ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 44 — 1996

Estimation of SUR Model with Non-nested Missing Observations

Hae-shin HWANG, Craig SCHULMAN*

ABSTRACT. – This paper considers alternative two-step estimators and their small sample properties for the seemingly unrelated regression (SUR) model with ***non-nested*** missing observations. A Monte Carlo experiment indicates that alternative estimators have more profound differences in their efficiency, compared to the case of ***nested*** missing observations. In particular, the two-step application of the Hartley-Hocking maximum likelihood estimator can realize a significant gain in efficiency. There are substantial losses in efficiency when only the subset of data that has complete observations is used in estimation.

L'estimation du modèle SUR avec des observations manquantes non-imbriquées

RÉSUMÉ. – Cet article prend en considération des estimateurs de remplacement à deux échelons et leurs petites propriétés d'échantillon pour le modèle "Régression en apparence sans rapports" (Seeminly Unrelated Regression, SUR) avec des observations manquantes non-imbriquées. Une expérience à la Monte Carlo indique que des estimateurs de remplacement ont des différences plus notables quant à leur efficacité comparé aux cas d'observations manquantes imbriquées. En particulier, l'application à deux échelons de l'estimateur à probabilité maximale Hartley-Hocking peut réaliser d'appréciables gains en efficacité. Il n'y a perte considérable de gains d'efficacité uniquement lorsqu'on emploie dans l'estimateur un sous-ensemble de données ayant des observations complètes.

* H. S. HWANG: Department of Economics, Texas A&M University; C. SCHULMAN: Department of Economics, University of Arkansas.

1 Introduction

This paper considers alternative estimators and their small sample properties for the seemingly unrelated regression (SUR) model with ***non-nested*** missing observations. A data set with missing observations is said to be ***non-nested*** if the observed sample periods of the variables cannot be arranged in a sequence of nested subsets. Estimation of the SUR model with ***nested*** missing observations has been examined by SCHMIDT [1977], CONNIFFE [1985], HWANG [1990], and BALTAGI et al. [1988]. Schmidt conducts a Monte Carlo experiment to compare the small sample properties of the maximum likelihood estimator (MLE) and four alternative two-step GLS estimators in a two equation SUR model where the regression coefficients are subject to linear restrictions. The alternative estimators of the error covariance matrix Ω that Schmidt considered are the estimators proposed by WILKS [1932], HOCKING and SMITH [1968], and SRIVASTAVA and ZAATAR [1973], and the "usual" estimator. Conniffe derives the small sample properties of the MLE analytically when no restrictions are imposed on the coefficients.

SCHMIDT [1977] reports a "surprising" result in his Monte Carlo study of a two-equation SUR model. The HOCKING and SMITH [1968] estimator of the error covariance matrix Ω does not produce more efficient coefficient estimates than other alternative estimators of Ω, although the Hocking and Smith estimator appears to use more sample information. This sampling result proves to be remarkably robust in a two-equation model with ***nested*** missing observations.

HWANG [1990] notes, however, that Schmidt's interpretation of the sample information contained in each estimator of Ω is misleading. The Hocking and Smith estimator does not necessarily use more sample information than other estimators of Ω. Furthermore, Hwang identifies a sample statistic α that differentiates alternative estimators of Ω. If α is equal to one, there is no difference among alternative estimators. The statistic α is the ratio of two estimators for the variance of the error term in the equation which has no missing observations: One estimator uses the full sample, and the other uses the jointly observed sample only. This ratio is expected to be close to one on the average and hence, all estimators of Ω exhibit a similar property. HWANG [1990] reports that the Hocking-Smith estimator can be significantly more efficient than other estimators in the subsample of sufficiently high value of α, if the correlation between the error terms in the two equations is very high.

In the case of non-nested missing observations, alternative estimators of Ω are differentiated by more than one sample statistic, and hence are expected to have more diverse properties. This can lead to more distinct two-step GLS estimators of the coefficients in a small sample. On this premise, this paper examines the relative efficiencies of alternative estimators in a general SUR model with linear restrictions and non-nested missing observations.

There are numerous studies on the methodology for analyzing incomplete data situations [1]. The methodology varies with the characteristics of the missing observations and the purpose of the analysis. In this paper, we apply the direct Maximum Likelihood approach of HARTLEY and HOCKING [1971] and the "Missing Information Principle" of ORCHARD and WOODBURY [1972] to the estimation of SUR models with non-nested missing observations and linear restrictions of the coefficients. We present the general case of missing observations in a compact matrix form by using the elimination and duplication matrices, and derive in section 3 the normal equations for the coefficients and covariance matrix that arise from these two approaches. Since the normal equations are nonlinear in unknown parameters, iterative algorithms to solve the equations are presented. The case of the nested missing observations is a special case. In particular, we show that the estimation algorithm for the coefficients becomes recursive when missing observations are nested and no restrictions are imposed on the regression coefficients. This result generalizes Conniffe's analysis of the two group case. We also generalize in section 4 the estimators of Ω that Schmidt considered in a two equation model to the case of more than two equations. These estimators of Ω are different from the estimators of Hartley-Hocking and Orchard-Woodbury, and their relationships are illustrated in a two equation model.

The small sample properties of alternative two-step estimators are compared in section 5 through Monte Carlo experiments in a two-equation model with non-nested missing observations. The results confirm our conjectures. There are more profound differences in the efficiency of the coefficient estimates between the "usual" estimator and other estimators in the non-nested case than in the nested case. The effects of the sample statistics α_1 and α_2 on the relative efficiencies of the alternative estimators are also more significant in the non-nested case. The experiments also reveal that the estimates of coefficients using only the data of complete observations are substantially inferior to other estimators that utilize the missing observations. Section 6 concludes the paper with a summary of the findings.

2 Model Specification

A standard SUR model with G equations and T observations is specified in a general multiple regression form

(1) $$Y = XB + U,$$

1. There is an extensive literature on the estimation of population parameters of a statistical model from a sample with missing observations on some variables. See a brief review in the appendix of ORCHARD and WOODBURY [1972], the extensive references in HARTLEY and HOCKING [1971] and in DEMPSTER, LAIRD and RUBIN [1977]. See also TRAWINSKI and BARGMANN [1964], KELEGIAN [1969], DAGENAIS [1973] and BEALE and LITTLE [1975].

where $Y(T \times G)$ and $X(T \times K)$ are the observation matrices on the endogenous and exogenous variables, respectively. The coefficient matrix $B(K \times G)$ is subject to a set of linear restrictions

$$R\beta = r, \tag{2}$$

where $\beta = \operatorname{vec}(B)$, R is a matrix of known constants with full row rank, and r is a vector of known constants. Each row of the disturbance matrix U is assumed to be an i.i.d. normal random vector with zero mean and a nonsingular covariance matrix Ω.

In the case of incomplete data, some elements of y_t, the t-th row of Y, are missing [2]. The observation matrix Y is classified into N subgroups such that T_n observations in the n-th subgroup exhibit the same pattern of incompleteness. The n-th subgroup has complete observations on G_n endogenous variables and missing observations on $(G - G_n)$ variables.

Let P_n be the $T_n \times T$ observation selection matrix for the n-th subgroup: P_n consists of T_n distinct rows of a T-dimensional identity matrix. An allocation matrix that selects the observed variables in the n-th subgroup is denoted by Q_n, which consists of G_n distinct columns of a G-dimensional identity matrix. The remaining columns of I_G corresponding to the variables with missing observations is denoted by $\overline{Q}_n$. Using these notations, the regression equations for the observed and unobserved variables in the n-th subgroup can be written as

$$\begin{aligned} Y_n Q_n &= X_n B_n + V_n, \\ Y_n \overline{Q}_n &= X_n \overline{B}_n + \overline{V}_n, \end{aligned} \tag{3}$$

where $Y_n = P_n Y$, $X_n = P_n X$, $U_n = P_n U$, $V_n = U_n Q_n$, $\overline{V}_n = U_n \overline{Q}_n$, $B_n = BQ_n$ and $\overline{B}_n = B\overline{Q}_n$. The covariance matrices of the rows of V_n and $\overline{V}_n$ are given by $\Omega_n = Q_n' \Omega Q_n$ and $\overline{\Omega}_n = \overline{Q}_n' \Omega \overline{Q}_n$, respectively. To facilitate matrix calculus, the following notations are used. For a symmetric matrix A of dimension n, $\nu(A)$ is an $n(n+1)/2$ dimensional column vector obtained from $\operatorname{vec}(A)$ by eliminating all supradiagonal elements of A. The elimination matrix Ξ has the dimension $n(n+1)/2$ by n^2 such that $\nu(A) = \Xi \operatorname{vec}(A)$. The duplication matrix D is defined to be an n^2 by $n(n+1)/2$ matrix such that $D\nu(A) = \operatorname{vec}(A)$. Using these notations, the distinctive elements of Ω and Ω_n can be expressed as

$$\begin{aligned} \delta &= \nu(\Omega) = \Xi \operatorname{vec}(\Omega), \qquad & D\delta &= \operatorname{vec}(\Omega), \\ \delta_n &= \nu(\Omega_n) = \Xi_\eta \operatorname{vec}(\Omega_n), \qquad & D_n \delta_n &= \operatorname{vec}(\Omega_n). \end{aligned}$$

---

2. When some of the regressors are also missing, one may apply the "zero order regression method" or the "first order regression method" to replace the missing regressors with their predicted values (see AFIFI and ELASHOFF [1968]). If the observations on a regressor x_k are missing only when the corresponding dependent variable, for which x_k enters with a nonzero coefficient, has missing observations, then x_k may be set to zero (or any arbitrary value) without affecting the estimation algorithm. The model that Schmidt considered belongs to this special case.

The relationship between δ and δ_n is defined by the selection matrix C_n such that

$$\delta_n = C_n\,\delta = \Xi_n\,(Q_n' \otimes Q_n')\,D\delta.$$

The log likehood function for the n-th group on the basis of the observed values only is given by

(4) $$L_n\,(V_n;\,\theta) = \text{constant} - (T_n/2)\log|\Omega_n| \; - 0.5\,\text{tr}\,(V_n'\,V_n\,\Omega_n^{-1}),$$

and for the entire set of observations for all N groups by

(5) $$L\,(V_1,\,\ldots,\,V_N;\,\theta) = \Sigma\,L_n\,(V_n;\,\theta).$$

3 Maximum Likelihood Estimator

The maximum likelihood estimators of β and δ are obtained from the maximization of the log likehood function (5) subject to restrictions (2). Thus the function to be maximized is

(6) $$L^R\,(\theta) = \Sigma\,L_n\,(V_n;\,\theta) - \lambda'\,(R\,\beta - r),$$

where λ is a vector of Lagrangean multipliers. The first order conditions for a maximum are

(7) $$\partial L^R/\partial\beta = \Sigma\,[\partial L_n\,(V_n;\,\theta)/\partial\beta] - R'\,\lambda = 0,$$

(8) $$\partial L^R/\partial\lambda = \; -(R\beta - r) = 0,$$

(9) $$\partial L^R/\partial\delta = \Sigma\,(\partial\delta_n/\partial\delta)\;(\partial L_n\,(V_n;\,\theta)/\partial\delta_n) = 0.$$

Two alternative algorithms for the solutions of (7)-(9) are presented in this section. The first approach, employed by HARTLEY and HOCKING [1971], is a direct differentiation of (6) with respect to θ and λ, and the second approach is based on the "missing information principle" of ORCHARD and WOODBURY [1972], which is a special case of the more general EM algorithm.

(a) Hartley-Hocking (HH) algorithm

To derive the first order conditions we first differentiate (4) with respect to $\beta_n = \text{vec}(B_n)$:

$$(10) \quad \partial L_n\,(V_n;\,\theta)/\partial\beta_n = (\Omega_n^{-1} \otimes X_n')\,(Y_n\,Q_n)^c - (\Omega_n^{-1} \otimes X_n'\,X_n)\,\beta_n,$$

where superscript c denotes the vec operator. Since $\beta_n = (Q_n' \otimes I)\,\beta$ and $\partial\beta_n/\partial\beta = Q_n \otimes I$, we have

$$(11) \quad \begin{aligned}\partial L_n\,(V_n;\,\theta)/\partial\beta &= (Q_n\,Q_n^{-1} \otimes X_n')\,(Y_n\,Q_n)^c - \Psi_n\,\beta \\ &= (Q_n\,\Omega_n^{-1}\,Q_n' \otimes X_n'\,Y_n)\,I_G^c - \Psi_n\,\beta\end{aligned}$$

where $\Psi_n = (Q_n\,\Omega_n^{-1}\,Q_n' \otimes X_n'\,X_n)$. The derivative of (2.4) with respect to δ is given by

$$(12) \quad \begin{aligned}\partial L_n\,(V_n;\,\theta)/\partial\delta &= (\partial\delta_n/\partial\delta)\,(\partial L_n\,(V_n;\,\theta)/\partial\delta_n) \\ &= (\partial\delta_n/\partial\delta)\,(T_n/2)\,D_n'\,(\Omega_n^{-1} \otimes \Omega_n^{-1})\,D_n\,(d_n - \delta_n) \\ &= C_n'\,W_n\,(d_n - \delta_n) = C_n'\,W_n\,d_n - C_n'\,W_n\,C_n\,\delta,\end{aligned}$$

where $d_n = \nu\,(V_n'\,V_n/T_n)$ and $W_n = (T_n/2)\,D_n'\,(\Omega_n^{-1} \otimes \Omega_n^{-1})\,D_n$. Substituting (11) and (12) into (7) and (9), respectively, the system of equations to be solved for β, λ and δ can be summarized as

$$(13a) \quad \Psi\,\hat\beta + R'\,\hat\lambda = \Sigma\,(Q_n\,Q_n^{-1}\,Q_n' \otimes X_n'\,Y_n)\,I_G^c$$

$$(13b) \quad R\,\hat\beta = r$$

$$(14) \quad W\,\hat\delta = \Sigma\,C_n'\,W_n\,d_n,$$

where $\Psi = \Sigma\Psi_n$ and $W = \Sigma\,C_n'\,W_n\,C_n$.

The iterative procedure begins with an initial consistent estimate of δ. The new estimate of β is obtained from (13), and d_n is computed from the residuals $V_n = Y_n\,Q_n - X_n\,B_n$. The new estimate of δ is then computed by

$$(15) \quad \hat\delta = W^{-1}\,(\Sigma\,C_n'\,W_n\,d_n),$$

evaluating W_n and W by the previous estimate of δ.

It is easy to verify (see SILVEY [1975] that, if $T_n/T \to \alpha_n$ and $X_n'\,X_n/T_n \to M_{XX}$ as $T \to \infty$, where α_n is a finite positive constant and M_{XX} is the positive definite moment matrix of X, then the asymptotic distribution of $\hat\beta$ is given by

$$(16) \quad \sqrt{T}\,(\hat\beta - \beta) \sim N\,\{0,\ \Psi_0^{-1} - \Psi_0^{-1}\,R'\,[R\,\Psi_0^{-1}\,R']^{-1}\,R\,\Psi_0^{-1}\}$$

where $\Psi_0 = \Sigma\,\alpha_n\,(Q_n\,\Omega_n^{-1}\,Q_n' \otimes M_{XX})$.

Two remarks about the maximum likelihood algorithm are in order. First, the estimator of δ in (15) requires the inverse of W whose dimension is $G(G+1)/2$. When the number of equations is large, the following algorithm that does not require the inverse of W may prove useful for computational speeds as well as for numerical accuracy. Suppose that the first group has no missing observations on all G variables, so that $C_1 = I$, $\Omega_1 = \Omega$, $W_1 = (T_1/2)\, D' (\Omega^{-1} \otimes \Omega^{-1})\, D$, and $\Xi_1 = \Xi$. Then, using Lemma 4.4 (vi) in MAGNUS and NEUDECKER [1980], we can write $W_1^{-1} = (2/T_1)\, \Xi\, (\Omega \otimes \Omega)\, \Xi'$. Therefore, the normal equation (14) for δ can be rewritten as

$$(17) \qquad \hat{\delta} = d_1 + (2/T_1)\, \Xi(\Omega \otimes \Omega)\, \Xi' \left(\sum_{n=2}^{N} C_n'\, W_n\, (d_n - \delta_n) \right).$$

The right hand side of (17) is now evaluated by the previous estimate of δ in the iterative procedure.

Second, the estimation algorithm for β becomes recursive when the missing observations are ***nested*** and no restrictions are imposed on β. CONNIFFE [1985] examined this case and presented the small sample distribution of $\hat{\beta}$ when there are two subgroups ($N = 2$). For more than two subgroups ($N > 2$), suppose that the first g_1 equations have complete observations, the next g_2 equations have T_2 missing observations in the beginning, and the next g_3 equations have $T_3 > T_2$ missing observations in the beginning, and etc. Let $I_G = (J_1,\ J_2,\ \ldots,\ J_N)$ be a partition of an identity matrix, where J_k has the dimension $G \times g_k$. It is easy to verify in this case that $Q_n = (J_1,\ J_2,\ \ldots,\ J_n)$, $n = 1,\ 2,\ \ldots,\ N$, and that $Q_n\, \Omega_n^{-1}\, Q_n'\, \Omega\, J_k = J_k$ for all $n \geq k$. When there are no restrictions on β, its estimator in (13a) becomes

$$\Sigma\, [X_n'\, X_n\, B\, (Q_n\, \Omega_n^{-1}\, Q_n')] = \Sigma\, [X_n'\, Y_n\ (Q_n\, \Omega_n^{-1}\, Q_n')].$$

Post-multiplying $\Omega\, J_k$ and using the relationship $Q_n\, \Omega_n^{-1}\, Q_n'\, \Omega\, J_k = J_k$ for all $n \geq k$, we find the estimator of $b_k = B J_k$, the coefficient for the k-th group of equations which have T_k number of missing observations, as

$$(18) \quad \hat{b}_k = \tilde{b}_k + \left(\sum_{n=k}^{N} X_n'\, X_n \right)^{-1} \left(\sum_{n=1}^{k-1} X_n'\, (Y_n - X_n\, B)\, Q_n\, \Omega_n^{-1}\, Q_n'\, \Omega\, J_k \right)$$

where

$$\tilde{b}_k = \left(\sum_{n=k}^{N} X_n'\, X_n \right)^{-1} \left(\sum_{n=k}^{N} X_n'\, Y_n\, J_k \right)$$

is the OLS estimator of the coefficients in the k-th group of equations on the basis of their observed samples only. For the first group ($k = 1$) of equations with complete observations, $\hat{b}_1 = \tilde{b}_1$. The estimator $\hat{b}_k$, $k \geq 2$, for the k-th group of equations is its own OLS estimator $\tilde{b}_k$, adjusted for the missing observations by the second term on the right hand side of (18). The adjustment factor involves coefficient estimators in the first $k-1$ groups of equations. Note that $\Omega_n^{-1}\, Q_n'\, \Omega\, J_k$ in (18) is the regression coefficient matrix between $U_n\, J_k$ and $U_n\, Q_n$ for $n < k$.

(b) Orchard-Woodbury (OW) algorithm

Using the definition of the conditional density of $\overline{V}_n$ given V_n, the derivative of the log likelihood function (4) can be written as

$$\partial L_n (V_n;\, \theta)/\partial\theta = \partial L_n (U_n;\, \theta)/\partial\theta - \partial L_n (\overline{V}_n | V_n;\, \theta)/\partial\theta.$$

Taking the conditional expectations (given V_n) of both sides, and using $E\,[\partial L_n (\overline{V}_n|V_n;\, \theta)/\partial\theta|V_n] = 0$, we obtain

$$\partial L_n (V_n;\, \theta)/\partial\theta = E\,[\partial L_n (U_n;\, \theta)/\partial\theta|V_n]. \tag{19}$$

This equation suggests an alternative way to find the normal equations. The derivatives that are required to evaluate the right hand side of (19) are

$$\partial L_n (U_n;\, \theta)/\partial\beta = (\Omega^{-1} \otimes X_n')\, Y_n^c - (\Omega^{-1} \otimes X_n' X_n)\, \beta, \tag{20}$$

$$\partial L_n (U_n;\, \theta)/\partial\delta = (1/2)\, D' \,(\Omega^{-1} \otimes \Omega^{-1})\, D\nu\, (U_n' U_n - T_n\, \Omega). \tag{21}$$

Thus we need to find the conditional expectations of Y_n and $U_n' U_n$, given V_n. It is easy to verify that

$$E\,(Y_n\, \bar{Q}_n|V_n) = X_n\, B\, \overline{Q}_n + \bar{V}_n^*, \tag{22}$$

$$E\,(V_n' \,\overline{V}_n|V_n) = V_n' \,\overline{V}_n^*, \tag{23}$$

$$E\,(\overline{V}_n' \,\overline{V}_n|V_n) = \overline{V}_n^{*\prime}\, \overline{V}_n^* + T_n\, \{\overline{\Omega}_n - (\overline{Q}_n' \,\Omega\; Q_n)\, \Omega_n^{-1}\, (Q_n' \,\Omega\, \overline{Q}_n)\}, \tag{24}$$

where $\bar{V}_n^* = V_n\, \Omega_n^{-1}\, (Q_n' \,\Omega\, \bar{Q}_n)$ is the predicted value of $\bar{V}_n$ given by the regression function of a multivariate normal random vector. Since $E\,(Y_n\, Q_n|V_n) = Y_n\, Q_n$ and $E\,(V_n' \,V_n|V_n) = V_n' \,V_n$, placing these conditional expectations in the proper positions, we can write

$$\hat{Y}_n = E\,(Y_n|V_n) = X_n\, B + V_n\, \Omega_n^{-1}\, Q_n' \,\Omega \tag{25}$$

$$E\,(U_n' \,U_n|V_n) = F_n' \begin{bmatrix} V_n' \,V_n & V_n' \,\overline{V}_n^* \\ \bar{V}_n^{*\prime}\, V_n & E\,(\overline{V}_n' \,\overline{V}_n \,|\, V_n) \end{bmatrix} F_n \tag{26}$$

where $F_n = (Q_n,\, \overline{Q}_n)^{-1}$.

Therefore, the system of first order conditions (7)-(9) can now be written as

$$[\Sigma\, (\Omega^{-1} \otimes X_n' \,X_n)]\, \hat{\beta} + R' \,\hat{\lambda} = \Sigma(\,\Omega^{-1} \otimes X_n' \,\hat{Y}_n)\, I_G^c \tag{27a}$$

$$R\, \hat{\beta} = r \tag{27b}$$

$$\hat{\Omega} = (1/T)\, \Sigma\, E\,(U_n' \,U_n|V_n). \tag{28}$$

Starting with the initial estimates of β and Ω the iterative procedure first computes (22)-(24), then constructs $\hat{Y}_n$ and $E\,(U_n' U_n|V_n)$ according to (25) and (26). The new estimates of β and Ω will then follow from (27) and (28). The asymptotic distribution of the OW estimator $\hat{\beta}$ is the same as that in (16). Unlike the HH algorithm, however, the estimated covariance matrix of $\hat{\beta}$ does not emerge in the process of solving (27). it has to be computed separately, which is a disadvantage of the OW algorithm.

4 Alternative Covariance Estimators

This section presents the generalization of four alternative estimators of Ω that SCHMIDT [1977] considered in his Monte Carlo experiments for the two-equation SUR model with nested missing observations. These estimators of Ω may be used in place of (15) for the Hartley-Hocking algorithm or in place of (28) for the Orchard-Woodbury algorithm without affecting the asymptotic properties of the coefficient estimators.

(a) "Usual" estimator

The "usual" estimator S of Ω is the moment matrix of residuals in a subgroup in which no variable has missing observations, i.e., $\overline{V}_n$ is empty.

(b) Wilks estimator

Let u_i, $i = 1, 2, \ldots, G$ be the i-th column of U (i.e., the residual vector of the i-th equation), which has m_i observed values and $(T - m_i)$ missing observations. Let A_i be an $m_i \times T$ selection matrix, which consists of m_i distinct rows of a T-dimensional identity matrix, such that $u_i^* = A_i\, u_i$ is the vector of observed values of the i-th variable. The Wilks estimator S_{ij}^* of ω_{ij} is defined by

(29) $$S_{ij}^* = [u_i^{*\prime}\, A_i\, A_j'\, u_j^*]/[\mathrm{tr}\,(A_i'\; A_i\, A_j'\; A_j)].$$

The variance of a variable is computed by using all its observations. The covariance is computed for a pair of variables by using only their joint observations. FAREBROTHER [1975] proposed to use this estimator in the estimation of SUR model. Its drawback is that the estimated covariance matrix may not be positive definite.

(c) Srivastava-Zaatar (SZ) estimator

Let S_{ij} denote the ij-th element of the "usual" estimator S, and define a diagonal matrix Λ with $(S_{ii}^*/S_{ii})^{1/2}$ as its i-th diagonal element, where S_{ii}^*

is given in (29). The Srivastava-Zaatar estimator of Ω is defined by

$$\hat{\Omega} = \Lambda\, S\ \Lambda, \tag{30}$$

which is a positive definite matrix with probability 1. The elements of this estimator can be expressed as

$$\omega_{ij} = S_{ij}\,[(S^*_{ii}\,S^*_{ij})/(S_{ii}\,S_{jj})]^{1/2}, \tag{31}$$

which is proportional to the "usual" estimator S_{ij}, the proportionality factor being the geometric means of S^*_{ii}/S_{ii} and S^*_{jj}/S_{jj}. This estimator does not fully utilize the sample cross-moments in the estimation of the covariance terms when there are more than two variables.

(d) Hocking-Smith estimator

HOCKING and SMITH [1968] proposed a sequential procedure, which is an alternative algorithm to solve (14) for δ. Suppose that the last subgroup has no missing observations on all variables. Let

$$H_i = \sum_{n=i}^{N} C'_n\, W_n\, C_n, \qquad \xi_i = \sum_{n=i}^{N} C'_n\, W_n\, d_n, \quad i = 1, 2, \ldots, N.$$

Then, using the recursive relationships

$$\begin{aligned} &H_i = H_{i+1} + C'_i\, W_i\, C_i, && \xi_i = \xi_{i+1} + C'_i\, W_i d_i \\ &H_i^{-1} = H_{i+1}^{-1} - H_{i+1}^{-1}\, C'_i\, \Phi_{i+1}^{-1}\, C_i\, H_{i+1}^{-1}, && \Phi_{i+1} = C_i\, H_{i+1}^{-1}\, C'_i + W_i^{-1}, \end{aligned}$$

we can write

$$\hat{\delta}_i \equiv H_i^{-1}\, \xi_i = \hat{\delta}_{i+1} - H_{i+1}^{-1}\, C'_1\, \Phi_{i+1}^{-1}\, [C_i\, \hat{\delta}_{i+1} - d_i], \quad i = 1, 2, \ldots, N-1. \tag{32}$$

The procedure starts with an initial consistent estimator for $\hat{\delta}_N$, and compute $\hat{\delta}_i$ recursively backward, evaluating the unknown matrices on the right hand side of (32) by $\hat{\delta}_{i+1}$, until the final estimate $\hat{\delta}_l = \hat{\delta}$ is computed. Since this procedure is an alternative algorithm to solve (14) for δ, it will yield the same estimate as the HH algorithm when convergence is obtained.

5 Sampling Experiments

In this section we report the results of a Monte Carlo experiment investigating the small sample properties of the alternative estimators of Ω. For comparability with the studies of the case of *nested* missing observations in SCHMIDT [1977], BALTAGI et al. [1988] and HWANG [1990], we consider the same two equations model:

$$(y_1,\, y_2) = X\,(\beta_1,\, \beta_2) + (u,\, \varepsilon),$$

with the same zero restrictions on the coefficients. The second equation has T_1 missing observations on y_2 in the beginning, the first equation has T_3 missing observations on y_1 at the end, and there are no missing observations in the middle T_2 observations. Let u_i and ε_i be $T_i \times 1$ subvector of u and ε, respectively. u_3 and ε_1 are residuals in missing observations.

Let $\hat{u}_1$ and $\hat{u}_2$ be the OLS residuals from the first equation using all $(T_1 + T_2)$ observations, and $\hat{\varepsilon}_2$ and $\hat{\varepsilon}_3$ be the OLS residuals from the second equation using all $(T_2 + T_3)$ observations. Let

$$\begin{aligned}
&S_{11} = \hat{u}_2'\,\hat{u}_2/T_2, \qquad S_{12} = \hat{u}_2'\,\hat{\varepsilon}_2/T_2, \quad S_{22} = \hat{\varepsilon}_2'\,\hat{\varepsilon}_2/T_2,\\
&S_{11}^* = (\hat{u}_1'\,\hat{u}_1 + \hat{u}_2'\,\hat{u}_2)/(T_1 + T_2), \qquad S_{22}^* = (\hat{\varepsilon}_2'\,\hat{\varepsilon}_2 + \hat{\varepsilon}_3'\,\hat{\varepsilon}_3)/(T_2 + T_3)\\
&\alpha_1 = S_{11}^*/S_{11}, \qquad \alpha_2 = S_{22}^*/S_{22}, \qquad \hat{\rho} = S_{12}/\sqrt{S_{11}\,S_{22}}.
\end{aligned}$$

Note that α_i is the ratio of the estimate of the variance S_{ii}^* from all observations to the estimate S_{ii} from the joint observations in the i-th equation. $\hat{\rho}$ is an estimate of the correlation coefficient. The "usual" estimators of ω_{ij} are S_{ij} and the Wilks estimators are given by S_{11}^*, S_{12}, and S_{22}^*. The SZ estimators are S_{11}^*, $S_{12}\,\sqrt{\alpha_1\,\alpha_2}$, and S_{22}^*.

For the Hartley-Hocking maximum likelihood estimator of this two equation model, the notations specified in section 3 are

$$\begin{aligned}
&\delta_1 = \sigma_{11}, \qquad \delta_2 = \delta, \qquad \delta_3 = \sigma_{22},\\
&C_1 = (1,\, 0,\, 0), \qquad C_2 = I_3, \qquad C_3 = (0,\, 0,\, 1),\\
&D_1 = 1, \quad D_2 = D, \quad D_3 = 1,\\
&d_1 = u_1'\,u_1/T_1, \quad d_2 = d = (u_2'\,u_2/T_2,\, u_2'\,\varepsilon_2/T_2, \varepsilon_2'\,\varepsilon_2/T_2)', \quad d_3 = \varepsilon_3'\,\varepsilon_3/T_3,\\
&W_1 = T_1/(2\,\omega_{11}^2), \quad W_2 = (T_2/2)\,D'\,(\Omega^{-1} \otimes \Omega^{-1})\,D, \quad W_3 = T_3/(2\,\omega_{22}^2),
\end{aligned}$$

and

$$\delta = \begin{pmatrix}\omega_{11}\\ \omega_{21}\\ \omega_{22}\end{pmatrix} \qquad D = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} \qquad \Xi = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}.$$

Consider first the *nested* case with $T_3 = 0$, and let $\tilde{W} \equiv (C_1'\,W_1\,C_1 + W_2)$. In this case, the solution for δ in (3.9) becomes

$$\begin{aligned}
(33) \qquad \tilde{\delta} &= (C_1'\,W_1\,C_1 + W_1)^{-1}\,(C_1'\,W_1\,d_1 + W_2\,d_2)\\
&= \tilde{W}^{-1}\,(C_1'\,W_1\,d_1 + W_2\,d_2)\\
&= d_2 - [T_1/(T_1 + T_2)]\,[u_2'\,u_2/T_2 - u_1'\,u_1/T_1]\,(1,\, \theta_1,\, \theta_1^2)'
\end{aligned}$$

where $\theta_1 = \omega_{21}/\omega_{11}$. If the right hand side of (33) is evaluated by the usual estimator S_{ij}, then it becomes the Hocking-Smith estimator for the nested case.

$$(34)\qquad \begin{aligned}\tilde{\omega}_{11} &= S^*_{11} = \alpha_1\, S_{11},\\ \tilde{\omega}_{12} &= S^*_{11}\,(S_{12}/S_{11} = \alpha_1\, S_{12},\\ \tilde{\omega}_{22} &= S_{22} - (S_{11} - S^*_{11})\,(S_{12}/S_{11})^2 = S_{22}\,[1 - (1-\alpha_1)\,\hat{\rho}^2].\end{aligned}$$

For the *non-nested* case, the Hartley-Hocking maximum likelihood estimator for δ is given by

$$(35)\quad \begin{aligned}\hat{\delta} &= (\hat{W} + C_3'\, W_3\, C_3)^{-1}\,(C_1'\, W_1\, d_1 + W_2\, d_2 + C_3'\, W_3 d_3)\\ &= \tilde{\delta} - (\tilde{\omega}_{22} - d_3)\,(C_3\, \tilde{W} C_3' + W_3^{-1})\,(C_1'\, W_1\, C_1 + W_2)^{-1}\, C_3'\\ &= \tilde{\delta} - (\tilde{\omega}_{22} - d_3)\, h\end{aligned}$$

where $\tilde{\delta}$ and $\tilde{\omega}_{22}$ are given in (33), and h is a 3×1 vector:

$$h_1 = T_2\, T_3\, \theta_2^2/\tau, h_2 = T_3\, \theta_2\, (T_1{+}T_2{-}T_1\, \theta_1\, \theta_2)/\tau, h_3 = T_3(T_1{+}T_2{-}T_1\, \theta_1^2\, \theta_2^2)/\tau$$

where $\tau = (T_1 + T_2)\,(T_2 + T_3) - T_1\, T_3\, \theta_1^2\, \theta_2^2$, $\theta_1 = \omega_{12}/\omega_{11}$, and $\theta_2 = \omega_{12}/\omega_{22}$.

The non-nested estimator is obtained by adjusting the nested estimator. The adjustment factor is proportional to the difference $(\tilde{\omega}_{22} - d_3)$, which can be rewritten by using (34) as

$$\tilde{\omega}_{22} - d_3 = S_{22}\, F\,(\alpha_1,\ \alpha_2),$$

where

$$F\,(\alpha_1,\ \alpha_2) = (1-\alpha_2)\,(1 + T_2/T_3) - (1-\alpha_1)\,\hat{\rho}^2.$$

The sequential procedure of the Hocking and Smith estimator evaluates h_i by using the nested estimate $\tilde{\delta}$, while the Hartley-Hocking maximum likehood iterative procedure will evaluate h_i by using the initial "usual" estimate d_2. In the latter case we can write

$$(36)\quad \begin{aligned}\hat{\omega}_{11} &= S_{11}\,[\alpha_1 - F\,(\alpha_1,\ \alpha_2)\,\hat{\rho}^2\, T_2\, T_3/\tau],\\ \hat{\omega}_{12} &= S_{12}\,\{\alpha_1 - F\,(\alpha_1,\ \alpha_2)\,[(1-\hat{\rho}^2)\,T_1 + T_2]\,T_3/\tau\},\\ \hat{\omega}_{22} &= S_{22}\,\{1 - (1-\alpha_1)\,\hat{\rho}^2 - F\,(\alpha_1,\ \alpha_2)\,[(1-\hat{\rho}^4)\,T_1 + T_2]\,T_3/\tau\}.\end{aligned}$$

The Orchard-Woodbury algorithm computes the estimates of the missing residuals u_3 and ε_1 by $\hat{u}_3 = \theta_2\, \varepsilon_3$ and $\hat{\varepsilon}_1 = \theta_1\, u_1$. Let u^+ and ε^+ be the residual vectors appended by $\hat{u}_3$ and $\hat{\varepsilon}_1$ for the missing parts. Then, the OW estimate of Ω is given by

$$(37)\qquad \begin{aligned}\hat{\omega}_{11} &= u^{+\prime}\, u^+/T + (T_3/T)\,\omega_{11.2},\\ \hat{\omega}_{12} &= u^{+\prime}\, \varepsilon^+/T,\\ \hat{\omega}_{22} &= \varepsilon^{+\prime}\, \varepsilon^+/T + (T_1/T)\,\omega_{22.1}.\end{aligned}$$

where $\omega_{11.2} = \omega_{11} - \omega_{12}\,\theta_2$ and $\omega_{22.1} = \omega_{22} - \omega_{12}\,\theta_1$. Evaluation of the right hand side expressions in (37) leads the following results:

$$
\begin{aligned}
&\hat{\omega}^0_{11} = S_{11}\,\{\alpha_1 + (T_3/T)\,(1-\alpha_1) - [\hat{\rho}^2\,(T_2+T_3)/T]\,(1-\alpha_2)\},\\
(38)\quad &\hat{\omega}^0_{12} = S_{12}\,\{\alpha_1 + (T_3/T)\,(1-\alpha_1) - [\,(T_2+T_3)/T]\,(1-\alpha_2)\},\\
&\hat{\omega}^0_{22} = S_{22}\,\{\alpha_2 + (T_1/T)\,(1-\alpha_2) - [\hat{\rho}^2\,(T_1+T_2)/T]\,(1-\alpha_1)\}.
\end{aligned}
$$

The two-step OW estimator for the case of *nested* missing observations is equivalent to the case of $T_3 = 0$ and $\alpha_2 = 1$ in (38), and therefore, it is identical to the Hocking-Smith estimator in (34) [3].

The two-step estimators of ω_{ij} presented in (36) and (38) are proportional to the "usual" estimator S_{ij}. The proportionality factors are the functions of α_1 and α_2 an estimate of correlation coefficient $\hat{\rho}$, and the extent of missing observations. If α_i is equal to one in both equations, then the two-step estimators $\hat{\omega}_{ij}$ and $\hat{\omega}^0_{ij}$ are identical to the "usual" estimator S_{ij}. Therefore, if α_1 and α_2 are close to one in a given sample, differences among these estimators of β are expected to be small. However, it is not obvious how a value of α_i that is sufficiently different from one will affect the relative performance of the alternative estimators of β. In the case of the *nested* missing observations, HWANG [1990] observed that the HS estimator of Ω yields significantly more efficient coefficient estimates than the "usual" estimator of Ω when α_i is significantly larger than one and ρ is high.

We conduct sampling experiments to examine the effects of α_i, ρ and the extent of missing observations on the relative efficiencies of the coefficient estimates among alternative estimators of Ω. The model is the two equation model that Schmidt and Hwang considered:

$$
\begin{aligned}
y_1 &= 10 + 2\,X_1 - 5\,X_2 + u\\
y_2 &= -10 + 6\,X_3 + 3\,X_4 + \varepsilon
\end{aligned}
$$

where the regressors are those of Experiment I in KMENTA and GILBERT [1968]. The error terms u_t and ε_t are standard normal with correlation coefficient $\rho = \{0.0,\,0.3,\,0.6,\,0.8,\,0.9\}$. In the experiments, the number of extra observations T_1 in equation 1 is fixed to be 10. The number of joint observations T_2 and the number of extra observations T_3 in equation 2 are varied: $T_2 = \{10, 20\}$ and $T_3 = \{10, 20,\ 30\}$.

Given each set of parameters, random samples are generated, and two-step estimates of the coefficients are computed by using alternative estimates of Ω [4]. In addition to the estimators described in the previous sections, we also compute the GLS estimates of the coefficients by using only the T_2 joint observations. This is to assess the efficiency loss of the common practice that throws away the data in the period where one or more variables have missing observations. Relative efficiencies of alternative estimators are measured by the ratios of the mean-squared erros (MSE) of their coefficient estimates to

3. This can also be verified directly from (37).
4. The Wilks estimator is not used because the estimated covariance matrix is not always positive definite.

that of the “usual” estimator. The mean-squared errors of the coefficient estimates are computed out of 1000 replications. Since MSE ratios of the individual coefficient estimates vary from one coefficient to another, the average of the MSE ratios of the three coefficients in each equation are computed. The average ratios are hoped to give a better measure of overall efficiency in each equation. The results for $\rho = 0$ are very close to the results for $\rho = 0.3$, and the results for $\rho = 0.8$ are in between the results for $\rho = 0.6$ and $\rho = 0.9$. Therefore, the results for $\rho = 0$ and 0.8 are not reported in the following tables.

Table 1 reports the relative efficiencies of alternative estimators for the case of *nested* missing observations $(T_3 = 0)$. As noted earlier, the two-step Hocking-Smith (HS), Hartley-Hocking (HH) and Orchard-Woodbury (OW) maximum likelihood algorithms are identical in the nested case. The column under JT in Table 1 represents the relative efficiency of the GLS estimator that uses only the T_2 joint observations in both equations. The ratios of the MSE’s from the entire sample of 1000 replications indicate that the SZ and HS estimators of Ω give slightly more efficient coefficient estimates than the “usual” estimator when the correlation coefficient ρ is high, but the differences are practically negligible. This is in line with the results reported in earlier studies of the nested case. On the other hand, the column under JT indicates substantial losses in efficiency when the extra observations are thrown away. These losses are increasing in ρ.

TABLE 1

Relative Efficiencies of Alternative Estimators
Nested Missing Observations: (Normal).

	Full sample			$\alpha_1 < 1$			$\alpha_1 > 1$			$\alpha_1 > \alpha_1^*$			
ρ	JT	SZ	HS	JT	SZ	HS	JT	SZ	HS	JT	SZ	HS	α_1^*
Equation 1													
$(T_2 = 10)$													
0.30	2.01	1.00	1.00	2.00	0.99	0.99	2.01	1.01	1.02	1.86	1.01	1.03	1.58
0.60	2.23	1.00	1.00	2.24	1.00	1.00	2.23	1.00	1.00	2.02	0.99	0.99	1.58
0.90	3.23	1.00	0.99	3.38	1.03	1.05	3.11	0.98	0.94	2.57	0.99	0.89	1.58
$(T_2 = 20)$													
0.30	1.55	1.00	1.00	1.49	1.00	1.00	1.62	1.00	1.00	1.62	0.99	0.99	1.27
0.60	1.80	0.99	0.99	1.73	1.00	1.00	1.88	0.99	0.98	1.89	0.97	0.96	1.27
0.90	3.30	0.98	0.98	3.32	1.00	1.02	3.30	0.96	0.94	3.20	0.90	0.87	1.27
Equation 2													
$(T_2 = 10)$													
0.30	1.00	1.00	1.00	1.04	1.01	1.00	0.97	0.99	1.00	0.96	0.98	1.00	1.58
0.60	1.25	0.99	1.00	1.35	0.99	1.00	1.16	1.00	1.00	1.10	0.98	1.00	1.58
0.90	2.24	0.99	1.00	2.42	0.94	1.01	2.05	1.05	0.99	1.81	1.09	0.96	1.58
$(T_2 = 0)$													
0.30	1.04	1.00	1.00	1.06	1.00	1.00	1.02	1.00	1.00	1.04	1.00	1.01	1.27
0.60	1.34	1.00	1.00	1.40	0.99	1.00	1.27	1.00	1.00	1.28	1.00	1.00	1.27
0.90	2.90	1.00	0.99	3.19	0.96	1.01	2.65	1.04	0.98	2.48	1.08	0.93	1.27

The effects of α_1 on the relative efficiency of alternative estimators are shown by the ratios of the MSE's in the subsamples of $\alpha_1 < 1$, $\alpha_1 > 1$, and $\alpha_1 > \alpha_1^*$, respectively, where α_1^* is the sample mean of α_1 plus its standard deviation. In the subsample of small α_1 there is little difference in the efficiency between the HS and the "usual" estimators of the second equation that has missing observations, but the HS estimator of the first equation that has no missing observations tends to be less efficient for a high ρ. In the subsample of large $\alpha_1\,(\alpha_1 > 1)$, the HS estimator of the second equation is only slightly more efficient than the "usual" estimator, but its estimate of the first equation is significantly more efficient when ρ is high. Higher efficiency of the HS estimator is even more profound for the subsample of very large $\alpha_1\,(\alpha_1 > \alpha_1^*)$. The SZ estimator has a quite different properties. When ρ is high, the SZ estimator of the second equation tends to be more efficient than the "usual" estimator in the subsample of small α_1, and tends to be less efficient in the subsample of large α_1. This relation ship is reversed for the estimates of the first equation.

TABLE 2.A

Relative Efficiencies of Alternative Estimators
Non-Nested Missing Observations: Full Sample (Normal).

	Equation 1				Equation 2			
ρ	JT	SZ	HH	OW	JT	SZ	HH	OW
$(T_2 = 10,\ T_3 = 10)$								
0.30	2.03	1.00	1.00	1.00	1.91	1.00	1.00	1.00
0.60	2.31	1.00	0.99	0.99	2.13	0.99	0.99	0.99
0.90	3.59	0.98	0.96	0.98	3.35	0.98	0.96	0.98
$(T_2 = 10,\ T_3 = 20)$								
0.30	2.03	1.00	1.00	1.00	2.92	1.00	1.00	1.00
0.60	2.34	1.00	0.99	0.99	3.07	0.99	0.99	0.99
0.90	3.77	0.98	0.95	0.97	4.35	0.98	0.96	0.97
$(T_2 = 10,\ T_3 = 30)$								
0.30	2.03	1.00	1.01	1.00	3.84	1.00	1.00	1.00
0.60	2.35	1.00	1.00	0.99	4.05	0.99	0.99	0.99
0.90	3.94	0.98	0.95	0.97	5.51	0.98	0.96	0.97
$(T_2 = 20,\ T_3 = 10)$								
0.30	1.54	1.00	1.00	1.00	1.57	1.00	1.00	1.00
0.60	1.83	0.99	0.99	0.99	1.85	1.00	1.00	1.00
0.90	3.52	0.98	0.97	0.98	3.47	0.99	0.98	0.98
$(T_2 = 10,\ T_3 = 20)$								
0.30	1.55	1.00	1.00	1.00	2.05	1.00	1.00	1.00
0.60	1.87	0.99	0.99	0.99	2.36	0.99	0.99	1.00
0.90	3.73	0.98	0.97	0.98	4.10	1.00	0.98	0.98
$(T_2 = 10,\ T_3 = 10)$								
0.30	1.55	1.00	1.00	1.00	2.57	1.00	1.00	1.00
0.60	1.90	0.99	0.99	0.99	2.92	0.99	0.99	0.99
0.90	3.88	0.99	0.97	0.98	4.85	0.99	0.98	0.99

Table 2.B

Relative Efficiencies of Alternative Estimators Non-Nested Missing Observations: (Normal) Equation 1.

	$\alpha_1 < 1$				$\alpha_1 > 1$				$\alpha_1 > \alpha_1^*$					
ρ	JT	SZ	HH	OW	JT	SZ	HH	OW	JT	SZ	HH	OW	α_1^*	α_2^*
$(T_2 = 10,\ T_3 = 10)$														
0.30	2.05	1.00	0.99	0.99	2.02	1.00	1.01	1.01	1.87	1.01	1.02	1.01	1.58	1.53
0.60	2.37	1.01	1.00	1.00	2.27	0.99	0.99	0.99	2.08	0.99	0.97	0.97	1.58	1.54
0.90	3.92	1.01	1.03	1.02	3.34	0.95	0.91	0.94	2.81	0.93	0.83	0.88	1.58	1.58
$(T_2 = 10,\ T_3 = 20)$														
0.30	2.05	1.00	0.99	1.00	2.02	1.01	1.01	1.00	1.87	1.00	1.02	1.00	1.58	1.69
0.60	2.39	1.01	1.00	1.00	2.30	0.99	0.98	0.99	2.09	0.98	0.96	0.97	1.58	1.72
0.90	4.10	1.02	1.03	1.01	3.52	0.94	0.90	0.94	2.92	0.92	0.81	0.89	1.58	1.74
$(T_2 = 10,\ T_3 = 30)$														
0.30	2.05	1.00	1.00	1.00	2.02	1.01	1.01	1.00	1.87	1.00	1.02	1.00	1.58	1.78
0.60	2.42	1.00	1.01	1.00	2.31	0.99	0.98	0.99	2.11	0.97	0.95	0.97	1.58	1.83
0.90	4.32	1.01	1.03	1.01	3.67	0.95	0.88	0.94	3.06	0.91	0.79	0.89	1.58	1.81
$(T_2 = 20,\ T_3 = 10)$														
0.30	1.48	1.00	0.99	1.00	1.60	1.00	1.00	1.00	1.63	0.99	0.99	0.99	1.27	1.25
0.60	1.76	1.00	1.00	1.00	1.90	0.99	0.98	0.98	1.91	0.95	0.94	0.96	1.27	1.25
0.90	3.52	1.02	1.02	1.01	3.52	0.95	0.92	0.94	3.35	0.86	0.83	0.87	1.27	1.25
$(T_2 = 20,\ T_3 = 20)$														
0.30	1.49	1.00	1.00	1.00	1.61	1.00	1.00	1.00	1.63	0.99	0.99	0.99	1.27	1.34
0.60	1.81	1.00	1.00	1.00	1.94	0.99	0.98	0.98	1.94	0.95	0.94	0.96	1.27	1.35
0.90	3.72	1.02	1.01	1.01	3.74	0.95	0.92	0.95	3.49	0.85	0.81	0.88	1.27	1.34
$(T_2 = 20,\ T_3 = 30)$														
0.30	1.50	1.00	1.00	1.00	1.61	1.00	1.00	1.00	1.64	0.98	0.98	0.99	1.27	1.39
0.60	1.83	1.00	1.00	1.00	1.97	0.98	0.97	0.99	1.98	0.95	0.93	0.96	1.27	1.40
0.90	3.86	1.02	1.02	1.01	3.91	0.95	0.92	0.95	3.65	0.86	0.82	0.90	1.27	1.38

Table 2 presents the results for the case of *non-nested* missing observations. The HS and the HH two-step algorithms yield almost identical results, and only the HH results are reported. The full sample result in Table 2.A show that higher relative efficiencies of the SZ, HH and OW estimators compared to the "usual" estimator are more noticeable than in the nested case. In particular, the HH estimator is about $4 \sim 5\,\%$ more efficient than the "usual" estimator for $\rho = 0.9$ when $T_2 = 10$. The superiority of the HH estimator is less significant when the degree of missing observations is less severe (i.e., the case of $T_2 = 20$), but it is still more noticeable than the case of nested missing observations. This confirms our initial conjecture that there will be more significant differences among alternative estimators in the non-nested case than in the nested case. We also observe that the efficiency losses from throwing away extra observations are more severe in the non-nested case than in the nested case.

TABLE 2.C

Relative Efficiencies of Alternative Estimators Non-Nested Missing Observations: (Normal) Equation 2.

	$\alpha_1 < 1$				$\alpha_1 > 1$				$\alpha_1 > \alpha_1^*$					
ρ	JT	SZ	HH	OW	JT	SZ	HH	OW	JT	SZ	HH	OW	α_1^*	α_2^*
$(T_2 = 10, T_3 = 10)$														
0.30	2.15	1.00	0.99	0.99	1.73	0.99	1.00	1.00	1.67	0.98	1.01	1.00	1.58	1.53
0.60	2.54	0.99	0.99	0.99	1.81	0.99	0.99	1.00	1.73	0.99	0.99	1.00	1.58	1.54
0.90	4.16	0.96	0.99	1.00	2.68	0.99	0.94	0.96	2.21	1.00	0.90	0.93	1.58	1.58
$(T_2 = 10, T_3 = 20)$														
0.30	3.18	1.00	0.99	0.99	2.70	0.99	1.01	1.01	2.32	1.00	1.01	1.00	1.58	1.69
0.60	3.60	0.99	0.98	0.99	2.63	0.99	1.00	1.0	2.22	1.00	1.00	1.00	1.58	1.72
0.90	5.48	0.97	0.97	0.99	3.43	0.99	0.94	0.96	2.53	0.99	0.89	0.93	1.58	1.74
$(T_2 = 10, T_3 = 30)$														
0.30	4.27	1.00	1.00	0.99	3.50	0.99	1.01	1.00	2.92	1.00	1.02	1.00	1.58	1.78
0.60	4.84	1.00	0.99	0.99	3.42	0.99	0.99	0.99	2.82	0.99	0.99	0.98	1.58	1.83
0.90	7.01	0.99	0.98	0.99	4.31	0.98	0.94	0.96	3.12	0.95	0.87	0.92	1.58	1.81
$(T_2 = 20, T_3 = 10)$														
0.30	1.56	1.00	1.00	1.00	1.59	1.00	1.00	1.00	1.60	1.00	1.00	1.00	1.27	1.25
0.60	1.90	0.99	1.00	1.00	1.79	1.00	0.99	1.00	1.74	1.01	0.98	0.98	1.27	1.25
0.90	3.77	0.98	0.99	1.00	3.16	1.01	0.97	0.97	2.79	1.05	0.91	0.92	1.27	1.25
$(T_2 = 20, T_3 = 20)$														
0.30	1.98	1.00	1.00	1.00	2.14	0.99	1.00	1.00	1.91	0.99	1.00	1.00	1.27	1.34
0.60	2.38	1.00	1.00	1.00	2.33	0.99	0.99	0.99	2.08	0.99	0.97	0.98	1.27	1.35
0.90	4.46	1.00	1.00	1.00	3.72	0.99	0.96	0.96	3.19	1.01	0.90	0.92	1.27	1.34
$(T_2 = 20, T_3 = 30)$														
0.30	2.54	1.00	1.00	1.00	2.62	1.00	1.00	1.00	2.48	1.00	1.00	1.00	1.27	1.39
0.60	2.97	0.99	0.99	àçç	2.88	0.99	0.99	1.00	2.75	0.99	0.99	1.00	1.27	1.40
0.90	5.15	1.00	0.99	1.00	4.53	0.99	0.97	0.98	4.05	1.00	0.94	0.96	1.27	1.38

The effects of α_1 on the relative efficiency of alternative estimators of equation 1 are reported in Table 2.B. In the subsample of small α_1, there are practically no differences in the overall perforamnces, even though the "usual" estimator performs slightly better or worse than the other estimators, depending on the degree of correlation. However, in the subsample of large α_1, the SZ, HH and OW estimators significantly outperform the "usual" estimator when ρ is high. In particular, the HH estimator is about 10 % more efficient than the usual estimator in the subsample $\alpha_1 > 1$ when $\rho = 0.90$. The gain in relative efficiency of the HH estimator is more impressive in the subsample of very large $\alpha_1 > \alpha_1^*$. The efficiency gain of the HH estimator reaches 20 % when ρ is high. The gain in the relative efficiency is certainly greater in the non-nested case than in the nested case. Among the three alternative estimators, the HH estimator tends to perform better than the SZ and the OW estimators. In the subsample of very large values of both α_1 and α_2, the relative efficiencies of the HH and OW estimators are similar

TABLE 3.A

Relative Efficiencies of Alternative Estimators
Non-nested Missing Observationss: Full Sample (Uniform).

	Equation 1				Equation 2			
ρ	JT	SZ	HH	OW	JT	SZ	HH	OW
$(T_2 = 10, T_3 = 10)$								
0.30	1.89	1.00	1.00	1.00	2.04	1.00	1.00	1.00
0.60	2.10	0.99	0.99	0.99	2.20	1.00	1.00	1.00
0.90	3.31	0.97	0.97	0.98	3.30	0.99	0.98	0.99
$(T_2 = 10, T_3 = 20)$								
0.30	1.88	0.99	1.00	1.00	3.07	1.00	1.00	1.00
0.60	2.12	0.99	0.99	0.99	3.19	1.00	1.01	1.00
0.90	3.51	0.97	0.96	0.98	4.43	0.99	0.98	0.99
$(T_2 = 10, T_3 = 30)$								
0.30	1.88	0.99	1.00	1.00	3.82	1.00	1.01	1.00
0.60	2.12	0.98	0.99	0.99	3.86	1.00	1.01	1.00
0.90	3.59	0.97	0.96	0.98	5.06	0.99	0.98	0.99
$(T_2 = 20, T_3 = 10)$								
0.30	1.43	1.00	1.00	1.00	1.55	1.00	1.00	1.00
0.60	1.70	1.00	1.00	1.00	1.85	1.00	1.00	1.00
0.90	3.33	0.99	0.99	0.99	3.53	1.00	0.99	0.99
$(T_2 = 20, T_3 = 20)$								
0.30	1.44	1.00	1.00	1.00	1.99	1.00	1.00	1.00
0.60	1.71	0.99	0.99	1.00	2.18	1.00	1.00	1.00
0.90	3.44	1.00	0.99	0.99	3.80	1.00	0.99	0.99
$(T_2 = 20, T_3 = 30)$								
0.30	1.44	1.00	1.00	1.00	2.43	1.00	1.00	1.00
0.60	1.74	0.99	0.99	1.00	2.69	1.00	1.00	1.00
0.90	3.58	1.00	0.99	0.99	4.37	0.99	0.99	0.99

to those in the subsample $\alpha_1 > \alpha_1^*$, but the relative efficiency of the SZ estimator is lower in this subsample.

The sample value of α_1 has similar effects on the relative efficiency of alternative estimators of equation 2 as reported in Table 2.C. The HH estimator has a significant gain in its relatives efficiency for a large value of α_1 when ρ is high, though the gain is slightly less than in equation 1. The effects of α_2 also follow a similar pattern, and hence are not reported.

The results in Table 2 are based on the normal error terms. To examine the effects of non-normal distributions we conducted the same experiment with uniformly distributed error terms, and the results are reported in Table 3. The change of the error distribution from a normal to a uniform distribution has significant impacts on the relative efficiency of the JT estimator. Its overall relative performance improves in the full sample and in the subsample of small α_1, but it becomes worse for large α_1. Regardless of these changes, losses in efficiency from throwing away extra observations remain high.

TABLE 3.B

Relative Efficiencies of Alternative Estimators Non-Nested Missing Observations: (Uniform) Equation 1.

	$\alpha_1 < 1$				$\alpha_1 > 1$				$\alpha_1 > \alpha_1^*$					
ρ	JT	SZ	HH	0W	JT	SZ	HH	0W	JT	SZ	HH	OW	α_1^*	α_2^*
$(T_2 = 10,\ T_3 = 10)$														
0.30	1.54	0.99	0.99	0.99	2.24	1.00	1.01	1.01	2.31	1.01	1.03	1.02	1.37	1.46
0.60	1.68	0.99	0.99	0.99	2.54	0.99	0.99	0.99	2.56	0.97	0.98	0.99	1.37	1.52
0.90	2.78	1.01	1.03	1.02	3.84	0.93	0.91	0.94	3.81	0.85	0.83	0.90	1.37	1.49
$(T_2 = 10,\ T_3 = 20)$														
0.30	1.53	0.99	0.99	0.99	2.24	1.00	1.01	1.01	2.30	1.00	1.03	1.02	1.37	1.54
0.60	1.69	0.99	0.99	0.99	2.56	0.98	0.99	1.00	2.60	0.97	0.98	0.99	1.37	1.61
0.90	2.93	1.01	1.03	1.01	4.08	0.93	0.90	0.95	4.05	0.85	0.83	0.92	1.37	1.58
$(T_2 = 10,\ T_3 = 30)$														
0.30	1.53	0.99	0.99	0.99	2.24	1.00	1.01	1.01	2.31	1.00	1.04	1.02	1.37	1.64
0.60	1.70	0.99	0.99	0.99	2.57	0.98	0.99	1.00	2.60	0.97	0.98	0.99	1.37	1.73
0.90	3.02	1.01	1.02	1.01	4.14	0.93	0.90	0.96	4.10	0.86	0.82	0.92	1.37	1.66
$(T_2 = 20,\ T_3 = 10)$														
0.30	1.25	1.00	1.00	1.00	1.61	1.00	1.00	1.00	1.89	1.00	1.00	1.00	1.16	1.18
0.60	1.46	1.00	1.00	1.00	1.95	1.00	0.99	0.99	2.31	0.98	0.97	0.98	1.16	1.22
0.90	2.86	1.01	1.03	1.02	3.81	0.98	0.95	0.96	4.52	0.93	0.90	0.92	1.16	1.20
$(T_2 = 20,\ T_3 = 20)$														
0.30	1.26	0.99	1.00	1.00	1.61	1.00	1.00	1.00	1.90	1.00	1.00	1.00	1.16	1.21
0.60	1.47	0.99	1.00	1.00	1.96	1.00	0.99	0.99	2.36	0.98	0.97	0.98	1.16	1.27
0.90	2.95	1.01	1.02	1.02	3.96	0.98	0.95	0.97	4.74	0.93	0.89	0.93	1.16	1.24
$(T_2 = 20,\ T_3 = 30)$														
0.30	1.26	0.99	1.00	1.00	1.62	1.00	1.00	1.00	1.90	0.90	1.00	1.00	1.16	1.24
0.60	1.50	0.99	1.00	1.00	1.99	0.99	0.99	0.99	2.38	0.98	0.96	0.98	1.16	1.31
0.90	3.09	1.01	1.02	1.01	4.09	0.99	0.95	0.97	4.92	0.94	0.89	0.94	1.16	1.28

The effects of the uniform distribution on the relative efficiency of the SZ, HH and OW estimators are very minimal for the full sample and for the subsample of small α_1. Although the effects are more noticeable for the subsamples of large α_1, the ranking of the estimators and the gain in efficiency are qualitatively similar to the case of normal error terms.

TABLE 3.C

Relative Efficiencies of Alternative Estimators Nested Missing Observations: (Uniform) Equation 2.

	$\alpha_1 < 1$				$\alpha_1 > 1$				$\alpha_1 > \alpha_1^*$					
ρ	JT	SZ	HH	OW	JT	SZ	HH	OW	JT	SZ	HH	OW	α_1^*	α_2^*
$(T_2 = 10,\ T_3 = 10)$														
0.30	1.98	1.00	1.00	1.00	2.09	0.99	1.00	1.00	1.92	1.00	1.01	1.01	1.37	1.46
0.60	2.23	1.01	1.01	1.00	2.18	1.00	1.00	1.00	1.90	1.02	1.03	1.02	1.37	1.52
0.90	3.69	0.99	1.01	1.01	2.98	0.99	0.95	0.96	2.63	1.04	0.96	0.96	1.37	1.49
$(T_2 = 10,\ T_3 = 20)$														
0.30	2.84	1.00	1.00	1.00	3.31	1.00	1.00	1.00	3.05	1.00	1.02	1.01	1.37	1.54
0.60	3.07	1.00	1.00	1.01	3.33	1.00	1.01	1.01	2.99	1.00	1.05	1.03	1.37	1.61
0.90	4.66	0.99	1.00	1.00	4.21	0.99	0.97	0.98	3.96	1.02	0.99	0.99	1.37	1.58
$(T_2 = 10,\ T_3 = 30)$														
0.30	3.55	1.00	1.00	1.00	4.11	1.00	1.01	1.00	3.57	1.00	1.02	1.01	1.37	1.64
0.60	3.71	1.00	1.01	1.00	4.02	1.00	1.01	1.01	3.41	1.00	1.03	1.02	1.37	1.73
0.90	5.25	0.99	0.99	1.00	4.86	0.98	0.97	0.98	4.24	0.98	0.97	0.98	1.37	1.66
$(T_2 = 20,\ T_3 = 10)$														
0.30	1.50	1.00	1.00	1.00	1.60	1.00	1.00	1.00	1.59	0.99	1.00	1.00	1.16	1.18
0.60	1.81	1.00	1.00	1.00	1.89	1.00	1.00	1.00	1.76	0.99	1.00	1.00	1.16	1.22
0.90	3.55	0.99	1.01	1.01	3.50	1.00	0.97	0.98	3.08	1.00	0.95	0.95	1.16	1.20
$(T_2 = 20,\ T_3 = 20)$														
0.30	1.87	1.00	1.00	1.00	1.99	1.00	1.00	1.00	1.96	0.99	1.00	1.00	1.16	1.21
0.60	2.15	1.00	1.00	1.00	2.21	1.00	1.00	1.00	2.07	1.00	1.00	1.00	1.16	1.27
0.90	3.84	1.00	1.00	1.00	3.75	1.00	0.97	0.97	3.34	1.00	0.94	0.95	1.16	1.24
$(T_2 = 20,\ T_3 = 30)$														
0.30	2.46	1.00	1.00	1.00	2.41	1.00	1.00	1.00	2.27	1.00	1.00	1.00	1.16	1.24
0.60	2.73	1.00	1.00	1.00	2.65	1.00	1.00	1.00	2.46	1.00	1.00	1.00	1.16	1.31
0.90	4.44	0.99	1.00	1.00	4.29	0.99	0.98	0.98	3.93	0.99	0.95	0.96	1.16	1.28

6 Conclusion

This paper considers alternative two-step estimators of a general SUR model with linear restrictions on the coefficients when the missing observations are not nested. Their small sample properties are compared through sampling experiments. As we conjectured, the differences among the alternative estimators are more profound in the case of non-nested missing observations than in the nested case. The HH estimator in particular can give a significant efficiency gain over the "usual" estimator when the sample statistics α_1 and/or α_2 are sufficiently large and the error correlation is high. The statistics α_1 and α_2 summarize the essential differences among alternative estimators of the covariance matrix Ω. The sampling experiments also indicate that discarding the set of incomplete observations results in a substantial loss in efficiency.

• References

AFIFI, A. A., ELASHOFF, R. M. (1966). – "Missing Observations in Multivariate Statistics", *Journal of American Statistical Association*, 61, pp. 595-604.

BALTAGI, BADI, H., GARVIN, S., KERMAN S. (1989). – "Further Monte Carlo Evidence on Seemingly Unrelated Regressions with Unequal Number of Observations", *Annales d'Économie et de Statistique*, 14, pp. 103-15.

BEALE, E. M. L., LITTLE, R. J. L. (1975). – "Missing Values in Multivariate Analysis", *Journal of the Royal Statistical Society*, (B), 37, pp. 129-145.

CONNIFFE D. (1985). – "Estimating Regression Equations with Common Explanatory Variables but Unequal Numbers of Observations", *Journal of Econometrics*, 27, pp. 179-196.

DAGENAIS M. G. (1973). – "The Use of Incomplete Observations in Multiple Regression Analysis: A Generalized Least Squares Approach", *Journal of Econometrics*, 1, pp. 317-328.

DEMPSTER, A. P., LAIRD, N. M., RUBIN, D. B. (1977). – "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society*, (B), 39, pp. 1-38.

FAREBROTHER, R. W. (1975). – "The Estimation Problem in Econometrics", Ph. D. Dissertation, the Victoria University of Manchester.

HARTLEY, H. O., HOCKING, R. R. (1971). – "The Analysis of Incomplete Data", *Biometrics*, 27, pp. 783-823.

HOCKING, R. R., SMITH, W. B. (1968). – "Estimation of Parameters in the Multivariate Normal Distribution with Missing Observations", *Journal of American Statistical Association*, 63, pp. 159-173.

HWANG H. (1990). – "Estimation of Linear SUR Model with Unequal Numbers of Observations", *Review of Economics and Statistics*, pp. 510-515.

KELEGIAN H. H. (1969). – "Missing Observations in Multivariate Regression – Efficiency of a First Order Method", *Journal of American Statistical Association*, 64, pp. 1609-1616.

MAGNUS, J. R., NEUDECKER H. (1980). – "The Elimination Matrix: Some Lemmas and Applications", *SIAM Journal*, 1, pp. 422-449.

MEISNER J. F. (1979). – "The Sad Fate of Asymptotic Slutsky Asymmetry Test for Large Systems", *Economics Letters*, 2, pp. 231-233.

MIYAZAKI, S., GRIFFITHS, W. E., "The Properties of Some Covariance Matrix Estimators in Linear Models with AR(1) Errors", *Economics Letters*, 14, pp. 351-356.

ORCHARD, T. WOODBURY, H. A. (1972). – "A Missing Information Principle: Theory and Applications", in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, pp. 697-715.

SCHMIDT P. (1977). – "Estimation of Seemingly Unrelated Regressions with Unequal Numbers of Observations", *Journal of Econometrics*, 5, pp. 365-377.

Silvey, S. D. (1975). – *Statistical Inference*, London: Chapman and Hall.

Srivastava, J. N., Zaatar, M. K. (1973). – "A Monte Carlo Comparison of Four Estimators of the Dispersion Matrix of a Bivariate Normal Population, Using Incomplete Data", *Journal of American Statistical Association*, 68, pp. 180-183.

Trawinski, I. M., Bargmann, R. E. (1964). – "Maximum Likelihood Estimation with Incomplete Multivariate Data", *Annals of Mathematical Statistics*, 35, pp. 647-657.

Wilks S. S. (1992). – "Moments and Distributions of Estimates of Population Parameters from Fragmentary Samples", *Annals of Mathematical Statistics,* 3, pp. 167-195.

Zellner A. (1962). – "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", *Journal of American Statistical Association*, 57, pp. 348-368.