R-102 R-102

Center for Research on Economic and Social Theory
Research Seminar in Quantitative Economics

Discussion Paper

DEPARTMENT OF ECONOMICS

University of Michigan

Ann Arbor, Michigan 48109

[redacted]

On the Problem of Missing Measurements in the Estimation of Economic Relationships

Jan Kmenta*

R-102

May 1978

*Professor of Economics, The University of Michigan

On the Problem of Missing Measurements in the Estimation of Economic Relationships*

Jan Kmenta

University of Michigan

This paper deals with the problem of estimating the parameters of economic relationships when some of the observations are incomplete but the sample selection rules are not violated. The models considered include classical multiple regression, generalized regression (with special attention to autoregressive disturbances and seemingly unrelated regressions), and recursive systems. In each case we examine the possibility of extracting information from observations with missing measurements and then analyze suitable estimation methods.

1. INTRODUCTION

The problem of estimating the parameters that characterize economic or other relationships when some sample measurements for certain variables are missing has been attracting the attention of statisticians and econometricians for a number of years. Some of this attention has been directed at situations in which the omission of incomplete or missing observations leads to a selectivity bias in estimation.[1] Most authors, however, have been concerned with situations in which it is reasonably safe to assume that discarding incomplete or missing observations leaves the sample selection rules unaffected. In this paper we also restrict ourselves to the situations in which the underlying stochastic structure of each model that we consider is not changed by the fact that some measurements are missing.

Past studies dealing with the general topic of missing measurements under an unchanged sample selection rule can be divided into two broad categories. In the first category we find attempts to fill existing gaps

---

*The research reported here was supported in part by a Faculty Research Fellowship of the University of Michigan, and in part by the U. S. Department of Labor, Contract No. B9M6-1057.

[1] See, e. g., Heckman (1977) who provides a very insightful formulation of this problem.

in a series of observations on some variable of interest. Typically--though not exclusively--such observations are thought to be in the form of a time series, and the gaps are filled by the use of some approximation scheme. Commonly the approximations of the missing values are derived by invoking ad hoc the existence of some related series for which all observations are available. The one or more related series are then used to predict the missing values of the variable of interest. The estimation of a relationship in which this variable may be involved is of only secondary concern to the authors of the papers in this category. The seminal paper of this kind appears to be that of Friedman (1962), later followed by Chow and Lin (1971). An approach utilizing spectral analysis suggested by Doran (1974) was criticized in a further paper by Chow and Lin (1976) who consider the absence of a related series to be a highly artificial situation. Dagenais (1973) considers the problem of filling gaps in the measurements of one or more explanatory variables of a multiple regression model, and makes the simplifying assumption that there exist auxiliary relationships involving the explanatory variables of the model.[2]

The authors of the second category of papers are primarily concerned with incomplete observations in the context of the problem of estimating the parameters of some given relationship without any additional prior information in the form of auxiliary relations. An early leader in this category is the paper by Anderson (1957) ~~who~~ in which the author chooses to treat the missing measurements as unknown parameters to be estimated along with the other parameters of a multivariate normal distribution. This or a similar approach is common to a number of papers whose content is summarized and further extended in a series of papers by Afifi and Elashoff (1966, 1967, and 1969). Among other topics, the authors discuss certain "zero-order" and "first-order" methods of estimating the regression coefficients on the basis of various substitutions for the missing measurements. A formal study of some of these methods is presented in Kelejian (1969); computational results and some further derivations are given in Beale and Little

---

[2]A further simplication introduced by Dagenais (1973) is discussed below.

(1975). Further, Sargan and Drettakis (1974) study the problem of incomplete observations in the context of an autoregressive model and prove an interesting lemma concerning maximum likelihood estimation. Finally, [illegible] Gilbert (1977) ~~contains~~ presents a discussion about estimating a regression model (or an equation belonging to a system of simultaneous equations) from a mixture of annual and quarterly observations.

In what follows we consider the problem of estimating the parameters of an economic relationship when some observations are incomplete or missing, when the omission of incomplete or missing observations leaves the sample selection rule unaffected, and when no auxiliary relations are assumed to exist unless justified by theory. In Section 2 we discuss the estimation of a multiple regression model for which all classical assumptions are satisfied. Section 3 extends the discussion to the generalized regression model. The results obtained in this section are then used to deal with regression models characterized by autoregressive disturbances (Section 4) and with seemingly unrelated regressions (Section 5). Finally, in Section 6 we consider the estimation of a recursive system, which is interesting since it represents a situation where the existence of an "auxiliary relationship" is clearly sanctioned by prior theoretical reasoning.

2. CLASSICAL MULTIPLE REGRESSION MODEL

Consider a regression model

$$y = X\beta + \varepsilon \qquad (2.1)$$

where y is a $(n\times 1)$ vector of values of the dependent variable, X is a $(n\times K)$ matrix of values of the explanatory variables, β is a $(K\times 1)$ vector of unknown parameters, and ε is a $(n\times 1)$ vector of unobservable stochastic disturbances. Further, in accordance with the specification of the classical regression model it is assumed that ε is normally distributed with

$$E(\varepsilon) = 0$$

$$E(\varepsilon\varepsilon') = \sigma^2 I$$

The explanatory variables are thought to be non-stochastic or, if stochastic, distributed independently of ε.

Missing measurements on X

Let us now suppose that of the total of n observations only n_c ($n_c < n$) are complete while in the remaining $n_m = n - n_c$ observations the values of one or more of the explanatory variables are missing. In this case we may partition X and y as follows:

$$(2.2) \qquad X = \begin{bmatrix} X^* \\ X^o \end{bmatrix} , \quad y = \begin{bmatrix} y^* \\ y^o \end{bmatrix}$$

where $X^* \to (n_c \times K)$ is a matrix of measurements on X corresponding to the complete observations, and $X^o \to (n_m \times K)$ is a matrix of measurements on X with at least one value in each row missing. The partition of y conforms to that of X but there are no missing values. In the first part of our analysis X is taken to be non-stochastic.

To estimate β, we may apply the least squares method to the set of complete observations to obtain

$$(2.3) \qquad \hat{\beta}^* = (X^{*\prime}X^*)^{-1}(X^{*\prime}y^*) .$$

Under the conditions of the classical regression model this estimator has all desirable properties given the sample information used. Its variance-covariance matrix is

$$(2.4) \qquad \text{Var-Cov}(\hat{\beta}^*) = \sigma^2(X^{*\prime}X^*)^{-1} .$$

If all the values of X were available, the least squares estimator of β would be

$$(2.5) \qquad \hat{\beta} = (X'X)^{-1}(X'y)$$

with the variance-covariance matrix given as

$$(2.6) \qquad \text{Var-Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1} = \sigma^2(X^{*\prime}X^* + X^{o\prime}X^o)^{-1} .$$

The loss of observations clearly involves a loss of efficiency. Formally,

(2.7) $$\text{Var-Cov}(\hat{\beta}^*) - \text{Var-Cov}(\hat{\beta}) = \sigma^2[(X^{*\prime}X^*)^{-1} - (X'X)^{-1}]$$

which is a non-negative definite matrix.[3]

We may wish to examine the loss of efficiency in greater detail by considering a regression model with two explanatory variables:

(2.8) $$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad , \qquad i = 1, 2, \ldots, n.$$

Suppose that n_m measurements on X_3 are missing. Let $\hat{\beta}_2^*$ be the least squares estimator of β_2 based on the complete observations only, and let $\hat{\beta}_2$ represent the least squares estimator of β_2 that would be obtained if all n observations were complete. Then, using (2.4) and (2.6), we find

(2.9) $$\frac{\text{Var}(\hat{\beta}_2^*)}{\text{Var}(\hat{\beta}_2)} = \frac{(m_{33}^* + c_{33})(1 - r_{23}^2)}{m_{33}^*(1 - r_{23}^{*2})}$$

where $$m_{33}^* = \sum_c (X_{i3} - \bar{X}_3^*)^2 ,$$

$$c_{33} = \frac{n_c n_m}{n} (\bar{X}_3^* - \bar{X}_3^o)^2 + \sum_m (X_{i3} - \bar{X}_3^o)^2 .$$

The summation subscripts "c" and "m" refer to complete and missing values, respectively, and $\bar{X}_3^*$ is the mean of the available values and $\bar{X}_3^o$ the mean of the missing values of X_3. Further, r_{23}^* is the value of the coefficient of correlation between the two explanatory variables based on complete observations only, while r_{23} is the value of this coefficient if no measurements on X_3 were missing. It follows then that the loss of efficiency due to the missing measurments will be smaller,

(1) the smaller the number of incomplete observations;

(2) the smaller the gap between the mean of the observed and the mean of the missing values of X_3;

(3) the smaller the dispersion of the missing values; and

(4) the higher the correlation between the two explanatory variables in the incomplete set of observations.

---

[3] See Goldberger (1964, p. 38).

The first three of the above points apply also in the case of a simple regression model[4] and thus offer no surprise. The last point reflects the fact that an increase in the correlation between the explanatory variables increases the variance of the estimated coefficients and thus mitigates the loss of efficiency as a result of discarded observations. Of course, the exact extent of the loss of efficiency depends on the missing values--which are not known but which could be estimated as indicated below.

The obvious question to be answered is whether or not the incomplete observations contain any information about the regression parameters that could be used in their estimation. To find this out, we take the approach of viewing the missing values of the explanatory variables as unknown parameters to be estimated along with the regression parameters. This can be most conveniently accomplished in the framework of maximum likelihood estimation. Let us consider the model with two explanatory variables presented in (2.8) in a situation where n_m of the n values of one of the explanatory variables, say X_3, are missing. If we denote the missing values X_{i3} by ξ_{i3}, the likelihood function for the n sample observations can be written as

$$(2.10) \qquad L = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_c (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 X_{i3})^2/(2\sigma^2) - \sum_m (Y_i - \beta_1 - \beta_2 X_{i2} - \beta_3 \xi_{i3})^2/(2\sigma^2)$$

where, as before, the summation subscripts c and m refer to complete and incomplete observations, repsectively. The first-order conditions for maximization of L are:

$$(2.11a) \qquad \sum_c (Y_i - \hat\beta_1 - \hat\beta_2 X_{i2} - \hat\beta_3 X_{i3}) + \sum_m (Y_i - \hat\beta_1 - \hat\beta_2 X_{i2} - \beta_3 \hat\xi_{i3}) = 0 ;$$

$$(2.11b) \qquad \sum_c X_{i2}(Y_i - \hat\beta_1 - \hat\beta_2 X_{i2} - \hat\beta_3 X_{i3}) + \sum_m X_{i2}(Y_i - \hat\beta_1 - \hat\beta_2 X_{i2} - \hat\beta_3 \hat\xi_{i3}) = 0 ;$$

$$(2.11c) \qquad \sum_c X_{i3}(Y_i - \hat\beta_1 - \hat\beta_2 X_{i2} - \hat\beta_3 X_{i3}) + \sum_m X_{i3}(Y_i - \hat\beta_1 - \hat\beta_2 X_{i2} - \hat\beta_3 \hat\xi_{i3}) = 0 ;$$

---

[4] See Kmenta (1971, p. 339).

(2.11d) $\hat{\beta}_3(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{i2} - \hat{\beta}_3 \hat{\xi}_{ie}) = 0$, $(i = n_c+1, \ldots, n)$.

From (2.11d) we obtain

(2.12) $\hat{\xi}_{i3} = (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{i2})/\hat{\beta}_3$.

By substitution for $\hat{\xi}_{i3}$ in (2.11a) through (2.11c), we find that the maximum likelihood estimates of the regression coefficients obtained in this way are exactly the same as the least squares estimates (and, equivalently, the maximum likelihood estimates) based on the complete observations only. Thus we have to conclude that the incomplete observations provide no additional information about the regression parameters in this case. All that can be gained is an estimate of the loss of efficiency obtained from (2.9) by utilizing the maximum likelihood estimates of the missing measurements.

Let us now turn to the situations where the explanatory variables are stochastic and independent of ε. Since the mathematical expectation of each of the missing values is equal to the mean of the respective explanatory variable, we could replace the missing values by the respective sample mean of the available values. If we then apply the least squares estimation to all n observations thus completed, we obtain the so-called <u>zero-order</u> regression estimator.[5] Alternatively, we would replace each of the missing values by parameters that would differ from one explanatory variable to another but would remain unchanged from observation to observation. These parameters would then be estimated along with the regression parameters. The resulting estimation is called <u>modified zero-order</u> regression estimator.[6] Neither method utilizes any new information in addition to that provided by the complete observations, and thus neither is of special interest to us.

An entirely different approach to the problem of incomplete observations in estimating the parameters of a multiple regression model is adopted by Dagenais (1973). Dagenais extracts additional information

---

[5]See Afifi and Elashoff (1967).

[6]Ibid.

from incomplete observations by postulating the existence of an auxiliary relationship between the explanatory variables with missing values and those with all values recorded. The procedure suggested by Dagenais involves replacing the missing values by the least squares predictors obtained from the auxiliary equations, and estimating the regression coefficients by the generalized least squares method. The author claims that the resulting estimates have all the desirable asymptotic properties.

There is an obvious difficulty in Dagenais' approach in that, without any theoretical underpinnings, the auxiliary relation between the explanatory variables estimated from complete observations may be peculiar to the available sample values and need not extend to the observations that happen to be incomplete. But even if the auxiliary relation did in fact extend beyond the set of available measurements, the procedure recommended by Dagenais yields estimates that are, under all reasonable circumstances, biased. The demonstration of this proposition, along with a simplified exposition of Dagenais' approach, is given below.

The classical regression model with incomplete observations presented in (2.1) and (2.2) can also be written as follows:

$$(2.13)\qquad y = [X \quad Z]\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} + \varepsilon$$

or

$$(2.13a)\qquad \begin{bmatrix}y^*\\ y^o\end{bmatrix} = \begin{bmatrix}X^* & Z^*\\ X^o & Z^o\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} + \begin{bmatrix}\varepsilon^*\\ \varepsilon^o\end{bmatrix}$$

where $X^* \rightarrow (n_c \times K_1)$, $Z^* \rightarrow (n_c \times K_2)$, $X^o \rightarrow (n_m \times K_1)$, $Z^o \rightarrow (n_m \times K_2)$, $\beta_1 \rightarrow (K_1 \times 1)$, $\beta_2 \rightarrow (K_2 \times 1)$, $\varepsilon^* \rightarrow (n_c \times 1)$, and $\varepsilon^o \rightarrow (n_m \times 1)$. Note that, in accordance with our earlier specification, $\beta' = [\beta_1' \quad \beta_2']$, $\varepsilon' = [\varepsilon^{*\prime} \quad \varepsilon^{o\prime}]$, and $K = K_1 + K_2$. All elements of X^o are missing; all other sample values of the regressors and of the dependent variable are recorded.

The auxiliary relationship involving the explanatory variables is postulated as

$$(2.14)\qquad X = Z\theta + v$$

or

$$\text{(2.14a)} \quad \begin{bmatrix} X^* \\ X^o \end{bmatrix} = \begin{bmatrix} Z^* \\ Z^o \end{bmatrix} \theta + \begin{bmatrix} v^* \\ v^o \end{bmatrix}$$

where $\theta \rightarrow (K_2 \times K_1)$, $v \rightarrow (n \times K_1)$, $v^* \rightarrow (n_c \times K_1)$, and $v^o \rightarrow (n_m \times K_1)$. The relationship in (2.14) is thought to satisfy all assumptions of the classical regression model. Further, ε and v are assumed to be mutually independent To complete the missing values of X, we replace X^o by $\hat{X}^o$, which is defined as

$$\text{(2.15)} \qquad \hat{X}^o = Z^o \hat{\theta}$$

where

$$\text{(2.16)} \qquad \hat{\theta} = (Z^{*\prime} Z^*)^{-1} (Z^{*\prime} X^*)$$

Substituting for X^* from (2.14a) we get

$$\text{(2.17)} \qquad \hat{X}^o = Z^o \theta + Z^o (Z^{*\prime} Z^*)^{-1} Z^{*\prime} v^*$$

$$= Z^o \theta + A v^*$$

where the definition of A is self-evident.

The regression parameters are now to be estimated from

$$\text{(2.18)} \qquad y = [\hat{X} \quad Z] \beta + \tilde{\varepsilon}$$

or, in partitioned form,

$$\text{(2.18a)} \quad \begin{bmatrix} y^* \\ y^o \end{bmatrix} = \begin{bmatrix} X^* & Z^* \\ \hat{X}^o & Z^o \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon^* \\ \tilde{\varepsilon}^o \end{bmatrix}$$

Note that replacing X^o by $\hat{X}^o$ results in replacing ε^o by $\tilde{\varepsilon}^o$. Since

$$\text{(2.19)} \qquad y^o = \hat{X}^o \beta_1 + Z^o \beta_2 + \tilde{\varepsilon}^o$$

$$= (X^o - \hat{v}^o) \beta_1 + Z^o \beta_2 + \tilde{\varepsilon}^o$$

$$= X^o \beta_1 + Z^o \beta_2 + \tilde{\varepsilon}^o - \hat{v}^o \beta_1$$

where

$$(2.20)\qquad \hat{v}^o = X^o - \hat{X}^o$$
$$= Z^o\theta + v^o - \hat{X}^o$$
$$= v^o - Av^* ,$$

we have, from (2.13a) and (2.18a),

$$(2.21)\qquad \tilde{\varepsilon}^o = \varepsilon^o + v^o\beta_1 - Av^*\beta_1$$

To apply the generalized least squares estimation method to (2.18), we have to determine $\Omega = E(\tilde{\varepsilon}\tilde{\varepsilon}')$.

Now

$$(2.22)\qquad \Omega = \begin{bmatrix} E(\varepsilon^*\varepsilon^{*\prime}) & E(\varepsilon^*\tilde{\varepsilon}^{o\prime}) \\ E(\tilde{\varepsilon}^o\varepsilon^{*\prime}) & E(\tilde{\varepsilon}^o\tilde{\varepsilon}^{o\prime}) \end{bmatrix}$$

where

$$E(\varepsilon^*\varepsilon^{*\prime}) = \sigma^2 I$$
$$E(\varepsilon^*\tilde{\varepsilon}^{o\prime}) = 0$$
$$E(\tilde{\varepsilon}^o\tilde{\varepsilon}^{o\prime}) = \sigma^2 I + E(Av^*\beta_1\beta_1'v^{*\prime}A') + E(v^o\beta_1\beta_1'v^{o\prime})$$

or, in Dagenais' terminology,

$$E(\tilde{\varepsilon}^o\tilde{\varepsilon}^{o\prime}) = \sigma^2 I + L(\beta) ,$$

Then, writing (2.18) as

$$(2.23)\qquad y = \hat{W}\beta + \tilde{\varepsilon}$$

where

$$\hat{W} = [\hat{X} \quad Z] ,$$

the generalized least squares estimator of β is

$$(2.24)\qquad \tilde{\beta} = (\hat{W}'\Omega^{-1}\hat{W})^{-1}(\hat{W}'\Omega^{-1}y) .$$

The bias of $\tilde{\beta}$ arises from the fact that $\hat{W}$ and $\tilde{\varepsilon}$ of (2.23) are not uncorrelated. In fact

$$(2.25)\qquad E(\hat{W}'\tilde{\varepsilon}) = E\begin{bmatrix} \hat{X}'\tilde{\varepsilon} \\ Z'\tilde{\varepsilon} \end{bmatrix}$$

$$= E \begin{bmatrix} X^{*\prime}\varepsilon^{*} + \hat{X}^{o\prime}\tilde{\varepsilon}^{o} \\ Z^{*\prime}\varepsilon^{*} + Z^{o\prime}\tilde{\varepsilon}^{o} \end{bmatrix}$$

Now,

$$E(X^{*\prime}\varepsilon^{*}) = 0$$

$$E(\hat{X}^{o\prime}\tilde{\varepsilon}^{o}) = E[Z^{o}\theta + Av^{*}]'[\varepsilon^{o} + v^{o}\beta_1 - Av^{*}\beta_1]$$

$$= E(v^{*\prime}A'Av^{*}\beta_1) .$$

$$E(Z^{*\prime}\varepsilon^{*}) = 0$$

$$E(Z^{o\prime}\tilde{\varepsilon}^{o}) = 0$$

Therefore

$$(2.26) \qquad E(\hat{W}'\tilde{\varepsilon}) = \begin{bmatrix} E(v^{*\prime}A'Av^{*}\beta_1) \\ 0 \end{bmatrix}$$

which is different from zero.

Dagenais' derivation of his proposed estimator is based on the assertion that[7]

$$(2.27) \qquad X^{o} = Z^{o}\hat{\theta} + v_0$$

instead of

$$X^{o} = Z^{o}\hat{\theta} + \hat{v}^{o}$$

as we suppose. But the assertion in (2.27) would be true only if either $\hat{\theta} = \theta$ or if the orthogonality of X and v went <u>beyond</u> the available values of X and Z. Both eventualities are clearly very unrealistic.

Missing measurements on Y

Let us again consider the model given in (2.1), partitioned as

$$\begin{bmatrix} y^{*} \\ y^{o} \end{bmatrix} = \begin{bmatrix} X^{*} \\ X^{o} \end{bmatrix} \beta + \begin{bmatrix} \varepsilon^{*} \\ \varepsilon^{o} \end{bmatrix} .$$

Suppose now that X^{*} and X^{o} are available but y^{o} is missing. An obvious way of replacing y^{o} would be by using the least squares predictor of y^{o} obtained from the available complete observations. The equation to be

[7]See Dagenais (1973, p. 320).

estimated would then be given as

$$(2.28)\qquad \begin{bmatrix} y^* \\ \hat{y}^o \end{bmatrix} = \begin{bmatrix} X^* \\ X^o \end{bmatrix} \beta + \begin{bmatrix} \varepsilon^* \\ \tilde{\varepsilon}^o \end{bmatrix}$$

where

$$\hat{y}^o = X^o\hat{\beta}$$

and

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

If we apply the ordinary least squares method to (2.28), we obtain

$$(2.29)\qquad \hat{\hat{\beta}} = (X'X)^{-1}X' \begin{bmatrix} y^* \\ \hat{y}^o \end{bmatrix}$$

$$= (X'X)^{-1}[X^{*\prime}y^* + X^{o\prime}\hat{y}^o]$$

Using the definition of $\hat{\beta}$ and $\hat{y}^o$ above, we get

$$\hat{\hat{\beta}} = (X'X)^{-1}[(X^{*\prime}X^*)\hat{\beta} + (X^{o\prime}X^o)\hat{\beta}]$$

$$= \hat{\beta} .$$

Thus the application of the least squares method to (2.28) yields the same result as the least squares esimation based on the complete observations only.

Since replacing y^o by $\hat{y}^o$ results in replacing ε^o by $\tilde{\varepsilon}^o$, we may consider estimating (2.28) by generalized rather than simple least squares. To this end, we have to determine

$$\Omega = E \begin{bmatrix} \varepsilon^*\varepsilon^{*\prime} & \varepsilon^*\tilde{\varepsilon}^{o\prime} \\ \tilde{\varepsilon}^o\varepsilon^{*\prime} & \tilde{\varepsilon}^o\tilde{\varepsilon}^{o\prime} \end{bmatrix} .$$

Now since

$$\tilde{\varepsilon}^o = \hat{y}^o - X^o\beta$$

$$= X^o(\hat{\beta} - \beta)$$

$$= X^o(X^{*\prime}X^*)^{-1}X^{*\prime}\varepsilon^*$$

$$= \mathbf{B}\varepsilon^* ,$$

it follows that

$$\Omega = \begin{bmatrix} \sigma^2 I & \sigma^2 B' \\ \sigma^2 B & \sigma^2 BB' \end{bmatrix} .$$

It is easy to show that Ω is a singular matrix. Therefore, the generalized least squares method cannot be applied to (2.28).

3. GENERALIZED REGRESSION MODEL

The model to be considered in this section is the same as the regression model of the preceding section except for the specification of the variance-covariance matrix of the disturbances. Specifically, we postulate

(3.1) $$y = X\beta + \varepsilon$$

where the dimensions of the vectors and matrices involved are the same as in (2.1) but now

$$\begin{aligned} E(\varepsilon\varepsilon') &= \Omega \\ &= \sigma^2 V . \end{aligned}$$

Again, we consider the case of missing measurements on X and on y in turn.

Missing Measurements on X

If some of the measurements on X are missing, we use the partitioned formulation

(3.2) $$\begin{bmatrix} y^* \\ y^o \end{bmatrix} = \begin{bmatrix} X^* \\ X^o \end{bmatrix} \beta + \begin{bmatrix} \varepsilon^* \\ \varepsilon^o \end{bmatrix}$$

where the asterisk refers to the n_c complete observations and the circle superscript to the n_m incomplete observations. The matrix X^o has at least one element in each row missing. The corresponding partitioning of the Ω matrix is

$$\Omega = E \begin{bmatrix} \varepsilon^*\varepsilon^{*'} & \varepsilon^*\varepsilon^{o'} \\ \varepsilon^o\varepsilon^* & \varepsilon^o\varepsilon^{o'} \end{bmatrix}$$

$$= \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

The estimation of the generalized regression model can frequently be simplified by noting that there exists a nonsingular matrix P → (n×n) such that

(3.3) $$PVP' = I$$

so that (3.1) can be rewritten as

(3.4) $$Py = PX\beta + P\varepsilon$$

where

$$E(P\varepsilon\varepsilon'P') = \sigma^2 I \ .$$

Further, in accordance with the partitioning of the V matrix, P can be partitioned as[8]

(3.5) $$P = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}$$

so that (3.2) can be rewritten as

(3.6) $$\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} y^* \\ y^o \end{bmatrix} = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} X^* \\ X^o \end{bmatrix} \beta + \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} \varepsilon^* \\ \varepsilon^o \end{bmatrix}$$

or

(3.6a) $$P_{11}y^* = P_{11}X^*\beta + P_{11}\varepsilon^*$$

(3.6b) $$P_{21}y^* + P_{22}y^o = (P_{21}X^* + P_{22}X^o)\beta + (P_{21}\varepsilon^* + P_{22}\varepsilon^o)$$

In accordance with the above result, the application of the generalized least squares method to the complete observations is then equivalent to the application of the simple least squares method to (3.6a).

---

[8] See Riddell (1977).

This gives

$$(3.7)\qquad \tilde{\beta}^* = (X^{*\prime}P_{11}'P_{11}X^*)^{-1}(X^{*\prime}P_{11}'P_{11}y^*)$$
$$= (X^{*\prime}V_{11}^{-1}X^*)^{-1}(X^{*\prime}V_{11}^{-1}y^*)$$

and, for a nonstochastic X,

$$(3.8)\qquad \text{Var-Cov}(\tilde{\beta}^*) = \sigma^2(X^{*\prime}P_{11}'P_{11}X^*)^{-1}$$
$$= \sigma^2(X^{*\prime}V_{11}^{-1}X^*)^{-1}$$

If all observations were complete, the generalized least squares would be

$$(3.9)\qquad \tilde{\beta} = (X'P'PX)^{-1}(X'P'Py)$$
$$= (X'V^{-1}X)^{-1}(X'V^{-1}y)$$

and its variance-covariance matrix would be

$$(3.10)\qquad \text{Var-Cov}(\tilde{\beta}) = \sigma^2(X'P'PX)^{-1}$$
$$= \sigma^2(X'V^{-1}X)^{-1}$$

The loss of efficiency as a result of omitting the incomplete observations could be examined by applying the analysis of Section 2 to the transformed equations (3.6a) and (3.6b).

The question concerning the potential information about the regression parameters can again be approached by viewing the missing values of X as unknown parameters to be estimated. The estimation can be carried out by maximizing the following log-likelihood function (conditional on initial values of y if applicable)

$$(3.11)\qquad L = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log\det(V)$$
$$-\frac{1}{2\sigma^2}(y - X\beta)'V^{-1}(y - X\beta)$$

or equivalently

$$(3.11a)\qquad L = -\frac{n}{2}\log(2\sigma^2\pi) + \frac{1}{2}\log\det.(P'P)^{-1}$$
$$-\frac{1}{2\sigma^2}[(P_{11}y^* - P_{11}X^*\beta)'(P_{11}y^* - P_{11}X^*\beta)$$
$$+ (P_{21}y^* - P_{21}X^*\beta)'(P_{21}y^* - P_{21}X^*\beta)$$

$$+ 2(P_{21}y^* - P_{21}X^*\beta)'(P_{22}y^o - P_{22}X^o\beta)$$

$$+ (P_{22}y^o - P_{22}X^o\beta)'(P_{22}y^o - P_{22}X^o\beta)] .$$

By differentiating L with respect to the unknown parameters, including the missing elements in X^0, we could derive the appropriate maximum likelihood estimator of β and check whether or not it differs from the maximum likelihood estimator based only on the complete observations. This will be done in connection with the seemingly unrelated regression model discussed in Section 5.

Missing measurements on y

Let us consider now the specification of the generalized regression model in (3.2) in the case where there are no missing values in the X^o matrix but the values in y^o are not available. Under these circumstances we may replace y^o by its generalized least squares predictor $\tilde{y}^o$ derived from the complete observations. Now from (3.6b) we have

$$P_{21}y^* + P_{22}\tilde{y}^o = (P_{21}X^* + P_{22}X^o)\tilde{\beta}^* , \tag{3.12}$$

Solving for $\tilde{y}^o$, we obtain the following formula for the generalized least squares predictor of y^o, which is the best linear unbiased predictor under the specification of the model[9]:

$$\begin{aligned} \tilde{y}^o &= X^o\tilde{\beta} - P_{22}^{-1}P_{21}(y^* - X^*\tilde{\beta}^*) \\ &= X^o\tilde{\beta} - P_{22}^{-1}P_{21}\tilde{\varepsilon}^* \\ &= X^o\tilde{\beta} + V_{21}V_{11}^{-1}\tilde{\varepsilon}^* \end{aligned} \tag{3.13}$$

where $\tilde{\varepsilon}^*$ is the vector of generalized least squares residuals pertaining to the complete observations.

We may now try to obtain an improved estimator of β by applying the generalized least squares method to

---

[9] Our formula represents a slight generalization of the result obtained by Goldberger (1962) and Riddell (1977). The same result is also given in Schim van der Loeff and Leclerg (1975).

$$(3.14)\qquad \begin{bmatrix} y^* \\ \tilde{y}^o \end{bmatrix} = \begin{bmatrix} X^* \\ X^o \end{bmatrix} \beta + \begin{bmatrix} \varepsilon^* \\ \tilde{\varepsilon}^o \end{bmatrix}$$

First we note that

$$\begin{aligned}(3.15)\qquad \tilde{\varepsilon}^o &= \tilde{y}^o - X^o\beta \\ &= X^o\tilde{\beta}^* + V_{21}V_{11}^{-1}(y^* - X^*\tilde{\beta}^*) - X^o\beta \\ &= X^o(\tilde{\beta}^* - \beta) + V_{21}V_{11}^{-1}[-X^*(\tilde{\beta}^* - \beta) + \varepsilon^*] \\ &= [(X^o - V_{21}V_{11}^{-1}X^*)(X^{*\prime}V_{11}^{-1}X^*)^{-1}X^{*\prime}V_{11}^{-1} + V_{21}V_{11}^{-1}]\varepsilon^* \\ &= G\varepsilon^*\end{aligned}$$

Therefore

$$\begin{aligned}(3.16)\qquad \Omega &= E\begin{bmatrix} \varepsilon^*\varepsilon^{*\prime} & \varepsilon^*\tilde{\varepsilon}^{o\prime} \\ \tilde{\varepsilon}^o\varepsilon^{*\prime} & \tilde{\varepsilon}^o\tilde{\varepsilon}^{o\prime} \end{bmatrix} \\ &= \sigma^2\begin{bmatrix} V_{11} & V_{11}G' \\ GV_{11} & GV_{11}G' \end{bmatrix}\end{aligned}$$

which is singular. Thus, generalized least squares estimation of (3.14) is not possible. The same conclusion would be reached if y^o were replaced by any predictor that is linear in ε^*. We could, of course, estimate β in (3.14) by ordinary least squares, but then the resulting estimator would be identical to the least squares estimator obtained from the complete observations only.

4. REGRESSION MODEL WITH AUTOREGRESSIVE DISTURBANCES

In this section we deal with incomplete or missing observations in the context of the following model:

$$(4.1)\qquad y = X\beta + \varepsilon$$

where the dimensions are the same as in (2.1) but the stochastic disturbance is characterized as

(4.2) $\varepsilon(t) = \rho\varepsilon(t - 1) + u(t) \qquad (0 \leq \rho^2 < 1,\ t = 1, 2, \ldots, n)$,

$u(t) \sim N[0, \sigma_u^2]$,

$E[\varepsilon(t - 1)u(t)] = 0$,

$\varepsilon(0) \sim N[0, \sigma_u^2/(1 - \rho^2)]$.

Note that in this case

(4.3) $E(\varepsilon\varepsilon') = \sigma^2 V$

$$= \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & & & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}$$

and the transformation matrix P, for which $PVP' = I$, is given as

(4.4) $$P = \begin{bmatrix} \sqrt{1 - \rho^2} & 0 & \cdots & 0 & 0 \\ -\rho & 1 & \cdots & 0 & 0 \\ 0 & -\rho & \cdots & 0 & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

so that the application of the least squares method to

(4.5) $Py = PX\beta + P\varepsilon$

would lead to estimates of β that are equal to those obtained by the application of the generalized least squares method to (4.1).

Missing observations

Consider now a situation where some observations are altogether missing so that the series of values of y and X have gaps. Suppose that of the n possible sample observations only n_c $(n_c < n)$ are available. We define a "deletion matrix" $D \rightarrow (n_c \times n)$, which is formed by deleting from a unit matrix of order n those rows that correspond to the missing observations.[10] Premultiplying both sides of (4.1) by D, we obtain a regression equation that is applicable to the available observations. The best linear unbiased estimator of β then is

$$(4.6)\qquad \tilde{\beta} = [X'D'(DVD')^{-1}DX]^{-1}[X'D'(DVD')^{-1}Dy]$$

and its variance-covariance matrix is

$$(4.7)\qquad \text{Var-Cov}(\tilde{\beta}) = \sigma^2[X'D'(DVD')^{-1}DX] .$$

When all observations are available, the transformation in (4.5) leads to a simple formulation in non-matrix notation as follows:

$$(4.8)\qquad \sqrt{(1-\rho^2)}\; y(1) = \sum_{k=1}^{K} \sqrt{(1-\rho^2)}\; X_k(1)\beta_k + \sqrt{(1-\rho^2)}\; \varepsilon(1)$$

$$y(t) - \rho y(t-1) = \sum_{k=1}^{K} [X(t) - \rho X(t-1)]\beta_k + u(t)$$

$$t = 2, 3, \ldots, n .$$

When some of the observations are missing, the above simple transformation has to be modified. Suppose that there is a single gap of p periods after the r-th observation. Then the appropriate transformation is

$$(4.9)\qquad \sqrt{(1-\rho^2)}\; y(1) = \sum_{k=1}^{K} \sqrt{(1-\rho^2)}\; X_k(1)\beta_k + \sqrt{(1-\rho^2)}\; \varepsilon(1)$$

$$y(2) - \rho y(1) = \sum_{k=1}^{K} [X(2) - \rho X(1)]\beta_k + u(2)$$

$$\vdots$$

$$y(r) - \rho y(r-1) = \sum_{k=1}^{K} [X(r) - \rho X(r-1)]\beta_k + u(r)$$

$$y(r+p+1) - \rho^{r+1} y(r) = \sum_{k=1}^{K} [X_k(r+p+1) - \rho^{r+1} X_k(r)]\beta_k + u^*(r+p+1)$$

---

[10] See Kapteyn and Wansbeek (1976).

$$y(r+p+2) - \rho y(r+p+1) = \sum_{k=1}^{K} [X_k(r+p+2) - \rho X_k(r+p+1)]\beta_k + u(r+p+2)$$

$$\vdots$$

$$y(n) - \rho y(n-1) = \sum_{k=1}^{K} [X_k(n) - \rho X_k(n-1)]\beta_k + u(n)$$

where

$$u^*(r+p+1) = u(r+p+1) + \rho u(r+p) + \ldots + \rho^p u(r) .$$

Since

$$\text{Var}[u^*(r+p+1)] = \sigma_u^2[1 - \rho^{2(p+1)}]/[1 - \rho^2)$$

the system in (4.9) is heteroskedastic. To rectify that, we multiply both sides of the equation containing $u^*(r+p+1)$ by

$$\sqrt{(1 - \rho^2)/[1 - \rho^{2(p+1)}]}$$

In the preceding discussion we presumed that the value of ρ is either known or can be readily estimated. This is, of course, the case when all observations are made at regular intervals and none are missing. When there is a gap in the series of observations, estimation of ρ becomes somewhat more complicated. Perhaps the easiest way of getting a consistent estimate of ρ under these circumstances is to use the ordinary least squares residuals, $e(t)$, and then use the following formula:

$$(4.10) \qquad \hat{\rho} = [\sum_{t=2}^{r} e_t e_{t-1} + \sum_{t=r+p+1}^{n} e_t e_{t-1}]/[\sum_{t=2}^{r} e_{t-1}^2 + \sum_{t=r+p+1}^{n} e_{t-1}^2] .$$

Alternatively, we could estimate ρ along with β and σ^2 by maximizing the log-likelihood function

$$(4.11) \qquad L = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log\det.(DVD')$$
$$- \frac{1}{2\sigma^2}(Dy - DX\beta)'(DVD')^{-1}(Dy - DX\beta) .$$

An associated problem when some observations are missing is that of testing for the absence of autoregression. With missing observations the traditional Durbin-Watson test is no longer strictly applicable. However, recent results of Savin and White (1976) indicate that the distortion seems to be relatively mild.

Missing measurements on y

When some measurements on y are missing but the corresponding values of X are available, the missing values of y could be replaced by their respective predictors based on the complete observations. The formula for the best linear unbiased predictor of y in the generalized regression model is given in (3.13) above. Its adaptation to the regression model with autoregressive disturbances is quite straightforward. If only one value of y is missing, say the one indexed by (r+1), then the best linear unbiased predictor of y(r+1) is

$$\tilde{y}(r+1) = \sum_{k=1}^{K} X_k(r+1)\tilde{\beta}_k + \rho\tilde{\varepsilon}(r) \tag{4.12}$$

where the $\tilde{\beta}$'s are obtained from (4.6)--or (4.9) adjusted for heteroskedasticity--and $\tilde{\varepsilon}(r)$ is the generalized least squares residual corresponding to the r-th observation. If there are p consecutive values of y missing, the best linear unbiased predictor will be

$$\tilde{y}(r+i) = \sum_{k=1}^{K} X_k(r+i)\tilde{\beta}_k + \rho^i\tilde{\varepsilon}(r) \qquad (i = 1, 2, \ldots p) . \tag{4.13}$$

The predicted values of y could be used to replace the corresponding missing values of y, and the resulting set of n observational points could then be used to estimate the regression parameters. However, as shown in the preceding section, the generalized least squares estimates based on these observations are indeterminate, and the least squares estimates are identical to those obtained from complete observations only.

5. SEEMINGLY UNRELATED REGRESSIONS

A set of two seemingly unrelated regressions can be presented as

$$\begin{aligned} y_1 &= X_1\beta_1 + \varepsilon_1 \\ y_2 &= X_2\beta_2 + \varepsilon_2 \end{aligned} \tag{5.1}$$

where $y_1 \rightarrow (n\times 1)$, $X_1 \rightarrow (n\times K_1)$, $\beta_1 \rightarrow (K_1\times 1)$, $\varepsilon_1 \rightarrow (n\times 1)$, $y_2 \rightarrow (n\times 1)$, $X_2 \rightarrow (n\times K_2)$, $\beta_2 \rightarrow (K_2\times 1)$, and $\varepsilon_2 \rightarrow (n\times 1)$. Each equation is thought to satisfy all assumptions of the classical regression model, and in particular

$$E(\varepsilon_1\varepsilon_1') = \sigma_{11}I$$

$$E(\varepsilon_2\varepsilon_2') = \sigma_{22}I \ .$$

Further,

$$E(\varepsilon_1\varepsilon_2') = \sigma_{12}I \ .$$

The system can be compactly written as

$$(5.2) \qquad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

To allow for possible incomplete or missing observations, we further partition (5.2) to form

$$(5.3) \qquad \begin{bmatrix} y_1^* \\ y_1^o \\ y_2^* \\ y_2^o \end{bmatrix} = \begin{bmatrix} X_1^* & 0 \\ X_1^o & 0 \\ 0 & X_2^* \\ 0 & X_2^o \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1^* \\ \varepsilon_1^o \\ \varepsilon_2^* \\ \varepsilon_2^o \end{bmatrix}$$

where, as before, the vectors or matrices marked by an asterisk are complete whereas those with a circle subscript may have some or all elements missing. For each equation we specify the number of complete observations as n_c and those possibly incomplete or missing as n_m. The variance-covariance matrix of the disturbances in (5.3) can be presented as

$$(5.4) \qquad \Omega = \begin{bmatrix} \sigma_{11}I & 0 & \sigma_{12}I & 0 \\ 0 & \sigma_{11}I & 0 & \sigma_{12}I \\ \sigma_{12}I & 0 & \sigma_{22}I & 0 \\ 0 & \sigma_{12}I & 0 & \sigma_{22}I \end{bmatrix}$$

Missing observations

Suppose all values in y_1^o and X_1^o are missing so that the system in (5.3) reduces to

$$(5.5) \qquad \begin{bmatrix} y_1^* \\ y_2^* \\ y_2^o \end{bmatrix} = \begin{bmatrix} X_1^* & 0 \\ 0 & X_2^* \\ 0 & X_2^o \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \varepsilon_2^o \end{bmatrix}$$

or, in compact notation,

$$(5.5a) \qquad y = X\beta + \varepsilon$$

where now

$$\Omega = E(\varepsilon\varepsilon')$$

$$= \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & 0 \\ \sigma_{21}I & \sigma_{22}I & 0 \\ 0 & 0 & \sigma_{22}I \end{bmatrix}$$

and the best linear unbiased estimator of β is given by

$$(5.6) \qquad \tilde{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y)$$

$$\text{Var-Cov}(\tilde{\beta}) = (X'\Omega^{-1}X)^{-1}$$

Thus the unequal number of observations for the two equations appears to present no difficulty in applying the generalized least squares estimation.[11]

Missing measurements on X

We consider now the situation where one column of values of X^o is missing. The question then arises as to the amount of information concerning β contained in the incomplete observations. To examine this question, we restrict ourselves to the model with two explanatory variables (and no constant term) in each equation, since the generalization

[11] This is pointed out in Schmidt (undated).

to models with a larger number of explanatory variables is obvious. The system in (5.3) then specializes to

$$(5.7) \qquad \begin{aligned} y_1^* &= X_{11}^*\beta_{11} + X_{12}^*\beta_{12} + \varepsilon_1^* , \\ y_1^o &= X_{11}^o\beta_{11} + X_{12}^o\beta_{12} + \varepsilon_1^o , \\ y_2^* &= X_{21}^*\beta_{21} + X_{22}^*\beta_{22} + \varepsilon_2^* , \\ y_2^o &= X_{21}^o\beta_{21} + X_{22}^o\beta_{22} + \varepsilon_2^o . \end{aligned}$$

Suppose now that the values of X_{12}^o are not available. We may regard the missing values of X_{12}^o as unknown parameters to be estimated along with the β's, and set up the log-likelihood function as

$$(5.8) \qquad L = -n \log (2\pi) - \frac{1}{2} \log \det.(\Omega) - \frac{1}{2}(y - X\beta)'\Omega^{-1}(y - X\beta)$$

where X_{12}^o is considered to have been replaced by ξ_{12}. Now, maximizing L with respect to β and ξ_{12} is equivalent to minimizing

$$\begin{aligned} (5.9) \qquad S &= (y - X\beta)'\Omega^{-1}(y - X\beta) \\ &= \Delta[\sigma_{22}(y_1^* - X_{11}^*\beta_{11} - X_{12}^*\beta_{12})'(y_1^* - X_{11}^*\beta_{11} - X_{12}^*\beta_{12}) \\ &\quad -2\sigma_{12}(y_1^* - X_{11}^*\beta_{11} - X_{12}^*\beta_{12})'(y_2^* - X_{21}^*\beta_{21} - X_{22}^*\beta_{22}) \\ &\quad -2\sigma_{12}(y_1^o - X_{11}^o\beta_{11} - \xi_{12}\beta_{12})'(y_2^o - X_{21}^o\beta_{21} - X_{22}^o\beta_{22}) \\ &\quad + \sigma_{11}(y_2^* - X_{21}^*\beta_{21} - X_{22}^*\beta_{22})'(y_2^* - X_{21}^*\beta_{21} - X_{22}^*\beta_{22}) \\ &\quad + \sigma_{11}(y_2^o - X_{21}^o\beta_{21} - X_{22}^o\beta_{22})'(y_2^o - X_{21}^o\beta_{21} - X_{22}^o\beta_{22}) \\ &\quad + \sigma_{22}(y_1^o - X_{11}^o\beta_{11} - \xi_{12}\beta_{12})'(y_1^o - X_{11}^o\beta_{11} - \xi_{12}\beta_{12}) \end{aligned}$$

where

$$\Delta = 1/(\sigma_{11}\sigma_{22} - \sigma_{12}^2) .$$

Differentiating S with respect to ξ_{12} and putting the result equal to zero yields

$$(5.10)\qquad y_1^o - X_{11}^o\beta_{11} - \hat{\xi}_{12}\beta_{12} = \left(\frac{\sigma_{12}}{\sigma_{22}}\right)(y_2^o - X_{21}^o\beta_{21} - X_{22}^o\beta_{22}) \ .$$

Substituting for $\hat{\xi}_{12}$ from (5.10) into (5.9), differentiating the resulting expression with respect to the β's and putting each derivative equal to zero, we obtain the following:

$$(5.11)\qquad \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sigma_{22}(X_1^{*\prime}X_1^*) & -\sigma_{12}(X_1^{*\prime}X_2^*) \\ -\sigma_{12}(X_2^{*\prime}X_1^*) & \sigma_{11}(X_2^{*\prime}X_2^* + X_2^{o\prime}X_2^o) - \frac{\sigma_{12}^2}{\sigma_{12}}(X_2^{o\prime}X_2^o) \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \sigma_{22}(X_1^{*\prime}y_1^*) - \sigma_{12}(X_1^{*\prime}y_2^*) \\ -\sigma_{12}(X_2^{*\prime}y_2^*) + \sigma_{11}(X_2^{*\prime}y_2^* + X_2^{o\prime}y_2^o) - \frac{\sigma_{12}^2}{\sigma_{22}}(X_2^{o\prime}y_2^o) \end{bmatrix}$$

But this is exactly the same result as that obtained from the application of the generalized least method to the complete observations only. Thus we are led to the conclusion that the incomplete observation contain no additional information concerning the regression parameters and may be discarded.

6. RECURSIVE SYSTEM

A simple form of a recursive system may be presented as follows:

$$(6.1)\qquad y_i = \beta x_i + \varepsilon_{1i}$$

$$(6.2)\qquad x_i = \alpha z_i + \varepsilon_{2i} \qquad\qquad (i = 1, 2, \ldots, n)$$

where

$$\varepsilon_{1i} \sim N(0, \sigma_{11})$$

$$\varepsilon_{2i} \sim N(0, \sigma_{22})$$

$$E(\varepsilon_{1i}\varepsilon_{1j}) = E(\varepsilon_{2i}\varepsilon_{2j}) = 0 \qquad\qquad (i \neq j)$$

and the disturbances in the two equations are mutually independent. Further, z is nonstochastic or, if stochastic, distributed independently of the stochastic disturbances. We suppose that some sample values of x are missing and we have only n_c ($n_c < n$) complete observations. Our problem is to estimate β.

As the first estimator of β we consider the simple least squares estimator based on the complete observations only, i. e.,

(6.3) $$\hat{\beta} = \sum_c x_i y_i / \sum_c x_i^2$$

where the summation subscript c indicates summation over complete observations only. The mean and variance of this estimator are

(6.4) $$E(\hat{\beta}) = \beta + E[\sum_c x_i \varepsilon_{1i} / \sum_c x_i^2]$$
$$= \beta$$

since x and ε_1 are independent.

Further

(6.5) $$\mathrm{Var}(\hat{\beta}) = E[\sum_c x_i \varepsilon_{1i} / \sum_c x_i^2]^2$$
$$= \sigma_{11} E[1/\sum_c x_i^2]$$

An alternative estimator of β can be constructed by replacing the missing values of x by their least squares predictors based on equations (6.2) and obtained from the available observations. Thus

(6.6) $$\hat{x}_i = \hat{\alpha} z_i$$
$$= (\sum_c x_i z_i / \sum_c z_i^2) z_i$$

and

(6.7) $$\hat{\hat{\beta}} = [\sum_c x_i y_i + \sum_m \hat{x}_i y_i] / [\sum_c x_i^2 + \sum_m \hat{x}_i^2]$$

where the summation subscript m indicates summation over all observations for which x is not measured. Here

(6.8) $$E(\hat{\hat{\beta}}) = \beta + E[\beta \sum_m \hat{x}_i(x_i - \hat{x}_i)]/[\sum_c x_i^2 + \sum_m \hat{x}_i^2]$$

This estimator is biased but asymptotically unbiased. Further

(6.9) $$E(\hat{\hat{\beta}} - \beta)^2 = \beta^2 E[\sum_m \hat{x}_i(x_i - \hat{x}_i)]^2/[\sum_c x_i^2 + \sum_m \hat{x}_i^2]^2$$
$$+ \sigma_{11}E[1/\sum_c x_i^2 + \sum_m \hat{x}_i^2]$$

Asymptotically $E(\hat{\hat{\beta}} - \beta)^2$ will tend to approach $E(\hat{\beta} - \beta)^2$ but in small samples comparisons between the two mean square errors are difficult to make.

As the third estimator we consider the indirect least squares estimator. By substituting for x from (6.2) into (6.1) we obtain

(6.10) $$y_i = \alpha\beta z_i + (\varepsilon_{1i} + \beta\varepsilon_{2i})$$
$$= \delta z_i + \varepsilon_i$$

Then the indirect least squares estimator of β is defined as

(6.11) $$\tilde{\beta} = \hat{\delta}/\hat{\alpha}$$

where $\hat{\delta}$ is a least squares estimator of δ in (6.10) and $\hat{\alpha}$ is a least squares estimator of α in (6.2). Specifically,

(6.12) $$\hat{\delta} = (\sum_c y_i z_i + \sum_m y_i z_i)/(\sum_c z_i^2 + \sum_m z_i^2)$$

and

(6.13) $$\hat{\alpha} = (\sum_c x_i z_i)/(\sum_c z_i^2)$$

The mean of this estimator is

(6.14) $$E(\tilde{\beta}) = \beta[1 + E(\sum_m x_i z_i)/(\sum_c x_i z_i)](\sum_c z_i^2)/(\sum_c z_i^2 + \sum_m z_i^2)$$

This estimator is thus also biased although asymptotically unbiased. Its small-sample mean square error is difficult to determine.

The results on the three estimators, though not conclusive, nevertheless indicate that the simple least squares estimator based on complete observations only is not necessarily inferior to the other two estimators which use the incomplete observations as well. Further study is needed to determine the circumstances under which any of the three estimators considered would prove to be superior to the other two.

REFERENCES

Afifi, A. A. and R. M. Elashoff (1966), "Missing values in multivariate statistics-I. Review of the Literature," Journal of the American Statistical Association, Vol. 61, September 1966, pp. 595-604.

Afifi, A. A. and R. M. Elashoff (1967), "Missing observations in multivariate statistics-II. Point estimation in simple linear regression," Journal of the American Statistical Association, Vol. 62, March 1967, pp. 10-29.

Afifi, A. A. and R. M. Elashoff (1969a), "Missing observations in multivariate statistics-III. Large sample analysis of simple linear regression," Journal of American Statistical Association, Vol. 64, March 1969, pp. 337-358.

Afifi, A. A. and R. M. Elashoff (1969b), "Missing observations in multivariate statistics-IV. A note on simple linear regression," Journal of American Statistical Association, Vol. 64, March 1969, pp. 359-365.

Anderson, T. W. (1957), "Maximum likelihood estimates for a multivariate normal distribution when some observations are missing," Journal of the American Statistical Association, Vol. 52, June 1957, pp. 200-203.

Beale, E. M. L. and R. J. A. Little (1975), "Missing values in multivariate analyais," Journal of the Royal Statistical Society, Series B, Vol. 37, No. 1, 1975, pp. 129-145.

Chow, G. C. and A. Lin (1971), "Best linear unbiased interpolation, distribution and extrapolation of time series by related series," Review of Economics and Statistics, Vol. 53, November 1971, pp. 372-375.

Chow, G. C. and A. Lin (1976), "Best linear unbiased estimation of missing observations in an economic time series," Journal of the American Statistical Association, Vol. 71, September 1976, pp. 719-721.

Dagenais, M. G. 91973), "The use of incomplete observations in multiple regression analysis," Journal of Econometrics, Vol. 1, December 1973, pp. 317-328.

Doran, H. E. (1974), "Prediction of missing observations in the time series of an economic variable," Journal of the American Statistical Association, Vol. 69, June 1974, pp. 546-554.

Friedman, M. (1962), "The interpolation of time series by related series," Journal of the American Statistical Association, Vol. 57, December 1962, pp. 729-757.

Gilbert, C. L. (1977), "Regression using mixed annual and quarterly data," Journal of Econometrics, Vol. 5, March 1977, pp. 221-240.

Goldberger, A. S. (1962), "Best linear unbiased prediction in the generalized regression model," Journal of the American Statistical Association, Vol. 57, June 1962, pp. 369-375.

Goldberger, A. S. (1964), Econometric Theory, New York: John Wiley and Sons, 1964.

Heckman, J. J. (1977), "Sample selection bias as a specification error," University of Chicago, The Rand Corporation, and the National Bureau of Economic Research, February 1977 (unpublished).

Kapteyn, A. and T. J. Wansbeek (1976), "Autocorrelation in the linear model when observations are missing," Economic Institution, Leyden, November 1976 (unpublished).

Kelejian, H. H. (1969), "Missing observations in multivariate regression: Efficiency of a first-order method," Journal of the American Statistical Association, Vol. 64, December 1969, pp. 1609-1616.

Kmenta, J. (1971), Elements of Econometrics, New York: Macmillan Company, 1971.

Riddell, W. C. (1977), "Prediction in generalized least squares," The American Statistician, Vol. 31, May 1977, pp. 88-90.

Sargan, J. D. and E. G. Drettakis (1974), "Missing data in an autoregressive model," International Economic Review, Vol. 15, February 1974, pp. 39-58.

Savin, N. E. and K. J. White (1976), "Testing for autocorrelation with missing observations," University of British Columbia, Department of Economics, Discussion Paper 76-25, August 1976 (unpublished).

Schim van der Loeff, S. and L. Leclerq (1975), "A note on Goldberger's best linear unbiased predictor in the generalized regression model," Erasmus University Rotterdam, Econometric Institute, Report 7503/ES, January 1975 (unpublished).

Schmidt, P. J. (undated), "A note on the estimation of seemingly unrelated regressions with unequal numbers of observations," University of North Carolina (unpublished).