4$$

$$\text{(4)(b)} \qquad \bar{w}_5 = (1+r_5)w_4 + x_5 + \varepsilon_5,$$

where, to emphasize their fixed or given character, we have placed bars over the available observations. Because w_4 enters into the formation of (4)(b), and $\bar{w}_5$ is fixed, the values of ε_4 and ε_5 will, conditional upon $\bar{w}_3$, $\bar{w}_5$, no longer be independent. Indeed they must vary along the straight line obtained by eliminating w_4 from (4)(a) and (b), namely

$$\text{(5)} \quad \varepsilon_5 + (1+r_5)\varepsilon_4 = \bar{w}_5 - (1+r_5)(1+r_4)\bar{w}_3 - x_5 - (1+r_5)x_4.$$

Letting f(.) denote the ordinates of the normal distribution with mean 0 and variance σ^2, the above consideration enables us to write the conditional density:

$$\text{(6)} \quad p(w_4:\bar{w}_3,\bar{w}_5) = \frac{f(w_4-(1+r_4)\bar{w}_3-x_4)\text{x}f(\bar{w}_5-(1+r_5)w_4-x_5)}{\int_{-\infty}^{\infty} f\text{x}f \, dw_4}$$

where in the bottom line fxf denotes the top line. Since the process is Markovian, the conditional density (6) is the same as $p(w_4 : \bar{w}_1, \bar{w}_3, \ldots, \bar{w}_T)$.

The term fxf can now be simplified by completing the square on w_4, a familiar process in the manipulation of normal distributions. Once this is done, the integration on the bottom line of (6) follows immediately from the property that the integral of any pdf over its range is unity. The whole can then be expressed as:

$$(7)\quad p(w_4 : \bar{w}_1, \bar{w}_3, \ldots, \bar{w}_T) = \frac{1}{\left(2\pi \frac{\sigma^2}{(1+(1+r_5)^2}\right)^{\frac{1}{2}}} \exp\left(\frac{-1}{\frac{2\sigma^2}{(1+(1+r_5)^2}}(w_4 - \hat{w}_4)^2\right),$$

where $\hat{w}_4$ has already been defined in (3)(b). Thus the conditional distribution of w_4 is normal with mean equal to the QL estimator and variance given by $\frac{\sigma^2}{1+(1+r_5)^2}$.

The joint conditional distribution of all the missing observations, $p(w_2, w_4, \ldots, w_{T-1} : \bar{w}_1, \bar{w}_3, \ldots, \bar{w}_T)$ is obtained by multiplying together terms like (7), for under the Markovian assumption all the cells like (4) are independent. More germane to our argument, it also equal to the product of the original terms like (6). The denominator of these terms consist only in functions of the given observations $\bar{w}_1$, $\bar{w}_3$, ..., $\bar{w}_T$, for the variables to be estimated have been integrated out. The result is therefore proportional to the quasi-likelihood function (2). Thus maximising the latter with respect to w_2, $w_4, \ldots,$ w_{T-1} is equivalent to finding the maximum ordinate, i.e. the mode of the conditional density $p(w_2, w_4, \ldots, w_{T-1}; \bar{w}_1, \bar{w}_3, \ldots, \bar{w}_T)$. The result is rather reminiscent of a Bayesian estimator based on the mode of the posterior distribution, although the usual Bayesian calculus has not been employed to obtain the posterior. In the present instance, the linearity of (1) ensures that the resulting conditional density is itself normal. Thus the mode coincides with the mean, as we have just discovered in equation (7).

In section II(c) however, we exhibit an example where this may not be the case.

In view of the above interpretation, it is natural to choose as the variance of the QL estimator the variance of the conditional density (or "posterior"). We would thus take, for example

$$\mathrm{Var}(\hat{w}_4) = \frac{\sigma^2}{1+(1+r_5)},$$

which can be estimated by replacing σ^2 by its own QL estimator as outlined earlier. Such a procedure is acceptable where the conditional density is symmetric, but where this is not the case difficulties arise which are postponed until section II(c).

It is convenient at this point to note a short method for obtaining the variance. Starting with equations (4)(a) and (b), take deviations from the conditional means of each variable to obtain:

(8)(a) $$\Delta w_4 = \Delta\varepsilon_4$$

(8)(b) $$0 = (1+r_5)\Delta w_4 + \Delta\varepsilon_5.$$

The conditional variance of $\Delta\varepsilon_4$ is the same as that of ε_4. This fact, together with the normality and independence of the unconditional distribution for ε_4 and ε_5, means that to obtain the conditional variance of Δw_4, we add $\Delta\varepsilon_4^2 + \Delta\varepsilon_5^2$ to obtain the coefficient of Δw_4^2 in the exponent of e in (6) as $1+(1+r_5)^2$. The variance of w_4 is thus $\frac{\sigma^2}{1+(1+r_5)^2}$. This shortcut is quite general and will later be useful.

This completes our preliminary discussion of the QL method. At this point it is useful to ask if there exists a true maximum-likelihood estimator. The observable data is w_1, $w_3,\ldots,w_T$. We ask what values $w_2,w_4,\ldots,w_{T-1}$ would maximize the probability of obtaining this observed series. This involves two steps: first, finding the conditional density for $w_1, w_3, \ldots, w_T$ given $w_2, w_4, \ldots, w_{T-1}$, and second, choosing values of the latter to maximize the probability of obtaining the values $w_1, w_3, \ldots w_T$ actually observed.

The first step is the mirror image of the development for the QL estimator. A typical cell corresponding to (4)(a) and (b) is now

(9)(a) $$w_3 = (1+r_3)w_2^* + x_3 + \varepsilon_3$$

(9)(b) $$w_4^* = (1+r_4)w_3 + x_4 + \varepsilon_4.$$

The asterisk denotes values that are now assumed fixed, and we ask for the conditional density of w_3, given w_2^* and w_4^*. Just as in the dual treatment for the QL estimate, we conclude that w_3 is conditionally normally distributed with mean:

$$E(w_3 : w_2^*, w_4^*) = \frac{(1+r_3)w_2^* + (1+r_4)w_4^* + x_3 - (1+r_4)x_4}{1+(1+r_4)^2}$$

and

$$\text{Var}(w_3 : w_2^*, w_4^*) = \frac{\sigma^2}{1+(1+r_4)^2} \quad .$$

Similarly we obtain $p(w_5 : w_4^*, w_6^*)$ and so on. We can then write the likelihood function in terms of such elementary components as:

(10) $$L(w_1, w_3, w_5 \ldots w_T : w_2^*, w_4^* \ldots w_{T-1}^*) = p(w_1 : w_2^*, w_4^*) \text{x} p(w_3 : w_2^*, w_4^*) \text{x} \ldots$$
$$\ldots \text{x} p(w_T : w_{T-1}^*, w_{T+1}^*),$$

where we have had to introduce an additional nuisance parameter w_{T+1}^*. The likelihood function has now to be maximized with respect to the unknown quantities - now regarded as true parameters - $w_2^*, w_4^*, \ldots, w_{T-1}^*$. Things are now a little more complicated, for it will be noted that a term such as w_4^* appears in two of the individual multiplicands on the RHS of (10). Upon deriving the normal equations we find that the resulting interpolation formulas are multipoint. For example, differentiating (10) with respect to w_4^*, we derive the following 4-point formula:

$$(11)\quad \hat{w}_4 = \frac{1}{\frac{(1+r_4)^2}{1+(1+r_4)^2} + \frac{(1+r_5)^2}{1+(1+r_5)^2}} \Big(- \frac{(1+r_4)(1+r_5)}{1+(1+r_4)^2} . w_2 + (1+r_4) w_3 + $$

$$+ (1+r_5) w_5 - \frac{(1+r_5)(1+r_6)}{1+(1+r_6)^2} w_6 - \frac{(1+r_4)}{1+(1+r_4)^2} x_3 +$$

$$+ \frac{(1+r_4)^2}{1+(1+r_4)^2} x_4 - \frac{(1+r_5)}{1+(1+r_6)^2} x_5 + \frac{(1+r_5)(1+r_6)}{1+(1+r_6)^2} x_6 \Big).$$

Since values of w_2 and w_6, themselves to be estimated, appear in (11), the whole system has to be solved simultaneously. This is, however, easily done with a simple Seidel iteration. A more serious objection is that although the asymptotic variance-covariance matrix of the estimators can in principle be obtained by differentiating twice the information matrix, this is in practice hardly feasible.

Nevertheless is is interesting to compare the performance on a test problem of the two estimators, QL and true ML. Accordingly, we set up a small simulation experiment. Starting from a given value for w_1, seven further observations $w_2, w_3, \ldots, w_8$ were generated from equation(1), employing a random number generator for the series ε_t, $t=2\ldots8$. We assumed that $c_t = 0.8y_t$. The figures employed for w_1, $y_2 \ldots y_8$, $r_2 \ldots r_8$ are listed in footnote 4. We set σ, the standard deviation of the disturbance term at 200. Having thus obtained $w_1, w_2, w_3, \ldots, w_8$, we now assumed that w_4 and w_6 were unknown and were to be interpolated. The deviations $\hat{w}_4 - w_4$ or $\hat{w}_6 - w_6$ could then be computed. Each such experiment was replicated 19times, so that 20 runs were available. Mean deviations, mean absolute deviations and mean square deviations were calculated over the 20 runs. For the ML estimator, we investigated two cases. In the first, both w_4 and w_6 were unknown; in the second only w_4 was unknown. The object in this distinction was to investigate whether the simultaneity problem referred to above made a great deal of difference to the performance of the estimator. Comparative results are reproduced in Table I.

Table I

Estimator:	of	Mean Deviation	Mean Absolute Deviation	$(\text{Mean Square Deviation})^{\frac{1}{2}}$
QL	w_4	-39.0	139.7	170.7
	w_6	20.2	93.6	119.8
ML	w_4	-611.3	611.3	637.3
	w_6	-1639.8	1639.8	1650.6
ML	w_4 (w_6 known)	-1029.0	1029.0	1051.6

It is at once apparent that the ML estimator is so badly biased that it can be ruled out of consideration as a serious candidate. Its performance may be particularly poor in the context of an explosive system such as (1), since as (11) shows, the estimator reaches forward further than does the QL estimator.

Assuming, then, that QL remains as the less objectionable estimator, we can pause to take stock of some general considerations. We have earlier remarked that if the distribution of the disturbance terms is normal, and the model linear, the QL estimate will be identical with the conditional expectation of the observation to be interpolated, given the available observations. (To be sure, we have so far considered only first-order interpolation in first-order autoregressive processes, but as we shall later show, these assumptions are not critical.) Thus in such a case, the QL estimate will be identical to the least-squares interpolator (based on, say, Wiener-Hopf theory), for the latter has the property of being identical with the conditional expectation. The advantages of the QL formulation are twofold. First they allow an easy, not to say automatic, derivation of the estimating formula. Second, they can be obtained in non-linear situations where the least-squares estimates may be difficult or impossible to derive. In such a circumstance it may no longer be true that QL estimates are identical with the conditional expectations and thus with the least-squares estimator. But they do not lose thereby the property of being a sensible and meaningful interpolation procedure.

II. The QL-Estimator: Further Considerations

(a) Higher-Order Interpolation

Let us continue with the example of § 1, but suppose instead that every second and third observation has to be interpolated. Thus, $\bar{w}_1, \bar{w}_4, \bar{w}_7, ..$ are known, and the object is to interpolate $(w_2, w_3), (w_5, w_6), ...$ Consider the cell $\bar{w}_1, w_2, w_3, \bar{w}_4$ as representative. Setting up the QL function yields the following estimators:

$$\hat{w}_2 = \frac{(1+r_2)\bar{w}_1 + (1+r_3)\hat{w}_3 + x_2 - (1+r_3)x_3}{1+(1+r_3)^2}$$

$$\hat{w}_3 = \frac{(1+r_3)\hat{w}_2 + (1+r_4)\bar{w}_4 + x_3 - (1+r_4)x_4}{1+(1+r_4)^2} .$$

It is evident that solution for $\hat{w}_2$, $\hat{w}_3$ is now simultaneous. Since the equations are linear no problems are involved but there is an associated point of identification. In the present instance, for a solution to exist, we must have

$$\det \begin{bmatrix} 1 & \dfrac{-(1+r_3)}{1+(1+r_3)^2} \\ \dfrac{-(1+r_3)}{1+(1+r_4)^2} & 1 \end{bmatrix} \neq 0,$$

which is always true.

Once again it turns out that the conditional density of the interpolations given the available observations is multivariate normal, in this case with a covariance matrix consisting in 2x2 blocks down the diagonal. The covariance matrix of the representative elements $\hat{w}_2$, $\hat{w}_3$ can be obtained by the short-cut mentioned in section I. Thus, taking deviations around conditional expectations, we obtain from the cell $\bar{w}_1$, w_2, w_3, $\bar{w}_4$:

$$\Delta w_2 = \Delta\varepsilon_2$$

$$\Delta w_3 = (1+r_3)\Delta w_2 + \Delta\varepsilon_3$$

$$0 = (1+r_4)\Delta w_3 + \Delta\varepsilon_4.$$

Thus $\Delta\varepsilon_2^2 + \Delta\varepsilon_3^2 + \Delta\varepsilon_4^2 = (1+(1+r_3)^2)\Delta w_2{}^2 - 2(1+r_3)\Delta w_2\Delta w_3 +$

$$+ (1+(1+r_4)^2)\Delta w_3^2,$$

from which the required covariance matrix for $\hat{w}_2$, $\hat{w}_3$ follows immediately as:

$$(12) \quad \sigma^2 \begin{bmatrix} 1+(1+r_3)^3 & -(1+r_3) \\ -(1+r_3) & 1+(1+r_4)^2 \end{bmatrix}^{-1} =$$

$$= \frac{\sigma^2}{1+(1+r_4)^2(1+(1+r_3)^2)} \begin{bmatrix} 1+(1+r_4)^2 & 1+r_3 \\ 1+r_3 & 1+(1+r_3)^2 \end{bmatrix}.$$

It is interesting to compare the higher-order variances with those obtained when the interpolation of only one observation is considered. Thus suppose we knew $\bar{w}_1$, $\bar{w}_3$. We would then have

$$\mathrm{Var}(\hat{w}_2:\bar{w}_1,\bar{w}_3) = \frac{\sigma^2}{1+(1+r_3)^2}.$$

When only $\bar{w}_1$, $\bar{w}_4$ are known we obtain from (12):

$$\mathrm{Var}(\hat{w}_2:\bar{w}_1,\bar{w}_4) = \frac{\sigma^2}{1+(1+r_3)^2} \cdot \frac{1}{1-\dfrac{(1+r_3)^2}{1+(1+r_4)^2}}.$$

The new variance is therefore a little more than twice the variance of first-order interpolation. For example if we are interpolating quarterly observations on a monthly basis with $r_3 = r_4 = 0.625$ % (an annual rate of 7.5 %), the factor involved is 2.0125.

(b) Higher-Order Autoregressions

We illustrate some general considerations with the model:

$$(13) \qquad y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 X_t + \varepsilon_t ,$$

with the usual specification on ε_t. Observations are available for y_1, y_3, y_5, ..., y_T, and it is desired to interpolate y_2, y_4, ..., y_{T-1}. Start this time with the end-cell:

. . .

$$(15)(c) \quad \bar{y}_{T-2} = a_1 y_{T-3} + a_2 \bar{y}_{T-4} + a_3 X_{T-2} + \varepsilon_{T-2}$$

$$(15)(b) \quad y_{T-1} = a_1 \bar{y}_{T-2} + a_2 y_{T-3} + a_3 X_{T-1} + \varepsilon_{T-1}$$

$$(15)(a) \quad \bar{y}_T = a_1 y_{T-1} + a_2 \bar{y}_{T-2} + a_3 X_T + \varepsilon_T .$$

A cell of this kind is no longer self contained. For y_{T-3}, which appears in (15)(b) depends upon previous values of the disturbance terms. There is now an implied constraint on disturbance terms which can be obtained as follows. Write

$$(16) \quad K_T = \bar{y}_T - (a_1^2+2a_2)\bar{y}_{T-2} + a_2^2\bar{y}_{T-4} - a_3(X_T+a_1X_{T-1}+a_2X_{T-2}),$$

with corresponding expressions for K_{T-2}, K_{T-4}, ..., K_3. Considering the cell given by (15) above, we obtain the constraint:

$$(17)(a) \qquad \varepsilon_T - a_1\varepsilon_{T-1} - a_2\varepsilon_{T-2} = K_T$$

Similarly, considering the cell prior to this, we obtain

(17)(b) $\varepsilon_{T-2} - a_1\varepsilon_{T-3} - a_2\varepsilon_{T-4} = K_{T-2}$.

We thus obtain a set of planes, given by (17)(a), (b)... . By successive elimination of ε_{T-2} from (a), ε_{T-4} from (b),..., we obtain the equation of their intersection as:

$$(18)\quad \varepsilon_T - a_1(\varepsilon_{T-1}+a_2\varepsilon_{T-3}+a_2^2\varepsilon_{T-5}+\ldots+a_2^{\frac{T-3}{2}}\varepsilon_2) =$$

$$= K_T+a_2K_{T-2}+a_2^2K_{T-4}+\ldots+a_2^{\frac{T-3}{2}}K_3.$$

The derivation of the conditional density now follows along the lines of section I, with equation (18) playing a similar role to that of equation (5) of that section. We obtain

$$(19)\quad p(y_2,y_4\ldots y_{T-1}:\bar{y}_1,\bar{y}_3,\ldots,\bar{y}_T) = \frac{f(\varepsilon_2^o)xf(\varepsilon_3^o)x\ldots xf(\varepsilon_T^o)}{\int_{-\infty}^{\infty}\ldots\int(fxfx\ldots xf)dy_2dy_4\ldots dy_{T-1}},$$

where $\varepsilon_t^o = y_t-a_1y_{t-1}-a_2y_{t-2}-a_3X_t$. Since the term on the bottom is a constant depending only upon the known values $\bar{y}_1,\bar{y}_3,\ldots,\bar{y}_T$, the QL estimate will once again yield the mode of the conditional density; and in the linear model (13) the mode will once again coincide with the mean.

The only difficulty involved is that the estimators, and the above conditional density, depend upon an initial value y_o, which we have assumed fixed. This will not usually be available, and must be arbitrarily assigned (although a Bayesian approach assigning it a prior distribution may be possible). If the difference equation (13) is stable, this will be of importance only at the start of the series, since the effect of initial errors will be gradually damped in the solution for $\hat{y}_2$, $\hat{y}_4$,... . If, however, the model is unstable this may not be the case and it may be desirable to try several values for y_o in what may _de facto_ comprise a Bayesian approach to the resolution of the effects of this nuisance parameter.

(c) Nonlinearities

Let us return to the problem of interpolating wealth observations on a 6-monthly basis. To suppose, as we did in section I, a homoscedastic error term is a little unrealistic. For if the rate of interest variable were an imperfect indicator of the true rate (see footnote 3), the resulting additive error term would be heteroscedastic, depending on w_t. Likewise additive errors in y_t or c_t may increase steadily in time, in line with w_t and the growth of the economy. The following nonlinear transformation enables us to circumvent this problem. As we shall see, however, some general considerations emerge which must be discussed.

Let us assume that, although reliable incomes data may be available on the basis of income tax returns, consumption data on the required basis bannot be obtained. We might accordingly specify that $c_t = \gamma y_t$, where γ is an unknown constant. On the other hand, individual components of wealth may be readily available on the required basis, such as the stock of outside money. Starting with the relationship

$$w_t = (1+r_t)w_{t-1} + \rho_t y_t , \qquad \rho_t = 1-\gamma(1+r_t)$$

we obtain the approximation

$$\log w_t \simeq \log w_{t-1} + r_t + \frac{\rho_t y_t}{w_{t-1}} .$$

Since the index r_t is imperfectly known but appears in the factor ρ_t, we separate it from the latter by means of a Taylor approximation. At this stage we can also add other variables (such as $\log M$, where M is the stock of money, appropriately defined), represented by the factor X_t. The final estimating equation might thus read:

$$(20)\quad \log w_t = \log w_{t-1} + \alpha_1 r_t + \frac{\alpha_2 y_t}{w_{t-1}} + \alpha_3 X_t + \varepsilon_t ,$$

where the α_i are constants to be estimated and ε_t is the error term, assumed homoscedastic. The above equation is not linear, either in w_t on in $\log w_t$. We shall assume that the problem is to interpolate values of $\log w_t$, since wealth often appears in multiplicative fashion, as in equations for portfolio proportions. We can now form the quasi-likelihood function in terms of $\log w_t$. The transformation $\varepsilon_t \rightarrow \log w_t$ has Jacobian unity, so that the form of the QL function is as in the preceding sections. Differentiating with respect to $\log w_2$, $\log w_4$ ... $\log w_{T-1}$, and with respect to the remaining parameters α_1, α_2, α_3, σ^2, we obtain as usual the normal equations. Those for α_1, α_2 and α_3 are identical with ordinary least squares applied to (20) on the presumption that all w_t are known. The equation for σ^2 yields the usual likelihood estimate.

For the interpolations themselves, we obtain equations like:

$$(21)\quad \widehat{\log w_2} = \frac{1}{(2-\frac{\alpha_2 y_3}{w_2})} (\log w_1+\alpha_1 r_2+\alpha_2\frac{y_2}{w_2}+\alpha_3 X_2+(1-\alpha_2\frac{y_3}{w_2}).$$

$$.(\log w_3-\alpha_1 r_3-\alpha_2\frac{y_3}{w_2} -\alpha_3 X_3)).$$

Such an equation is easy to solve iteratively for $\log w_2$. Note, however, that the resulting estimate will no longer be equal to the mean of the conditional density for $\log w_2$. For it no longer follows that if the ε_t are normal, so too are the $\log w_t$. Nevertheless, estimates like that defined by (21) continue to represent the mode of the conditional density. Obtaining the mean of this density - equivalently, the least squares estimator - may be difficult if not impossible, but the QL method continues to provide a sensible estimation procedure.

The possible non-coincidence of mode with mean complicates the treatment of variance questions. Ideally what is now required is the mean square error about the mode (eg $\widehat{\log w_2}$). As a first step, one can approximate this with the variance about the mean obtained as follows: Consider the cell

$$(22) \qquad \log w_2 = \log\bar{w}_1 + \alpha_1 r_2 + \alpha_2 \frac{y_2}{w_1} + \alpha_2 X_2 + \varepsilon_2$$

$$\log \bar{w}_3 = \log w_2 + \alpha_1 r_3 + \alpha_2 \frac{y_3}{w_2} + \alpha_3 X_3 + \varepsilon_3 .$$

Using the relationship $\frac{d}{d(\log w_2)} = w_2 \frac{d}{dw_2}$, we can take deviations about the conditional mean to obtain, approximately:

$$\Delta(\log w_2) = \Delta\varepsilon_2$$

$$0 = \Delta(\log w_2) - \alpha_2 \frac{y_3}{w_2^*} \Delta \log w_2 + \Delta\varepsilon_3 ,$$

where $w_2^* = \exp(E(\log w_2))$. We thus obtain

$$(23) \quad \mathrm{Var}(\log w_2) = - \frac{\sigma^2}{1+(1-\frac{\alpha_2 y_3}{w_2^*})^2} .$$

The factor w_2^*, which depends upon the true mean is unknown, but can be estimated in terms of the mode as $\exp(\widehat{\log w_2})$. However the use of the variance as obtained above in situations which may call for the mean square error will involve underestimation to the extent of the square bias. This said, the use of the variance does at least have the virtue of providing a lower bound for the true mean square error.

(d) Sampling Problems: Consistency

The example treated in section I exhibits an unusual amount of information regarding the structural parameters of the underlying generating mechanism. As our discussion under (c) above suggested, the more common situation is where in order to apply the interpolating formula representing say the appropriate conditional expectation or mode, one has first to estimate some of these structural parameters. This immediately introduces the new question of sampling variability of such estimators, and its effects on the interpolations themselves. The subject is, however, a very large one and the following represents no more than a sketch of some of the salient points.

To illustrate those points, it is useful to take the following example:

(24) $$y_t = \alpha y_{t-1} + \beta x_t + u_t.$$

The parameters α and β are not known. Assuming that we have to interpolate every second observation, the appropriate interpolator (I) is:

(25) $$y_t^I = \frac{1}{1+\alpha^2}(\alpha(\bar{y}_{t+1}+\bar{y}_{t-1}) + \beta(x_t - \alpha x_{t-1})).$$

Since α, β are unknown, the QL method calls for a two-stage routine. Denote by $\hat{y}_t$ an estimate of the complete interpolated series. In stage one, we solve the normal equations:

(26) $$\begin{bmatrix} \sum_{t=2}^{T} \hat{y}_{t-1}^2 & \Sigma \hat{y}_{t-1} x_t \\ \Sigma \hat{y}_{t-1} x_t & \Sigma x_t^2 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \sum_{t=2}^{T} \hat{y}_t \hat{y}_{t-1} \\ \Sigma \hat{y}_t x_t \end{bmatrix}.$$

These will be recognised as those of ordinary least squares applied to (24) as if y_t were known to be $\hat{y}_t$. In stage two we return with the calculated values $\hat{\alpha}$, $\hat{\beta}$ to obtain from (25) a new series $\hat{y}_t$. Some other standard iteration procedure, such as the Newton-Raphson method, may alternatively be employed. However, we have found that the above simple two stage procedure has acceptable convergence properties.

Nevertheless, the nonlinearity of the estimation procedure gives rise to new difficulties. Basically, in estimating α and β above we are attempting to estimate the conditional expectation given by (25). Is it true that $\hat{\alpha}$ and $\hat{\beta}$ as defined by the above procedure are consistent estimates - and thus the resulting $\hat{y}_t$ a consistent estimate of the conditional expectation as given by (25)? In the context of our example, it appears[5)] that the answer is in the affirmative, provided that $|\alpha|<1$ and the error process u_t serially uncorrelated. On the other hand, we might step outside the confines of our particular example and raise the question of the unbiasedness of the interpolative procedure, in the sense of preceding discussion. We conjecture that a biased interpolating procedure will not yield consistent estimates either for structural parameters or else for the object of basic interst, namely the mode of the conditional distribution for the observations to be interpolated.

Related to the above issue is the question of obtaining a sampling estimate of the variance of the interpolator. In general, we can write the following error decomposition:

$$(27) \qquad \hat{y}_t - y_t = (\hat{y}_t - y_t^I) + (y_t^I - Ey_t) - (y_t - Ey_t),$$

where Ey_t refers to the expectation of y_t conditional upon available information. Assuming that the procedure is unbiased, (27) becomes

$$(28) \qquad \hat{y}_t - y_t = (\hat{y}_t - Ey_t) - (y_t - Ey_t).$$

It is possible to obtain working approximations to the variances of the two terms on the RHS of this equation. Thus in the above example we could obtain an (under-) estimate of VAR($\hat{y}_t$-Ey_t) by considering corresponding formulas for the variances and covariance of $\hat{\alpha}$ and $\hat{\beta}$ in conjunction with (25). To obtain the covariance matrix of $\hat{\alpha}$ and $\hat{\beta}$, one can consider the OLS normal equations (26) based on (25) under the assumption that the true series y_t is approximated by the final estimate $\hat{y}_t$. The second component, Var(y_t-Ey_t), can be estimated by methods similar to those of preceding discussion. What is at the moment lacking is an expression for the covariance of the two error components of (28). The simplest working hypothesis is to adopt the linear hypothesis result, and assume that this covariance is zero. The resulting expression is the sum of the two components corresponding to the elements on the RHS of (28).

III. Discounting

Suppose that point interpolators, together with estimates of their variances and perhaps covariances have been obtained, and we are considering the use of such a series in running a set of ordinary least squares regressions to test alternative hypotheses. It is intuitively an appealing principle that the use of interpolated observations should result in a loss of power of these tests. In this section we consider the problem of how to discount the observations which have been interpolated in order to provide more correct testing procedures. We shall suppose that we are to regress a dependent variable y on a set $X_1 \ldots X_K$ of independent variables, including possibly the constant term. One of these variables, say variable i, has to be interpolated. For the moment we shall suppose that every second observation has to be estimated. It is assumed that an unbiased estimation procedure has been employed in the sense that the interpolated values are equal to the means of their respective conditional density functions, given the available observations.

We shall first have to consider fairly carefully the specification of the substantive regression. We shall make the conventional assumption that $X_1 \ldots X_{i-1}$, $X_{i+1} \ldots X_K$, and the available observations on variable X_i, are all fixed in repeated realizations of the system; alternatively, the error term is distributed independently of these variables, and estimation is conditional upon fixed values of the latter. With regard to the interpolated observations, we can write

(29) $$\hat{X}_{2i} = \hat{X}_{2i}(X_{1i}, X_{3i}, \ldots, X_{Ti}; z),$$

$$\hat{X}_{4i} = \hat{X}_{4i}(X_{1i}, X_{3i}, \ldots, X_{Ti}; z),$$

$$\cdots$$

where z indicates other variables employed to estimate the missing observations, which are likewise supposed fixed. Thus under our specification on the available observations, the estimates $\hat{X}_{2i}$, $\hat{X}_{4i}$ ... will themselves be fixed in repeated realizations.

Assume now that every time a drawing is taken from the distribution of the disturbance vector u_t of the regression, so too is a new drawing taken from the conditional distribution of X_{2i}, X_{4i} ... given the (fixed) available observations. In the latter case we can thus write $X_{ti} = \hat{X}_{ti} + \varepsilon_{ti}$, for $t = 2,4,\ldots,T-1$, and $\text{Var}\ \varepsilon_{ti} = \sigma^2_{it}$, the variance of the estimator $\hat{X}_{ti}$. Such an interpretation accords, for example, with the QL estimators discussed in sections I and II.
We can now write the substantive regression model as:

$$(30)\quad y_t = \beta_1 X_{t1} + \ldots + \beta_i X_{ti} + \ldots + \beta_k X_{tk} + u_t\ , \qquad t=1,3,5,\ldots,T$$

$$y_t = \beta_1 X_{t1} + \ldots + \beta_i \hat{X}_{ti} + \ldots + \beta_k X_{tk} + (\beta_i \varepsilon_{it} + u_t)\ , \qquad t=2,4,\ldots,(T-1)$$

In vector terms, we shall express this as simply

$$(31)\qquad \underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{v}\ ,$$

where the definitions of X and $\underset{\sim}{v}$ follow from (30). We shall assume ε_{it} and u_t are uncorrelated.

We have thus formulated the problem as one in which the disturbance has different variance components. This is, indeed, forced on us if we assume that all available independent observations are fixed. In particular, the situation does not allow of an interpretation in the classical errors-in-variables framework. For suppose we argued that a true, underlying X_{ti} were fixed, but that the observed $\hat{X}_{ti}$ were subject to error. This would convert $\hat{X}_{ti}$ to a random variable. It is implicit in equations (29) that this is not the case. The entire matrix in (31) has therefore the specification of remaining fixed in repeated samples.

Let us, then, consider equation (31). We write:

$$Evv' = \sigma^2\Omega = \sigma^2 \begin{bmatrix} 1 & & & & & & \\ & 1+\frac{\beta_i^2\sigma_{i2}^2}{\sigma^2} & & & & 0 & \\ & & 1 & & & & \\ & & & 1+\frac{\beta_i^2\sigma_{i4}^2}{\sigma^2} & & & \\ & & & & \cdot & & \\ & 0 & & & & \cdot & \\ & & & & & & \cdot \\ & & & & & & & 1 \end{bmatrix}$$

If we write

$$D = \begin{bmatrix} 1 & & & & & & \\ & \frac{1}{\surd(1+\frac{\beta_i^2\sigma_{i2}^2}{\sigma^2}} & & 0 & & & \\ & & 1 & & & & \\ & 0 & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & & & & & 1 \end{bmatrix},$$

equation (31) can be recast as

$$(32) \qquad D\underset{\sim}{y} = DX\underset{\sim}{\beta} + D\underset{\sim}{v},$$

where $E(Dv)(D\underset{\sim}{v})' = \sigma^2 D\Omega D' = \sigma^2 I$. Equation (32) conforms to the spherical regression model. Thus the usual tests of significance can be applied to an OLS model in which every second observation on every variable is discounted by dividing by factors like

$$(1+\frac{\beta_i^2\sigma_{i2}^2}{\sigma^2})^{\frac{1}{2}} .$$

Of course, it is quite apparent that we are in fact performing a generalized least squares regression calculation. However, to transform the observations as in (32) is likely to lead to less confusion than working directly in terms of Ω, and may offer programming advantages.

However, before the observations can be transformed, we have still to determine σ^2, the spherical constant. This can be done to any desired degree of accuracy by starting with the OLS estimate $\sigma^2 = \frac{1}{T-K}\Sigma e_t^2$, where e is the least-squares residual. Write

$$\Omega = I + R.$$

Then $E(s^2) = \frac{1}{T-K}E\underset{\sim}{y}'(I-P)\underset{\sim}{y}$,

where $P = ((p_{ij})) = X(X'X)^{-1}X'$ is the least-squares projection matrix. Since $(I-P)X = 0$, it follows that

$$E(s^2) = \frac{\sigma^2}{T-K} \text{ trace } (I-P)\Omega,$$

$$= \sigma^2(1 + \frac{1}{T-K} \text{ trace } (I-P)R). \quad (33)$$

Equation (33) is quite general. In our particular case,

$$R = \frac{\beta_i^2}{\sigma^2}\begin{bmatrix} 0 & & & & & & \\ & \sigma_{i2}^2 & & & 0 & & \\ & & 0 & & & & \\ & & & \sigma_{i4}^2 & & & \\ & 0 & & & \cdot & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ & & & & & & & 0 \end{bmatrix},$$

and we derive

(34) $$E(s^2) = \sigma^2(1 + CF),$$

where the correction factor

$$CF = \frac{1}{T-K}\left((1-p_{22})\sigma_{i2}^2 + (1-p_{44})\sigma_{i4}^2 + \ldots + (1-p_{T-1,T-1})\sigma_{iT-1}^2\right).$$

Thus an unbiased estimator of the spherical parameter can be obtained as

$$\frac{s^2}{I+CF}$$, where s^2 is the OLS estimate of the

residual variance. Having obtained the estimate for σ^2, the data can now be transformed with the use of the discount matrix D. If necessary, the whole process can be iterated.

The technique is easily generalized to higher-order interpolation. Suppose now that every second and third observations on X_i have to be interpolated. In this case,

$$\Omega = \begin{bmatrix} 1 & & & & & & \\ & 1+\frac{\beta_i^2\sigma_{i2}^2}{\sigma^2} & \frac{\beta_i^2\sigma_{23}^i}{\sigma^2} & & & & \\ & \frac{\beta_i^2\sigma_{23}^i}{\sigma^2} & 1+\beta_i^2\sigma_{i3}^2 & & & & \\ & & & 1 & & & \\ & & & & \ddots & & \\ & & & & & & 1 \end{bmatrix}$$

where σ_{23}^i denotes the covariance between the interpolation estimates for the second and third observations. We seek now a discount matrix D which renders Ω spherical, i.e. such that $\Omega = (D'D)^{-1}$. This can now be done in at least two separate ways, so that the discount factors are no longer uniquely determined. The easiest procedure is to find a lower triangular matrix D such that $D'D = \Omega^{-1}$. Standard subroutines are available for this calculation, although it can be done analytically. Equation (33), with R defined as $\Omega - I$ will yield the appropriate correction factor to apply to the OLS estimate of the residual variance in order to obtain $\hat{\sigma}^2$.

References

Afifi, A.A. and Elashoff, E.M., "Missing Observations in Multivariate Statistics, I - A Review of the Literature". Journal of the American Statistical Association, Vol. 61 (1966), 595-604.

Chow, G.C. and Lin, An-Loh, "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series", Review of Economics and Statistics, Vol. LIII (1972), 372-375.

Drettakis, M.,"Missing Data in Econometric Estimation", Review of Economic Studies, Vol. 40 (1973), 537-552.

Friedman, M. "The Interpolation of Time Series by Related Series." Journal of the American Statistical Association, Vol. 57 (1962), 729-757.

Goldberger, A.S., Econometric Theory. J. Wiley, New York, 1964.

Footnotes

1. This paper was written while the author was on leave at the Institute for Advanced Studies, Vienna. I am grateful for their support. It is a pleasure, too, to thank Klaus Plasser whose patient assistance with my computing diagnostics enabled me to coexist peacefully with the computer.

2. The continuous-time analogue of (1) is: $\dot{W} = rW + y - (1+r)c$, where r is the continuous discount rate. This has solution over the interval (t-1,t) as follows:

$$W_t = (1+a(t))W_{t-1} + (1+a(t))\int_{t-1}^{t} \frac{(y(s)-(1+r(s))c(s)}{1+a(s)}\,ds,$$

where $1+a(u) = \exp\int_{t-1}^{u} r(s)ds$ for $t-1 < u \leq t$. If $r(t)=r$ constant over the interval (t-1,t), this reduces to (1) with $1+r_t=e^r$, but if not an approximation is involved.

3. Suppose an individual invests z_{it} of his capital in asset i with return r_{it}. Assuming discrete discounting,

$$W_t = \sum_i (1+r_{it})z_{it} + y_t$$

where $\sum_i z_{it} = W_{t-1} - c_t$. Equation (1) will be strictly valid only if r_t is defined by $r_t = \sum_i w_{it}r_{it}$, where $w_{it} = \frac{z_{it}}{\sum_i z_{it}}$. Since these true weights are by definition unknown, a question of approximation is once again involved. The more perfect the capital market the more synchronized will be the movement of yields, thus the better any one yield or index of such will reflect the movement of the true index.

4. $w_1 = 7376.1$
$r_2 \ldots r_8$ = 5.323, 5.573, 5.718, 5.810, 6.125, 6.375, 6.180
$y_2 \ldots y_8$ = 2011.8, 2131.8, 2289.9, 2516.0, 2606.3, 2692.7, 2760.2

The income and wealth figures were chosen from some unofficial estimates for the N.Z. economy ($000).

5. The proof we have constructed of this assertion is not only informal but very lengthy. However, a brief sketch may be useful. Express $\hat{y}_t$ in terms of the error decomposition (28) below, and take limits in probability of each side of the normal equations (26). Compare the result with the limiting "true" normal equations:

$$\text{(i)} \quad \text{plim}\ \frac{1}{T} \begin{bmatrix} \sum y_{t-1}^2 & \sum y_{t-1}x_t \\ \sum y_{t-1}x_t & \sum x_t^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \text{plim}\ \frac{1}{T} \begin{bmatrix} \sum y_t y_{t-1} \\ \sum y_t x_t \end{bmatrix} .$$

We know (i) to be true, since if the error process u_t is uncorrelated, OLS consistently estimates α and β. The above comparison then yields (to terms of the first order in a Taylor expansion) two equations, of full rank, homogeneous in $(\text{plim}(\hat{\alpha}-\alpha), \text{plim}(\hat{\beta}-\beta))$. It follows that $\text{plim}\ \hat{\alpha} = \alpha$, $\text{plim}\ \hat{\beta} = \beta$.