MISSING OBSERVATIONS IN MULTIVARIATE REGRESSION: EFFICIENCY OF A FIRST ORDER METHOD[*/]

H. H. Kelejian
Princeton University

. INTRODUCTION AND SUMMARY

In many situations the experimenter encounters the problem of estimating the parameters of a regression model when some observations on certain of the regressors are missing. Among the solutions to this problem suggested in the literature are the so-called first order methods.[1/] That is, the missing observations are first estimated by various regression techniques, and then the regression parameters are estimated in terms of the completed sample by the least squares procedure. Unfortunately, results concerning the efficiency of such estimates are not available--for an exception, see Buck (1960).

The purpose of this paper is, first, to examine the nature of the information contained in the incomplete portion of the sample that is relevant in the estimation of the regression parameters. Consistent with this information, a variance-covariance matrix (henceforth, V-C matrix) is derived which, under usual assumptions, represents an unattainable upper limit on the efficiency of regression parameter estimates based upon

---

[*/] This article is based upon a research study prepared under a contract with the Office of Manpower Policy, Evaluation, and Research, U. S. Dept. of Labor, under the authority of the Manpower Development and Training Act, and jointly sponsored by the Departments of Commerce and Health, Education, and Welfare and the National Science Foundation. Researchers undertaking such projects are encouraged to express their own judgment freely. Therefore, points of view or opinion stated in this document do not necessarily represent the official position or policy of these agencies.

[1/] For a review of the literature and some interesting extensions see Afifi and Elashoff (1966, 1967).

both the complete and incomplete portions of the sample. Then, this matrix and the V-C matrix corresponding to ordinary least squares estimates based only on the complete portion of the sample are used to evaluate the asymptotic efficiency of regression parameter estimates based on a first order method. It is assumed that the missing observations relate to only one of the regressors.

2. THE MODEL

Consider the following regression model

$$y_t = x_t A + z_t B + u_t, \quad t = 1, \ldots N, \tag{1}$$

where y_t is the t^{th} observation on the regressand; x_t is the t^{th} observation on the first regressor; A is the corresponding parameter; z_t is a $1 \times K$ vector of observations at time t on the last K regressors, one of which can be taken to be the constant term; B is a $K \times 1$ vector of parameters; u_t is the t^{th} disturbance term. We assume that the regressors and the disturbance term have been generated by a stationary multivariate stochastic process with finite moments. The dependence in this process is assumed to be sufficiently weak so that the sample moments converge in probability to the corresponding population moments. We also assume that the limit, with respect to the sample size, of the mathematical expectation of a sample moment is the corresponding population moment. Concerning the relationship between the disturbance term and the regressors, we assume

$$Eu_t | (x_s, z_s) = 0 \text{ and } E\, u_t^2 | (x_s, z_s) = \sigma_u^2$$

for all t and s, and $E\, u_t u_s = 0$ for $t \neq s$. Concerning the pattern of missing observations, we assume that all the observations on the regressand, on the last K regressors, and $N_1 \geq K + 1$ observations on the first regressor are available. Finally, we assume that the process selecting the time periods for which observations on the first regressor are missing is either deterministic (e.g., every j^{th} period) or stochastic but independent of the process generating x_t and z_t.

3. ASYMPTOTICALLY EFFICIENT ESTIMATES

Let the time periods for which observations on x_t are available be numbered in the order of their occurrence by $j = 1,\ldots N_1$. Let the complementary set of time periods be numbered similarly by $r = 1,\ldots N_2$, where $N_2 = N-N_1$. Then (1) can be decomposed as

$$y_j = x_j A + z_j B + u_j, \; j = 1,\ldots N_1, \tag{2}$$

$$y_r = x_r A + z_r B + u_r, \; r = 1,\ldots N_2. \tag{3}$$

The procedure now is to eliminate x_r from (3) and then inquire as to the maximum of information contained in the resulting relationship that is relevant in the estimation of A and B.

Because the expectation of one variable conditional upon a set of others is, in general, a function of those conditioning variables, we have, assuming initially the usual linear relationship,

$$Ex_r | z_r = z_r \gamma, \tag{4}$$

where γ is a $K \times 1$ vector of parameters. It follows from (4) that x_r can be expressed as

$$x_r = z_r \gamma + v_r, \tag{5}$$

where v_r is a disturbance term such that $Ev_r | z_r = 0$ and, in general,

$Ev_r^2 | z_r = h_r$ where h_r is a function of the elements of z_r. It is evident that if x_r is autocorrelated so, in general, will be v_r: $Ev_{r_1}v_{r_2} = w_{r_1r_2} \neq 0$. A point to note is that $w_{r_1r_2}$ depends upon the particular pattern of missing observations.

Substituting (5) into (3), we obtain what might be termed an unconditional regression relationship between y_r and z_r,

$$y_r = z_r(\gamma A + B) + m_r, \quad r = 1,\ldots N_2, \tag{6}$$

where $m_r = v_rA + u_r$. Therefore, $Em_r | z_r = 0$. From (5) we see that v_r can be expressed in terms of x_r and z_r. Therefore, the assumption underlying (1) that $Eu_t | (x_s, z_s) = 0$ for all t and s implies $Eu_r | v_r = 0$. We have, therefore, $Em_r^2 | z_r = A^2h_r + \sigma_u^2$.

Equations (2) and (6), respectively, can be expressed in matrix form as

$$Y_1 = X_1A + Z_1B + U_1 \text{ and} \tag{7}$$

$$Y_2 = Z_2\pi + M_2, \quad \pi = \gamma A + B, \tag{8}$$

where Y_1, X_1, Z_1, and U_1 are the vectors and matrices of values corresponding respectively to y_j, x_j, z_j, and u_j; and Y_2, Z_2, and M_2 are the vectors and matrices corresponding to y_r, z_r, and m_r. Therefore, from (8) we see that if the regression function relating x_r to z_r is linear, the available information contained in the incomplete portion of the sample that is relevant in the estimation of A and B is in the form of a linear restriction. It is clear that this information is at a maximum when γ is known and $N_2 \to \infty$. To demonstrate this, we first show that the least squares estimate of π in (8) is consistent.

The k^{th} element of the $K \times 1$ vector $N_2^{-1}Z_2^1M_2$ is

$$p_k = N_2^{-1}\sum_{r=1}^{N_2} z_{kr}m_r, \quad k = 1,\ldots K, \tag{9}$$

where z_{kr} is the k^{th} element of z_r. Then, the condition $Em_r|z_r = 0$ implies $Em_r|z_{kr} = 0$ and this, in turn, implies $Em_r z_{kr} = 0$. Therefore, $Ep_k = 0$. Consider now the variance of p_k:

$$Ep_k^2 = N_2^{-2} E\sum_{r=1}^{N_2} \psi_{kr}^2 + 2N_2^{-2} E\sum_{i<j} \psi_{ki}\psi_{kj} \tag{10}$$

where, in general, $\psi_{kr} = z_{kr}m_r$. Note first that $E\psi_{kr} = 0$. Then, because the regressors in (1) are assumed to be generated by a stationary stochastic process, $E\psi_{kr}^2 = \sigma_\psi^2$ where σ_ψ^2 is the variance of ψ_{kr}, $r=1,\ldots N_2$. Hence, the limit as $N_2 \rightarrow \infty$ of the first term in (10) is zero.

Consider now the second term in (10). Denote this term by T. Then,

$$T = 2N_2^{-2}E \sum_{i=1}^{N_2-1} \sum_{j=1}^{N_2-i} \psi_{ki}\psi_{ki+j} \tag{11}$$

$$= 2N_2^{-2} \sum_{i=1}^{N_2-1} \sum_{j=1}^{N_2-i} \sigma_i(j)$$

where $\sigma_i(j) = E\psi_{ki}\psi_{ki+j}$. The reader should note that $\sigma_i(j)$ is a function of j alone only if the missing observations occur at equal intervals. We now assume that $|\sigma(s_1)| \leq |\sigma(s_2)|$ if $s_1 > s_2$, where $\sigma(s) = E\psi_{kt}\psi_{kt+s}$, where t and s number the original time periods corresponding to (1). That is, the dependence between elements is negatively related to the number of time periods between them. With this assumption we have $|\sigma_i(j)| \leq |\sigma(j)|$, $i=1,\ldots N_2$. Therefore,

$$|T| \leq 2N_2^{-2} \sum_{i=1}^{N_2-1} \sum_{j=1}^{N_2-i} |\sigma(j)| = 2N_2^{-2} \sum_{j=1}^{N_2-1} (N_2-j)|\sigma(j)|. \tag{12}$$

Thus a sufficient condition for $|T| \to 0$ as $N_2 \to \infty$ is

$$\sum_{j=1}^{N_2-1} |\sigma(j)| = \text{const.} \tag{13}$$

Therefore, letting Plim denote probability limit, we have, by assuming (13), $\underset{N_2 \to \infty}{\text{Plim}}\ p_k = 0$. Hence, under these conditions we have $\underset{N_2 \to \infty}{\text{Plim}}\ N_2^{-1} Z_2' M_2 = 0.$

By a similar argument, and by making similar assumptions, we may show that $\underset{N_2 \to \infty}{\text{Plim}}\ N_2^{-1} (Z_2' Z_2)^{-1} = \bar{V}_z^{-1}$ where $E z_t' z_t = \bar{V}_z$. Henceforth, we assume

that whatever the pattern of missing observations, the probability limit of a statistic based either on the complete or on the incomplete portion of the sample is equal to the limit of the expectation of that statistic.

Using the above results, it can be shown that $\hat{\pi} = (Z_2' Z_2)^{-1} Z_2' Y_2$ is a consistent estimate of π in (8). Therefore, if γ is known and $N_2 \to \infty$, asymptotically efficient estimates of A and B in (7) are obtained by minimizing the corresponding error sum of squares, $U_1' U_1$, subject to the restriction $\hat{\pi} = \gamma A + B$. In particular, let $C' = (A\ B')$ $P_1 = (X_1 Z_1)$, and $R = (\gamma I_k)$ where I_k is the $K \times K$ unit matrix. Further, denote the least squares estimate of C which is based only on the complete portion of the sample by $\hat{C}_{LS} = (P_1' P_1)^{-1} P_1' Y_1$. Then, as $N_2 \to \infty$, the efficient estimate of C is C* where [2/]

[2/] For a discussion of restricted least squares estimation, see Goldberger (1964, pp. 256-57).

$$C^{*} = \hat{C}_{LS} + \left[(P_1'P_1)^{-1}R'\right]\left[R(P_1'\ P_1)^{-1}R'\right]^{-1}\left[\hat{\pi} - R\hat{C}_{LS}\right]. \qquad (14)$$

The V-C matrix of C^{*} is

$$V_{*} = \lim_{N_2 \to \infty} E(C^{*} - C)\ (C^{*} - C)' = V_{LS} - \sigma_u^2\ E(P_1'P_1)^{-1}R'\left[R(P_1'P_1)^{-1}R'\right]R(P_1'P)^{-1} \qquad (15)$$

where $V_{LS} = E(\hat{C}_{LS} - C)\ (\hat{C}_{LS} - C)'$. Under the assumed conditions that γ is known and $N_2 \to \infty$, it can be shown that C* is the minimum variance unbiased estimator of C - see Theil (1961, pp. 536-38). Therefore, because γ, in general, will not be known, the V-C matrix given in (15) can be considered as an unattainable upper limit on the efficiency of estimates of A and B.

Before leaving this section, we briefly outline the consequences of having a nonlinear regression function relate x_r to z_r. Assume

$$Ex_r | z_r = f(z_r)d \qquad (16)$$

where $f(z_r)$ is a nonlinear function of the elements of z_r, and d is a factor of proportionality. Then, by an analysis similar to the above, (3) can be written as

$$Y_2 = F(Z_2)\ dA + Z_2B + G \qquad (17)$$

where $F(Z_2)$ is a $N_2 \times 1$ vector of observations on $f(z_r)$, and G is an $N_2 \times 1$ vector of disturbances such that $Eg_r | z_r = 0$, where g_r is the r^{th} element of G. Therefore, assuming $f(z_r)$ is known, the ordinary least squares estimates of dA and B, and, if d is known, of A converge in probability to these parameters as $N_2 \to \infty$. Clearly, even if $f(z_r)$ contains unknown parameters, consistent estimates of dA and B, with respect to N_2, exist if the parameters of $f(z_r)$ can be consistently estimated along with dA and B.

4. EFFICIENCY OF A FIRST ORDER METHOD

Consider the usual first order method in which the missing observations on the first regressor are estimated in terms of the linear regression of that regressor on the last K regressors, and then A and B are estimated by the least squares procedure applied to the completed sample. In this case the N_2 missing observations in (3), x_r, $r = 1,\ldots N_2$, are first replaced by their respective values in the $N_2 \times 1$ vector $\hat{X}_2 = Z_2\hat{\gamma}$, $\hat{\gamma} = (Z_1'Z_1)^{-1}Z_1'X_1$, and then A and B are estimated by the least squares procedure applied to (2) and (3).

The normal equations for the estimates of A and B, say $\hat{A}$ and $\hat{B}$, based on the first order method can be expressed as

$$(H_1 + H_2)\hat{C} = P_1'Y_1 + P_2'Y_2, \tag{18}$$

where $H_1 = P_1'P_1$, $H_2 = P_2'P_2$, $P_1 = (X_1Z_1)$ $P_2 = (\hat{X}_2Z_2)$, and $\hat{C}' = (\hat{A}\ \hat{B}')$. We now use a theorem due to Tocher (1951, p. 41) and elaborated in Tiao and Zellner (1964, p. 283) which says that if Q_1 is a $m \times n$ matrix and Q_2 is a $n \times m$ matrix, then assuming the inverse to exist,

$$(I_m - Q_1Q_2)^{-1} = I_m + Q_1(I_n - Q_2Q_1)^{-1}Q_2. \tag{19}$$

Using Tocher's theorem, we note that

$$(H_1 + H_2)^{-1} = (I_{k+1} + H_1^{-1}H_2)^{-1}H_1^{-1}$$

$$= \left[I_{k+1} - H_1^{-1}(I_{k+1} + H_2H_1^{-1})^{-1}H_2\right]H_1^{-1}. \tag{20}$$

Combining the results in (20) with those in (18) and noting that $\hat{C}_{LS} = H_1^{-1}P_1'Y_1$, we have

$$\hat{C} = (I_{k+1} - H_1^{-1}(I_{k+1} + H_2H_1^{-1})^{-1} H_2)\,\hat{C}_{LS} + (H_1+H_2)^{-1} P_2'Y_2 \qquad (21)$$

$$= \hat{C}_{LS} - H_1^{-1}(I_{k+1} + H_2H_1^{-1})^{-1} H_2\hat{C}_{LS} + H_1^{-1}(I_{k+1} + H_2H_1^{-1})^{-1} P_2'Y_2$$

$$= \hat{C}_{LS} + H_1^{-1}(I_{k+1} + H_2H_1^{-1})^{-1} (P_2'Y_2 - H_2\hat{C}_{LS}).$$

We now note that P_2' and H_2 can be expressed as

$$P_2' = (Z_2\hat{\gamma}\ Z_2)' = \hat{R}'Z_2' \qquad (22)$$

$$H_2 = P_2'P_2 = \hat{R}'Z_2'Z_2\hat{R} \qquad (23)$$

where $\hat{R} = (\hat{\gamma}\ I)$. Using (22) and (23), the final expression in (21) can be written as

$$\hat{C} = \hat{C}_{LS} + H_1^{-1}\left[I_{k+1} + \hat{R}'Z_2'Z_2\hat{R}H_1^{-1}\right]^{-1}\hat{R}'Z_2'Z_2\left[\hat{\pi} - \hat{R}\hat{C}_{LS}\right]. \qquad (24)$$

We see, therefore, that $\hat{C}$ differs from $\hat{C}_{LS}$ by a linear function of the extent to which $\hat{C}_{LS}$ fails to satisfy the linear restriction $\hat{\pi} = \hat{R}\hat{C}_{LS}$.

We now investigate the asymptotic properties of $\hat{C}$ as N_1 and $N_2 \to \infty$ in such a way that $\lambda \to 0$ or $\lambda \to \infty$ where $\lambda = N_1/N_2$. More explicitly, we show that $\hat{C} \to \hat{C}_{LS}$ as $\lambda \to \infty$ and $\hat{C} \to \hat{C}^*$ as $\lambda \to 0$ where $\hat{C}^*$ is identical to C^* in (14) with the exception that, since γ is unknown, R is replaced everywhere in that expression by $\hat{R}$.

Applying Tocher's theorem to the first factor postmultiplying H_1^{-1} in (24) and taking $Q_1 = -\hat{R}'Z_2'Z_2$ we have, after simplifying,

$$\hat{C} = \hat{C}_{LS} + H_1^{-1}\hat{R}'(I_k - \hat{\Gamma}^{-1})\ Z_2'Z_2\ (\hat{\pi} - \hat{R}\hat{C}_{LS}), \qquad (25)$$

where $\hat{\Gamma} = \left[I_k + \hat{D}^{-1}(Z_2'Z_2)^{-1}\right]$, and $\hat{D} = \hat{R}H_1^{-1}\hat{R}'$. We now note that $H_1^{-1} = N_1^{-1}(N_1H_1^{-1}) = N_1^{-1}V_{xz}^{-1}$ and $Z_2'Z_2 = N_2(N_2^{-1}Z_2'Z_2) = N_2V_z$ where $\underset{N_1 \to \infty}{\text{Plim}}\, V_{xz} = \bar{V}_{xz}$ and $\underset{N_2 \to \infty}{\text{Plim}}\, V_z = \bar{V}_z$, where $\bar{V}_{xz} = E(x_t z_t)'(x_t z_t)$. Substituting these expressions into (25), we have

$$\hat{C} = \hat{C}_{LS} + V_{xz}^{-1}\hat{R}'\,\frac{1}{\lambda}\left[I_k - \frac{1}{\lambda}(I_k\,\frac{1}{\lambda} + \hat{D}_*^{-1}V_z^{-1})^{-1}\right]\; V_z(\hat{\pi}-\hat{R}\hat{C}_{LS}), \qquad (26)$$

where $\hat{D}_* = \hat{R}\,V_{xz}^{-1}\hat{R}'$.

From (26) we see that $\hat{C}$ can be expressed as

$$\hat{C} = \hat{C}_{LS} + \Delta_1,\ \Delta_1 = \frac{1}{\lambda}\Delta_2 \qquad (27)$$

where Δ_2 is a $(K + 1)\times 1$ vector. Recall now that $\hat{\pi}$ is a consistent estimate of π: $\underset{N_2 \to \infty}{\text{Plim}}\, \hat{\pi} = \pi$. Using this along with an assumption similar to (13), it can be shown that

$$\underset{N_2 \to \infty}{\text{Plim}}\, N_2(\hat{\pi} - RC)\ (\hat{\pi} - RC)' = \bar{V}_{\hat{\pi}} \qquad (28)$$

where $\bar{V}_{\hat{\pi}}$ is a matrix of finite constants - i.e., the asymptotic V-C matrix of $\hat{\pi}$ is $N_2^{-1}\bar{V}_{\hat{\pi}}$. Finally, note that $\hat{R}$ and $\hat{C}_{LS}$ are estimates of R and C which are based upon a sample of size N_1. Therefore, combining the results in (26) - (28) we see that Δ_1 is a vector of higher order elements in the sense that $\underset{(\lambda \to \infty)}{\text{Plim}}\, \Delta_1 = 0$ and $\underset{(\lambda \to \infty)}{\text{Plim}}\, N_1\Delta_1\Delta_1' = 0$ where $(\lambda \to \infty)$ is a summary statement for $N_1 \to \infty$, $N_2 \to \infty$, and $\lambda \to \infty$. A point to note is that Δ_1 contains two factors, namely, $\frac{1}{\lambda}$ and $(\hat{\pi} - \hat{R}\hat{C}_{LS})$, that approach zero as $(\lambda \to \infty)$.

The upshot of the above argument is that with the exception of the higher order vector Δ_1, $\hat{C}$ converges to $\hat{C}_{LS}$ as $(\lambda \to \infty)$. Hence, since $\hat{C}_{LS}$ is easier to compute, there is no reason to consider the first order estimate $\hat{C}$. Furthermore, one should note that this conclusion does not depend upon our lack of knowledge concerning γ. That is, even if γ were known, a priori, $\hat{C}$ would still converge to $\hat{C}_{LS}$ as $(\lambda \to \infty)$. The reader may prove this by noting that equations (18) - (28) do not depend upon the definition of $\hat{\gamma}$; they hold, therefore, if $\hat{\gamma}$ is replaced by γ.

Assume now that the constant term is the second regressor in (1) - i.e., $z_{1t} \equiv 1$. Then note that equations (18) - (28) also hold if $\hat{\gamma}'$ is replaced by the $1 \times k$ vector

$$\bar{\delta}' = (\bar{X}\ 0\ \dots\ 0),\ \bar{X} = N_1^{-1} \sum_{j=1}^{N_1} x_j \tag{29}$$

Therefore, if the sample is completed by taking $\bar{X}$ as the estimate of each of the N_2 missing observations z_r, $r = 1, \dots N_2$, and then C is estimated by applying the least squares procedure to the completed sample, the resulting "zero order" estimate of C would converge to $\hat{C}_{LS}$ as $(\lambda \to \infty)$ - see Afifi and Elashoff (1966, pp. 598-99) for a discussion of zero order estimates.

Consider again equation (26) but assume now that as N_1 and $N_2 \to \infty$, $\lambda \to 0$. Let $(\lambda \to 0)$ summarize these conditions. Then, applying a power series expansion to the inverse matrix postmultiplying $\frac{1}{\lambda}$ in (26), we see that $\hat{C}$ can be expressed as

$$\hat{C} = \hat{C}_{LS} + V_{xz}^{-1} \hat{R}' \frac{1}{\lambda} \left[I_k - (I_k - \lambda \hat{D}_*^{-1} V_z^{-1} + \lambda^2\ T_*) \right] V_z\ (\hat{\pi} - \hat{R}\hat{C}_{LS}) \tag{30}$$

$$= \hat{C}_{LS} + V_{xz}^{-1} \hat{R}' \left[\hat{D}_*^{-1} V_z^{-1} - \lambda\ T_* \right] V_z\ (\hat{\pi} - \hat{R}\hat{C}_{LS})$$

where T_* is a K × K matrix such that $\underset{N_1 \to \infty}{\text{Plim}}\, T_* = T^*$ where T^* is a matrix of constants. Noting that $\underset{(\lambda \to 0)}{\text{Plim}}\, (\hat{\pi} - \hat{R}\hat{C}_{LS}) = 0$, it follows from (30) and the definitions of V_{xz}^{-1} and D_*^{-1} that

$$\hat{C} = \hat{C}_{LS} + V_{xz}^{-1} \hat{R}' \hat{D}_*^{-1} (\hat{\pi} - \hat{R}\hat{C}_{LS}) + \Delta_3 \tag{31}$$

$$= \hat{C}^* + \Delta_3$$

where Δ_3 is a (K + 1) × 1 vector of higher order elements: $\underset{(\lambda \to 0)}{\text{Plim}}\, \Delta_3 = 0$ and $\underset{(\lambda \to 0)}{\text{Plim}}\, N_2 \Delta_3 \Delta_3' = 0$. Hence, with the exception of Δ_3, $\hat{C} = \hat{C}^*$.

We turn now to derive the asymptotic V-C matrix of $\hat{C}^*$. Our procedure is to express $\hat{C}^*$ as a function of C^* and then use results corresponding to C^*.

Recalling that $\hat{R} = (\hat{\gamma}\ I_k)$ and $\hat{D}_* = \hat{R} V_{xz}^{-1} \hat{R}'$, we note that

$$\hat{R} = R + e, \quad e = (\hat{\gamma} - \gamma \quad 0) \tag{32}$$

$$\hat{D}_* = D_* + \Phi, \quad D_* = R V_{xz}^{-1} R', \tag{33}$$

where Φ is a K ×K matrix such that $\underset{N_1 \to \infty}{\text{Plim}}\, \Phi = 0$. Combining the results in (32) and (33) with those in (31) we have

$$\hat{C}^* = \hat{C}_{LS} + V_{xz}^{-1} R' [D_* + \Phi]^{-1} [\hat{\pi} - R\hat{C}_{LS}] - \tag{34}$$

$$V_{xz}^{-1} R' [D_* + \Phi]^{-1} e\hat{C}_{LS} + \Delta_4$$

where Δ_4 is a vector of higher order elements such that $\underset{(\lambda \to 0)}{\text{Plim}}\, \Delta_4 = 0$ and $\underset{(\lambda \to 0)}{\text{Plim}}\, N_1 \Delta_4 \Delta_4' = 0$. Henceforth, all such higher order sectors are denoted by Δ^* - e.g., $\Delta_4 = \Delta^*$.

We now note that

$$[D_* + \Phi]^{-1} = D_*^{-1} \, [I_k + \Phi \, D_*^{-1}]^{-1} \tag{35}$$

$$= D_*^{-1} + \sum_{j=1}^{\infty} (-1)^j \, D_*^{-1} \, (\Phi D_*^{-1})^j.$$

Substituting (35) into (34) and noting that $\underset{(\lambda \to 0)}{\text{Plim}} \, D_*^{-1}\Phi \, D_*^{-1} = \underset{(\lambda \to 0)}{\text{Plim}} \, \Phi D_*^{-1} = 0$, we have

$$\hat{C}^* = \hat{C}_{LS} + V_{xz}^{-1}R' \, D_*^{-1}(\hat{\pi} - R\hat{C}_{LS}) - V_{xz}^{-1} \, R' \, D_*^{-1}e\hat{C}_{LS} + \Delta^* \tag{36}$$

$$= C^* - V_{xz}^{-1} \, R' \, D_*^{-1}e \, \hat{C}_{LS} + \Delta^*.$$

As expected, we see from (36) that with the exception of the higher order term Δ^*, the difference between $\hat{C}^*$ and C^* is a function of $e = (\hat{\gamma} - \gamma \;\; 0)$.

In order to derive the asymptotic V-C matrix of $\hat{C}^*$ as $(\lambda \to 0)$, we drop the higher order term and impose the consistency condition $\hat{\pi} = \pi$ since $\hat{\pi}$ is based upon a sample of size N_2. Then, the asymptotic V-C matrix of $\hat{C}^*$ is, from (36),

$$N_1^{-1} \, \bar{V}_{\hat{C}*} = N_1^{-1} \lim_{(\lambda \to o)} N_1 E(\hat{C}^* - C) \, (\hat{C}^* - C)'$$

$$= N_1^{-1} \, \bar{V}_* + N_1^{-1} \lim_{(\lambda \to o)} N_1 (ES_1 + ES_2 + ES_3),$$

where $\bar{V}_* = \lim_{N_1 \to o} E \, N_1 V_*$, $S_1 = (C^* - C) \, \hat{C}_{LS}' \, e' \, D_*^{-1} R V_{xz}^{-1}$, $S_2 = S_1'$, and $S_3 = V_{xz}^{-1}R'D_*^{-1}e \, \hat{C}_{LS}\hat{C}_{LS}' \, e' \, D_*^{-1} \, R \, V_{xz}^{-1}$. We now show that the limits corresponding to S_1 and S_2 are zero. To do this we first note that C^* and $\hat{C}_{LS}$ may be expressed as

$$C^* = C + S_4 \, U_1 \tag{38}$$

$$\hat{C}_{LS} = C + S_5 U_1 \qquad (39)$$

where S_4 and S_5 are $(K + 1)\times N_1$ matrices such that $ES_4U_1 = ES_5U_1 = 0$ and $\lim_{(\lambda\to 0)} N_1 ES_4 S_5' = \frac{1}{\sigma_u^2} \bar{V}_*$. Substituting (38) and (39) into the expression for S_1, we have

$$ES_1 = ES_4U_1C'e'D_*^{-1}R\, V_{xz}^{-1} + ES_4U_1U_1'S_5'e'D_*^{-1}R\, V_{xz}^{-1} \qquad (40)$$

$$= \sigma_u^2\, ES_4S_5'e'D_*^{-1}R\, V_{xz}^{-1},$$

since $EU_1|X_1,Z_1 = 0$ and $EU_1U_1'|X_1,Z_1 = \sigma_u^2\, I_{N_1}$. Therefore, in the light of (40) we have

$$\lim_{(\lambda\to 0)} N_1ES_1 = \sigma_u^2 (\operatorname{Plim}_{(\lambda\to 0)} N_1S_4S_5')\ (\operatorname{Plim}_{(\lambda\to 0)} e')\ (\operatorname{Plim}_{(\lambda\to 0)} D_*^{-1}RV_{xz}^{-1}) \qquad (41)$$

$$= 0,$$

since the probability limit of e' is zero while the probability limits of the two $(K + 1)\times (K + 1)$ matrices are matrices of finite constants.

Consider now the limit in (37) corresponding to S_3. To obtain this limit, we first note that $e\hat{C}_{LS} = (\hat{\gamma} - \gamma)\hat{A}_{LS}$ where $\hat{A}_{LS}$ is the least squares estimate of A based on the complete portion of the sample. Note further that $\hat{A}_{LS}$ is a scalar. Therefore, $E\hat{A}_{LS}^2 = A^2 + E\,\sigma^2(\hat{A}_{LS})$ where $\sigma^2(\hat{A}_{LS})$ is the variance of $\hat{A}_{LS}$ conditional upon X_1 and Z_1. Using these results, we have

$$ES_3 = E(A^2 + \sigma^2(\hat{A}_{LS}))\, V_{xz}^{-1}R'D_*^{-1}(\hat{\gamma} - \gamma)\ (\hat{\gamma} - \gamma)'\ D_*^{-1}R\, V_{xz}^{-1}. \qquad (42)$$

Therefore, since $\underset{(\lambda \to 0)}{\text{Plim}}\ \sigma^2(\hat{A}_{LS}) = 0$,

$$\underset{(\lambda \to 0)}{\lim} N_1 ES_3 = \underset{(\lambda \to 0)}{\text{Plim}} N_1 S_3 = A^2 \bar{V}_{xz}^{-1} R' \bar{D}_*^{-1} \bar{V}_\gamma \bar{D}_*^{-1} R \bar{V}_{xz}^{-1}, \tag{43}$$

where $\bar{V}_\gamma = \underset{(\lambda \to 0)}{\text{Plim}} N_1 (\hat{\gamma} - \gamma)(\hat{\gamma} - \gamma)'$ and $\bar{D}^{-1} = R \bar{V}_{xz}^{-1} R'$.

Combining the results in (40) - (43) with those in (37), we see

$$N_1^{-1} \bar{V}_{\hat{C}*} = N_1^{-1} \bar{V}_* + N_1^{-1} A^2 \bar{V}_{xz}^{-1} R' \bar{D}_*^{-1} \bar{V}_\gamma \bar{D}_*^{-1} R \bar{V}_{xz}^{-1} \tag{44}$$

$$= N_1^{-1} \bar{V}_* + A^2 L (N_1^{-1} \bar{V}_\gamma) L'$$

where $L = \bar{V}_{xz}^{-1} R' \bar{D}_*^{-1}$. In brief, the difference between the asymptotic V-C matrix of $\hat{C}^*$ and that of C^* depends upon A^2 as well as the asymptotic V-C matrix of $\hat{\gamma}$. The reader should note that $\bar{V}_\gamma$ depends upon the particular pattern of missing observations.

Since $A^2 > 0$ and $\bar{V}_\gamma$ is a nonsingular V-C matrix and so must be positive definite, we see from (44) that the asymptotic variances of the elements of $\hat{C}^*$ are larger than those of the corresponding elements of C^*. The question now arises as to whether or not the asymptotic variances of the elements of $\hat{C}^*$ are larger than those of the corresponding elements in $\hat{C}_{LS}$. In order to answer this question we use (15) to derive an expression for $N_1^{-1} \bar{V}_*$ in terms of the asymptotic V-C matrix of $\hat{C}_{LS}$, say, $N_1^{-1} \bar{V}_{LS}$, and then substitute into (44) to get

$$N_1^{-1} \bar{V}_C = N_1^{-1} \bar{V}_{LS} - \sigma_u^2 N_1^{-1} \bar{V}_{xz}^{-1} R' \bar{D}_*^{-1} R \bar{V}_{xz}^{-1} \tag{45}$$

$$+ A^2 L (N_1^{-1} \bar{V}_\gamma) L'$$

$$= N_1^{-1} \bar{V}_{LS} + N_1^{-1} \bar{V}_{xz}^{-1} R' \left[A^2 \bar{D}_*^{-1} \bar{V}_\gamma \bar{D}_*^{-1} - \sigma_u^2 \bar{D}_*^{-1} \right] R \bar{V}_{xz}^{-1} .$$

Therefore, a sufficient condition that the asymptotic variances of the elements in $\hat{C}^*$ be less than or equal to those of $\hat{C}_{LS}$ is that $(A^2 \bar{D}_*^{-1} \bar{V}_\gamma \bar{D}_*^{-1} - \sigma_u^2 \bar{D}_*^{-1})$ be negative definite.

List of References

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

2. Afifi, A.A. and Elashoff, R.M. (1967), "Missing Observations in Multivariate Statistics II. Point Estimation in Simple Linear Regression," Journal of the American Statistical Association, March 1967, pp. 10 - 29.

3. Buck, S.F. (1960), "A Method of Estimation of Missing Values in Multivariate Data Suitable for use with an Electronic Computer," Journal of the Royal Statistical Society, Series B, 22, 1960, pp. 302 - 306.

4. Draper, N.R. and Smith, H. (1966), Applied Regression Analysis, New York: John Wiley and Sons, 1966.

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

6. Hussain, A. and Wallace, T.D. (1967), "A Note on Some Misconceptions of Asymptotic Distribution Theory in Econometrics," Unpublished Paper, North Carolina State University, October 1967.

7. Loeve, M. (1955), Probability Theory, New York: D. Van Nostrand Company, 1955.

8. Theil, H. (1951), Economic Forecasts and Policy, Amsterdam: North-Holland, 1961.

9. Tiao, G.C. and Zellner, A. (1964), "On the Bayesion Estimation of Multivariate Regression," Journal of the Royal Statistical Society, Series B, 26, pp. 277-285.

List of References (continued)

10. Tocher, K.D. (1951), "Discussion on Mr. Box and Dr. Wilson's paper," Journal of the Royal Statistical Society, Series B, 13, pp. 39-42.

11. Wold, H. (1960), "A Generalization of Causal Chain Models: (Part III of a Triptych on Causal Chain Systems)" Econometrica, April 1960, pp. 443-463.