# BEST LINEAR UNBIASED INTERPOLATION, DISTRIBUTION, AND EXTRAPOLATION OF TIME SERIES BY RELATED SERIES

Gregory C. Chow
An-Loh Lin

Econometric Research Program
Research Memorandum No. 124
April 1971

The research described in this paper was supported by NSF Grant GS 2799.

PRINCETON UNIVERSITY
Econometric Research Program
207 Dickinson Hall
Princeton, New Jersey

# BEST LINEAR UNBIASED INTERPOLATION, DISTRIBUTION, AND EXTRAPOLATION OF TIME SERIES BY RELATED SERIES

Gregory C. Chow
Princeton University

An-loh Lin
National Bureau of Economic Research

## 1. INTRODUCTION

In a very informative paper on the interpolation of time series by a related series, Milton Friedman [2] called attention to the two related problems of distribution and extrapolation. Given the value of a time series at the beginning of each quarter for $n$ quarters, say, and given the value of a related series at the beginning of each month for these $3n$ months, the problem of interpolation is to estimate the first series for the remaining $2n$ months. Given the value of a series of flows during each quarter for $n$ quarters, and the value of a related series for each month, the problem of distribution is to estimate the first series for the $3n$ months - presumably the sum of the three monthly estimates for each quarter ought to equal the observed value for the quarter. In either of the above situations, there is the associated problem of extrapolation, namely, of estimating a monthly value of the first series (be it a stock or a flow) outside the sample period of $n$ quarters. Although Friedman recognized that these problems are related, he wrote [2, p. 730], "they differ enough so that I have been unable to encompass them in full in the same analysis." It has now appeared, as will be shown in this paper, that a unified treatment of the three problems can be obtained by using explicitly a regression model which has remained implicit in Friedman's analysis of the problem of interpolation.

As far as the distribution problem is concerned, this paper has been influenced by the practice, as evidenced in the works of Chang and Liu [1] and of Liu [4], of obtaining monthly estimates of some components of GNP by applying the quarterly regressions of the components on some related series to monthly data of the latter series. There is no guarantee from this method that the monthly estimates will add up to the given quarterly totals. Furthermore, justification of this method has not been fully given. These deficiencies will be overcome in the following pages where a method is presented for obtaining best linear unbiased estimates of a monthly series by regression on related series. This method differs from and (under stated assumptions) improves upon the regression method described above.[1]

In section 2, a regression model is formulated for all three problems, and the best linear unbiased estimator is derived for all three purposes. As any solution involving generalized least-squares entails a problem of estimating the covariance matrix of the residuals, this problem will be treated in section 3. Section 4 contains some concluding remarks.

## 2. DERIVATION OF THE ESTIMATOR

It is assumed that monthly observations (if available) of the series to be estimated satisfy a multiple regression relationship with $p$ related series $x_1,\ldots,x_p$. During the sample period of $3n$ months, the relation is[1]

$$y = X\beta + u \tag{2.1}$$

where $y$ is $3n\times1$, $X$ is $3n\times p$, and $u$ is a random vector with mean 0 and covariance matrix $V$. How to select the related series is a question to be bypassed in this paper. Suffice it to say that they may include related economic variables for the same month, for the preceding or even the following month, trend variables, and dummy variables. The related series $X$ will be treated as fixed in the regression model (2.1) for the purpose of statistical analysis. Note also that one may choose to transform the variables (such as taking logarithms) before applying the model (2.1).

Let $C$ be the $n\times3n$ matrix that converts the $3n$ monthly observations into $n$ quarterly observations. For interpolation and distribution respectively, this matrix takes the forms

$$(2.2) \quad C_I = \begin{bmatrix} 1 & 0 & 0 & 0 & & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \cdots & 0 \\ & & & \cdots & & & \\ 0 & \cdots & & & & 1\,0 & 0 \end{bmatrix}; \; C_D = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 & 0 & & & \cdots & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \cdots & 0 \\ & & & & \cdots & & & \\ 0 & \cdots & & & & & 1\;1 & 1 \end{bmatrix} .$$

The vector of $n$ quarterly observations of the dependent variable, to be subscripted by a dot which signifies being quarterly, will satisfy the regression model

$$(2.3) \qquad y_{\bullet} = Cy = CX\beta + Cu = X_{\bullet}\beta + u_{\bullet}$$

with

$$(2.4) \qquad Eu_{\bullet}u_{\bullet}' = V_{\bullet} = CVC' \; .$$

The problem is to estimate a vector $z$ of $m$ observations on the dependent variables, where $z$ would be identical with $y$ in the cases of interpolation and distribution, and would consist of observations outside the sample period in the case of extrapolation. A linear unbiased estimator $\hat{z}$ of $z$ satisfies, for some mxn matrix $A$ ,

$$(2.5) \qquad \hat{z} = Ay_{\bullet} = A(X_{\bullet}\beta + u_{\bullet})$$

and, with $X_z$ and $u_z$ denoting variables in the regression model for $z$, ($X_z$ and $u_z$ identical with $X$ and $u$ in the cases of interpolation and distribution),

(2.6) $$E(\hat{z}-z) = E[A(X_{.}\beta + u_{.}) - (X_z\beta + u_z)] = (AX_{.} - X_z)\beta = 0.$$

The conditions (2.5) and (2.6) imply

(2.7) $$AX_{.} - X_z = 0 \ ;$$

(2.8) $$\hat{z} - z = Au_{.} - u_z \ .$$

The covariance matrix of $(\hat{z} - z)$ will therefore be

(2.9) $$\begin{aligned} \mathrm{Cov}(\hat{z} - z) &= E(Au_{.} - u_z)(Au_{.} - u_z)' \\ &= AV_{.}A' - AV_{.z} - V_{z.}A' + V_z \end{aligned}$$

with $V_{.z}$ denoting $Eu_{.}u_z'$ and $V_z$ denoting $Eu_zu_z'$ .

To find the best linear unbiased estimator $\hat{z}$ , we minimize the trace of (2.9) with respect to $A$ , subject to the $m \times p$ matrix equation $AX_{.} - X_z = 0$ of (2.7). Using an $m \times p$ matrix $M$ of Lagrange mutlipliers, we form the Lagrangian expression

(2.10) $$L = \tfrac{1}{2}\mathrm{tr}\,[AV_{.}A' - AV_{.z} - V_{z.}A' + V_z] - \mathrm{tr}[M'(AX_{.} - X_z)] \ ,$$

set its partial derivatives equal to zero, recalling the rule

$$\frac{\partial}{\partial A} \operatorname{tr} AB = \frac{\partial}{\partial A} \operatorname{tr} BA = B' ,$$

and obtain

(2.11) $$AV_{\cdot} - V_{z\cdot} = MX'_{\cdot} .$$

Solving (2.11) for $A$ gives $A = MX'_{\cdot}V_{\cdot}^{-1} + V_{z\cdot}V_{\cdot}^{-1}$ , which, when substituted in (2.7), gives the solution for $M$ :

(2.12) $$MX'_{\cdot}V_{\cdot}^{-1}X_{\cdot} + V_{z\cdot}V_{\cdot}^{-1}X_{\cdot} - X_z = 0 ,$$

or

$$M = X_z(X'_{\cdot}V_{\cdot}^{-1}X_{\cdot})^{-1} - (V_{z\cdot}V_{\cdot}^{-1})X_{\cdot}(X'_{\cdot}V_{\cdot}^{-1}X_{\cdot})^{-1} .$$

The solution for $A$ is then

(2.13) $$A = X_z(X'_{\cdot}V_{\cdot}^{-1}X_{\cdot})^{-1}X'_{\cdot}V_{\cdot}^{-1} + V_{z\cdot}V_{\cdot}^{-1}[I-X_{\cdot}(X'_{\cdot}V_{\cdot}^{-1}X_{\cdot})^{-1}X'_{\cdot}V_{\cdot}^{-1}] .$$

The resulting estimator is

(2.14) $$\hat{z} = Ay_{\cdot} = X_z\hat{\beta} + (V_{z\cdot}V_{\cdot}^{-1})\hat{u}_{\cdot}$$

where, in obvious notations,

$$(2.15) \qquad \hat{\beta} = (X_{\bullet}'V_{\bullet}^{-1}X_{\bullet})^{-1}X_{\bullet}'V_{\bullet}^{-1}y_{\bullet}$$

is the least squares estimate of the regression coefficients using the $n$ quarterly observations in the sample, and

$$(2.16) \qquad \hat{u}_{\bullet} = [I - X_{\bullet}(X_{\bullet}'V_{\bullet}^{-1}X_{\bullet})^{-1}X_{\bullet}'V_{\bullet}^{-1}]y_{\bullet} = y_{\bullet} - X_{\bullet}\hat{\beta}$$

is the $n\times 1$ vector of residuals in the regression using quarterly data. The estimate (2.14) consists of two components. The first, $X_z\hat{\beta}$, applies the estimated regression coefficients $\hat{\beta}$ to the monthly observations of the explanatory variables $X_z$ associated with the vector $z$ to be estimated. The second is an estimate of the $m\times 1$ vector $u_z$ of residuals, obtained by applying the $m\times n$ matrix $V_{z\bullet}V_{\bullet}^{-1} = (Eu_z u_{\bullet}')(Eu_{\bullet}u_{\bullet}')^{-1}$ of coefficients in the multivariate regression of $u_z$ on $u_{\bullet}$, to the estimated residuals $\hat{u}_{\bullet}$.

The estimator (2.14) is applicable to interpolation of a series of stocks, to distribution of a series of flows, and to extrapolation of series of stocks or flows. The treatments of stocks and flows are distinguished only by the matrix $C$ used in converting monthly to quarterly data, as defined by equation (2.2). For both interpolation and distribution, the vector $z$ to be estimated is the same as the vector $y$ of $3n$ monthly observations (mostly unavailable) during the sample period. For extrapolation, the vector $z$ consists of a few monthly observations after the sample period, but before a quarterly figure becomes available (otherwise, the sample period would be extended and the problem would be one of interpolation or distribution).

Note that, except for the defined relation between a quarterly series and a monthly series, which is really a matter of specifying the variables, our method treats interpolation and distribution identically. Distribution has been considered a different problem from interpolation probably because of the idea that the three monthly estimates ought to add up to the observed quarterly total, whereas no such requirement exists in the case of interpolation. This distinction is unjustified from the viewpoint of the theory of estimation where a good estimator is defined explicitly in terms of its sampling properties. It is certainly possible for one set of monthly estimates to have a smaller sum of squared errors than another, even if its quarterly totals do not match the observed figures as do the totals from the other. However, as a consequence of best linear unbiasedness, our method of estimation does produce monthly estimates whose quarterly totals equal the observed, for by (2.14) and with $V_{z\cdot} = Euu'_{\cdot} = VC'$ ,

$$(2.17) \qquad C\hat{y} = CX\hat{\beta} + CVC'V_{\cdot}^{-1}\hat{u}_{\cdot} = X_{\cdot}\hat{\beta} + \hat{u}_{\cdot} = y_{\cdot} \; .$$

For interpolation, (2.17) simply states that the quarterly estimates (obtained by selecting the first, fourth, seventh,..., monthly series) are the same as the observed quarterly data.

It may be useful to record an expression for the covariance matrix of estimation errors, derived by substituting the right-hand side of (2.13) for A in (2.9) and by some further algebraic manipulations,

$$(2.18)\quad \mathrm{Cov}(\hat{z}-z) = (X_z - V_{z.}V_{.}^{-1}X_{.})(X_{.}'V_{.}^{-1}X_{.})^{-1}(X_z' - X_{.}'V_{.}^{-1}V_{.z}) + (V_z - V_{z.}V_{.}^{-1}V_{.z}).$$

## 3. ESTIMATION OF THE COVARIANCE MATRIX OF RESIDUALS

Like any other generalized least-squares estimator, the estimator (2.14) requires knowledge of the covariance matrix V. In practice V is not given, and has to be estimated by assuming some struc ure to the residuals in the monthly regression (2.1). A few simple, but perhaps most useful, cases will be briefly treated in this section, since the subject is a familiar one in connection with the application of generalized least squares.

The simplest case is to assume that the monthly regression residuals are serially uncorrelated, each with variance $\sigma^2$. In this case, $V = I_{3n}\sigma^2$, and $V_{.} = CC'\sigma^2$. For extrapolation, $V_{z.} = Eu_z u_{.}' = 0$. For interpolation and distribution, $V_{z.} = Euu_{.}' = C'\sigma^2$, where the definitions of C were given in (2.2). For interpolation, one can easily see that the term $(V_{z.}V_{.}^{-1})\hat{u}_{.} = C_I'\hat{u}_{.}$ in (2.14) assigns the regression residual for (the beginning of) any quarter to (the beginning of) the first month of that quarter. For distribution, the corresponding term $3C_D'\hat{u}_.$ allocates a third of the quarterly residual (at annual rate, say) of the three months.

The second case is to assume that the monthly residuals follow a first-order autoregression

(3.1) $$u_t = a\,u_{t-1} + \epsilon_t \qquad (E\epsilon_t\epsilon_s = \delta_{ts}\,\sigma^2)\ .$$

The autocovariances of $u_t$ are well-known to be

(3.2) $$V = Euu' = \begin{bmatrix} 1 & a & a^2 \ldots & a^{3n-1} \\ a & 1 & a \ldots & a^{3n-2} \\ a^2 & a & 1 \ldots & a^{3n-3} \\ & & \ldots & \\ a^{3n-1} & & \ldots & 1 \end{bmatrix} \frac{\sigma^2}{1-a^2} \equiv A \cdot \frac{\sigma^2}{1-a^2}\ .$$

For the purposes of interpolation and distribution, one needs

(3.3) $$V_{z.}V_{.}^{-1} = VC'(CVC')^{-1} = AC'(CAC')^{-1}$$

and for the purpose of extrapolation, one needs

(3.4) $$V_{z.}V_{.}^{-1} = (Eu_z u_.')(CVC')^{-1} = \begin{bmatrix} Eu_{3n+1}u_1 \ldots Eu_{3n+1}u_{3n} \\ \ldots \\ Eu_{3n+m}u_1 \ldots Eu_{3n+m}u_{3n} \end{bmatrix} C'(CVC')^{-1}$$

$$= \begin{bmatrix} a^{3n} & \ldots & a^2 & a \\ & \ldots & & \\ a^{3n+m-1} & \ldots & a^{m+1} & a^m \end{bmatrix} C'(CAC')^{-1}\ .$$

Either (3.3) or (3.4) requires knowledge of $\underline{a}$ only, and not of $\sigma^2$. Once a consistent estimate of $\underline{a}$ is obtained, the estimate (2.14) can be computed, since both $\hat{\beta}$ and $\hat{u}_{\cdot}$ can then be computed by (2.15) and (2.16) respectively.

To obtain a consistent estimate of $\underline{a}$, observe first that the first-order autocorrelation coefficient of the quarterly residuals is the ratio of the second element to the first element on the first row of the matrix $V_{\cdot} = CVC'$. For interpolation, $C = C_I$, and this ratio is simply $a^3$. For distribution, $C = C_D$, and this ratio is easily calculated to be

$$(3.5) \qquad q = \frac{a^5 + 2a^4 + 3a^3 + 2a^2 + a}{2a^2 + 4a + 3} .$$

Given $q$, it can be seen that (3.5) as a polynomial equation in $\underline{a}$ will provide a unique solution for $\underline{a}$.[2] The following iterative method of estimating $\underline{a}$ can therefore be suggested. Given an initial guess of $q$, take its cube root as $\underline{a}$ in an interpolation problem, or solve (3.5) for $\underline{a}$ in a distribution problem. Then use (3.2) and (2.16) to calculate a set of quarterly regression residuals. From these residuals, compute the first-order autocorrelation coefficient as the next guess of $q$, and proceed as before.

As a third simple case for estimating V , assume that the monthly residuals are serially uncorrelated, but have variances proportional to a known function of an explanatory variable, or of a certain linear combination (possibly the principal component) of the explanatory variables. V will then be diagonal, and proportional to a given matrix. The estimate (2.14) can be computed without difficulty in this case.

## 4. CONCLUDING REMARKS

By a fairly straight forward application of the theory of best linear unbiased estimation to a linear regression model, the problems of interpolation, distribution, and extrapolation of a time series by related series can be solved in a unified fashion. The resulting estimator (2.14) applies to all these cases.

The usefulness of this method in practice, as far as the estimation of monthly economic time series is concerned, surely depends on the validity of the regression model assumed, or on the possibility of finding related series which will make the regression model a good approximation to reality. Therefore, empirical works are being conducted to apply the method of this paper to the estimation of observed quarterly and monthly series for the purpose of testing; comparisons are being made with estimates obtained by more naive methods of interpolation and distribution, including the method of using only the first term of (2.14). These empirical results will be reported in the near future.

## REFERENCES

[1] Chang, Ching-gwan and Liu, Ta-chung: "Monthly Estimates of Certain National Product Components, 1946-49," The Review of Economics and Statistics, Vol. XXXIII, No. 3 (August 1951), pp. 219-227.

[2] Friedman, Milton: "The Interpolation of Time Series by Related Series," Journal of the American Statistical Association, Vol. 57, No. 300 (December 1962), pp. 729-757.

[3] Goldberger, Arthur S.: "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," Journal of the American Statistical Association, Vol. 57, No. 298 (June 1962), pp. 369-375.

[4] Liu, Ta-chung: "A Monthly Recursive Econometric Model of United States: A Test of Feasibility," The Review of Economics and Statistics, Vol. LI, No. 1 (February 1969), pp. 1-13.

## FOOTNOTES

1. Insofar as our approach to the three related problems is that of best linear unbiased estimation by generalized least squares, the method is by no means novel. Among the many previous applications of best linear unbiased estimation, we would like to mention the work of Goldberger [3], since it deals with a prediction problem which is related to the "extrapolation" aspect of our problem. The main difference is that, in our problem, monthly observations on the dependent variable are not available during the sample period. Another difference, though minor, is that Goldberger provides prediction of a scalar random variable, while we extrapolate for a vector.

2. Throughout this paper, we will speak only of estimating a monthly series given its quarterly data and monthly data on related series. The theory, with very minor modifications, obviously applies to the estimation of quarterly series from annual data, etc. A time series, for a specified time, is a random variable. We speak of "estimating" the value of this random variable, and not of estimating the value of a fixed parameter as is often the case in statistics.

3. To see this, note that $q$ and $a$ will have the same sign, since the denominator of (3.5) is always positive, and its numerator, written as $a[(a^4 + 2a^3 + a^2) + a^2 + (a^2 + 2a + 1)]$, must have the same sign as $a$. Assume successively case (1) $1 > q > 0$ and, case (2) $0 > q > -1$, write (3.5) as a polynomial in $a$, and one can show, using Descartes' rule of signs, that there is only one real root between 0 and one in case (1), and only one real root between 0 and $-1$ in case (2). Hence, the solution for $a$ is unique.