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]

MISSING MEASUREMENTS IN A REGRESSION PROBLEM WITH NO AUXILIARY RELATIONS

P. BALESTRA
University of Geneva, 1211 Geneva 4, Switzerland

J. KMENTA
University of Michigan
Ann Arbor, Michigan 48109, USA

R-113.84

MISSING MEASUREMENTS IN A REGRESSION PROBLEM WITH NO AUXILIARY RELATIONS*

This paper deals with the problem of estimating the coefficients of classical and generalized regression models when some sample measurements are missing and no auxiliary relations are postulated. It is shown that all relevant information about the regression coefficients is contained in the complete portion of the sample and that there is no gain in using the incomplete portion.

1. INTRODUCTION

In considering the problem of missing measurement in the context of a regression model, most authors have attempted to fill the missing values by using some approximation scheme. Commonly the approximations of the missing values have been derived by invoking *ad hoc* the existence of an auxiliary relation between the variables for which some observations are missing and the variables for which all observations are available. [See, e.g., Dagenais (1973), Gourieroux and Monfort (1981), or Conniffe (1983).] This *ad hoc* relation is then used to predict the missing values of the variables of interest.

The problem with using an auxiliary relation is that for the predictions to work the relationship should also hold in previously unobserved situations, which means that there should be a reason for it. Such a reason would be provided by theory. Without any justification for the postulated auxiliary relation other than that provided by the observed correlations, its presumption is inappropriate. For instance, in an equation designed to explain changes over time in the household demand for wine, the explanatory variables would typically include household income and price of wine. Clearly, a prediction of household income on the basis of the price of wine is absurd. In the case where there is some theoretical justification for the auxiliary relation to hold in unobserved situations, this should be made explicit whether there are missing

*This paper represents a substantially revised and generalized version of an earlier paper presented at the Econometric Society European Meeting in Athens, 1979. [See Kmenta (1981)]. The generalizations were derived independently by Balestra.

measurements or not. In such a case the researcher would be dealing with a system of equations (typically a recursive one) rather than with a single regression equation.

Another approach to the problem of missing values in regression is to assume that the values of the dependent and of the explanatory variables all come from a multivariate normal distribution. The missing values are then treated as unknown parameters to be estimated along with the regression coefficients. This approach was initiated by Anderson (1957) and was followed by other authors whose work is summarized and extended in a series of papers by Afifi and Elashoff (1966, 1967, and 1969). A formal analysis of the problem is presented in Kelejian (1969); computational results and some further derivations are given in Beale and Little (1975). We regard the assumption of a multivariate normal distribution as unduly restrictive since it excludes purely nonstochastic variables, dummy variables, and variables generated by distributions other than normal.

In what follows we consider the problem of estimating the coefficients of a regression model with missing measurements when no auxiliary relations can be justified and when the omission of incomplete observations leaves the sample selection rule unaffected. In Section 2 we discuss the estimation of a multiple regression model for which all classical assumptions are satisfied, and in Section 3 we extend the analysis to the generalized regression model. For each model we consider the case of missing values of explanatory variables separately from the case of missing values of the dependent variable.

2. CLASSICAL MULTIPLE REGRESSION MODEL

Consider a regression model

$$Y = X\beta + \varepsilon \tag{2.1}$$

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.

Missing measurements on X

Let us now suppose that of the total of n observations only n_1 $(n_1 < n)$ are complete while in the remaining $n_2 = n = n_1$ observations the values of one or more of the explanatory variables are missing. In this case we may partition X and y as follows:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \qquad (2.2)$$

where $X_1 \rightarrow (n_1 \times K)$ is a matrix of measurements on X corresponding to the complete observations, and $X_2 \rightarrow (n_2 \times K)$ 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.

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

$$\hat{\beta}^* = (X_1'X_1)^{-1}X_1'Y_1. \qquad (2.3)$$

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

$$\text{Var-Cov}(\hat{\beta}^*) = \sigma^2(X_1'X_1)^{-1}. \qquad (2.4)$$

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

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{2.5}$$

with the variance-covariance matrix given as

$$\begin{aligned} \text{Var-Cov}(\hat{\beta}) &= \sigma^2(X'X)^{-1} \\ &= \sigma^2(X_1'X_1 + X_2'X_2)^{-1}. \end{aligned} \tag{2.6}$$

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

$$\begin{aligned} \text{Var-Cov}(\hat{\beta}^*) - \text{Var-Cov}(\hat{\beta}) &= \sigma^2[(X_1'X_1)^{-1} - (X'X)^{-1}] \\ &= \sigma^2(X_1'X_1)^{-1}X_2'[I + X_2(X_1'X_1)^{-1}X_2']^{-1}X_2(X_1'X_1)^{-1} \end{aligned} \tag{2.7}$$

which is clearly a non-negative definite matrix.

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. The log-likelihood function may be written as

$$\begin{aligned} L(Y|\beta,X_2,\sigma^2) = &-(n/2)\log 2\pi - (n/2)\log \sigma^2 - (1/2\sigma^2)\{Y'Y - 2\beta'X_1'Y_1 - 2\beta'X_2'Y_2 \\ &+ \beta'X_1'X_1\beta + \beta'X_2'X_2\beta\} \end{aligned} \tag{2.8}$$

Setting the two sets of partial derivatives[1] with respect to β and X_2 equal to zero, we get the following two normal equations:

$$X_1'Y_1 + X_2'Y_2 = X_1'X_1\overset{\Delta}{\beta} + X_2'X_2\overset{\Delta}{\beta} \tag{2.9}$$

$$Y_2\overset{\Delta}{\beta}' = X_2\overset{\Delta}{\beta}\overset{\Delta}{\beta}' \tag{2.10}$$

[1] See Balestra (1977)

Disregarding the trivial case $\overset{\Delta}{\beta} = 0$, from (2.10) we obtain

$$Y_2 - X_2\overset{\Delta}{\beta} = 0 \qquad (2.11)$$

so that X_2 must be chosen in such a way that each error corresponding to the missing observations is identically zero. This can always be done. Substituting for Y_2 from (2.11) into (2.9) and solving for $\overset{\Delta}{\beta}$, we obtain

$$\overset{\Delta}{\beta} = (X_1'X_1)^{-1}X_1'Y \qquad (2.12)$$

which is exactly the same as the least squares estimator of β based on complete observations only. Thus we have to conclude that the incomplete observations provide no additional information about the regression parameters in this case.

Although the actual estimates of X_2 are not needed in the maximum likelihood estimation of β, it is worth noting that X_2 is not in general identified. That is, equation (2.11) does not yield a unique solution for X_2 except in special circumstances. Given Y_2 and $\overset{\Delta}{\beta}$, each row of equation (2.11) represents a system of one equation in K unknowns. Therefore, a unique solution obtains if and only if there is one single unknown per row. This condition is satisfied when only one value of the explanatory variables is missing per row. This would occur, for instance, in the case of a simple regression model.

Missing measurements on Y

Let us now consider the model given in (2.1) and partitioned as in (2.2). Suppose that X_1 and X_2 are available but Y_2 is missing. A natural way to proceed is as follows. First, we use the available information on Y_1 and X to estimate Y_2; next, we expand the sample using the predicted values for Y_2; and finally, we estimate the regression coefficients from the full, expanded sample. A linear predictor for Y_2 is defined by:

$$\hat{Y}_2 = AY_1 \qquad (2.13)$$

where A is a $n_2 x n_1$ matrix of constants. For the predictor to be unbiased (in the sense that the prediction errors have zero expectation), it is necessary that:

$$AX_1 = X_2 \tag{2.14}$$

The expanded sample thus becomes:

$$\tilde{Y} = X\beta + \tilde{\varepsilon} \tag{2.15}$$

where

$$\tilde{Y} = \begin{bmatrix} Y_1 \\ \hat{Y}_2 \end{bmatrix} = \begin{bmatrix} Y_1 \\ AY_1 \end{bmatrix} = \begin{bmatrix} I \\ A \end{bmatrix} Y_1 = FY_1 \tag{2.16}$$

and

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \tilde{\varepsilon}_2 \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ A\varepsilon_1 \end{bmatrix} = \begin{bmatrix} I \\ A \end{bmatrix} \varepsilon_1 = F\varepsilon_1 \tag{2.17}$$

Note that the matrix F has full column rank equal to n_1. Note also that in the expanded sample ε must be replaced by $\tilde{\varepsilon}$, since $\hat{Y}_2$ takes the place of Y_2. From the easily checked fact that

$$FX_1 = \Lambda \tag{2.18}$$

the expanded sample may be written as

$$FY_1 = FX_1\beta + F\varepsilon_1 \tag{2.19}$$

which represents a linear transformation of the model corresponding to the complete observations only. It is therefore evident that no additional information can be gained by expanding the sample using any linear unbiased predictor for Y_2. The rest of this section provides a detailed proof of the above assertion in the case of a classical regression model.

When the disturbances are homoskedastic and independent, the best linear unbiased predictor of Y_2 is $\hat{Y}_2 = AY_1$ where

$$A = X_2(X_1'X_1)^{-1}X_1' \tag{2.20}$$

Using (2.20), we may estimate β in (2.15) by ordinary least squares, although this is an inappropriate procedure since $\tilde{\varepsilon}$ does not satisfy the classical assumptions. Noting that

$$X'F = X_1 + X_2'X_2(X_1'X_1)^{-1}X_1' \tag{2.21}$$

$$= X_1'X_1(X_1'X_1)^{-1}X_1' + X_2'X_2(X_1'X_1)^{-1}X_1'$$

$$= (X_1'X_1 + X_2'X_2)(X_1'X_1)^{-1}X_1'$$

$$= (X'X)(X_1'X_1)^{-1}X_1',$$

the ordinary least squares estimator of β based on (2.15) and (2.20) is

$$\hat{\beta} = (X'X)^{-1}X'\tilde{Y} = (X'X)^{-1}X'FY_1 = (X'X)^{-1}X'X(X_1'X_1)^{-1}X_1'Y_1 = (X_1'X_1)^{-1}X_1'Y \tag{2.22}$$

which is just the ordinary least squares estimator based on the complete observations.

Since the variance-covariance matrix of $\tilde{\varepsilon}$ is (2.15) is

$$\text{Var-Cov }(\tilde{\varepsilon}) = \sigma^2FF' \tag{2.23}$$

$$= \sigma^2R$$

which is a positive semi-definite matrix of rank n_1, the generalized least squares estimator of β in (2.15) becomes

$$= (X'R^+X)^{-1}X'R^+Y = (X'R^+X)^{-1}X'R^+FY_1 \tag{2.24}$$

where R^+ is a generalized inverse of R such that

$$RR^+R = R \text{ and } R^+RR^+ = R^+. \tag{2.25}$$

From (2.21) and (2.18) we first observe that

$$X'R = X'FF' = (X'X)(X_1'X_1)^{-1}X_1F' = (X'X)(X_1'X_1)^{-1}X' \tag{2.26}$$

or, solving for X',

$$X' = (X_1'X_1)(X'X)^{-1}X'R. \tag{2.27}$$

Using (2.25) it can be shown that $RR^+F = F$, so that by (2.21)

$$X'R^+F = (X_1'X_1)(X'X)^{-1}X'RR^+F = (X_1'X_1)(X'X)^{-1}X'F = X_1' \tag{2.28}$$

and therefore

$$X'R^+X = X'R^+FX_1 = X_1'X_1 \tag{2.29}$$

where reference is made to (2.18) and (2.28). From (2.28) and (2.29) it is evident that the generalized least squares estimator of β is simply the ordinary least squares estimator, i.e., that

$$\tilde{\beta} = (X_1'X_1)^{-1}X_1'Y_1 \tag{2.30}$$

3. GENERALIZED REGRESSION MODEL

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

(3.1) $\quad Y = X\beta + \varepsilon$

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

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

where Ω is a known positive definite matrix which can be partitioned in accordance with (2.2) as

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \tag{3.2}$$

We shall call its inverse by V and write accordingly

$$V = \Omega^{-1} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \tag{3.3}$$

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

<u>Missing Measurements on X</u>

To estimate β in (3.1), we may apply the generalized least squares method to the set of complete observations to obtain

$$\tilde{\beta}^* = (X_1'V_{11}X_1)^{-1}X_1'V_{11}Y_1. \tag{3.4}$$

The variance-covariance matrix of β^* is

$$\text{Var-Cov}(\tilde{\beta}^*) = (X_1'V_{11}X_1)^{-1}. \tag{3.5}$$

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

$$\tilde{\beta} = (X'VX)^{-1}X'VY \tag{3.6}$$

and its variance-covariance matrix would be

$$\text{Var-Cov}(\tilde{\beta}) = (X'VX)^{-1}. \tag{3.7}$$

The loss of efficiency as a result of omitting the incomplete observations can be determined as follows. First we note that there exists a nonsingular matrix $P \rightarrow (n \times n)$ such that

$$P\Omega P' = \sigma^2 I \tag{3.8}$$

so that (3.1) can be rewritten as

$$PY = PX\beta + P\varepsilon \tag{3.9}$$

where

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

Further, in accordance with the partitioning of the Ω matrix, P can be partitioned as [2]

$$P = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \tag{3.10}$$

Using (3.9) and (3.10) in conjunction with (2.7), the exact formula for the loss of efficiency becomes

$$\text{Var-Cov}(\tilde{\beta}^*) - \text{Var-Cov}(\tilde{\beta}) = (X_1'V_{11}X_1)^{-1} - (X'VX)^{-1} \tag{3.11}$$

$$= \sigma^2(X_1'P_{11}'P_{11}X_1)^{-1}X_2'P_{22}'[I + P_{22}X_2(X_1'P_{11}'P_{11}X_1)^{-1}X_2'P_{22}']P_{22}X_2(X_1'P_{11}'P_{11}X_1)^{-1}$$

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

[2] See Riddell (1977).

maximizing the following log-likelihood function (conditional on initial values of Y if applicable)

$$L(Y|\beta,X_2) = (n/2)\log 2\pi + (1/2)\log|V| \quad - (1/2)q$$

where

$$q = Y'VY - 2\beta'X_1'V_{11}Y_1 - 2\beta'X_2'V_{22}Y_2 - 2\beta'X_2'V_{12}'Y_1 - 2\beta'X_1'V_{12}Y_2 + \beta'X_1'V_{11}X_1\beta + \beta'X_2'V_{22}X_2\beta + 2\beta'X_2'V_{12}'X_1\beta \qquad (3.12)$$

The first order conditions for the maximization of L with respect to β and X_2 are

$$-X'V_{11}Y_1 - X_2'V_{22}Y_2 - X_2'V_{12}'Y_1 - X_1'V_{12}Y_2 + X_1'V_{11}X_1\hat{\beta} + X_2'V_{22}X_2\hat{\beta} + $$
$$+ X_2'V_{12}'X_1\hat{\beta} + X_1'V_{12}X_2\hat{\beta} = 0 \qquad (3.13)$$

$$-V_{22}Y_2\hat{\beta}' - V_{12}'Y_1\hat{\beta}' + V_{22}X_2\hat{\beta}\hat{\beta}' + V_{12}'X_1\hat{\beta}\hat{\beta}' = 0 \qquad (3.14)$$

Disregarding the trivial case $\hat{\beta} = 0$, from (3.14) we get

$$(Y_2 - X_2\hat{\beta}) = - V_{22}^{-1}V_{12}'(Y_1 - X_1\hat{\beta}) \qquad (3.15)$$

In this case X_2 must be chosen so that the errors corresponding to the missing observations are linear combinations of the errors pertaining to the complete observations. We may write (3.13) as

$$-X_1'V_{11}(Y_1 - X_1\hat{\beta}) - X_2'V_{22}(Y_2 - X_2\hat{\beta}) - V_2'V_{12}'(Y_1 - X_1\hat{\beta}) -$$
$$-X_1'V_{12}(Y_2 - X_2\hat{\beta}) = 0 \qquad (3.16)$$

and, using (3.15), we get

$$X_1'[V_{11} - V_{12}V_{22}^{-1}V_{12}']X_1\hat{\beta} = X_1'[V_{11} - V_{12}V_{22}^{-1}V_{12}']Y_1 \qquad (3.17$$

From the expression we derive the maximum likelihood estimator of β as

$$\hat{\beta} = (X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1}Y_1 \qquad (3.18)$$

which is exactly the same as the generalized least squares estimator of β based on the complete observations. Again, no additional information is gained by treating X_2 as unknown parameters. As for the uniqueness of the solution for X_2, the same remarks apply as in the classical regression case.

Missing measurements on Y

Let us consider now the specification of the generalized regression model in (3.1) in the case where there are no missing values in the X matrix but the values Y_2 are not available. Under these circumstances we may replace Y_2 by its generalized least squares predictor Y_2 derived from the complete observations. It is well known that Y_2 is given as $\hat{Y}_2 = AY_1$ where[3]

$$A = X_2(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1} + \Omega_{12}'\Omega_{11}^{-1}M \qquad (3.19)$$

and

$$M = I - X_1(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1\Omega_{11}^{-1} \qquad (3.20)$$

Replacing the missing values Y_2 by Y_2 as defined above, we obtain an extended specification of the model of the form given in (2.15)-(2.17). Assuming (incorrectly) that the variance-covariance matrix of the disturbance in (2.15)-(2.17) is Ω, we can use generalized least to obtain the following estimator of β:

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}FY_1 \qquad (3.21)$$

Now

$$\begin{aligned} X'\Omega^{-1}F &= (X_1'V_{11} + X_2'V_{12}') + (X_1'V_{12} + X_2'V_{22})[X_2(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1} + \Omega_{12}'\Omega_{11}^{-1}M] \\ &= (X_1'V_{11} + X_2'V_{12}') + (X_1'V_{12}X_2 + X_2'V_{22}X_2)(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1} + \\ &\quad + (X_1'V_{12} + X_2'V_{22})\Omega_{12}'\Omega_{11}^{-1}M \\ &= (X_1'V_{11} + X_2'V_{12}') + (X'\Omega^{-1}X - X_1'V_{11}X_1 - X_2'V_{12}'X_1)(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1} + \\ &\quad + (X_1'V_{12} + X_2'V_{22})\Omega_{12}'\Omega_{11}^{-1}M \end{aligned}$$

[3] The predictor may be expressed as $\hat{Y}_2 = X_2\tilde{\beta}_1 + \Omega_{12}'\Omega_{11}^{-1}(Y_1 - X_1\tilde{\beta}_1)$ where $\tilde{\beta}_1$ is the generalized least squares estimator based on the complete observations.

$$= (X_1'V_{11} + X_2'V_{12}')[I - X_1(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1}] + (X_1'V_{12} + X_2'V_{22})\Omega_{12}'\Omega_{11}^{-1}M +$$

$$+ X'\Omega^{-1}X(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1}$$

$$= [X_1'(V_{11}\Omega_{11} + V_{12}\Omega_{12}') + X_2'(V_{12}'\Omega_{11} + V_{22}\Omega_{12}')]\Omega_{11}^{-1}M + X'\Omega^{-1}X(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1}$$

Noting that, from the properties of an inverse of a partitioned matrix,

$$V_{11}\Omega_{11} + V_{12}\Omega_{12}' = I \tag{3.22}$$

$$V_{12}'\Omega_{11} + V_{22}\Omega_{12}' = 0 \tag{3.23}$$

and that

$$X'\Omega_{11}^{-1}M = 0 \tag{3.24}$$

we see that

$$X'\Omega^{-1}F = X'\Omega^{-1}X(X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1} \tag{3.25}$$

Substituting this result into (3.21) we obtain

$$= (X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega^{-1}Y_1 \tag{3.26}$$

which is the generalized least squares estimator based on the complete observations only.

Since the variance-covariance matrix of the disturbance in (2.15)-(2.17) is not Ω but

$$F\Omega_{11}F' = R,$$

which is a positive definite matrix of rank n_1, the appropriate generalized least squares estimator should be based on the generalized-Penrose inverse of R. This is

$$R^+ = F(F'F)^{-1}\Omega_{11}^{-1}(F'F)^{-1}F' \tag{3.27}$$

From the fact that $FX_1 = X$, see (2.18), we get

$$X_1 = (F'F)^{-1}F'X \tag{3.28}$$

so that:

$$X'R^+ = X'F(F'F)^{-1}\Omega_{11}^{-1}(F'F)^{-1} = X'\Omega^{-1}(F'F)^{-1}F' \tag{3.29}$$

$$X'R^+F = X_1'\Omega_{11}^{-1} \tag{3.30}$$

$$X'R^+X = X_1'\Omega_{11}^{-1}(F'F)^{-1}F'X = X_1'\Omega_{11}^{-1}X_1 \tag{3.31}$$

The proper generalized least squares estimator of β is thus given by

$$\tilde{\beta} = (X'R^+X)^{-1}X'R^+Y = (X'R^+X)^{-1}X'R^+FY_1 = (X_1'\Omega_{11}^{-1}X_1)^{-1}X_1'\Omega_{11}^{-1}Y_1 \tag{3.32}$$

which again is the same as the generalized least squares estimator based only on the complete observations. No information is gained by first estimating Y_2. We may note, finally, that this conclusion is valid for any unbiased predictor $\tilde{Y}_2$, since in the proof only the fact that $FX_1 = X$ is used.

The preceding result pertaining to the generalized regression model applies not only to the standard cases of heterskedostic or autoregressive disturbances, but also to seemingly unrelated regressions and to models of pooled cross-section and time-series data typically characterized by a rather complicated structure of the disturbance variance-covariance matrix.

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.

Balestra, Pietro (1977), La Derivation Matricielle, Collection de l'IME (Sirey, Paris).

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

Conniffe, Denis (1983), "Small-Sample Properties of Estimators of Regression Coefficients Given a Common Pattern of Missing Data," Review of Economic Studies, Vol. 50, pp. 111-120.

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

Gourieroux, C. and A. Monfort (1981), "On the Problem of Missing Data in Linear Models," Review of Economic Studies, Vol. 48, pp. 579-586.

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. (1981), "On the Problem of Missing Measurements in the Estimation of Economic Relationships," in E.G. Charatsis (ed.), Proceedings of the Econometric Society European Meeting 1979 (North-Holland, Amsterdam).

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

WORKING PAPERS

R-101 Albert A. Hirsch, Saul H. Hymans, Harold T. Shapiro

ECONOMETRIC REVIEW OF ALTERNATIVE FISCAL AND MONETARY POLICIES, 1971-75. October 1977.

R-102 Jan Kmenta

ON THE PROBLEM OF MISSING MEASUREMENTS IN THE ESTIMATION OF ECONOMIC RELATIONSHIPS. May 1978.

R-103 Harold T. Shapiro, David M. Garman

PERSPECTIVES ON THE ACCURACY OF MACRO-ECONOMETRIC FORECASTING MODELS. January 1979.

R-104 Saul H. Hymans

SAVING, INVESTMENT, AND SOCIAL SECURITY. August 1980.

R-105 Karl Lin, Jan Kmenta

SOME NEW RESULTS ON RIDGE REGRESSION ESTIMATION.

R-106 E. Philip Howrey, Saul H. Hymans, Mark N. Greene

THE USE OF OUTSIDE INFORMATION IN ECONOMETRIC FORECASTING. June 1982.

R-107.81 Saul H. Hymans

MEDIAN VOTER MODELS AND THE GROWTH OF GOVERNMENT SERVICES. September 1981.

R-108.81 Terrence Belton, Saul H. Hymans, Cara Lown

THE DYNAMICS OF THE MICHIGAN QUARTERLY ECONOMETRIC MODEL OF THE U.S. ECONOMY. December 1981.

R-109.81 E. Philip Howrey, Saul H. Hymans

A CENTRAL LIMIT THEOREM WITH APPLICATIONS TO ECONOMETRICS. December 1981.

R-110.83 E. Philip Howrey, Saul H. Hymans

A CENTRAL LIMIT THEOREM WITH APPLICATIONS TO ECONOMETRICS. Revised April, 1983.

R-111.83 Saul H. Hymans

MACROECONOMETRICS AMIDST SENSE AND NONSENSE. August 1983.

R-112.83 Charles S. Roehrig

OPTIMAL CRITICAL REGIONS FOR PRE-TEST ESTIMATORS USING A BAYES RISK CRITERION. November 1983.

R-113.84 P. Balestra, J. Kmenta

MISSING MEASUREMENTS IN A REGRESSION PROBLEM WITH NO AUXILIARY RELATIONS.