ANNALES D'ÉCONOMIE ET DE STATISTIQUE. - N° 12 — 1988

Consistent Estimation of Regression Models with Incompletely Observed Exogenous Variables

Theodore E. NIJMAN, Franz C. PALM *

ABSTRACT. – We consider consistent estimation of regression models in which the exogenous variables are incompletely observed assuming that the response mechanism is random. In the literature on imputed data, several estimators have been proposed which are based on approximations substituted for the missing data. We discuss conditions under which these proxy variables estimators are asymptotically more efficient than the estimator based on complete observations and we show how an optimal proxy variables estimator can be obtained. For simple models, some proxy variables estimators are almost as efficient as the Gaussian maximum likelihood (ML) estimator and sometimes more efficient than the pseudo ML estimator.

Estimation convergente de modèles de régression où les variables exogènes ne sont pas complètement observées

RÉSUMÉ. – On étudie l'estimation convergente de modèles de régression où les variables exogènes ne sont pas complètement observées, les non réponses à l'origine des données manquantes étant aléatoires. Il existe plusieurs techniques d'estimation où l'on remplace les données manquantes par des approximations bien choisies. On dérive des conditions sous lesquelles ces techniques sont asymptotiquement plus efficaces que l'estimation fondée sur les observations complètes seules, et on montre comment construire au mieux les approximations. Dans des modèles simples, les estimateurs par approximation sont presque aussi efficaces que le maximum de vraisemblance gaussien et parfois plus efficaces que le pseudo maximum de vraisemblance.

---

* TH. E. NIJMAN : Tilburg University, Department of Econometrics, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; F. C. PALM : University of Limburg, Department of Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands. The authors wish to thank two unknown referees for helpful comments.

1 Introduction

In applied research, it is common practice to impute the missing values of variables which are incompletely observed (see e. g. LITTLE and RUBIN [1987]). Imputation is applied to cross section data which suffer from partial non-response (for a survey of the literature in the context of non-response in sample surveys, see e. g. LITTLE [1982]) and to time series which are available on a high temporal aggregation level only. There is a large literature on maximum likelihood (ML) estimation for incomplete data. For the EM-algorithm to obtain ML estimates and for a review of the literature, we refer to DEMPSTER, LAIRD and RUBIN [1977].

In a common imputation procedure, the observations on the incompletely observed variable are regressed on auxiliary variables. The missing values are then approximated by the predictions from this auxiliary regression equation. In this paper, it is shown that a regression using imputed observations does not necessarily yield more efficient parameter estimates than a regression based on data points for which all variables are observed (in the sequel called complete observations). We discuss conditions under which an estimator based on approximations for unobserved variables is asymptotically more efficient than an estimator based on complete observations only and we show how an optimal proxy variables estimator can be obtained. We also consider the estimation of standard errors of proxy variables estimators. For a simple case, we derive the relative efficiency of several proxy variables estimators compared with the maximum likelihood estimator under the normality assumption. Finally extensions of the results to cases where only aggregates of the exogenous variables are observed and to dynamic models are considered.

We consider the following regression model

$$(1) \qquad y_i = \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i, \qquad i = 1, \ldots, N,$$

$$(2) \qquad x_{ik} = \sum_{j=1}^{J} \alpha_{jk} z_{ij} + v_{ik}, \qquad i = 1, \ldots, N, \quad k = 1, \ldots, K,$$

where the regression disturbances ε_i and v_{ik} are i. i. d. with mean zero and variances σ^2 and σ_{kk} respectively, have finite fourth moments, are independent of the corresponding regressors and satisfy $E\,\varepsilon_i v_{ik} = 0$, for $i = 1, \ldots N$, and $k = 1, \ldots K$, $E\, v_{ik} v_{ij} = \sigma_{kj}$, for $i = 1, \ldots, N$, and $j, k = 1, \ldots, K$, $j \neq k$. Equation (1) corresponds to the conditional distribution of y given the x's and z's.

Assume moreover that plim $Z'Z/N$ is finite and non-singular where the matrix Z has typical element z_{ij}. We consider the case where y_i and z_{ij}, ($j = 1, \ldots, J$), are observed for $i = 1, \ldots, N$ whereas x_{ik} is observed if and only if the random variable δ_{ik} takes the value 1. The random variables δ_{ik} are assumed to be independent of $\varepsilon_{j'}$, $z_{i'j'}$ and $v_{i'k'}$ for j', $i' = 1, \ldots N$,

$k'=1, \ldots K$ i.e. we assume that the sample design is completely at random (in RUBIN's [1976] terminology who distinguishes it from ignorable when missingness depends on observed variables). Note that we do not exclude that some of the regressors in (1) are completely observed and are used as regressors in (2).

If one assumes that a fraction of the observations is complete in large samples, a first consistent estimator of $\beta'=[\beta_1, \ldots, \beta_K]$ can be obtained from the regression (1) using complete observations only. Evidently if only a few of the right hand side variables in (1) are complete and these variables can be closely approximated using equation (2) an estimator based on complete observations only will not be very efficient. But the missing values can be approximated by

$$\hat{x}_{ik}=\sum_{j=1}^{J} z_{ij}\,\hat{\alpha}_{jk},$$

where $\hat{\alpha}_{jk}$ is an estimate of α_{jk}. If one defines

$$\hat{x}_{ik}=x_{ik} \qquad \text{if} \quad \delta_{ik}=1,$$

an estimate of β can subsequently be obtained by regressing y_i on $\hat{x}_{ik}$, $k=1, \ldots, K$. This procedure is known as the first order method of AFIFI and ELASHOFF [1966]. NIJMAN and PALM [1986] refer to it as a proxy variables estimator. Special cases have been considered by e.g. GOURIÉROUX and MONFORT [1981] who derived the large sample distribution of several proxy variables estimators and by CONNIFFE [1983 a] who considered small sample properties.

The plan of this paper is as follows. In section 2 we analyze the model in (1) and (2) assuming that $K=J=1$, and for convenience that x_{i1} is observed if $i \leqslant N/2$ only. This special case illustrates very well the main issues related to proxy variable estimators. Numerical results on the relative efficiency of these estimators compared with the Gaussian ML estimator are presented for this model. In section 3 we consider the general case and show how the use of proxies can lead to an efficiency gain over the estimator based on complete observations only. In section 4 the analysis of proxy variables estimators is extended to observations of temporal aggregates of the exogenous variables and to dynamic models. Again numerical results on the relative efficiency of a number of estimators are presented for a simple model. Finally some concluding remarks are given in section 5. Three appendices contain the technical details.

2 An Example

In this section we analyze the model in (1) and (2) assuming that $K=J=1$ and that the exogenous variable in (1) is observed if $i \leqslant N/2$ only. Deleting redundant subscripts, the model can be written as

(3) $$y_i = \beta x_i + \varepsilon_i$$

(4) $$x_i = \alpha z_i + v_i.$$

The variance of v_i will be denoted by σ_V^2. It is useful to notice that the model (3)-(4) is a restricted version of a model analyzed by GOURIÉROUX and MONFORT [1981] who assume that z_i is also included in the regression equation for y,

(3′) $$y_i = \beta_1 x_i + \beta_2 z_i + \varepsilon_i.$$

If normality of ε_i and v_i is assumed, the asymptotically efficient ML estimator of the parameter in (3) and (4) can be obtained by maximizing the likelihood function

$$L(\alpha, \beta, \sigma^2, \sigma_V^2) = \prod_{i=1}^{N} L_i(\alpha, \beta, \sigma^2, \sigma_V^2)$$

with

$$L_i(\alpha, \beta, \sigma^2, \sigma_V^2) = C.\,(\sigma_V \sigma)^{-\delta_i} \tilde{\sigma}^{\delta_i - 1} \exp\{-\delta_i (y_i - \beta x_i)^2/2\sigma^2 - \delta_i (x_i - \alpha z_i)^2/2\sigma_V^2 - (1-\delta_i)(y_i - \alpha\beta z_i)^2/2\tilde{\sigma}^2\},$$

where $\tilde{\sigma}^2 = \sigma^2 + \beta^2 \sigma_V^2$, $\delta_i = 1$ if x_i is observed and $\delta_i = 0$ otherwise and C is a constant independent of the unknown parameters. Note that a computationally convenient reparametrization proposed by GOURIÉROUX and MONFORT [1981] for the model (3′)-(4), no longer applies when β_2 is known to be zero. Following an approach similar to that of ANDERSON [1957], GOURIÉROUX and MONFORT reparametrize the joint distribution for y_i and x_i given z_i as a product of the marginal distribution of y_i given z_i and the conditional distribution of x_i given y_i and z_i and they show that his reparametrization provides an immediate solution for the ML estimator. When $\beta_2 = 0$, the computational advantage of this approach is lost.

If the normality assumption is satisfied, the ML estimator will be asymptotically efficient but in general ML estimation will be computationally cumbersome for other than simple models. If the normality does not hold, the Gaussian ML estimator is still generally consistent but no longer efficient, a point to which we will return below. Notice also that generalized nonlinear least squares applied to the model $y_i = \beta x_i + \varepsilon_i$, $i = 1, \ldots N/2$, $y_i = \alpha\beta z_i + \tilde{v}_i$, $i = N/2+1, \ldots N$, $x_i = \alpha z_i + v_i$, $i = 1, \ldots N/2$, is not fully efficient as the slope coefficients of the model are functionally related to some disturbance covariance parameters.

Alternatively, the parameter β can be consistently estimated by OLS using the complete observations only

(5) $$\hat{\beta}_c = \Sigma_c x_i y_i / \Sigma_c x_i^2,$$

where Σ_c denotes summation over complete observations ($i \leqslant N/2$). In the sequel we will also use the notation Σ_I and Σ_A to denote summation over incomplete and all, complete and incomplete, observations respectively. Intuitively there seems to be a case for using imputed data and considering the following proxy variables estimator

(6) $$\hat{\beta}_p = \Sigma_A \hat{x}_i y_i / \Sigma_A \hat{x}_i^2,$$

where $\hat{x}_i = x_i$ if $i \leqslant N/2$ and $\hat{x}_i$ is some approximation for x_i if $i > N/2$. As mentioned in the introduction, a natural choice for the approximation is

$$\hat{x}_i = \hat{\alpha} z_i \quad \text{if} \quad i > N/2, \quad \text{where } \hat{\alpha} = \Sigma_c z_i x_i / \Sigma_c z_i^2.$$

The condition for consistency of the resulting estimator $\hat{\beta}_p$ is that

(7) $$\operatorname{plim} \Sigma_A \hat{x}_i w_i / \Sigma_A \hat{x}_i^2 = 0 \qquad \text{with} \quad w_i = y_i - \hat{x}_i \beta = \varepsilon_i + \beta (x_i - \hat{x}_i).$$

In applied work it is not only important to have a consistent estimator, but also to be able to estimate its large sample variance consistently. Substituting the expression for w_i in (7) into (6), we have for the first order method

$$\hat{\beta}_p - \beta = (\Sigma_c x_i^2 + \Sigma_I \hat{\alpha}^2 z_i^2)^{-1} \{ \Sigma_c x_i \varepsilon_i + \hat{\alpha} \Sigma_I z_i \varepsilon_i + \beta \hat{\alpha} \Sigma_I z_i v_i + \beta \hat{\alpha} \Sigma_I z_i^2 (\alpha - \hat{\alpha}) \}.$$

The large sample variance (avar) of $\sqrt{N}\, \hat{\beta}_p$ can be derived via substitution of the expression for $\hat{\alpha}$ into (6) and the use of the appropriate limiting theory,

(8) $$\operatorname{avar}(\sqrt{N}\, \hat{\beta}_p) = (\sigma_{\tilde{x}}^2 \sigma^2 + \alpha^2 \beta^2 \sigma_z^2 \sigma_V^2) \sigma_{\tilde{x}}^{-4}$$

with $\sigma_{\tilde{x}}^2 = \alpha^2 \sigma_z^2 + 1/2\, \sigma_V^2$. Three remarks have to be made.

First, although the distinction between the case where α is known and that when α is estimated could be neglected in proving consistency, it is essential for the computation of the large sample variance of $\hat{\beta}_p$. When α is known, the asymptotic variance of $\tilde{\beta}_p$ ($\sim$ denotes that the true value of α is used) is given by

(9) $$\operatorname{Avar}(\sqrt{N}\, \tilde{\beta}_p) = (\sigma_{\tilde{x}}^2 \sigma^2 + 1/2\, \alpha^2 \beta^2 \sigma_z^2 \sigma_V^2) \sigma_{\tilde{x}}^{-4}.$$

This point is often missed in the econometric literature, but has recently been stressed in the context of using approximations for unobserved expectations by PAGAN [1984] and by MURPHY and TOPEL [1985]. Second, as is obvious from a comparison of their asymptotic variances in appendix A, $\hat{\beta}_p$ can be more efficient as well as less efficient than $\hat{\beta}_c$, a finding which also holds for the unrestricted model considered by GOURIÉROUX and MONFORT [1981] as noted by GRILICHES [1986].

Third, the formula for the standard errors in a least squares regression does not yield a consistent estimator of the asymptotic standard errors for $\hat{\beta}_p$ and $\hat{\beta}_c$ as

(10) $$\operatorname{plim} \Sigma_A (y_i - \hat{\beta}_p \hat{x}_i)^2 (\Sigma_A \hat{x}_i^2)^{-1} = \left(\sigma^2 + \frac{1}{2} \beta^2 \sigma_V^2 \right) \sigma_{\tilde{x}}^{-2}.$$

It is obvious from (8) and (10) that the order of magnitude and the sign of the bias of the standard errors depend on the value of $\alpha^2 \sigma_z^2$. Some information on the order of magnitude of the bias will be provided in

TABLE 1

Relative Efficiency of Alternative Consistent Estimators for β Compared with the ML Estimator

		Alternative estimators				
1	2	3	4	5	6	7
R_x^2	R_y^2	$\hat{\beta}_p$	$\hat{\beta}_d$	$\hat{\beta}_g$	$\hat{\beta}_c$	$\frac{\text{plim } \hat{\sigma}_w^2/\Sigma \hat{x}_i^2}{\text{avar } (\hat{\beta}_p)}$
(1)	(2)	(6)	(11)	(11)	(5)	
			V: Dagenais	V: optimal		
0.20	0.20	1.1306	1.1161	1.1129	1.2719	1.0312
0.20	0.40	1.3068	1.2251	1.2140	1.3315	1.0755
0.20	0.60	1.5430	1.2675	1.2491	1.3226	1.1429
0.20	0.80	2.0741	1.1894	1.1726	1.2043	1.2581
0.20	0.95	5.3364	1.0546	1.0489	1.0556	1.4176
0.40	0.20	1.0737	1.0617	1.0587	1.3845	0.9901
0.40	0.40	1.2076	1.1382	1.1259	1.3761	0.9767
0.40	0.60	1.4517	1.2000	1.1744	1.3421	0.9575
0.40	0.80	2.0770	1.1786	1.1471	1.2262	0.9277
0.40	0.95	5.7154	1.0612	1.0473	1.0649	0.8916
0.60	0.20	1.0320	1.0258	1.0240	1.5360	0.9767
0.60	0.40	1.1033	1.0658	1.0573	1.4710	0.9444
0.60	0.60	1.2644	1.1184	1.0962	1.3952	0.8966
0.60	0.80	1.7420	1.1481	1.1085	1.2669	0.8182
0.60	0.95	4.5344	1.0706	1.0442	1.0828	0.7164
0.80	0.20	1.0078	1.0060	1.0055	1.7368	0.9814
0.80	0.40	1.0283	1.0172	1.0142	1.6547	0.9536
0.80	0.60	1.0859	1.0393	1.0288	1.5432	0.9079
0.80	0.80	1.3055	1.0807	1.0501	1.3733	0.8182
0.80	0.95	2.7544	1.0805	1.0361	1.1325	0.6624
0.95	0.20	1.0005	1.0004	1.0003	1.9274	0.9941
0.95	0.40	1.0019	1.0011	1.0009	1.8922	0.9847
0.95	0.60	1.0068	1.0030	1.0019	1.8296	0.9668
0.95	0.80	1.0335	1.0107	1.0048	1.6866	0.9206
0.95	0.95	1.3274	1.0481	1.0125	1.3442	0.7660

Table 1. When using a proxy variable $\hat{x}_i$ for the missing values of x_i, the disturbance $w_i = \varepsilon_i + \beta(x_i - \hat{x}_i)$ is no longer homoscedastic. It is natural therefore to consider generalized least squares (GLS) estimation, as GOURIÉROUX and MONFORT [1981] did for an exactly identified model, which can be denoted as

$$(11) \qquad \hat{\beta}_V = (\hat{X}' V^{-1} \hat{X})^{-1} \hat{X}' V^{-1} y,$$

where $y = (y_1, y_2, \ldots, y_N)'$, $X = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N)'$ and V is a weighting matrix.

DAGENAIS [1973] proposed to take V diagonal with $v_{ii} = \hat{\sigma}^2$ for $i \leqslant N/2$ and $v_{ii} = \hat{\sigma}^2 + \hat{\beta}^2 \hat{\sigma}_V^2$, $i > N/2$, and "^" indicating a consistent estimate of the corresponding parameter. This estimator will be referred to as $\hat{\beta}_d$. Although every element of the matrix V proposed by Dagenais converges in probability to the corresponding element of the covariance matrix of

w_i, Ω, the matrix $N^{-1}\hat{X}'V^{-1}\hat{X}$ does not converge to the same limit as $N^{-1}\hat{X}'\Omega^{-1}\hat{X}$. That the choice of the weights by DAGENAIS [1973] is not optimal has been pointed out by CONNIFFE [1983 b] who proposed another weighting matrix with constant elements (identical diagonal elements and identical off-diagonal ones), that differs from the optimal weighting matrix Ω. Assuming for the ease of simplicity that z_i is nonstochastic and writing

$$\begin{aligned} (12) \qquad w_i &= \varepsilon_i + \beta(x_i - \alpha z_i) + \beta z_i(\alpha - \hat{\alpha}) \qquad \text{if} \quad i > N/2 \\ &= \varepsilon_i \qquad \text{if} \quad i \leqslant N/2, \end{aligned}$$

we have for the $i-j$ th element of Ω

$$\begin{aligned} \omega_{ij} &= \sigma^2 + \beta^2\sigma_V^2 + \beta^2 z_i z_j (\Sigma_c z_i^2)^{-1}\sigma_V^2 \qquad i=j, \quad i>N/2 \\ (13) \qquad &= \beta^2 z_i z_j (\Sigma_c z_i^2)^{-1}\sigma_V^2 \qquad i \neq j, \quad i,j > N/2 \\ &= \sigma^2 \qquad i=j, \quad i \leqslant N/2 \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

A feasible GLS estimator $\hat{\beta}_g$ is obtained if we substitute consistent estimates for the unknown parameters in (3) and (4) and use $\hat{\Omega}$ instead of V in (12). To invert $\hat{\Omega}$, we use the binomial inversion theorem that will be needed in more complex cases as well. First we write

$$(14) \qquad \hat{\Omega} = G + ZHZ',$$

where G is a diagonal with $\hat{\sigma}^2$ and $\hat{\sigma}^2 + \hat{\beta}^2\hat{\sigma}_V^2$ in position i of the main diagonal for $i \leqslant N/2$ and $i > N/2$ respectively, $H = \hat{\sigma}_V^2\hat{\beta}^2(\Sigma_c z_i^2)^{-1}$ is a scalar, Z is a $N \times 1$ vector with i-th element being equal to zero and z_i, for $i \leqslant N/2$ and $i > N/2$ respectively. The inverse of $\hat{\Omega}$ can be obtained straightforwardly as

$$(15) \qquad (G + ZHZ')^{-1} = G^{-1} - G^{-1}Z(H^{-1} + Z'G^{-1}Z)^{-1}Z'G^{-1}.$$

The asymptotic variance of $\hat{\beta}_g$ can be consistently estimated by $(\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1}$. The estimator $\hat{\beta}_g$ is more efficient than the first order method in (6), the DAGENAIS [1973] estimator and the CONNIFFE [1983 *b*] estimator. The asymptotic variance of the estimators considered in this section will be given in appendix A.

In Table 1 we report the ratio of the variance of alternative consistent estimators compared with the variance of the ML estimator assuming normality of ε_i and v_i. From the results in appendix A, it follows that the relative efficiency only depends on $R_x^2 = \alpha^2\sigma_z^2\sigma_x^{-2}$ and $R_y^2 = \beta^2\sigma_x^2(\beta^2\sigma_x^2 + \sigma^2)^{-1}$, where $\sigma_x^2 = \alpha^2\sigma_z^2 + \sigma_V^2$. From the results in Table 1, it appears that the OLS estimator using the complete observations only, $\hat{\beta}_c$, is roughly as efficient as the proxy variables estimators $\hat{\beta}_d$ and $\hat{\beta}_g$, when R_x^2 is small. When R_x^2 is large, $\hat{\beta}_c$ is inferior to all proxy variables estimators considered. This finding is plausible. When a large fraction of the variance of x_i is explained by z_i, αz_i is a fairly accurate approximation of x_i and it pays to use this information. However, when R_x^2 is low compared to R_y^2, $\hat{\beta}_p$ can well be less efficient than $\hat{\beta}_c$.

In column 7 the ratios of the variances computed using the OLS formula for standard errors (10) and the correct asymptotic variance for $\hat{\beta}_p$ in (8) is presented. In a few occasions the asymptotic bias for the standard errors involved in using the OLS formula appears to be quite important.

In order to explain the results on the relative efficiency of the four estimators considered in columns 3 to 6 of Table 1 we express the proxy variables estimators as a linear combination of $\hat{\beta}_c$ and a consistent estimator of β, $\hat{\beta}_{mj}$, based on incomplete observations only (except for the estimate $\hat{\alpha}$),

$$\hat{\beta}_j = \hat{\lambda}_j \hat{\beta}_c + (1 - \hat{\lambda}_j)\, \hat{\beta}_{mj} \qquad \text{with} \quad j \in \{p, d, g\}.$$

The expressions for $\hat{\beta}_{mj}$ and $\hat{\lambda}_j$ are given below

j	$\hat{\beta}_{mj}$	$\hat{\lambda}_j$
p	$(\Sigma_I \hat{x}_i^2)^{-1} \Sigma_I \hat{x}_i y_i$	$(\Sigma_c x_i^2 + \Sigma_I \hat{x}_i^2)^{-1} \Sigma_c x_i^2$
d	$(\Sigma_I \hat{x}_i^2)^{-1} \Sigma_I \hat{x}_i y_i$	$(\hat{\sigma}^{-2} \Sigma_c x_i^2 + \hat{\bar{\sigma}}^{-2} \Sigma_I \hat{x}_i^2)^{-1} \hat{\sigma}^{-2} \Sigma_c x_i^2$
g	$(\Sigma_I \omega_{ii'}^* \hat{x}_i \hat{x}_{i'})^{-1} \Sigma_I \omega_{ii'}^* \hat{x}_i y_{i'}$	$(\hat{\sigma}^{-2} \Sigma_c x_i^2 + \Sigma_I \omega_{ii'}^* \hat{x}_i \hat{x}_{i'})^{-1} \hat{\sigma}^{-2} \Sigma_c x_i^2$

where $\omega_{ii'}^*$ denotes the (i, i')-th element of $\hat{\Omega}^{-1}$. The large sample variance of $\hat{\beta}_j$ is

$$\text{(16)} \qquad \operatorname{Avar}(\sqrt{N}\, \hat{\beta}_j) = \lambda_j^2\, v_c + (1 - \lambda_j)^2\, v_{mj}$$

with $\lambda_j = \operatorname{plim} \hat{\lambda}_j$, $v_c = \operatorname{Avar}(\sqrt{N}\, \hat{\beta}_c)$ and $v_{mj} = \operatorname{Avar}(\sqrt{N}\, \hat{\beta}_{mj})$. It is straightforward to verify that $\hat{\beta}_j$ is asymptotically more efficient than $\hat{\beta}_c$ if

$$\text{(17)} \qquad (v_{mj} - v_c)/(v_{mj} + v_c) < \lambda_j < 1$$

and that the choice of λ_j which minimizes the asymptotic variance of $\sqrt{N}\, \hat{\beta}_j$ is $\lambda_j^{\text{opt}} = v_{mj}(v_c + v_{mj})^{-1}$ which satisfies (17). As $\lambda_g = \lambda_g^{\text{opt}}$, $\hat{\beta}_g$ is efficient relative to $\hat{\beta}_c$. As $\lambda_p \neq \lambda_p^{opt}$ and $\lambda_d \neq \lambda_d^{\text{opt}}$, the estimators $\hat{\beta}_p$ and $\hat{\beta}_d$ are more efficient than $\hat{\beta}_c$ only if inequality (17) is satisfied.

This will not be the case if β (or R_y^2) is sufficiently large, as the lower bound in (17) is close to 1 if β is large, while λ_p and λ_d do not depend on β. In this case, due to suboptimal weighting, the additional information on β contained in $\hat{\beta}_{mj}$ leads to an efficiency loss of $\hat{\beta}_j$ compared with $\hat{\beta}_c$. If on the contrary R_x^2 is large so that αz_t tends to be a better proxy, v_{mj} gets close to v_c and the proxy variables estimators $\hat{\beta}_p$ and $\hat{\beta}_d$ become more efficient than $\hat{\beta}_c$ in large samples.

The efficiency of the ML estimator in Table 1 arises from the assumption that the distributions of ε_i and v_i are known to be normal. If normality is assumed but does not hold, the Gaussian ML estimator which maximizes $L(\alpha, \beta, \sigma^2, \sigma_V^2)$ above will still be consistent and the asymptotic distribution can be determined (see GOURIÉROUX *et al.* [1984]). This estimator is however not necessarily more efficient than the proxy variables estimators if ε_i and v_i are not normal. In Table 2 we present the relative efficiency of the pseudo ML estimator compared with the optimal proxy variables estimator $\hat{\beta}_g$. In appendix A, we give the formulae for the asymptotic variance that have been used to obtain the results in Table 2. The relative efficiency with respect to other proxy variables estimators can easily be derived from the results in Tables 1 and 2 because the relative efficiency of one proxy variables estimator to another is unaffected when normality does

TABLE 2

Relative Efficiency of the Optimal Proxy Variables Estimator $\hat{\beta}_g$ Compared with the Gaussian ML Estimator

R_x^2, R_y^2	Values of $E\varepsilon_i^4/(E\varepsilon_i^2)^2=(Ev_i^4)/(Ev_i^2)^2$				
	2	3	4	6	10
0.2.	1.1623	1.1129	1.0675	0.9869	0.8575
0.4.	1.1748	1.1259	1.0809	1.0011	0.8721
0.6.	1.1305	1.0962	1.0638	1.0045	0.9036
0.8.	1.0672	1.0501	1.0336	1.0019	0.9441
0.95.	1.0167	1.0125	1.0083	1.0001	0.9841

not hold. In Table 2 we restrict ourselves to cases where $R_x^2=R_y^2$ and $(E\varepsilon_i^4)/(E\varepsilon_i^2)^2=(Ev_i^4)/(Ev_i^2)^2$. Evidently the normality assumption does not have a very large effect on the relative efficiency unless only small fractions of the variances of y_i and x_i are explained by (3) and (4) and the true distributions of ε_i and v_i have very fat tails. Simulations have shown ML to be fairly robust to non-normaly. BEALE and LITTLE [1975] reject non-ML methods based on simulations using pseudo-normal data. Our findings suggest that this may not be true in general. The question whether ML for non-normal errors (e.g. a multivariate t-distribution, see LITTLE and RUBIN [1987], chap. 10) yields better results in the present case has not been investigated.

3 The General Model

In this section we consider the general model introduced in equations (1) and (2). As in the simple case considered in the previous section a consistent estimate of β can be computed from complete observations only. Define y_c, X_c and Z_c as the vector and matrices obtained after deletion of rows of y, X and Z respectively for which some variable is missing. The regression estimator based on complete observations only can be written as

$$\hat{\beta}_c=(X_c'X_c)^{-1}X_c'y_c. \tag{18}$$

In the model (1) and (2) several proxies can be considered. A first possibility is to obtain estimates of the α_{jk} from regressions using complete observations only

$$\hat{\alpha}=(Z_c'Z_c)^{-1}Z_c'X_c \tag{19}$$

and subsequently to approximate missing exogenous variables in (1) by

$$\hat{x}_{ik}^{(1)} = \sum_{j=1}^{J} z_{ij}\hat{\alpha}_{jk} \qquad \text{if} \quad \delta_{ik}=0.$$

If $\hat{x}_{ik}^{(1)} = x_{ik}$ if $\delta_{ik}=1$ is defined for notational convenience, β can subsequently be estimated from the regression model

$$\begin{bmatrix} y_c \\ y_{\mathrm{I}} \end{bmatrix} = \begin{bmatrix} X_c \\ \hat{X}_{\mathrm{I}}^{(1)} \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_c \\ w_{\mathrm{I}}^{(1)} \end{bmatrix} \tag{20}$$

or

$$y = \hat{X}^{(1)}\beta + w^{(1)}, \tag{21}$$

where the subscript I in (20) refers to incomplete observations. From the example in the previous section we know that the ordinary least squares estimator

$$\hat{\beta}_p = \{X_c'X_c + \hat{X}_{\mathrm{I}}^{(1)\prime}\hat{X}_{\mathrm{I}}^{(1)}\}^{-1}\{X_c'y_c + \hat{X}_{\mathrm{I}}^{(1)\prime}y_{\mathrm{I}}\} \tag{22}$$

is not necessarily more efficient than $\hat{\beta}_c$. As in section 2 we have to analyze the structure of the covariance matrix of the disturbances in order to derive a generalized least squares estimator. Because

$$\begin{aligned} y_i - \sum_{k=1}^{K} \hat{x}_{ik}^{(1)}\beta_k = \varepsilon_i + \sum_{k=1}^{K} (1-\delta_{ik})\beta_k v_{ik} \\ + \sum_{j=1}^{J}\sum_{k=1}^{K} (1-\delta_{ik})\beta_k z_{ij}(\alpha_{jk}-\hat{\alpha}_{jk}) \end{aligned} \tag{23}$$

and $\alpha_{jk}-\hat{\alpha}_{jk}$ is linear in the v_{ik} we have $w^{(1)}=\varepsilon+AV$ for a suitably chosen $(N\times N)$ matrix A and the GLS estimator $\hat{\beta}_g^{(1)}$ can straightforwardly be computed using (15). Moreover it is evident that $\hat{\beta}_g^{(1)}$ will be more efficient than $\hat{\beta}_c$ because $\hat{\beta}_c$ coincides with the IV estimator of β from (20), with $(Z_c', 0)$ being the matrix of instuments.

To see how the efficiency of $\hat{\beta}_g$ is affected if relevant regressors are excluded from the auxiliary regressions, partition Z as $Z=(Z_1, Z_2)$ where Z_1 and Z_2 are $(N\times J_1)$ and $(N\times(J-J_1))$ matrices respectively and assume that the regression model

$$x_{ik} = \sum_{j=1}^{J_1} \eta_{jk} z_{ij} + v_{ik}^{*} \tag{24}$$

still satisfies the assumptions that were made with respect to (2). This model suggests the use of the proxies

$$\begin{aligned} \hat{x}_{ik}^{(2)} &= \sum_{j=1}^{J_1} \hat{\eta}_{jk} z_{ij} \qquad \text{if} \quad \delta_{ik}=0 \\ &= x_{ik} \qquad \text{if} \quad \delta_{ik}=1 \end{aligned}$$

where $\hat{\eta}_{jk}$ is the regression estimate from (24). Substitution of this proxy yields the model

$$y = \hat{X}^{(2)} \beta + w^{(2)}, \tag{25}$$

from which β can again be estimated e. g. by generalized least squares yielding $\hat{\beta}_g^{(2)}$. The following theorem will be useful in determining the effect of the choice of a proxy variable on the associated estimators $\hat{\beta}_g^{(1)}$ and $\hat{\beta}_g^{(2)}$.

THEOREM: Assume that $y = X\beta + \varepsilon$ holds with plim $N^{-1} Z' \varepsilon = 0$ and let $\hat{X}$ and $\tilde{X}$ be two proxies for X. Consider the GLS-estimator $\hat{\beta}_{GLS} = (\hat{X}' \hat{\Sigma}^{-1} \hat{X})^{-1} \hat{X}' \hat{\Sigma}^{-1} y$ and the instrumental variables estimator $\tilde{\beta}_{IV} = (Z' \tilde{X})^{-1} Z' y$ and define $\hat{w} = y - \hat{X}\beta$ and $\tilde{w} = y - \tilde{X}\beta$.

Assume that

(i) $\sqrt{N}(\hat{\beta}_{GLS} - \beta) \overset{d}{\simeq} N(0, V^{-1})$, where $V^{-1} = \text{plim}\, N(\hat{X}' \hat{\Sigma}^{-1} \hat{X})^{-1}$ is finite and positive definite;

(ii) $p \lim N^{-1} Z' \tilde{X} = Q$ is finite and positive definite;

(iii) $\frac{1}{\sqrt{N}} \begin{bmatrix} Z' \hat{w} \\ Z'(\tilde{w} - \hat{w}) \end{bmatrix} \overset{d}{\simeq} N(0, D)$ where $D = \text{plim}\, N^{-1} \begin{bmatrix} Z' \hat{\Sigma} Z & 0 \\ 0 & Z' S Z \end{bmatrix}$

for some S, where $\overset{d}{\simeq}$ denotes convergence in distribution.

Then $\hat{\beta}_{GLS}$ is asymptotically at least as efficient as $\tilde{\beta}_{IV}$.

Proof: See appendix B.

The third requirement is most crucial. If two proxies $\hat{X}$ and $\tilde{X}$ are available, an IV estimator based on $\tilde{X}$ cannot be more efficient than a GLS estimator based on $\hat{X}$ if $Z' \hat{w}$ and $Z'(\tilde{w} - \hat{w})$ are asymptotically orthogonal provided the regularity conditions of the theorem are met. Notice that the theorem is concerned with the comparison of the asymptotic variance of generalized method of moments estimators based on different sets of orthogonality conditions.

Returning to the analysis of the relative efficiency of $\hat{\beta}_g^{(1)}$ and $\hat{\beta}_g^{(2)}$ let us first consider the case where the parameters α_{jk} and η_{jk} are known a priori. Then it is very simple to use the theorem to show that conditioning on the larger information set will yield more efficient estimates. Define $\hat{X} = \hat{X}^{(1)}$ and $\tilde{X} = \hat{X}^{(2)}$. The disturbances in theorem are $\hat{w} = \varepsilon + (X - \hat{X})\beta$ and $\tilde{w} = \varepsilon + (X - \tilde{X})\beta$ respectively. As $\hat{x}_{ik}^{(1)} = E[x_{ik} | I_1]$ and $\hat{x}_{ik}^{(2)} = E[x_{ik} | I_2]$ with $I_2 \subset I_1$, $\hat{w} - \tilde{w} = (\hat{X} - \tilde{X})\beta$ will be orthogonal to $\hat{w}$ by the properties of conditional expectations and the theorem immediately implies efficiency of the proxy variables estimator based on the larger conditioning set. Unfortunately this result does not hold true in general if α_{jk} and η_{jk} have to be estimated. A counter-example in a slightly different model is presented in the next section. In appendix C we show that the result holds for $\hat{\beta}_g^{(1)}$ and $\hat{\beta}_g^{(2)}$ if δ_{ik} does not depend on k that is if all exogenous variables in (1) are missing when one of them is. We conjecture that more general results can be proved along the same lines. The relative efficiency of $\hat{\beta}_g^{(1)}$ with respect to $\hat{\beta}_g^{(2)}$ implies among other things that if a constant is included in (2), the use of that auxiliary regression model will yield more efficient estimates of β than simple imputation of mean values for missing observations (which is equivalent to a regression on a constant only) as is often done in practice.

The theorem above can also be used to demonstrate the effect of more efficient estimation of the α_{jk} in (2). If prior restrictions on these parameters are available, or if observations for which some but not all exogenous variables in (1) are observed are also used to estimate these parameters, ZELLNER's [1962] SUR estimator will be more efficient than a regression on the set of complete observations only. If the SUR estimates are denoted by $\hat{\hat{\alpha}}_{jk}$ and $\hat{x}_{ik}^{(3)}$ is defined as

$$\hat{x}_{ik}^{(3)} = \sum_{j=1}^{J} \hat{\hat{\alpha}}_{jk} z_{ij} \qquad \text{if} \quad \delta_{ik}=0$$
$$= x_{ik} \qquad \text{if} \quad \delta_{ik}=1,$$

we obtain

$$(26) \qquad \sum_{k=1}^{K} (\hat{x}_{ik}^{(1)} - \hat{x}_{ik}^{(3)}) \beta_k = \sum_{j=1}^{J} \sum_{k=1}^{K} (1-\delta_{ik}) \beta_k z_{ij} (\hat{\alpha}_{jk} - \hat{\hat{\alpha}}_{jk}).$$

Using (23) with $\hat{x}_{ik}^{(1)}$ and $\hat{\alpha}_{jk}$ replaced by $\hat{x}_{ik}^{(3)}$ and $\hat{\hat{\alpha}}_{jk}$ respectively, the well-known fact that $\sqrt{N}(\hat{\hat{\alpha}}_{jk} - \alpha_{jk})$ and $\sqrt{N}(\hat{\alpha}_{j'k'} - \hat{\hat{\alpha}}_{j'k'})$ ($j, j'=1, \ldots, J$; $k, k'=1, \ldots, K$) are asymptotically orthogonal because $\hat{\hat{\alpha}}_{jk}$ is efficient (see HAUSMAN [1978]) implies that the requirements of the lemma are satisfied. Therefore the GLS proxy variables of β will in general be more efficient if the auxiliary regression coefficients are estimated by SUR instead of OLS.

4 Extensions

In this section we will indicate extensions of the results in sections 2 and 3 to cases in which aggregates of x_i in (3) are observed and to a dynamic auxiliary model. For simplicity we consider two examples. First assume that x_t is a flow variable which is observed every second period only, that is observations are available on $\bar{x}_t = x_t + x_{t-1}$ if $t \in T_2 = \{2, 4, 6, \ldots, T\}$. Throughout this section "$\bar{\ }$" will denote similar temporal aggregates. Because aggregates are observed more frequently for time series than for cross-sections, we change the notation for the subscripts. Assume that the analogues of (3) and (4) hold,

$$(27) \qquad y_t = \beta x_t + \varepsilon_t$$

$$(28) \qquad x_t = \alpha z_t + v_t.$$

Ordinary least squares applied to the aggregate data ("a" denotes that aggregates are observed) yields

$$(29) \qquad \hat{\beta}_{ac} = (\Sigma_{T_2} \bar{x}_t^2)^{-1} \Sigma_{T_2} \bar{x}_t \bar{y}_t$$

which is consistent. Using the proxy $\hat{x}_t = \hat{\alpha} z_t + 1/2 (\bar{x}_t - \hat{\alpha} \bar{z}_t)$ for $t \in T_2$ and $\hat{x}_t = \hat{\alpha} z_t + 1/2 (\bar{x}_{t+1} - \hat{\alpha} \bar{z}_{t+1})$ for $t \notin T_2$, where

$$(30) \qquad \hat{\alpha} = (\Sigma_{T_2} \bar{z}_t^2)^{-1} \Sigma_{T_2} \bar{z}_t \bar{x}_t,$$

we have

$$(31) \qquad y_t = \hat{x}_t \beta + w_t$$

with

$$(32) \quad \begin{cases} w_t = \varepsilon_t + \beta (v_t - 1/2 \bar{v}_t) + \beta (z_t - 1/2 \bar{z}_t)(\alpha - \hat{\alpha}) & \text{if} \quad t \in T_2 \\ w_t = \varepsilon_t + \beta (v_t - 1/2 \bar{v}_{t+1}) + \beta (z_t - 1/2 \bar{z}_{t+1})(\alpha - \hat{\alpha}) & \text{if} \quad t \notin T_2. \end{cases}$$

OLS applied to (31), GLS applied to (31) with V being the covariance matrix of w_t assuming $\hat{\alpha} = \alpha$, and GLS with optimal weights, i.e. V being the covariance matrix of w_t in (32), yield the consistent estimators $\hat{\beta}_{ap}$, $\hat{\beta}_{ad}$ and $\hat{\beta}_{ag}$ respectively. Expressions for the asymptotic variance of the estimators $\hat{\beta}_{ac}$ and $\hat{\beta}_{ad}$ and the ML estimator have been given by Palm and Nijman [1982] where $\hat{\beta}_{ad}$ is called the GLS estimator. For the sake of completeness the formulae are given in appendix A.

A simple transformation of equation (31) yields

$$(33) \quad \begin{cases} y_t + y_{t-1} = (x_t + x_{t-1}) \beta + \varepsilon_t + \varepsilon_{t-1}, & t \in T_2 \\ y_t - y_{t-1} = (\hat{x}_t - \hat{x}_{t-1}) \beta + w_t - w_{t-1}, & t \notin T_2. \end{cases}$$

From the theorem in the previous section it follows that because of the inclusion of $(\hat{x}_t - \hat{x}_{t-1}) \beta$ in the regressor, $\hat{\beta}_{ag}$ is asymptotically more efficient than $\hat{\beta}_{ac}$.

In Table 3 some numerical results on the ratio of the asymptotic variance of alternative consistent estimators compared with the large sample variance of the ML estimator are reported. For simplicity we only consider the case where the disturbances ε_t and v_t are normally distributed. In that case the relative efficiency depends on R_x^2, R_y^2 and $\rho = \sigma_z^{-2} E z_t z_{t-1}$.

For cross-sections $\rho = 0$. Column 8 of Table 3 contains the relative efficiency of the ML estimator for a complete sample with respect to that for the incomplete sample. In column 9 we compare the standard errors for $\hat{\beta}_{ap}$ computed by means of $\operatorname{plim} T \hat{\sigma}_w^2 (X' \hat{X})^{-1}$ with the correct formula for the variance of $\hat{\beta}_{ap}$.

From Table 3 we can conclude that $\hat{\beta}_{ag}$ is fairly accurate in most instances. The estimator $\hat{\beta}_{ac}$ seems to have a reasonable precision too. However, $\hat{\beta}_{ap}$ becomes very inaccurate when the autocorrelation of z_t is negative. This is not surprising as the estimator of α will in this case be very imprecise and no correction for heteroskedasticity is used. The estimator $\hat{\beta}_{ad}$ is sometimes less accurate than $\hat{\beta}_{ac}$. In these cases using additional information in a suboptimal way leads to a loss of efficiency. The

TABLE 3

Relative Efficiency of Alternative Consistent Estimators for β Compared with the ML Estimator

1 R_x^2	2 R_y^2	3 ρ	Alternative estimators 4 $\hat\beta_{ap}$	5 $\hat\beta_{ad}$	6 $\hat\beta_{ag}$	7 $\hat\beta_{ac}$	8 $\hat\beta_{ML}$ complete sampling	9 $\frac{\text{plim } \hat\sigma_w^2/\Sigma \hat{x}_t^2}{\text{avar } (\hat\beta_{ap})}$
0.40	0.40	−0.95	5.8050	4.3690	1.2310	1.3221	0.4099	0.1210
0.40	0.40	−0.80	2.1353	1.7254	1.1867	1.4380	0.4889	0.3925
0.40	0.40	−0.40	1.3514	1.2090	1.1425	1.4689	0.6169	0.7826
0.40	0.40	0.00	1.2076	1.1382	1.1259	1.3761	0.6880	0.9767
0.40	0.40	0.40	1.1502	1.1189	1.1172	1.2643	0.7333	1.0929
0.40	0.40	0.80	1.1200	1.1119	1.1119	1.1585	0.7646	1.1703
0.40	0.40	0.95	1.1122	1.1103	1.1103	1.1218	0.7740	1.1930
0.40	0.90	−0.95	61.9462	6.1713	1.1461	1.1528	0.3574	0.0305
0.40	0.90	−0.80	16.1162	2.0562	1.1315	1.1532	0.3921	0.1286
0.40	0.90	−0.40	5.6263	1.2280	1.1048	1.1436	0.4803	0.4512
0.40	0.90	0.00	3.2832	1.1114	1.0881	1.1250	0.5625	0.9056
0.40	0.90	0.40	2.1212	1.0799	1.0767	1.1023	0.6393	1.5931
0.40	0.90	0.80	1.3645	1.0685	1.0684	1.0777	0.7113	2.7553
0.40	0.90	0.95	1.1363	1.0659	1.0659	1.0682	0.7371	3.4287
0.90	0.40	−0.95	1.1481	1.1399	1.0100	4.3439	0.3149	0.2984
0.90	0.40	−0.80	1.0386	1.0317	1.0050	4.4937	0.6291	0.6588
0.90	0.40	−0.40	1.0128	1.0080	1.0037	2.6205	0.8386	0.9006
0.90	0.40	0.00	1.0074	1.0044	1.0035	1.8003	0.9002	0.9719
0.90	0.40	0.40	1.0050	1.0035	1.0034	1.3671	0.9296	1.0061
0.90	0.40	0.80	1.0037	1.0033	1.0033	1.1011	0.9469	1.0261
0.90	0.40	0.95	1.0034	1.0033	1.0033	1.0261	0.9517	1.0317
0.90	0.90	−0.95	4.0678	3.5825	1.1726	1.5562	0.1128	0.0423
0.90	0.90	−0.80	2.1694	1.7660	1.0751	1.6972	0.2376	0.1672
0.90	0.90	−0.40	1.5549	1.1848	1.0348	1.5441	0.4941	0.4850
0.90	0.90	0.00	1.3209	1.0592	1.0251	1.3546	0.6773	0.7827
0.90	0.90	0.40	1.1710	1.0258	1.0208	1.1981	0.8147	1.0619
0.90	0.90	0.80	1.0620	1.0185	1.0183	1.0716	0.9216	1.3245
0.90	0.90	0.95	1.0281	1.0177	1.0177	1.0305	0.9558	1.4190

estimator $\hat\beta_{ag}$ is of course more efficient than $\hat\beta_{ac}$. Finally the bias due to using $\hat\sigma_w^2(\hat X'\hat X)^{-1}$ to estimate the asymptotic variance of $\hat\beta_{ap}$ can be quite important.

In the second extension, we consider a dynamic equation for the exogenous variables x_t. In dynamic models, the prediction of the missing observations will usually depend on auxiliary variables and on the observed values of the variable itself. Simple examples have been considered e. g. by CHOW and LIN [1971, 1976], and by LITTERMAN [1983]. In more complex models the classical Wiener-Kolmogorov filtering theory or the Kalman filter can be used to derive the best approximations for missing observations, see e. g. NIJMAN and PALM [1986].

Here we restrict ourselves to a discussion of the relative efficiency of proxy variables estimators for the model

(34) $$y_t = \beta x_t + \varepsilon_t$$

(35) $$x_t = \gamma x_{t-1} + \alpha z_t + v_t, \qquad |\gamma| < 1,$$

where the assumptions on ε_t and v_t are as above. Assume that x_t is observed if $t \in T_2$ only, e.g. because the model is semi-annual but only annual data on x_t are available. When x_t is unobserved, we can use the proxy

$$(36) \qquad \hat{x}_t = (1+\hat{\gamma}^2)^{-1} [\hat{\gamma} x_{t-1} + \hat{\gamma} x_{t+1} + \hat{\alpha} z_t - \hat{\alpha}\hat{\gamma} z_{t+1}],$$

which is the expectation of x_t given past, present and future information on x_t and z_t where consistent estimates have been substituted for α and γ. OLS applied to (34) after substitution of this proxy for x_t is consistent for β because (36) is an estimate of the conditional expectation which implies that (7) is satisfied. Note that a regression on ad hoc interpolated values, e.g. using the method proposed by BOOT, FEIBES and LISMAN [1967] can yield estimates which are strongly biased asymptotically as is shown in PALM and NIJMAN [1984] and NIJMAN and PALM [1985].

Estimates of α and γ cannot be obtained by direct regression because x_t and x_{t-1} are not observed simultaneously. We consider the following three consistent estimators of α and γ:

First the ML applied to equation (35) after elimination of the unobserved values of x_t which can be written as

$$(37) \qquad x_t = \gamma^2 x_{t-2} + \alpha z_t + \alpha\gamma z_{t-1} + (v_t + \gamma v_{t-1}), \qquad t \in T_2$$

with one nonlinear restriction on the parameters (M1).

Second OLS applied to the unrestricted version of (37),

$$(38) \qquad x_t = \psi_1 x_{t-2} + \psi_2 z_t + \psi_3 z_{t-1} + (v_t + \gamma v_{t-1}),$$

and $\hat{\gamma} = \pm \hat{\psi}_1^{1/2}$, $\hat{\alpha} = \hat{\psi}_2$ (M2). The sign of $\hat{\gamma}$ is determined by $\hat{\psi}_3 \hat{\psi}_2^{-1}$.
Third, again using (38), with $\hat{\gamma} = \hat{\psi}_3 \hat{\psi}_2^{-1}$ and $\hat{\alpha} = \hat{\psi}_2$ (M3).

In order to illustrate what the effect is of neglecting the indicator variable z_t in the computation of the proxy for x_t, we also consider an estimate of the expectation of unobserved x_t's conditional on observed x_t's only. We restrict ourselves to the case where z_t is a white noise independent of x_{t-1} in which case $E[x_t | x_s, s \in T_2] = (1+\gamma^2)^{-1}(\gamma x_{t-1} + \gamma x_{t+1})$ if $t \notin T_2$. If the sign of γ is assumed to be known a priori, this parameter can be estimated by ML from (37) with $\alpha = 0$ and the corresponding proxy can be substituted into (34) (M4).

As argued above, OLS applied to (34) after substitution of one of these four proxies $\hat{x}_t$ will yield a consistent estimator of β. This estimator will be denoted $\hat{\beta}_p$. The error term $w_t = \varepsilon_t + \beta(x_t - \hat{x}_t)$ is heteroscedastic and serially correlated, so that one is again naturally led to consider GLS-estimators, such as proposed by DAGENAIS [1973], GOURIÉROUX and MONFORT [1981] and CONNIFFE [1983 *b*] for static regression models. The estimator using the weights suggested by DAGENAIS [1973] will be denoted as $\hat{\beta}_d$. The estimator using the optimal weights given by the inverse of the covariance matrix of w_t will be denoted by $\hat{\beta}_g$. Finally, β in (34) can be consistently estimated from the complete observations only by $\hat{\beta}_c$.

Numerical results on the relative asymptotic efficiency of the consistent estimators of β discussed above are given in Table 4. In the last column of Table 4, we compare the SE's of $\hat{\beta}_p$, with α and γ estimated by ML (*see* M1) with the correct asymptotic standard errors. The parameter γ, the first order autocorrelation of z_t, ρ, and the variance ratio's R_x^2 and R_y^2 determine the relative efficiency of the different estimators with respect to the ML-estimator. Computational details are given in appendix A.

From the results in Table 4, it is quite obvious that all proxy variables estimators are fairly efficient when γ and α are estimated by ML. Also, OLS applied to complete data only is reasonably efficient in most instances. The estimator $\hat{\beta}_p$ can be more efficient as well as less efficient than $\hat{\beta}_c$. Notice that the asymptotic variance of $\hat{\beta}_c$ is twice that of the ML or GLS estimate for the case where all values of x_t are observed. When a moment estimator (M2 or M3) is used for γ and α, the relative efficiency is very sensitive to the parameter values. In particular, a negative value for γ combined with negative first order autocorrelation of z_t often leads to a large relative efficiency of ML compared with the proxy variables estimators based on M2 or M3. The Jacobian of the transformation of the moments to $\gamma=\psi_1^{1/2}$ equals $.5\gamma^{-1}$, so that when $\gamma=0$ (which is ignored in the estimation), the large sample variance of these estimators cannot be evaluated. This is indicated by INF.

Evidently, more efficient estimation of α and γ yields more efficient proxy variables estimators of β, in accordance with our theorem. The inclusion of the observations on z_t in the conditional expectation of x_t appears to improve the efficiency if (34) is estimated by GLS and (37) is efficiently estimated, which is not surprising given the results in the previous section. Note that it can be more efficient to use the smaller information set if moment estimators of α and γ are used instead of efficient estimators. Finally the last column of Table 4 indicates that the commonly used formula for standard errors can be severely biased when proxies are used. Sign and magnitude of the bias depend on the true parameter values.

5 Concluding remarks

To summarize we considered several consistent estimators for regression models with missing exogenous variables. It is not difficult to obtain proxies for the missing obversations such that the resulting proxy variables estimators will be consistent. To assure consistency one should preferably use conditional expectations to construct the proxies. We have shown how to obtain proxy variables estimators that are more efficient than estimators based on complete observations only. The use of more information, when constructing a proxy and estimating the parameters of the auxiliary equation, will usually yield more efficient estimators. The asymptotic efficiency of some proxy variables estimators is much lower than that of the Gaussian

TABLE 4

Relative efficiency of the alternative consistent estimators for* β *in (34) compared with ML, when* x_t *is generated by (35) and the ratio of commonly used and true asymptotic standard errors.

				M₁			M₂		M₃		M₄			
γ 1	ρ 2	R_x^2 3	R_y^2 4	$\hat\beta_p$ 5	$\hat\beta_d$ 6	$\hat\beta_g$ 7	$\hat\beta_p$ 8	$\hat\beta_g$ 9	$\hat\beta_p$ 10	$\hat\beta_g$ 11	$\hat\beta_p$ 12	$\hat\beta_g$ 13	$\hat\beta_c$ 14	SE 15
−0.90	−0.90	0.90	0.90	1.29	1.26	1.20	6.06	1.75	635.13	1.89			1.90	2.19
−0.40	−0.90	0.90	0.90	2.38	1.73	1.23	76.63	1.45	25.54	1.40			1.54	0.80
0.00	−0.90	0.90	0.90	1.32	1.06	1.03	INF	INF	1.98	1.12			1.35	0.97
0.40	−0.90	0.90	0.90	1.67	1.26	1.02	17.69	1.14	2.20	1.07			1.26	0.24
0.90	−0.90	0.90	0.90	1.21	1.18	1.00	1.45	1.04	1.23	1.01			1.64	0.04
−0.90	0.00	0.90	0.90	1.15	1.14	1.12	3.62	1.57	6.34	1.60	6.66	1.51	1.90	0.47
−0.40	0.00	0.90	0.90	1.24	1.07	1.04	5.65	1.10	1.30	1.05	4.50	1.42	1.47	0.77
0.00	0.00	0.90	0.90	1.32	1.06	1.03	INF	INF	1.32	1.03	1.35	1.35	1.35	0.97
0.40	0.00	0.90	0.90	1.23	1.07	1.03	1.68	1.05	1.28	1.04	2.84	1.42	1.47	0.78
0.90	0.00	0.90	0.90	1.00	1.00	1.00	1.01	1.01	1.03	1.02	1.45	1.30	1.90	0.55
−0.90	0.90	0.90	0.90	1.22	1.19	1.05	1.57	1.17	1.24	1.08			1.64	0.04
−0.40	0.90	0.90	0.90	1.77	1.31	1.02	32.26	1.10	2.28	1.07			1.26	0.23
0.00	0.90	0.90	0.90	1.32	1.06	1.03	INF	INF	1.98	1.12			1.35	0.97
0.40	0.90	0.90	0.90	1.30	1.13	1.08	1.57	1.17	1.49	1.10			1.54	1.47
0.90	0.90	0.90	0.90	1.00	1.00	1.00	1.01	1.01	1.41	1.15			1.90	2.82
−0.90	−0.90	0.40	0.90	2.23	1.71	1.33	20.36	1.58	7326.75	1.61			1.62	0.47
−0.40	−0.90	0.40	0.90	8.48	1.61	1.14	170.52	1.17	263.76	1.16			1.19	0.34
0.00	−0.90	0.40	0.90	3.28	1.11	1.09	INF	INF	3.76	1.09			1.13	1.10
0.40	−0.90	0.40	0.90	4.76	1.27	1.09	13.07	1.10	6.57	1.10			1.14	0.32
0.90	−0.90	0.40	0.90	1.86	1.46	1.01	2.06	1.02	1.91	1.01			1.50	0.08
−0.90	0.00	0.40	0.90	1.96	1.55	1.28	15.91	1.54	71.08	1.55	5.65	1.28	1.62	0.34
−0.40	0.00	0.40	0.90	4.38	1.26	1.11	61.13	1.13	8.87	1.12	3.59	1.13	1.17	0.52
0.00	0.00	0.40	0.90	3.28	1.11	1.09	INF	INF	3.28	1.09	1.13	1.13	1.13	1.10
0.40	0.00	0.40	0.90	2.37	1.10	1.08	2.45	1.09	2.41	1.08	2.26	1.13	1.17	0.95
0.90	0.00	0.40	0.90	1.07	1.01	1.01	1.08	1.02	1.13	1.02	1.23	1.10	1.61	0.62
−0.90	0.90	0.40	0.90	2.37	1.77	1.23	10.20	1.41	4.18	1.36			1.50	0.06
−0.40	0.90	0.40	0.90	5.88	1.36	1.11	11.01	1.12	10.68	1.12			1.14	0.26
0.00	0.90	0.40	0.90	3.28	1.11	1.09	INF	INF	3.76	1.09			1.13	1.10
0.40	0.90	0.40	0.90	2.02	1.08	1.08	4.89	1.14	2.27	1.08			1.19	1.44
0.90	0.90	0.40	0.90	1.06	1.01	1.01	1.07	1.02	1.78	1.06			1.61	0.98
−0.90	−0.90	0.90	0.40	1.02	1.02	1.02	1.39	1.28	49.90	1.95			1.98	2.65
−0.40	−0.90	0.90	0.40	1.12	1.11	1.09	7.82	1.62	3.21	1.48			1.88	1.32
0.00	−0.90	0.90	0.40	1.01	1.00	1.00	INF	INF	1.07	1.05			1.80	0.95
0.40	−0.90	0.90	0.40	1.01	1.01	1.00	2.67	1.36	1.07	1.04			1.75	0.35
0.90	−0.90	0.90	0.40	1.00	1.00	1.00	1.02	1.02	1.00	1.00			1.93	0.06
−0.90	0.00	0.90	0.40	1.01	1.01	1.01	1.20	1.16	1.41	1.26	1.48	1.16	1.98	0.52
−0.40	0.00	0.90	0.40	1.01	1.01	1.01	1.42	1.08	1.01	1.01	1.77	1.57	1.86	0.76
0.00	0.00	0.90	0.40	1.01	1.00	1.00	INF	INF	1.01	1.00	1.80	1.80	1.80	0.95
0.40	0.00	0.90	0.40	1.01	1.01	1.01	1.05	1.03	1.01	1.01	1.62	1.57	1.86	0.76
0.90	0.00	0.90	0.40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.08	1.08	1.98	0.53
−0.90	0.90	0.90	0.40	1.00	1.00	1.00	1.03	1.03	1.01	1.00			1.93	0.06
−0.40	0.90	0.90	0.40	1.02	1.02	1.01	4.17	1.20	1.08	1.04			1.75	0.35
0.00	0.90	0.90	0.40	1.01	1.00	1.00	INF	INF	1.07	1.05			1.80	0.95
0.40	0.90	0.90	0.40	1.02	1.02	1.02	1.04	1.04	1.04	1.03			1.88	1.45
0.90	0.90	0.90	0.40	1.00	1.00	1.00	1.00	1.00	1.03	1.03			1.98	2.71
−0.90	−0.90	0.40	0.40	1.10	1.10	1.09	2.68	1.58	639.48	1.89			1.90	0.74
−0.40	−0.90	0.40	0.40	1.68	1.49	1.21	16.85	1.39	25.58	1.34			1.50	0.57
0.00	−0.90	0.40	0.40	1.21	1.14	1.13	INF	INF	1.25	1.14			1.38	0.80
0.40	−0.90	0.40	0.40	1.24	1.14	1.07	1.98	1.11	1.40	1.11			1.37	0.39
0.90	−0.90	0.40	0.40	1.04	1.03	1.00	1.05	1.01	1.04	1.00			1.82	0.11
−0.90	0.00	0.40	0.40	1.08	1.08	1.07	2.30	1.50	7.12	1.70	1.42	1.11	1.90	0.48
−0.40	0.00	0.40	0.40	1.29	1.20	1.13	6.61	1.22	1.71	1.16	1.41	1.26	1.48	0.58
0.00	0.00	0.40	0.40	1.21	1.14	1.13	INF	INF	1.21	1.13	1.38	1.38	1.38	0.80
0.40	0.00	0.40	0.40	1.10	1.07	1.06	1.11	1.07	1.10	1.06	1.29	1.25	1.48	0.68
0.90	0.00	0.40	0.40	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.04	1.03	1.90	0.52
−0.90	0.90	0.40	0.40	1.08	1.08	1.05	1.78	1.30	1.24	1.16			1.82	0.10
−0.40	0.90	0.40	0.40	1.34	1.22	1.11	1.79	1.18	1.76	1.17			1.37	0.36
0.00	0.90	0.40	0.40	1.21	1.14	1.13	INF	INF	1.25	1.14			1.38	0.80
0.40	0.90	0.40	0.40	1.08	1.06	1.06	1.35	1.18	1.10	1.06			1.50	0.89
0.90	0.90	0.40	0.40	1.00	1.00	1.00	1.00	1.00	1.06	1.03			1.90	0.81

ML estimator. However, the optimal proxy variables estimator, which can be obtained by GLS, appears to be almost as efficient as the ML estimator which is computationally unattractive in larger models and it can be more efficient than ML estimation if normality does not hold. This finding should be very useful for empirical work on data sets which are not complete. Although the computational complexity of ML estimation and the possible deviation of the data from normality are strong arguments in favor of using imputed data, one should be aware of the fact that consistent estimation of the large sample variance of the estimators discussed can sometimes be tricky.

APPENDIX A

In this appendix we shall give the large sample variance for several estimators presented in the paper. Consider first the model presented in (3) and (4). The asymptotic distribution of the Gaussian ML estimator of $\theta=(\alpha, \beta, \sigma^2, \sigma_V^2)'$ is given by (see e. g. GOURIÉROUX *et al.* [1984])

$$\sqrt{N}(\hat{\theta}-\theta) \overset{d}{\simeq} N(0, A^{-1} B A^{-1})$$

where

$$A=\lim E\left[-N^{-1}\sum_{i=1}^{N} \partial^2 \ln L_i/\partial\theta\,\partial\theta'\right]$$

and

$$B=\lim E\left[N^{-1}\sum_{i=1}^{N} (\partial \ln L_i/\partial\theta)(\partial \ln L_i/\partial\theta')\right],$$

where L_i has been defined in section 2. Defining

$$(E\,\varepsilon_i^4)/(E\,\varepsilon_i^2)^2=\eta_\varepsilon, (E\,v_i^4)/(E\,v_i^2)^2=\eta_v$$

and

$$E(\varepsilon_i+\beta v_i)^4/(E(\varepsilon_i+\beta v_i)^2)^2=\tilde{\eta}$$

it is straightforward to verify that the elements of B (symmetric) are given by

$$(39)\quad \left\{\begin{aligned}
b_{11}&=\beta^2\sigma_z^2/2\tilde{\sigma}^2+\sigma_z^2/2\sigma_V^2\\
b_{12}&=\alpha\beta\sigma_z^2/2\tilde{\sigma}^2\\
b_{13}&=b_{14}=0\\
b_{22}&=\sigma_x^2/2\sigma^2+\alpha^2\sigma_z^2/2\tilde{\sigma}^2+(\tilde{\eta}-1)\beta^2\sigma_v^4/2\tilde{\sigma}^4\\
b_{23}&=\beta\sigma_V^2(\tilde{\eta}-1)/4\tilde{\sigma}^4\\
b_{24}&=\beta^3\sigma_V^2(\tilde{\eta}-1)/4\tilde{\sigma}^4\\
b_{33}&=(\eta_\varepsilon-1)/8\sigma^4+(\tilde{\eta}-1)/8\tilde{\sigma}^4\\
b_{34}&=\beta^2(\tilde{\eta}-1)/8\tilde{\sigma}^4\\
b_{44}&=(\eta_v-1)/8\sigma_V^4+\beta^4(\tilde{\eta}-1)/8\tilde{\sigma}^4
\end{aligned}\right.$$

with $\tilde{\sigma}^2=\sigma^2+\beta^2\sigma_V^2$, $\sigma_z^2=E\,z_i^2$ and $\sigma_x^2=E\,x_i^2$. The matrix A has the same structure and is obtained if one puts $\eta_\varepsilon=\eta_v=\tilde{\eta}=3$ in (39). If normality holds B coincides with A and the large sample variance of $\sqrt{N}\,\hat{\beta}_{ML}$ is simply the (2, 2) element of A^{-1} which can be shown to be

$$(40)\quad \mathrm{Var}(\sqrt{N}\,\hat{\beta}_{ML})=\{\sigma_x^2/2\sigma^2+[\alpha^2\sigma_z^2/2\tilde{\sigma}^2][1-\beta^2\sigma_V^2/(\beta^2\sigma_V^2+\tilde{\sigma}^2)] + [\beta^2\sigma_V^2/\tilde{\sigma}^4][1-(\beta^4\sigma_V^4+\sigma^4)/(\tilde{\sigma}^4+\sigma^4+\beta^4\sigma_V^4)]\}^{-1}.$$

Along the lines followed by PALM and NIJMAN [1982] for aggregate observations, one can obtain the asymptotic variance of consistent estimators of β for skipped observations.

For the proxy variables estimator $\hat{\beta}_p$ in (6), we have:

$$\text{(41)} \quad \operatorname{Var}(\sqrt{N}\,\hat{\beta}_p) = 2(\sigma_x^2 + \alpha^2\sigma_z^2)^{-2}(\sigma^2\sigma_x^2 + \tilde{\sigma}^2\alpha^2\sigma_z^2 + \alpha^2\beta^2\sigma_V^2\sigma_z^2).$$

Similarly for $\hat{\beta}_d$ in (11), where the matrix of weights proposed by Dagenais is used, one gets:

$$\text{(42)} \quad \operatorname{Var}(\sqrt{N}\,\hat{\beta}_d) = p + p^2\beta^2\alpha^2\sigma_z^2\sigma_V^2/2\tilde{\sigma}^4,$$

with p being the large sample variance of the GLS estimator of β when α is known

$$\text{(43)} \quad p = \{\sigma_x^2/2\sigma^2 + \alpha^2\sigma_z^2/2\tilde{\sigma}^2\}^{-1}.$$

When the optimal weights in (14) are used, we get:

$$\text{(44)} \quad \operatorname{Var}(\sqrt{N}\,\hat{\beta}_g) = \{\sigma_x^2/2\sigma^2 + \alpha^2\sigma_z^2/2\tilde{\sigma}^2 - \beta^2\alpha^2\sigma_V^2\sigma_z^2/2(\tilde{\sigma}^4 + \sigma_V^2\beta^2\tilde{\sigma}^2)\}^{-1}.$$

Finally, when we apply OLS to the complete data, the variance is the double of the variance of OLS in the case that no data are missing

$$\text{(45)} \quad \operatorname{Var}(\sqrt{N}\,\hat{\beta}_c) = 2\sigma^2\sigma_x^{-2}.$$

For the static model with observed aggregates of the exogenous variable x_t, the large sample variances of the ML estimator and of the estimator $\sqrt{T}\,\hat{\beta}_{ad}$ are derived in PALM and NIJMAN [1982]. For the sake of completeness, we give all formulae for the asymptotic variances.

For the proxy variables estimator $\hat{\beta}_{ap}$, we have

$$\text{(46)} \quad \operatorname{Var}(\sqrt{T}\,\hat{\beta}_{ap}) = 4(\alpha^2\tilde{\sigma}_z^2 + \mathrm{E}\,\bar{x}^2)^{-2}(\tilde{\sigma}^2\alpha^2\tilde{\sigma}_z^2 + \sigma^2\mathrm{E}\,\bar{x}^2 + \beta^2\alpha^2\sigma_V^2\tilde{\sigma}_z^2/\mathrm{E}\,\bar{z}^2).$$

When a temporal aggregate of x_t is available, the variance of the Dagenais estimator in (11) is

$$\text{(47)} \quad \operatorname{Var}(\sqrt{T}\,\hat{\beta}_{ad}) = p + p^2\sigma^2\beta^2\tilde{\sigma}_z^4\sigma_V^2\tilde{\sigma}^{-4}(\mathrm{E}\,\bar{z}^2)^{-1},$$

where $p = \{\mathrm{E}\,\bar{x}^2/4\sigma^2 + \alpha^2\tilde{\sigma}_z^2/4\tilde{\sigma}^2\}^{-1}$ is the variance of $\hat{\beta}_{ad}$ given that α is known and $\tilde{\sigma}_z^2 = \mathrm{E}(z_t - z_{t-1})^2$. The asymptotic variance of the GLS estimator with optimal weights, $\hat{\beta}_{ag}$, is

$$\text{(48)} \quad \operatorname{Var}(\sqrt{T}\,\hat{\beta}_{ag}) = \{\alpha^2\tilde{\sigma}_z^2/4\tilde{\sigma}^2 + \mathrm{E}\,\bar{x}^2/4\sigma^2 - \alpha^2\beta^2\sigma_V^2\tilde{\sigma}^4/(4\tilde{\sigma}^4\mathrm{E}\,\bar{z}^2 + 4\beta^2\sigma_V^2\sigma^2\tilde{\sigma}_z^2)\}^{-1}.$$

The variance of the OLS estimator applied to the periods for which all variables are observed, $\hat{\beta}_{ac}$, is

$$\text{(49)} \quad \operatorname{Var}(\sqrt{T}\,\hat{\beta}_{ac}) = 4\sigma^2/\mathrm{E}\,\bar{x}^2.$$

It differs from the variance of the OLS estimator of β when no data are missing

(50) $$\mathrm{Var}(\sqrt{T}\,\hat{\beta}_{OLS}) = \sigma^2/\sigma_x^2.$$

with $\sigma_x^2 = \alpha^2\,\sigma_z^2 + \sigma_V^2$.
The asymptotic variance of the ML estimator $\hat{\beta}_{aML}$, is

(51) $$\mathrm{Var}(\sqrt{T}\,\hat{\beta}_{aML}) = \{\,\mathrm{E}\,\overline{x}^2/4\,\sigma^2 + [\alpha^2\,\tilde{\sigma}_z^2/4\,\tilde{\sigma}^2] \times [1 - \beta^2\,\tilde{\sigma}_z^2\,\sigma_V^2/(\beta^2\,\tilde{\sigma}_z^2\,\sigma_V^2 + \mathrm{E}\,\overline{z}^2\,\tilde{\sigma}^2)] + [\beta^2\,\sigma_V^4/\tilde{\sigma}^4]\,[1 - (\sigma^4 + \beta^4\,\sigma_V^4)/(\sigma^4 + \beta^4\,\sigma_V^4 + \tilde{\sigma}^4)]\,\}^{-1}.$$

Finally, we indicate briefly how table 4 was derived. If $\hat{x}_t$ defined in (36) is substituted in (34) the resulting error term has covariance matrix

(52) $$\Omega = \Omega_1 + \mathrm{W}\,\Omega_2\,\mathrm{W}',$$

where Ω_2 is the covariance matrix of the estimates $\hat{\gamma}$ and $\hat{\alpha}$, W is a $(\mathrm{T} \times 2)$ matrix which contains β times the derivatives of $\hat{x}_t$ with respect to $\hat{\alpha}$ and $\hat{\gamma}$ in the first and second column respectively and Ω_1 is a diagonal matrix with diagonal element σ^2 in case of an observed x_t and $\sigma^2 + \beta^2\sigma_V^2(1+\gamma^2)^{-1}$ if x_t is not observed. Again equation (15) can be used to get the inverse of Ω.

In order to derive the variance of the ML estimator for the dynamic regression model in (34) and (35) we write the model in recursive form as

(53) $$\left\{\begin{aligned} y_t &= \beta\,x_t + \varepsilon_t \\ y_{t-1} &= \beta\,(1+\gamma^2)^{-1}\,\{\gamma\,x_t + \gamma\,x_{t-2} - \alpha\gamma\,z_t + \alpha\,z_{t-1}\} + \varepsilon_{t-1} \\ &\quad - \beta\,(1+\gamma^2)^{-1}\,(\gamma\,v_t - v_{t-1}) \\ x_t &= \gamma^2\,x_{t-2} + \alpha\,z_t + \alpha\gamma\,z_{t-1} + v_t + \gamma\,v_{t-1} \end{aligned}\right.$$

for $t \in \mathrm{T}_2$. Notice that the disturbances in (53) are independent and orthogonal to the explanatory variables in the corresponding equation. The log-likelihood function can therefore be obtained in a straightforward manner, as well as the associated information matrix.

APPENDIX B

Proof of the Theorem

Using assumptions (ii) and (iii) one verifies that

$$\sqrt{N}(\tilde{\beta}_{IV}-\beta) \overset{d}{\sim} N(0, Q^{-1}(D_{11}+D_{22})Q'^{-1}),$$

where D_{11} and D_{22} are the upper-left and the lower-right blocks of D respectively. Furthermore, the asymptotic orthogonality of Z and ε and assumption (iii) imply that $\text{plim}\, N^{-1}Z'(\tilde{w}-\hat{w})=\text{plim}\, N^{-1}Z'(\hat{X}-\tilde{X})\beta=0$, for all β, so that $\text{plim}\, N^{-1}Z'\tilde{X}=\text{plim}\, N^{-1}Z'\hat{X}=Q$. Using this result, one obtains that

$$\begin{aligned} Q^{-1}(D_{11}+D_{22})Q'^{-1} = {} & \text{plim}\, N^{-1}(Q^{-1}Z'-V^{-1}\hat{X}'\hat{\Sigma}^{-1}) \\ & \times\hat{\Sigma}^{-1}(ZQ'^{-1}-\hat{\Sigma}^{-1}\hat{X}V^{-1}) \\ & +\text{plim}\, N^{-1}Q^{-1}Z'\hat{X}V^{-1} \\ & +\text{plim}\, N^{-1}V^{-1}\hat{X}'ZQ'^{-1}-V^{-1} \\ & +Q^{-1}D_{22}Q'^{-1}\geqslant V^{-1} \end{aligned}$$

which proves the result.

APPENDIX C

Proof of the Relative Efficiency of the Proxy Variables Estimator Based on the Larger Information Set for a Special Case

Assume that (4) and (24) both hold, which read in matrix notation as

(54) $$x = Z_1 \alpha_1 + Z_2 \alpha_2 + v$$

and

(55) $$x = Z_1 \eta + v^*$$

respectively. Evidently

(56) $$v^* = Z_1 (\alpha_1 - \eta) + Z_2 \alpha_2 + v$$

and

(57) $$\eta = \alpha_1 + \text{plim}\,(Z_1' Z_1)^{-1} Z_1' Z_2 \alpha_2$$

because $E[x \mid Z_1] = Z_1 \eta$. Using (56) and (57) we find that

(58) $$N^{-1/2}(Z_1' v^* - Z_1' v) = N^{-1/2}(Z_1' Z_1 (\alpha_1 - \eta) + Z_1' Z_2 \alpha_2)$$

converges in probability to zero.
It is not difficult to check using the theorem in section 3 that if $\hat{X} = \hat{X}^{(1)}$, $\tilde{X} = \tilde{X}^{(2)}$, $\hat{\beta}_{GLS} = \hat{\beta}_g^{(1)}$ and $\hat{\beta}_{IV} = \hat{\beta}_g^{(2)}$, a sufficient condition for relative efficiency of $\hat{\beta}_g^{(1)}$ is that $N^{-1/2} Z'(\tilde{w} - \hat{w}) \xrightarrow{p} 0$ which is satisfied when

(59) $$N^{-1/2}\{Z_{11}'(V_I^* - V_I)\beta + Z_{11}' Z_I (Z_c' Z_c)^{-1} Z_c' V_c \beta - Z_{11}' Z_{11} (Z_{1c}' Z_{1c})^{-1} Z_{1c}' V_c^* \beta\}$$

converges in probability to zero. The subscripts c and I refer to complete and incomplete observations as before. Condition (59) is satisfied because of (58).

• References

AFIFI, A. A. and ELASHOFF, R. W. (1966). — "Missing Observations in Multivariate Statistics, I. Review of the Literature", *Journal of the American Statistical Association*, 61, pp. 595-604.

ANDERSON, T. W. (1957). — "Maximum Likelihood Estimates for a Multivariate Normal Distribution when Some Observations are Missing", *Journal of the American Statistical Association*, 67, pp. 628-632.

BEALE, E. M. L. and LITTLE, R. J. A. (1975). — "Missing Values in Multivariate Analysis", *Journal of the Royal Statistical Society*, B, 37, pp. 129-146.

Boot, J. C. G., Feibes, W. and Lisman, J. H. C. (1967). – "Further Methods of Derivation of Quarterly Figures from Annual Data", *Applied Statistics*, 16, pp. 65-75.

Chow, G. C. and Lin, A. (1971). – "Best Linear Unbiased Interpolation and Extrapolation of Time Series by Related Series", *The Review of Economics and Statistics*, 53, pp. 372-375.

Chow, G. C. and Lin A. (1976). – "Best Linear Unbiased Estimation of Missing Observations in an Economic Time Series", *Journal of the American Statistical Association*, 71, pp. 719-721.

Conniffe, D. (1983 *a*). – "Small-sample Properties of Estimators of Regression Coefficients Given a Common Pattern of Missing Data", *Review of Economic Studies*, 50, pp. 111-120.

Conniffe, D. (1983 *b*). – "Comments on the Weighted Regression Approach to Missing Values", *The Economic and Social Review*, 14, pp. 259-272.

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. and Rubin, D. B. (1977). – "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society*, B, 39, pp. 1-38.

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

Gouriéroux, C., Monfort, A. and Trognon, A. (1984). – "Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, 54, pp. 681-700.

Griliches, Z. (1986). – "Economic Data Issues", in Z. Griliches and M. D. Intriligator, eds., *Handbook of Econometrics*, North-Holland Publ. Co, Amsterdam, pp. 1466-1514.

Hausman, J. (1978). – "Specification Tests in Econometrics", *Econometrica*, 46, pp. 1251-1272.

Litterman, R. B. (1983). – "A Random Walk, Markov Model for the Distribution of Time Series", *Journal of Business and Economic Statistics*, 1, pp. 169-173.

Little, R. J. A. (1982). – "Models for Non-Reponse in Sample Surveys", *Journal of the American Statistical Association*, 77, pp. 237-250.

Little, R. J. A. and Rubin, D. B. (1987). – *Statistical Analysis with Missing Data*, John Wiley & Sons, New York.

Murphy, K. M. and Topel, R. H. (1985). – "Estimation and Inference in Two-Step Econometric Models", *Journal of Business and Economic Statistics*, 3, pp. 370-379.

Nijman, T. E. and Palm, F. C. (1985). – "Séries temporelles incomplètes en modélisation macroéconomique", *Cahiers du Séminaire d'Économétrie*, 27. pp. 141-168.

Nijman, T. E. and Palm, F. C. (1986). – "The Construction and Use of Approximations for Missing Quarterly Observations: a Model-Based Approach", *Journal of Business and Economic Statistics*, 4, pp. 47-58.

Pagan, A. (1984). – "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review*, 25, pp. 221-248.

Palm, F. C. and Nijman, T. E. (1982). – "Linear Regression Using Both Temporally Aggregated and Temporally Disaggregated Data", *Journal of Econometrics*, 19, pp. 333-343.

Palm, F. C. and Nijman, T. E. (1984). – "Missing Observations in the Dynamic Regression Model", *Econometrica*, 52, pp. 1415-1435.

Rubin, D. B. (1976). – "Inference and Missing Data", *Biometrika*, 63, pp. 581-592.

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