A Method of Simulated Scores for Imputation of Continuous Variables Missing At Random*

GIORGIO CALZOLARI[a] Università di Firenze

LAURA NERI[b] Università di Siena

[a]Università di Firenze, Dipartimento di Statistica "G.Parenti", Viale Morgagni 59, I-50134 Firenze.

[b]Università di Siena, Dipartimento di Metodi Quantitativi, Piazza San Francesco 8, I-53100 Siena.

Preliminary version, not to be cited. January 2002

ABSTRACT

Given a set of continuous variables with missing data, we prove in this paper that the iterative application of a "least-squares estimation/multivariate normal imputation" procedure produces an efficient parameters estimator and is therefore an optimal parametric technique for imputation of missing data. There are two main assumptions behind our proof: (1) data are missing at random (MAR); (2) the data generating process is a multivariate normal linear regression. Disentangling the problem of convergence of the iterative procedure, we show that the estimator is a "method of simulated scores" (a particular case of McFadden's "method of simulated moments"), thus equivalent to maximum likelihood if the number of replications is conveniently large.

*We are grateful to Monica Billio, J. Lepkowski, Fabrizia Mealli, Donald B. Rubin and Federico M. Stefanini for suggestions and comments, but retain full responsibility for the contents of this paper.
We gratefully acknowledge financial support from CNR and MURST-MIUR through projects "Stochastic models and simulation methods for dependent data" and "Employment and unemployment in Italy: measurement issues and behavioural analyses".

1 Introduction

Missing data are a serious problem in almost all areas of empirical research. Sample surveys in economic, social and behavioral science frequently suffer from missing data due to nonresponse as well as biomedical applications involving missing data in surveys and experiments.

There are three major problems created by missing data. First, if the nonrespondents are systematically different from the respondents and we do not take into account the difference, analysis may be biased. Second, missing data imply loss of information, so estimates will be less efficient than planned. Finally, standard statistical methods are designed for complete data sets and missing data make the study more complicated.

Survey datasets often consist of large number of variables. Let be Y the $n \times p$ matrix of complete data, if missing values regard one and only one variable of the dataset, we deal with univariate nonresponse; otherwise, when missing values regard more than one variable of the dataset, we deal with multivariate nonresponse. Let R denote a $n \times p$ matrix of indicator variables whose elements are zero or ones depending on whether the corresponding elements of Y are missing or not; the matrix R describes the *pattern of missingness*. To be more specific, a missingness pattern is a unique combination of response status (observed or missing) for Y.

The pattern of missing data is a very important element if our aim is to deal with missing values, in some way. Such a pattern frequently depends on the variables being considered; in fact, different variables often have different rates of missingness (for example, income usually has higher nonresponse rate than gender).

A different problem is whether Y and R are associated, that is whether missingness depends on the values of the survey variables. For example, if income is a variable with missing data, we wonder if missingness depends on income or other variables in the dataset. Such a question regards the missing data mechanism and is the crucial issue in determining the extent of nonresponse bias.

Besides teoretical problems like bias, data analysts also meet a big practical problem when dealing with datasets affected by missing values: tools for effectively dealing with them are not readily available. The simpler and standard treatment of missing data in statistical-packages is the complete case analysis (Cc), where cases with any missing values are simply discarded, forcing the incomplete dataset into a a rectangular complete-data format. When the incomplete cases are a small fraction of all cases (say, five percent or less) then Cc method can be a reasonable solution to the missing data problem. However when missing values occur on more than one variable, the incomplete cases are often a substantial portion of the dataset and the Cc strategy may cause a large loss of information. Moreover, omitting a substantial portion of data from the analysis will tend to introduce bias, to the extent that the unobserved cases differ systematically from the completely observed cases.

Another way to estimate the unknown parameter θ of the probability distribution $P(Y|\theta)$, when Y is not completely observed, is to compute the maximum likelihood estimates (ML) of a model for the joint distribution of Y. Many papers refer to this approach, particularly Anderson (1957) introduced the idea of factoring the likelihood to obtain explicit ML solution for monotone missing data pattern and Gourieroux and Montfort (1981) applied Anderson's method to the regression with missing covariates. However, ML for a general pattern of missing data requires iterative methods (this topic is dealt with in Section 3.1). The general technique for finding ML estimates for parametric models when data are not completely observed is the EM algorithm (Dempster, Laird, Rubin, 1977). EM spawned a revolution in the analysis of incomplete data making possible to compute efficient parameter estimates and thus obviating the need for ad-hoc methods like CC in many statistical problems. However, EM provides only point estimates of the unknown parameters so, even if they are efficient, they are not useful to obtain valid inferential conclusion unless there is some measure of uncertainty associated with them. Such a disadvantage addresses the discussion towards the main applied approach to the general problem of obtaining valid inferences when facing missing data: the *Multiple Imputation* technique (MI).

The idea behind MI was explicitly proposed in Rubin (1978) and a decade later the basic reference textbook was published (Rubin, 1987). MI is a technique in which each missing values is replaced by $m > 1$ simulated values. After the multiple imputations are created, m plausible versions of the complete data exists, each one is analyzed by standard statistical methods. The results of the m analyses are then combined to produce a single inferential statement that takes account of the uncertainty due to missing data. However the task of generating multiple imputations is often a hard task, except in some simple cases such as datasets with only one variable affected by missing values or very special patterns of missingness; the main difficulty is to find a solution for imputing a general pattern of missing data preserving the original association structure of the data. The currently avaliable solution to this problem is to create multiple imputations specifying one "encompassing multivariate model" for the entire data set (at least conditional on completely observed variables), and then using fully principled likelihood/Bayesian techniques for analysis under that model. This generates a posterior distribution for the parameters of the model and a posterior predictive distribution for the missing values (given the model specifications and the observed data). The primary example of such approach is the Schafer's freeware (Schafer: www) based on Schafer (1997), which involves iterative Markov Chain Monte Carlo (MCMC) computations; the other similar concept software is the "IVE-*ware*" of Raghunathan (Raghunathan:www) to which we will refer thorough the paper.

Rubin (2000) explains the advantage and disadvantage of the previous mentioned approach; among the disadvantages, we point out on the fact that iterative version of software for creating multiple imputations are not always yet ready for real applications by the typical user dealing with missing data; it often needs experts to face with potentially misleading "non convergent" MCMC and any other possible difficulties.

In this paper we introduce a method, feasible for data analysts, for creating multiple imputations when we deal with a general missing data pattern of continuous variables.

We assume that:

1) missing data are *Missing At Random* (MAR) and the missing data mechanism is *ignorable*;

2) the data generating process is a multivariate normal linear regression;

3) all the covariates of these parametric regressions are fully observed.

We obtain the m imputed datasets by repeating m times (each time till convergence) the iterative "least-squares estimation/multivariate normal imputation" procedure. The properties of the defined estimator are analyzed showing that, at convergence, the estimator obtained using the iterative procedure is a *simulated scores estimator*.

In Section 2 we define the notation and the background for dealing with missing data problem.

In Section 3 we define the model and derive the maximum likelihood estimator, evidencing the practical difficulties involved in complex missingness patterns.

Section 4 defines the imputation function, given the parameters values.

Section 5 regards the "feasible" imputation technique, that is an iterative procedure that imputes missing values, given previously estimated parameters, and estimates parameters, given previously imputed values. It is first presented with a "reduced form" approach, that solves the problem in principle, but still encounters practical difficulties in case of complex missingness patterns; then it is presented with a "structural form" approach, which presents no difficulty in the practical application to complex cases.

In Section 6 we first prove that the two approaches of the previous section are "algebraically equal". This enables us at using in the subsequent proofs the reduced form approach, which is analytically more tractable, still keeping the structural form approach as the method to be used in practice. Then we derive the asymptotic properties of the parameters estimator. We show that it is a "method of simulated scores" estimator (MSS, see Hajivassiliou and McFadden, 1990, a particular case of McFadden's (1989) method of simulated moments MSM), we discuss its aymptotic variance-covariance matrix, and show how its efficiency can be made arbitrarily close to that of maximum likelihood.

Finally the problem of multiple imputation is again considered in Section 7. For precision's sake, we inform the reader that thorough the paper we always refer to a "single" imputation; this makes explanations easier, without loss of generality.

Appendices 1, 2, 3 give more analytical details on methods and proofs.

Y_1	Y_2	...	Y_p
			?
	?		
		?	?
	?	?	

Table 1: Dataset with missing values.

2 Statement of the problem of missing data

A schematic representation of an incomplete dataset is shown in Table 1, where the n rows represent the observational units and the p columns represent variables recorded for those units; question marks identify missing values; they can occur anywhere, in any *pattern*.

Let Y denote the $n \times p$ matrix of complete data and let y_i' denote the $i-$th row of Y $(i = 1, ..., n)$, assuming the rows as independent, identically distributed (i.i.d.) the probability density of the complete data may be written

$$P(Y|\theta) = \prod_{i=1}^{n} f(y_i|\theta) \tag{2.1}$$

where f is the probability density function for a single row and θ is the vector of unknown parameters.

Formally, let Y_{obs} denote the observed portion of Y and let Y_{mis} denote the missing portion, so that $Y = (Y_{obs}, Y_{mis})$.

2.1 The missing data mechanism

Let us now consider the missing data mechanism; our hypothesis, thorough this paper, is that the missing data are missing at random (MAR). We start explaining the concept related to the missing data mechanism adopting an informal way. If missing data on income are due to the fact that only a random subsample of the entire sample had to answer to question regarding income, we are sure that missingness is unrelated to the survey variables; in such a case the missing values are Missing Completely At Random (MCAR). In many practical instances the MCAR assumption is not realistic, so the important issue concerns whether differences in characteristics of nonrespondents and respondents can be captured in terms of observed values, the so called Missing At Random (MAR) case. For example, if income is missing and age is fully observed, missing data are MAR if missingness on income depends only on age; in such a case the nonresponse bias can be controlled by an analysis that stratifies on age or adjusts for age in some way; however, if for subjects of a given age missingness of income depends on the income values, then the missing data are not MAR since missingness depends on values of a variable that is sometimes missing.

Let now consider the formal definition in terms of probability model of the missingness (Rubin, 1976), introducing the MAR concept in terms of probability. Let R be the $n \times p$ matrix of indicator variables, and let ξ be the unknown parameters of the probability model for R, $P(R|Y, \xi)$. The missing data are MAR if the distribution $P(R|Y, \xi)$ does not depend on Y_{mis}that is:

$$P(R|Y_{obs}, Y_{mis}, \xi) = P(R|Y_{obs}, \xi). \tag{2.2}$$

A special case of MAR mechanism is the MCAR (Missing Completely at Random), in this case the missing data values are a simple random sample of all the data values so:

$$P(R|Y_{obs}, Y_{mis}, \xi) = P(R|\xi). \tag{2.3}$$

Obviously MAR is less restrictive than MCAR because it requires only that the missing values behave like a random sample of all the values within a subclass defined by observed data.

2.2 Maximum likelihood assuming ignorable nonresponse

The method of maximum likelihood for incomplete data, in general, requires specification of a model for the distribution of Y and R, so that

$$P(Y, R|\theta, \xi) = P(Y|\theta)P(R|Y, \xi). \tag{2.4}$$

If Y were completely observed, the likelihood of θ and ξ would be as in equation (2.4); in this case the unknown elements of the likelihood would be θ and ξ, because Y would be fixed at their observed values (and all elements of R are $= 1$); so the maximum likelihood estimates would be obtained maximizing (2.4) with respect to θ and ξ.

In presence of missing values in the Y matrix, the likelihood of θ and ξ is proportional to the marginal density of the observed part in Y and R, treated as a function of θ and ξ, with Y_{obs} and R fixed at their observed values. This likelihood is obtained formally by integrating the joint density of Y and R with respect to the missing part Y_{mis}, that is:

$$L(\theta, \xi|Y_{obs}, R) = \int P\left(Y_{obs}, Y_{mis}|\theta\right) P\left(R|Y_{obs}, Y_{mis}, \xi\right) dY_{mis}. \tag{2.5}$$

Maximum likelihood estimates of the unknown parameters θ and ξ are obtained by maximizing the function (2.5) with respect to θ and ξ.

Maximum likelihood estimation for θ is often considered without explicitly including a model for the missing data mechanism. The likelihood *ignoring the missing data mechanism* (Rubin 1976) is based on the marginal density of Y_{obs}, ignoring the contribution of

R to (2.5):

$$L(\theta|Y_{obs}) = \int P\left(Y_{obs}, Y_{mis}|\theta\right) dY_{mis}. \tag{2.6}$$

Maximum likelihood ignoring the missing data mechanism means maximizing (2.6) rather than the full likelihood (2.5). The missing data mechanism is called *ignorable* if inference about θ based on the likelihood (2.6) is equivalent to the inference about θ based on the likelihood (2.5).

Formally, Rubin (1987, pag.53) showed that the missing data mechanism is ignorable if *(i)* θ and ξ are distinct parameters, that is, the joint parameter space of (θ, ξ) is the Cartesian cross-product of the individual parameter space for θ and ξ; *(ii)* the missing data are MAR.

Once ML estimates of the parameters have been obtained, asymptotic standard errors, tests of hypotheses and confidence intervals can be derived by applying standard methods; they have appropriate statistical properties provided large sample size. However computational difficulties are frequent when the parameters are high-dimensional; even worse, sometimes it could be very difficult to write the likelihood function; this is particularly the case of multivariate pattern of missing data. In conclusion, ML estimates for general pattern of missing data requires long and boring application of iterative methods (Schafer, 1997; Chapter. 2, pag. 16).

3 Multivariate normal model

The multivariate normal distribution is a common baseline model for many data analysis. In presence of missing values, it may be reasonable to use the multivariate normal model to create imputations (Rubin, 1987), infact, even if the statistical analysis to be ultimately performed on the data does not assume normality, it may be reasonable to make this assumption for imputation purpose (often the assumption may be made more tenable by applying suitable trasformation to one or more variables). For these reasons it is convenient to consider the following model. Let be Y the $n \times p$ matrix and X a $n \times k$ given matrix of complete data, assuming for Y the following multivariate normal distribution

$$y_i \sim N(\Pi x_i, \Sigma), \qquad i = 1...n$$

where y_i is the $i-$column of the Y' matrix, x_i is the $i-$column of the X' matrix, Πx_i is the mean vector of the multivariate normal distribution, Π denotes the unknown matrix of coefficients $(p \times k)$ and Σ the unknown $(p \times p)$ covariance matrix.

3.1 Maximum likelihood estimation

Assuming missing values distributed according to a general pattern, on any or all p variables, we can group the rows of the matrix according to their missingness pattern as blocks. Having a p-variate distribution we can have 2^p possible missingness pattern; if p is large, it is plausible that not all possible patterns are represented in the sample. Index the unique missingness pattern that actually appear in the sample by δ, where $\delta = 1, 2, ..., \Delta$, and let $D(\delta)$ denote the subset of the rows $i = 1, 2, ..., n$ that exhibit pattern δ.

If the missing data mechanism is assumed ignorable, the likelihood of the parameters $\theta = (\Pi, \Sigma)$ is:

$$P(\theta|Y_{obs}) = \prod_{\delta=1}^{\Delta} \prod_{i \in D(\delta)} |\Sigma_\delta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_i - \Pi_\delta x_i)' \Sigma_\delta^{-1} (y_i - \Pi_\delta x_i)\right\} \tag{3.7}$$

where y_i' denotes the observed part of row i of the data matrix, and $\Pi_\delta x_i$and Σ_δ denote the subvector of the mean vector and the submatrix of the covariance matrix Σ, respectively, that pertain to the variables observed in pattern δ. If any rows of the data matrix are completely missing, those rows drop out of the likelihood function $P(\theta|Y_{obs})$; in fact under the ignorability assumption these rows do not contribute to the inference and can be ignored.

For monotone missing data patterns, explicit estimates of (Π, Σ) that maximize (3.7) can be derived (Rubin, 1974) using Anderson's (1957) method of *factored likelihoods*. The parameters (Π, Σ) are transformed in such a way that the likelihood factorizes into distinct factors corresponding to complete data problems (Little, Rubin, 1987).

For non monotone missing data patterns, maximization of (3.7) requires an iterative alghoritm, because we are not able to write an explicit closed form for the ML estimator of the means, variances and covariances.

To understand the problem in case of non monotone missing data patterns, we start from a simple case: instead of a p-variate normal distribution, we consider a bivariate normal distribution $Y = (Y_1, Y_2)$, with a single column matrix X of complete data. This is one of the smallest possible cases (in terms of dimensions), and we shall use it often in this paper, as it helps to simplify the study without substantial loss of generality. We specify the data generating process as follows:

$$Y = X\Pi' + E = X\Pi' + U\Sigma^{\frac{1}{2}} \tag{3.8}$$

where X is a given matrix X $(n \times 1)$ completely observed, $\Pi = [\Pi_1, \Pi_2]'$ denotes the unknown parameter matrix (2×1), $U = [u_1, u_2]$ is a $(n \times 2)$ random matrix whose rows have independent bivariate standard normal distribution, $\Sigma^{\frac{1}{2}}$ is a (2×2) matrix such that $\Sigma^{\frac{1}{2}\prime}\Sigma^{\frac{1}{2}} = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$ (for instance, Cholesky decomposition) and the rows of $E = U\Sigma^{\frac{1}{2}} = [e_1, e_2]$ have bivariate normal distribution with 0 mean and Σ variance-covariance matrix.

Let us consider the likelihood function $L(\Pi, \Sigma|Y_1, Y_2)$ in the case of completely observed

	X	Y_1	Y_2	
	X_A	Y_{A1}	Y_{A2}	$A-block$
$[X,\ Y_1,\ Y_2] =$	X_B	Y_{B1}	?	$B-block$
	X_C	?	Y_{C2}	$C-block$
	X_D	?	?	$D-block$

Table 2: Bivariate dataset with missing values.

data:

$$L(\Pi,\Sigma|Y_1,Y_2,X) \propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\Pi x_i)'\,\Sigma^{-1}\,(y_i-\Pi x_i)\right\}$$

and the log-likelihood function:

$$\begin{aligned}\log L(\Pi,\Sigma|Y_1,Y_2,X) &= -\frac{n}{2}\log|\Sigma| - \frac{1}{2}Tr\Sigma^{-1}\,(Y-X\Pi')'\,(Y-X\Pi') = \\ & -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(y_i-\Pi x_i)'\,\Sigma^{-1}\,(y_i-\Pi x_i)\,.\end{aligned}$$

In order to obtain the ML estimates of the unknown parameters, we maximize the log-likelihood function by differentiating the log-likelihood with respect to the unknown parameters using the rule of matrix differentiation (e.g. Amemiya, 1985, Appendix 1, Theorem 21; a slight simplification in the formulas is obtained if differentiation is performed with respect to Σ^{-1} rather than Σ), and equating derivatives to 0 as follows:

$$\Sigma^{-1}\,(Y-X\Pi')\,X = 0 \tag{3.9}$$

$$\frac{n}{2}\Sigma - \frac{1}{2}\,(Y-X\Pi')'\,(Y-X\Pi') = 0. \tag{3.10}$$

We suppose now that missing data affect Y_1 and Y_2 according to a general pattern, while X is completely observed. Grouping the rows of the matrix according to their missingness pattern as blocks, we have $2^2 = 4$ possible blocks. We indicate with A the block where Y_1 and Y_2 are both observed, with B the block where Y_1 is observed and Y_2 is missing, with C the block where Y_1 is missing and Y_2 is observed and finally with D the block where Y_1 and Y_2 are both missing as in Table 2.

If the missing data mechanism is assumed ignorable, the log-likelihood function is the following:

$$\begin{aligned}\log L(\Pi,\Sigma|Y_{obs},X) &= -\frac{n_A}{2}\log|\Sigma| - \frac{1}{2}\sum_{i\in A}(y_i-\Pi x_i)'\,\Sigma^{-1}\,(y_i-\Pi x_i) \\ & -\frac{n_B}{2}\log\sigma_{11} - \frac{1}{2\sigma_{11}}\sum_{i\in B}(y_{i1}-\Pi_1 x_i)^2 \\ & -\frac{n_C}{2}\log\sigma_{22} - \frac{1}{2\sigma_{22}}\sum_{i\in C}(y_{i2}-\Pi_2 x_i)^2\,, \end{aligned} \tag{3.11}$$

and the ML estimator of (Π, Σ) can be derived maximizing (3.11), that is by solving the normal equations

$$\frac{\partial \log L(\Pi, \Sigma | Y_{obs}, X)}{\partial \Pi} = 0 \tag{3.12}$$

and

$$\frac{\partial \log L(\Pi, \Sigma | Y_{obs}, X)}{\partial \Sigma^{-1}} = 0. \tag{3.13}$$

Expliciting the left hand side of (3.12) as a block matrix (the blocks are identified by square brackets), we obtain

$$\frac{\partial \log L(\Pi, \Sigma | Y_{obs}, X)}{\partial \Pi}$$

$$= \Sigma^{-1} \begin{pmatrix} [(Y_{A1} - X_A \Pi_1)]' & [(Y_{B1} - X_B \Pi_1)]' & \left[\frac{\sigma_{12}}{\sigma_{22}} (Y_{C2} - X_C \Pi_2)\right]' \\ [(Y_{A2} - X_A \Pi_2)]' & \left[\frac{\sigma_{12}}{\sigma_{11}} (Y_{B1} - X_B \Pi_1)\right]' & [(Y_{C1} - X_C \Pi_2)]' \end{pmatrix} X \tag{3.14}$$

and expliciting the left hand side of (3.13), we obtain:

$$\begin{aligned} \frac{\partial \log L(\Pi, \Sigma | Y_{obs}, X)}{\partial \Sigma^{-1}} &= \frac{n_A}{2}\Sigma + \frac{n_B}{2} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \frac{\sigma_{12}^2}{\sigma_{11}} \end{pmatrix} + \frac{n_C}{2} \begin{pmatrix} \frac{\sigma_{12}^2}{\sigma_{22}} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \\ & -\frac{1}{2} \sum_{i=A} (y_i - \Pi x_i)(y_i - \Pi x_i)' \\ & -\frac{1}{2} \begin{pmatrix} 1 & \frac{\sigma_{12}}{\sigma_{11}} \\ \frac{\sigma_{12}}{\sigma_{11}} & \frac{\sigma_{12}^2}{\sigma_{11}^2} \end{pmatrix} \sum_{i=B} (y_{i1} - \Pi_1 x_i)^2 \\ & -\frac{1}{2} \begin{pmatrix} \frac{\sigma_{12}^2}{\sigma_{22}^2} & \frac{\sigma_{12}}{\sigma_{22}} \\ \frac{\sigma_{12}}{\sigma_{22}} & 1 \end{pmatrix} \sum_{i=C} (y_{i2} - \Pi_2 x_i)^2 , \end{aligned} \tag{3.15}$$

thus the score is written in explicit form. Even if we are dealing with a very simple case, we are not able to obtain explicit estimates of (Π, Σ). Maximization of (3.11) requires the application of an iterative algorithm.

One can imagine, from the equations above, how complex the general case would be. The expression of the score (eq. 3.14 and 3.15) is already complex enough even if the variables y are only two, and so we have been able to group our data into the few blocks A, B, C (plus D that does not contribute to the likelihood, as y is not observed in block D). Should the number of variables y be $p > 2$, we should take into account explicitly up to 2^p blocks, each of which contributing to the likelihood in a different way, so that the estimation procedure would be practically intractable.

4 The imputation function

Imputation replaces missing values by suitable estimates of the values, so that complete-data methods can be applied to the filled-in data. For this reason the method is very attractive to practitioners because the obtained complete data set can be handled using standard methods.

In order to complete the data set we define an imputation function. This function is defined "as if" the parameters of the data generating process were known. Of course, in any real problem we do not know the true parameter values, so the next section will develop a feasible procedure to estimate the unknown parameters.

Considering the data generating process introduced in Sec. 3, we first assume $\Pi = [\Pi_1, \Pi_2]'$ and $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$ as known. The imputation function builds a new data matrix, expressed as a block matrix:

$$\widetilde{Y} = b(Y_{obs}) = \begin{bmatrix} Y_{A1} & Y_{A2} \\ Y_{B1} & \widetilde{Y}_{B2} \\ \widetilde{Y}_{C1} & Y_{C2} \\ \widetilde{Y}_{D1} & \widetilde{Y}_{D2} \end{bmatrix}, \tag{4.16}$$

such that, indicating $u_i = (u_{i1}, u_{i2})'$ as random variates from a bivariate standard normal distribution, the completed data are defined as follows:

$$\begin{cases} y_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}}\; u_{i1} \\ y_{i2} = \Pi_2 x_i + \frac{\sigma_{12}}{\sigma_{11}}(y_{i1} - \Pi_1 x_i) + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\; u_{i2} \end{cases} , i \in A \tag{4.17}$$

$$\begin{cases} y_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}}\; u_{i1} \\ \tilde{y}_{i2} = \Pi_2 x_i + \frac{\sigma_{12}}{\sigma_{11}}(y_{i1} - \Pi_1 x_i) + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\; \tilde{u}_{i2} \end{cases} , i \in B \tag{4.18}$$

$$\begin{cases} y_{i2} = \Pi_2 x_i + \sqrt{\sigma_{22}}\; u_{i2} \\ \tilde{y}_{i1} = \Pi_1 x_i + \frac{\sigma_{12}}{\sigma_{22}}(y_{i2} - \Pi_2 x_i) + \sqrt{\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}}\; \tilde{u}_{i1} \end{cases} , i \in C \tag{4.19}$$

$$\begin{cases} \tilde{y}_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}}\; \tilde{u}_{i1} \\ \tilde{y}_{i2} = \Pi_2 x_i + \frac{\sigma_{12}}{\sigma_{11}}\sqrt{\sigma_{11}}\; \tilde{u}_{i1} + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\; \tilde{u}_{i2} \end{cases} , i \in D \tag{4.20}$$

The imputation function defined is rather simple. When the variables are both observed (A-block) they are obviously left unchanged, and equation (4.17) writes their explicit expression as given by the data generating process (3.8). It must be noticed that the representation in equation (4.17) has been used to make easier the comparison with the other equations. It includes explicitly the error terms u_{i1} and u_{i2}, which are independent standard normal variables introduced by the data generating process (3.8). Thus the error term of y_{i1} is $\sqrt{\sigma_{11}}$ multiplied by u_{i1}. The expression of y_{i2} includes the conditional mean (given y_{i1}), which is $\Pi_2 x_i + (\sigma_{12}/\sigma_{11})(y_{i1} - \Pi_1 x_i)$ and the error term which is u_{i2} multiplied by the square root of the conditional variance $\sqrt{\sigma_{22} - \sigma_{12}^2/\sigma_{11}}$.

When one variable is observed and the other is missing (B and C-blocks), the observed variable is obviously left unchanged (and is represented according to the data generating process 3.8), while the other is replaced by its conditional mean (given the observed one) plus a zero mean pseudo-random error with the appropriate conditional variance (equations 4.18 and 4.19, where $\tilde{u}_{i1}$ and $\tilde{u}_{i2}$ are independent standard normal deviates produced by a pseudo-random number generator). Finally, when both variables are missing (D-block), imputation is performed using the unconditional mean plus a bivariate zero mean pseudo random error with covariance matrix Σ. Equivalently, as explicitly written in equation (4.20), we first impute the value $\tilde{y}_{i1}$, with the appropriate unconditional mean and variance, then impute $\tilde{y}_{i2}$ with the appropriate conditional mean and variance (given $\tilde{y}_{i1}$).

Summarizing, the equations above define the conditional distribution of the variables with missing data, given the observed data. Of course, this imputation technique is infeasible, since the parameters Π and Σ are unknown.

5 Feasible imputation methods

Let us consider the data generating process introduced in Sec. 3 , when missing data are distribuited according to a general pattern. Introducing the imputation function (Sec. 4) and assuming knowledge of the parameters $\theta = (\Pi, \Sigma)$, the imputation procedure is an easy task. When the parameters have been estimated, the imputation of missing data is still performed by means of (4.17-4.20), where the $\tilde{u}$'s are produced as standard normal variates from a pseudo-random number generator, while parameters Π and Σ are replaced by convenient estimates. The topic we deal with in this section is how to estimate such parameters in any real case, without considering explicitly the likelihood function stated in terms of the possible 2^p different blocks.

5.1 Feasible imputation for bivariate normal data: reduced form approach

The estimation procedure directly considers the data generating process (3.8) and estimates the unknown parameters by ordinary least squares (OLS) method. As there are no "dependent" variables on the right hand side of (3.8), the system is in "reduced form".

The estimation procedure starts building a system of two normal linear regression models on the observed part of the data matrix (A-block); the reduced form coefficients (Π) are estimated by OLS method and Σ (variance-covariance matrix) is estimated from OLS residuals; then missing data are imputed by the imputation function defined in Sec.4 using the estimates of Π and Σ (see Appendix 1, Iteration 0). Once the iteration 0 is over, the data matrix is completed and the procedure goes on, estimating by OLS the parameters

(Π, Σ) on the most recent completed data, and imputing again and re-estimating again until convergence on the estimates of Π and Σ is achieved (the complete iterative method is explained in Appendix 1). In practice, by this method we estimate iteratively the coefficients $\widehat{\Pi}^{(k)}$ and the residuals covariance matrix $\widehat{\Sigma}^{(k)}$ of a two equations reduced form system by OLS, and parameters estimates are used for imputing missing values. We indicate with $\widehat{\Pi}^{(k)}$ the OLS estimation of the unknown parameter Π at the k-iteration (thus using $\widetilde{Y}^{(k-1)}$ data completed at the end of iteration $k-1$) and with $\widehat{E}^{(k)} = \left(\widehat{e}_1^{(k)}, \widehat{e}_2^{(k)}\right) = \widetilde{Y}^{(k-1)} - X\widehat{\Pi}^{(k)}$ the corresponding residuals, from which we estimate $\widehat{\Sigma}^{(k)}$. Supposing that convergence is achieved at the k-iteration, we have (up to a reasonably large number of digits) $\widehat{\Pi}_1^{(k)} = \widehat{\Pi}_1^{(k-1)}$ and $\widehat{\Pi}_2^{(k)} = \widehat{\Pi}_2^{(k-1)}$, so the following conditions become true:

$$(X'X)^{-1} X' \begin{bmatrix} \widehat{e}_{A1}^{(k-1)} \\ \widehat{e}_{B1}^{(k-1)} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}} \widehat{e}_{C2}^{(k-1)} + \sqrt{\widehat{\sigma}_{11}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{22}^{(k-1)}}}\ \widetilde{u}_{C1} \\ \sqrt{\widehat{\sigma}_{11}^{(k-1)}}\ \widetilde{u}_{D1} \end{bmatrix} = 0 \tag{5.21}$$

$$(X'X)^{-1} X' \begin{bmatrix} \widehat{e}_{A2}^{(k-1)} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}} \widehat{e}_{B1}^{(k-1)} + \sqrt{\widehat{\sigma}_{22}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{11}^{(k-1)}}}\ \widetilde{u}_{B2} \\ \widehat{e}_{C2}^{(k-1)} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}} \sqrt{\widehat{\sigma}_{11}^{(k-1)}}\ \widetilde{u}_{D1} + \sqrt{\widehat{\sigma}_{22}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{11}^{(k-1)}}}\ \widetilde{u}_{D2} \end{bmatrix} = 0 \tag{5.22}$$

When convergence on Π is achieved, the expression for $\widehat{\Sigma}^{(k)}$, obtained from the k-th iteration residuals, is as follows:

$$\begin{aligned} n\ \widehat{\sigma}_{11}^{(k)} = {} & \widehat{e}_{A1}^{(k)\prime}\widehat{e}_{A1}^{(k)} + \widehat{e}_{B1}^{(k)\prime}\widehat{e}_{B1}^{(k)} + \left(\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}}\right)^2 \widehat{e}_{C2}^{(k-1)\prime}\widehat{e}_{C2}^{(k-1)} + \left(\widehat{\sigma}_{11}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{22}^{(k-1)}}\right) \widetilde{u}_{C1}'\ \widetilde{u}_{C1} \\ & + 2\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}} \sqrt{\widehat{\sigma}_{11}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{22}^{(k-1)}}}\, \widehat{e}_{C2}^{\prime(k-1)}\ \widetilde{u}_{C1} + \widehat{\sigma}_{11}^{(k-1)}\ \widetilde{u}_{D1}'\ \widetilde{u}_{D1} \end{aligned} \tag{5.23}$$

$$\begin{aligned} n\ \widehat{\sigma}_{12}^{(k)} = {} & \widehat{e}_{A1}^{(k)\prime}\widehat{e}_{A2}^{(k)} + \left(\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}}\right)^2 \widehat{e}_{B1}^{(k-1)\prime}\widehat{e}_{B1}^{(k-1)} + \left(\widehat{\sigma}_{22}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{11}^{(k-1)}}\right) \widehat{e}_{B1}^{(k-1)\prime}\ \widetilde{u}_{B2} \\ & + \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}} \widehat{e}_{C2}^{(k-1)\prime}\widehat{e}_{C2}^{(k-1)} + \sqrt{\widehat{\sigma}_{11}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{22}^{(k-1)}}}\, \widehat{e}_{C2}^{\prime(k-1)}\ \widetilde{u}_{C1} \\ & + \widehat{\sigma}_{12}^{(k-1)}\ \widetilde{u}_{D1}'\ \widetilde{u}_{D1} + .\sqrt{\widehat{\sigma}_{22}^{(k-1)}\widehat{\sigma}_{11}^{(k-1)} - \left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}\ \widetilde{u}_{D1}'\ \widetilde{u}_{D2} \end{aligned} \tag{5.24}$$

$$n\ \widehat{\sigma}_{22}^{(k)} = \widehat{e}_{A2}^{(k)\prime}\widehat{e}_{A2}^{(k)} + \widehat{e}_{C2}^{(k)\prime}\widehat{e}_{C2}^{(k)} + \left(\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}}\right)^2 \widehat{e}_{B1}^{(k-1)\prime}\widehat{e}_{B1}^{(k-1)} + \left(\widehat{\sigma}_{22}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{11}^{(k-1)}}\right) \widetilde{u}_{B2}' \ \widetilde{u}_{B2}$$

$$+2\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}}\sqrt{\widehat{\sigma}_{22}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{11}^{(k-1)}}}\widehat{e}_{B1}^{\prime(k-1)} \ \widetilde{u}_{B1} + \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{11}^{(k-1)}} \ \widetilde{u}_{D1}' \ \widetilde{u}_{D1}$$

$$+ \left(\widehat{\sigma}_{22}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{11}^{(k-1)}}\right) \widetilde{u}_{D2}' \ \widetilde{u}_{D2} + 2\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}}\sqrt{\widehat{\sigma}_{22}^{(k-1)}\widehat{\sigma}_{11}^{(k-1)} - \left(\widehat{\sigma}_{12}^{(k-1)}\right)^2} \ \widetilde{u}_{D1}' \ \widetilde{u}_{D2} \quad (5.25)$$

We may regard (5.23-5.25) as the convergence condition for the iterative estimation of Σ. The estimation/imputation procedure achieves convergence when equations (5.21-5.25) are jointly solved.

5.2 Feasible imputation for bivariate normal data: structural form approach

If the data set with missing values is a generic $n \times p$ matrix, and there are missing values in more than two variables, it is difficult to explicit the imputation function, as we did in case of bivariate normal data with missing values. As a consequence, the estimation/imputation method proposed has shifted the technical difficulty from the likelihood to the imputation function; thus, technical problem is still unsolved.

The way to solve such a technical problem is prompted to the "sequential regression multivariate imputation" (SRMI) approach adopted by the imputation software (*IVE-ware*). The method proposed by the authors of the software (Raghunathan, Lepkowski, Van Hoewyk, and Solenberger, 1997) builds the imputed values by fitting a sequence of regression models and drawing values from the corresponding predictive distribution, under the hypothesis of MAR mechanism, infinite sample size and simple random sampling. The method follows a bayesian paradigm. Each imputation consists of c "rounds". Round 1 starts regressing the variable with the fewest number of missing values, say Y_1, on X, and imputing the missing values with the appropriate regression model. Assuming a flat prior for the regression coefficient, the imputations for the missing values in Y_1 are drawn from the corresponding posterior predictive distribution. After Y_1 has been completed, the variable with the fewest number of missing values is considered, say Y_2; observed Y_2 values are regressed on (X, Y_1) and the missing values are imputed, and so on. The imputation process is then repeated in rounds 2 through c, modifying the predictor set to include all the Y variables except the one used as the dependent variable. Repeated cycles continue for a pre-specified number of rounds, or until stable imputed values occur (convergence in distribution).

In order to make our method "feasible in practice" even when the number of variables with missing values is grater than 2, we follow the SRMI method, but introducing a convenient modification of the variance covariance matrix estimator.

According to the SRMI method the step 0 is essentially the same as the one introduced in Section 5.1 (iteration 0). We estimate the coefficients of the linear regression model related to the variable with fewest missing values (let be Y_1), by OLS, using the Y_1 observed part ($Y_{obs,1}$). Suppose that $\widehat{\Pi}_1 = (X'_{obs}X_{obs})^{-1}X'_{obs}Y_{obs,1}$ is the regression coefficient and $\widehat{\sigma}_{11} = (1/n_{obs})\left(Y_{obs,1} - X_{obs}\widehat{\Pi}_1\right)'\left(Y_{obs,1} - X_{obs}\widehat{\Pi}_1\right)$ is the residual variance, then the imputed value set is

$$\widetilde{Y_1} = X_{mis,1}\widehat{\Pi}_1 + \sqrt{\widehat{\sigma}_{11}}\;\widetilde{u}_1$$

where $\widetilde{u}_1$ is the usual vector of independent pseudo-random standard normal deviates.

So we have a first set of completed values for Y_1 and we attach it as an additional column to X. We then regress the next variable with fewest missing values (say $Y_{obs,2}$) against X and the completed Y_1 and use the OLS estimated coefficients and variance for an imputation step that completes Y_2. Going on, the first round ends when all the missing values are completed. As the SRMI's authors put in evidence, the updating of the right hand side variables after imputing the missing values is dependent on the order in which we select the variables for imputation. Thus, the imputed values for Y_j, for example, involve only (X, $Y_1, ..., Y_{j-1}$), but not $Y_{j+1} ... Y_{p.}$ For this reason the procedure continues to overwrite the imputations for the missing values iteratively. After the first round, we have complete data for all variables, part of wich are observed, part have been imputed in the previous iterations. The system of regression equations has, as dependent variable for each equation, the variable to be "imputed if missing" and has on the right hand side all the others variables

$$\begin{aligned} Y_1 &= X\gamma_{11} + Y_2\gamma_{12} + Y_3\gamma_{13} + ... + Y_p\gamma_{1p} + \varepsilon_1 \\ Y_2 &= X\gamma_{21} + Y_1\gamma_{22} + Y_3\gamma_{23} + ... + Y_p\gamma_{2p} + \varepsilon_2 \\ &\;\;... \\ Y_p &= X\gamma_{p1} + Y_1\gamma_{p2} + Y_2\gamma_{p3} + ... + Y_{p-1}\gamma_{pp} + \varepsilon_p \end{aligned} \tag{5.26}$$

where $\gamma_{11}, \gamma_{21}, ..., \gamma_{p1}$ are scalars or $(k \times 1)$ vectors depending on X being a single column or a $(p \times k)$ matrix, while all the other γ are scalars and $\in = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_p)$ has a multivariate normal distribution.

Equations (5.26) represent a system of simultaneous equations in structural form, as the jointly dependent variables Y appear also on the right hand side of the equations. A convenient representation of the system in structural form (that will be used in the next section) is

$$BY' + \Gamma X' = \in', \tag{5.27}$$

that is, for the i-th observation $(i = 1, ...n)$

$$By_i + \Gamma x_i = \varepsilon_i, \tag{5.28}$$

where the matrices of the structural form coefficients are

$$B_{(p\times p)} = \begin{bmatrix} 1 & -\gamma_{12} & -\gamma_{13} & \cdots\cdots & -\gamma_{1p} \\ -\gamma_{22} & 1 & -\gamma_{23} & \cdots\cdots & -\gamma_{2p} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ -\gamma_{p2} & -\gamma_{p3} & \cdots\cdots & -\gamma_{pp} & 1 \end{bmatrix} \quad \text{and} \quad \Gamma_{(p\times k)} = \begin{bmatrix} -\gamma'_{11} \\ -\gamma'_{21} \\ \cdots\cdots \\ -\gamma'_{p1} \end{bmatrix}$$

(the dimensions of Γ being: $p \times k$, where k is the number of columns of X).

We remember that, from the stuctural form system given above, we can easily derive the reduced form "solving" the system (5.28) for y_i

$$y_i = -B^{-1}\Gamma x_i + B^{-1}\epsilon_i = \Pi x_i + e_i \tag{5.29}$$

A structural form model like (5.26) is underidentified, as it violates the order condition for identification (eg. Greene, 2000, Sec. 16.3.1): infinite sets of γ-values would be observationally equivalent. It is therefore useless (or impossible) to apply the estimation techniques suitable for simultaneous equation systems, like two or three stage least squares, full information maximum likelihood, etc. Nevertheless we can estimate each equation separatly by OLS as in SRMI approach. After coefficients have been estimated by OLS, we compute from the residuals the estimates of the variance covariance matrix as $\widehat{\Psi} = (1/n)\left(\widehat{B}Y' + \widehat{\Gamma}X'\right)(\widehat{B}Y' + \widehat{\Gamma}X')'$. Differently from the SRMI method, we use the Cholesky decomposition of the matrix $\widehat{\Psi}$ to produce vectors of pseudo-random numbers for imputation, thus considering also covariances besides variances.

When a value of Y_1 is missing, we impute the value obtained from the right hand side of the first equation in (5.26), where the γ -s are at the last estimated value, the value(s) of X is (are) observed, the value of the ε-s are produced by the pseudo-random generator with a variance-covariance matrix equal to the last estimated $\widehat{\Psi}$. The same is done for all missing values of the $Y_2, ...Y_p$ variables.

Repeated cycles continue until convergence on the estimated parameters has been achieved.

The transformation between structural form and reduced form helps to answer a question that naturally arises when dealing with an iterative simulation-based method: why and when does the the iterative estimation/imputation procedure converge? (Even if the MCMC context is different from our context, still recently Horton and Lipsitz, 2001, p. 246 points out that convergence "remains more of an art form than a science").

And, in fact, it is not at all obvious how convergence is achieved if we only consider the procedure as it has just been described. But we may think at the sequence of iterations in a different order, as if iterations were "grouped". Let's first see what happens if we keep parameter values fixed (the γ -s and the Cholesky decomposition of the matrix $\widehat{\Psi}$), and iterate substitutions of imputed values on the right hand side of equations (5.26). These iterated substitutions (e.g. Thisted, sec. 3.11.2) are "exactly" the steps of the well known Gauss-Seidel method for the "simultaneous solution" of the (5.26) system (also called "stochastic simulation" of the system, because of the presence of the ϵ terms). Thus, till we hold parametes fixed at some values, the iterated substitution of imputed values will converge to the reduced form derived from the structural form (or "restricted" reduced form, equation 5.29). Now we can re-estimate parameters (with OLS on the structural form) and start again a new cycle of iterated substitutions in (5.26), and so on.

The strictly thighted sequence of estimations and imputations for each structural equation is thus disentangled and converted into a sequence of iterations that are conceptually much more manageble. In each iteration, an OLS estimation of "all" the structural form equations (5.26), using observed and previously imputed values, is followed by the simultaneous solution of the equation system (or derivation of the reduced form 5.29) that produces "all" the values of the variables y to be imputed. Studying the convergence of this new sequence of estimation and imputation phases becomes more manageable, as it will be clear in Section 6.

The SRMI method and the one just proposed follow different paradigms, the former is based on the bayesian paradigm and the latter on the frequentist paradigm. Beyond this difference, it is important to put in evidence the main technical difference. The SRMI method draws the random normal deviates of the imputation step for each equation "independently"; the method we propose considers stochastic terms still drawn from a normal distribution, but having variance-covariance matrix $\widehat{\Psi}$.

6 Properties of the estimator obtained by the feasible imputation method

The properties of the estimated parameters obtained at convergence of the iterative estimation/imputation method are not evaluated considering the structural from, because the system is underidentified. To show the "good" properties of the estimator discussed above, we introduce and show in sequence the following propositions:

Proposition 1 *The reduced form parameters estimator, derived from the the* OLS *estimator of the strucural form parameters, is equal to the* OLS *direct estimator of the reduced form parameters.*

Proposition 2 *The* OLS *estimator of the reduced form parameters, at convergence of the estimation/imputation procedure, is a* MSS *(Method of Simulated Scores) estimator with "one" replication (Hajivassiliou and McFadden, 1990).*

The proofs for both propositions consider, without loss of generality, a bivariate normal case.

6.1 Proof - Proposition 1

The structural (underidentified) equation system we have to estimate is the following:

$$\begin{cases} Y_1 = X\gamma_{11} + Y_2\gamma_{12} + \varepsilon_1 \\ Y_2 = X\gamma_{21} + Y_1\gamma_{22} + \varepsilon_2 \end{cases}, \tag{6.30}$$

with $\Psi = E\left\{\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{bmatrix}\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{bmatrix}'\right\}$.

Note that each equation is not identified, as it violates the necessary order condition for identification (e.g. Greene, 2000, sec. 16.3.1); roughly speaking, in each equation there are more explanatory variables (X and Y_1 or Y_2) than exogenous variables in the whole system (X only) . The corresponding reduced form is:

$$\begin{cases} Y_1 = X\Pi_1 + e_1 \\ Y_2 = X\Pi_2 + e_2 \end{cases}.$$

The OLS direct estimator of the reduced form coefficients

$$\widehat{\Pi}' = (X'X)^{-1}X'Y \tag{6.31}$$

and the variance-covariance estimator based on the OLS residual ($\widehat{E} = Y - X\widehat{\Pi}'$) are consistent, and we can write:

$$\begin{cases} Y_1 = X\widehat{\Pi}_1 + \widehat{e}_1 \\ Y_2 = X\widehat{\Pi}_2 + \widehat{e}_2 \end{cases}. \tag{6.32}$$

The OLS estimates of the structural form coefficients of the equation having Y_1 as dependent variable (6.30) are:

$$\begin{bmatrix} \widetilde{\gamma}_{11} \\ \widetilde{\gamma}_{12} \end{bmatrix} = \begin{bmatrix} \sum x_i^2 & \sum x_i y_{i2} \\ \sum x_i y_{i2} & \sum y_{i2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_i y_{i1} \\ \sum y_{i2} y_{i1} \end{bmatrix},$$

now we can replace y_{i1}, y_{i2} with the expressions (6.32), obtaining:

$$\begin{aligned} \begin{bmatrix} \widetilde{\gamma}_{11} \\ \widetilde{\gamma}_{12} \end{bmatrix} &= \begin{bmatrix} \sum x_i^2 & \widehat{\Pi}_2 \sum x_i^2 + \sum x_i \widehat{e}_{i2} \\ \widehat{\Pi}_2 \sum x_i^2 + \sum x_i \widehat{e}_{i2} & \widehat{\Pi}_2{}^2 \sum x_i^2 + 2\widehat{\Pi}_2 \sum x_i \widehat{e}_{i2} + \sum \widehat{e}_{i2}^2 \end{bmatrix}^{-1} \\ &\cdot \begin{bmatrix} \widehat{\Pi}_1 \sum x_i^2 + \sum x_i \widehat{e}_{i1} \\ \widehat{\Pi}_1 \widehat{\Pi}_2 \sum x_i^2 + \widehat{\Pi}_1 \sum x_i \widehat{e}_{i2} + \widehat{\Pi}_2 \sum x_i \widehat{e}_{i1} + \sum \widehat{e}_{i1} \widehat{e}_{i2} \end{bmatrix}. \end{aligned}$$

Now posing $\sum_i x_i^2 = q^2$ and reminding that $\sum x_i \widehat{e}_{i1} = \sum x_i \widehat{e}_{i2} = 0$, because $\widehat{e}_{i1}$ and $\widehat{e}_{i2}$ are OLS residuals, we obtain for the first equation:

$$\begin{aligned}
\begin{bmatrix} \widetilde{\gamma}_{11} \\ \widetilde{\gamma}_{12} \end{bmatrix} &= \begin{bmatrix} q^2 & \widehat{\Pi}_2 q^2 \\ \widehat{\Pi}_2 q^2 & \widehat{\Pi}_2^2 q^2 + \sum \widehat{e}_{i2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\Pi}_1 q^2 \\ \widehat{\Pi}_1 \widehat{\Pi}_2 q^2 + \sum \widehat{e}_{i1} \widehat{e}_{i2} \end{bmatrix} \\
&= \frac{1}{q^2 \sum \widehat{e}_{i2}^2} \begin{bmatrix} \widehat{\Pi}_2^2 q^2 + \sum \widehat{e}_{i2}^2 & -\widehat{\Pi}_2 q^2 \\ -\widehat{\Pi}_2 q^2 & q^2 \end{bmatrix} \begin{bmatrix} \widehat{\Pi}_1 q^2 \\ \widehat{\Pi}_1 \widehat{\Pi}_2 q^2 + \sum \widehat{e}_{i1} \widehat{e}_{i2} \end{bmatrix} \\
&= \frac{1}{q^2 \sum \widehat{e}_{i2}^2} \begin{bmatrix} \widehat{\Pi}_1 \widehat{\Pi}_2^2 q^4 + \widehat{\Pi}_1 q^2 \sum \widehat{e}_{i2}^2 - \widehat{\Pi}_1 \widehat{\Pi}_2^2 q^4 - \widehat{\Pi}_2 q^2 \sum \widehat{e}_{i1} \widehat{e}_{i2} \\ -\widehat{\Pi}_1 \widehat{\Pi}_2 q^4 + \widehat{\Pi}_1 \widehat{\Pi}_2 q^4 + q^2 \sum \widehat{e}_{i1} \widehat{e}_{i2} \end{bmatrix} \\
&= \frac{1}{\sum \widehat{e}_{i2}^2} \begin{bmatrix} \widehat{\Pi}_1 \sum \widehat{e}_{i2}^2 - \widehat{\Pi}_2 \sum \widehat{e}_{i1} \widehat{e}_{i2} \\ \sum \widehat{e}_{i1} \widehat{e}_{i2} \end{bmatrix} = \begin{bmatrix} \widehat{\Pi}_1 - \frac{\sum \widehat{e}_{i1} \widehat{e}_{i2}}{\sum \widehat{e}_{i2}^2} \widehat{\Pi}_2 \\ \frac{\sum \widehat{e}_{i1} \widehat{e}_{i2}}{\sum \widehat{e}_{i2}^2} \end{bmatrix}
\end{aligned} \tag{6.33}$$

and analogously for the second equation of the structural form system:

$$\begin{bmatrix} \widetilde{\gamma}_{21} \\ \widetilde{\gamma}_{22} \end{bmatrix} = \begin{bmatrix} \widehat{\Pi}_2 - \frac{\sum \widehat{e}_{i1} \widehat{e}_{i2}}{\sum \widehat{e}_{i1}^2} \widehat{\Pi}_1 \\ \frac{\sum \widehat{e}_{i1} \widehat{e}_{i2}}{\sum \widehat{e}_{i1}^2} \end{bmatrix}. \tag{6.34}$$

The reduced form coefficients, derived from the structural form, are obtained as

$$\widetilde{\Pi} = -\widetilde{B}^{-1} \widetilde{\Gamma} \tag{6.35}$$

where

$$\widetilde{B} = \begin{bmatrix} 1 & -\widetilde{\gamma}_{12} \\ -\widetilde{\gamma}_{22} & 1 \end{bmatrix}, \qquad \widetilde{\Gamma} = \begin{bmatrix} -\widetilde{\gamma}_{11} \\ -\widetilde{\gamma}_{21} \end{bmatrix} \tag{6.36}$$

Substituting (6.33) and (6.34) into (6.36) and then into (6.35) we obtain exactly the $\widehat{\Pi}$ as in (6.31):

$$\widetilde{\Pi} = \widehat{\Pi}.$$

So that, estimating the structural form coefficients $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})$ of our underidentified system (6.30) by OLS and trasforming them into reduced form coefficients (by the proper trasformation) is exactly the same as estimating directly by OLS the reduced form coefficients.

We need to show now that the same concept is valid for the estimator of the variance-covariance matrix.

The structural form OLS residuals are:

$$\widetilde{\varepsilon}_{i1} = y_{i1} - \widetilde{\gamma}_{11} x_i - \widetilde{\gamma}_{12} y_{i2} = \widehat{\Pi}_1 x_i + \widehat{e}_{i1} - \widetilde{\gamma}_{11} x_i - (\widehat{\Pi}_2 x_i + \widehat{e}_{i2}) \widetilde{\gamma}_{12} = \widehat{e}_{i1} - \frac{\sum \widehat{e}_{i1} \widehat{e}_{i2}}{\sum \widehat{e}_{i2}^2} \widehat{e}_{i2}$$

$$\widetilde{\varepsilon}_{i2} = y_{i2} - \widetilde{\gamma}_{21} x_i - \widetilde{\gamma}_{22} y_{i1} = \widehat{\Pi}_2 x_i + \widehat{e}_{i2} - \widetilde{\gamma}_{21} x_i - (\widehat{\Pi}_1 x_i + \widehat{e}_{i1}) \widetilde{\gamma}_{22} = \widehat{e}_{i2} - \frac{\sum \widehat{e}_{i1} \widehat{e}_{i2}}{\sum \widehat{e}_{i1}^2} \widehat{e}_{i1}$$

so that the estimator of the variance-covariance matrix of the structural form error terms is:

$$\widetilde{\Psi} = \frac{1}{n}\sum\left\{\begin{bmatrix}\widetilde{\varepsilon}_{i1}\\ \widetilde{\varepsilon}_{i2}\end{bmatrix}\begin{bmatrix}\widetilde{\varepsilon}_{i1}\\ \widetilde{\varepsilon}_{i2}\end{bmatrix}'\right\}$$

$$= \frac{1}{n}\sum\left\{\begin{bmatrix}\widehat{e}_{i1} - \frac{\sum\widehat{e}_{i1}\widehat{e}_{i2}}{\sum\widehat{e}_{i2}^2}\widehat{e}_{i2}\\ \widehat{e}_{i2} - \frac{\sum\widehat{e}_{i1}\widehat{e}_{i2}}{\sum\widehat{e}_{i1}^2}\widehat{e}_{i1}\end{bmatrix}\begin{bmatrix}\widehat{e}_{i1} - \frac{\sum\widehat{e}_{i1}\widehat{e}_{i2}}{\sum\widehat{e}_{i2}^2}\widehat{e}_{i2}\\ \widehat{e}_{i2} - \frac{\sum\widehat{e}_{i1}\widehat{e}_{i2}}{\sum\widehat{e}_{i1}^2}\widehat{e}_{i1}\end{bmatrix}'\right\} \tag{6.37}$$

The corresponding reduced form variance-covariance matrix is obtained, as usual, computing:

$$\widetilde{\Sigma} = \widetilde{B}^{-1}\widetilde{\Psi}\left(\widetilde{B}^{-1}\right)' \tag{6.38}$$

Substituting (6.37) and (6.36) into (6.38) we obtain:

$$\widetilde{\Sigma} = \widehat{\Sigma}$$

So the proof is completed; in fact we have shown that the reduced form parameters estimator derived from the OLS estimator of the strucural form parameters is equal to the OLS direct estimator of the reduced form parameters.

6.2 Remark

The methods discussed in this paper and the proof given in the previous section hold also if no variable has complete data and therefore the matrix X is empty. For example, we might have two variables only, Y_1 and Y_2, none of them is completely observed, in some cases y_{i1} and y_{i2} are both observed, in other cases y_{i1} is observed and y_{i2} is missing, or y_{i2} is observed and y_{i1} is missing (still assuming MAR missing data mechanism). In such a case the reduced form would have no coefficients (Π), so the reduced form parameters would be only the covariance matrix Σ. The structural form (6.30) would be without the X variables, the matrix $\widetilde{B}$ would still be as in equation (6.36), thus still producing the equality (6.38). Notice however that this would be an extreme case. It would be much more usual to use a non-empty matrix X, containing at least the constant column.

6.3 Proof - Proposition 2

In order to prove the proposition 2, we need to introduce the method of simulated scores (MSS, Hajivassiliou and McFadden, 1990, Hajivassiliou, 1993, Hajivassiliou and Ruud, 1994, Stern, 2000), which is a particular MSM estimator (McFadden, 1989; Pakes and Pollard, 1989). For a brief review, see Appendix 2.

The intuition behind the method of simulated scores is the following. We may add, to the score function, a simulated term, and call the resulting expression a "simulated score".

If the additive simulated term has zero conditional expectation (given observations, and considering expectation with respect to the simulation process), then the resulting expression would be an "unbiased simulator of the score". Like the score function, also an unbiased simulator of the score should have a zero expected value at the "true" value of the parameter θ, and a nonzero expectation for different values of θ (identifiability). The estimator is the value of θ that sets to zero the simulated score in the sample.

In practice, it is not always easy to construct an unbiased estimator of the score (Stern, 2000, p. 25). But there are cases, like ours, where the "exact" score is much more difficult to produce than an unbiased simulated score.

We show in this section that if, given the parameter values $\theta = (\Pi, \Sigma)$,

1. we complete our data with the imputation function (4.16), so to have the set of "simulated" variables $\widetilde{Y} = b(Y_{obs})$;

2. we consider the log-likelihood of the completed data set (simulated log-likelihood) and its first derivatives (simulated score) that have the usual "simple" expression from the multivariate normal distribution since data are complete;

3. then, the conditional expectation of the simulated score, given the observed variables, is equal to the score (eq. 3.14 and 3.15).

In other words, we prove in this section that we have constructed an unbiased simulator of the score

$$E_\theta \left[\frac{\partial \log f(\widetilde{Y}|X;\theta)}{\partial \theta} | Y_{obs}, X \right] = \frac{\partial \log f(Y_{obs}|X;\theta)}{\partial \theta}, \tag{6.39}$$

where $\widetilde{Y} = b(Y_{obs})$ is the imputation function in (4.16).

We start expliciting, step by step, the left hand side of (6.39); we consider derivatives with respect to Π and, as in Sec. 3.1, derivatives with respect to Σ^{-1} rather than Σ.

Expliciting derivatives with respect to Π as a block matrix, we have

$$\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Pi}$$
$$= \Sigma^{-1} \begin{bmatrix} (Y_{A1} - X_A \Pi_1) & (Y_{A2} - X_A \Pi_2) \\ (Y_{B1} - X_B \Pi_1) & \frac{\sigma_{12}}{\sigma_{11}} (Y_{B1} - X_B \Pi_1) + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{B2} \\ \frac{\sigma_{12}}{\sigma_{22}} (Y_{C2} - X_C \Pi_2) + \sqrt{\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}}\ \widetilde{u}_{C1} & (Y_{C2} - X_C \Pi_2) \\ \sqrt{\sigma_{11}}\ \widetilde{u}_{D1} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}}\ \widetilde{u}_{D1} + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{D2} \end{bmatrix}' X$$

then, computing expectation conditional on Y_{obs} and X, we have:

$$E \left[\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Pi} | Y_{obs}, X \right]$$

$$= \Sigma^{-1}\begin{bmatrix} (Y_{A1}-X_A\Pi_1) & (Y_{A2}-X_A\Pi_2) \\ (Y_{B1}-X_B\Pi_1) & \frac{\sigma_{12}}{\sigma_{11}}(Y_{B1}-X_B\Pi_1) \\ \frac{\sigma_{12}}{\sigma_{22}}(Y_{C2}-X_C\Pi_2) & (Y_{C2}-X_C\Pi_2) \\ 0 & 0 \end{bmatrix}' X \tag{6.40}$$

In the same way, we explicit the derivatives with respect to Σ^{-1} as a block matrix, as follows:

$$\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Sigma^{-1}}$$

$$= \frac{n}{2}\Sigma - \begin{bmatrix} (Y_{A1}-X_A\Pi_1) & (Y_{A2}-X_A\Pi_2) \\ (Y_{B1}-X_B\Pi_1) & \left(\widetilde{Y}_{B2}-X_B\Pi_2\right) \\ \left(\widetilde{Y}_{C1}-X_C\Pi_1\right) & (Y_{C2}-X_C\Pi_2) \\ \left(\widetilde{Y}_{D1}-X_D\Pi_1\right) & \left(\widetilde{Y}_{D2}-X_D\Pi_2\right) \end{bmatrix}' \begin{bmatrix} (Y_{A1}-X_A\Pi_1) & (Y_{A2}-X_A\Pi_2) \\ (Y_{B1}-X_B\Pi_1) & \left(\widetilde{Y}_{B2}-X_B\Pi_2\right) \\ \left(\widetilde{Y}_{C1}-X_C\Pi_1\right) & (Y_{C2}-X_C\Pi_2) \\ \left(\widetilde{Y}_{D1}-X_D\Pi_1\right) & \left(\widetilde{Y}_{D2}-X_D\Pi_2\right) \end{bmatrix}$$

$$= \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i\in A}\left\{\begin{bmatrix} y_{i1}-\Pi_1 x_i \\ y_{i2}-\Pi_2 x_i \end{bmatrix}\begin{bmatrix} y_{i1}-\Pi_1 x_i \\ y_{i2}-\Pi_2 x_i \end{bmatrix}'\right\}$$

$$- \frac{1}{2}\sum_{i\in B}\left\{\begin{bmatrix} y_{i1}-\Pi_1 x_i \\ \frac{\sigma_{12}}{\sigma_{11}}(y_{i1}-\Pi_1 x_i)+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\ \tilde{u}_{i2} \end{bmatrix}\begin{bmatrix} y_{i1}-\Pi_1 x_i \\ \frac{\sigma_{12}}{\sigma_{11}}(y_{i1}-\Pi_1 x_i)+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\ \tilde{u}_{i2} \end{bmatrix}'\right\}$$

$$- \frac{1}{2}\sum_{i\in C}\left\{\begin{bmatrix} \frac{\sigma_{12}}{\sigma_{22}}(y_{i2}-\Pi_2 x_i)+\sqrt{\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}}\ \tilde{u}_{i1} \\ y_{i2}-\Pi_2 x_i \end{bmatrix}\begin{bmatrix} \frac{\sigma_{12}}{\sigma_{22}}(y_{i2}-\Pi_2 x_i)+\sqrt{\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}}\ \tilde{u}_{i1} \\ y_{i2}-\Pi_2 x_i \end{bmatrix}'\right\}$$

$$- \frac{1}{2}\sum_{i\in D}\left\{\begin{bmatrix} \sqrt{\sigma_{11}}\ \tilde{u}_{i1} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}}\ \tilde{u}_{i1}+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\ \tilde{u}_{i2} \end{bmatrix}\begin{bmatrix} \sqrt{\sigma_{11}}\ \tilde{u}_{i1} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}}\ \tilde{u}_{i1}+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\ \tilde{u}_{i2} \end{bmatrix}'\right\}$$

then, computing conditional expectation, we obtain:

$$E\left[\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Sigma^{-1}}|Y_{obs},X\right] = \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i=A}\left[(y_i-\Pi x_i)(y_i-\Pi x_i)'\right]$$

$$-\frac{1}{2}\begin{pmatrix} 1 & \frac{\sigma_{12}}{\sigma_{11}} \\ \frac{\sigma_{12}}{\sigma_{11}} & \frac{\sigma_{12}^2}{\sigma_{11}^2} \end{pmatrix}\sum_{i=B}(y_{i1}-\Pi x_i)^2 - \frac{n_B}{2}\begin{pmatrix} 0 & 0 \\ 0 & \frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{11}} \end{pmatrix}$$

$$-\frac{1}{2}\begin{pmatrix} \frac{\sigma_{12}^2}{\sigma_{22}^2} & \frac{\sigma_{12}}{\sigma_{22}} \\ \frac{\sigma_{12}}{\sigma_{22}} & 1 \end{pmatrix}\sum_{i=C}(y_{i2}-\Pi x_i)^2 - \frac{n_C}{2}\begin{pmatrix} \frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}} & 0 \\ 0 & 0 \end{pmatrix} - \frac{n_D}{2}\Sigma.$$

Remarking that

$$\frac{n}{2}\Sigma - \frac{n_D}{2}\Sigma = \frac{n_A+n_B+n_C}{2}\Sigma$$

we have:

$$E\left[\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Sigma^{-1}}|Y_{obs},X\right]$$

$$= \frac{n_A}{2}\Sigma + \frac{n_B}{2}\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \frac{\sigma_{12}^2}{\sigma_{11}} \end{pmatrix} + \frac{n_C}{2}\begin{pmatrix} \frac{\sigma_{12}^2}{\sigma_{22}} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} - \frac{1}{2}\sum_{i=A}\left[(y_i - \Pi x_i)(y_i - \Pi x_i)'\right]$$

$$- \frac{1}{2}\begin{pmatrix} 1 & \frac{\sigma_{12}}{\sigma_{11}} \\ \frac{\sigma_{12}}{\sigma_{11}} & \frac{\sigma_{12}^2}{\sigma_{11}^2} \end{pmatrix}\sum_{i=B}(y_{i1} - \Pi_1 x_i)^2 - \frac{1}{2}\begin{pmatrix} \frac{\sigma_{12}^2}{\sigma_{22}^2} & \frac{\sigma_{12}}{\sigma_{22}} \\ \frac{\sigma_{12}}{\sigma_{22}} & 1 \end{pmatrix}\sum_{i=C}(y_{i2} - \Pi_2 x_i)^2 . \quad (6.41)$$

This proves the identity (6.39). Infact, the explicit form of the right hand side of (6.39) are the expressions (3.14) and (3.15) and the left hand side of (6.39) are the (6.40) and (6.41); we reach the result observing that (3.14) is equal to (6.40) and (3.15) is equal to (6.41). Therefore it is natural to consider the equality (6.39) as an unbiasedness condition and to propose

$$\frac{\partial \log f(\widetilde{Y}|X;\theta)}{\partial \theta}$$

as unbiased simulator of the score.

6.4 Numerical convergence of the estimation/imputation procedure

The simulated scores estimator of $\theta = (\Pi, \Sigma)$ is obtained solving

$$\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Pi}|Y, X = 0 \quad (6.42)$$

and

$$\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Sigma^{-1}}|Y, X = 0 \quad (6.43)$$

Expliciting (6.42) we obtain

$$\begin{bmatrix} Y_{A1} - X_A\Pi_1 \\ Y_{B1} - X_B\Pi_1 \\ \frac{\sigma_{12}}{\sigma_{22}}\, e_{C2} + \sqrt{\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}}\widetilde{u}_{C1} \\ \sqrt{\sigma_{11}}\widetilde{u}_{D1} \end{bmatrix}' X = 0, \quad (6.44)$$

and

$$\begin{bmatrix} Y_{A2} - X_A\Pi_2 \\ \frac{\sigma_{12}}{\sigma_{11}}\, e_{B1} + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\widetilde{u}_{B2} \\ Y_{C2} - X_C\Pi_C \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}}\widetilde{u}_{D1} + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\widetilde{u}_{D2} \end{bmatrix}' X = 0. \quad (6.45)$$

Expliciting (6.43) we obtain

$$n\ \Sigma - \begin{bmatrix} Y_{A1} - X_A\Pi_1 & Y_{A2} - X_A\Pi_2 \\ Y_{B1} - X_B\Pi_1 & \widetilde{Y}_{B2} - X_B\Pi_2 \\ \widetilde{Y}_{C1} - X_C\Pi_1 & Y_{C2} - X_C\Pi_2 \\ \widetilde{Y}_{D1} - X_D\Pi_1 & \widetilde{Y}_{D2} - X_D\Pi_2 \end{bmatrix}' \begin{bmatrix} Y_{A1} - X_A\Pi_1 & Y_{A2} - X_A\Pi_2 \\ Y_{B1} - X_B\Pi_1 & \widetilde{Y}_{B2} - X_B\Pi_2 \\ \widetilde{Y}_{C1} - X_C\Pi_1 & Y_{C2} - X_C\Pi_2 \\ \widetilde{Y}_{D1} - X_D\Pi_1 & \widetilde{Y}_{D2} - X_D\Pi_2 \end{bmatrix} = 0 \quad (6.46)$$

In order to obtain the simulated scores estimator for Σ we have to solve (6.46) or equivalently the following sistem:

$$
\begin{aligned}
n\ \sigma_{11} \ &= \sum_{i\in A,B} (y_{i1} - \Pi_1 x_i)^2 + \left(\frac{\sigma_{12}}{\sigma_{22}}\right)^2 \sum_{i\in C} (y_{i2} - \Pi_2 x_i)^2 + \left(\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right) \sum_{i\in C}\ \tilde{u}_{i1}^2 \\
&+2\frac{\sigma_{12}}{\sigma_{22}}\sqrt{\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}} \sum_{i\in C} (y_{i2} - \Pi_2 x_i)\ \ \tilde{u}_{i1} + \sigma_{11} \sum_{i\in D}\ \tilde{u}_{i1}^2
\end{aligned}
$$

$$
\begin{aligned}
n\ \sigma_{12} \ &= \sum_{i\in A} (y_{i1} - \Pi_1 x_i)(y_{i2} - \Pi_2 x_i) + \frac{\sigma_{12}}{\sigma_{11}} \sum_{i\in B} (y_{i1} - \Pi_1 x_i)^2 \\
&+\sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}} \sum_{i\in B} (y_{i1} - \Pi_1 x_i)\ \ \tilde{u}_{i2} + \frac{\sigma_{12}}{\sigma_{22}} \sum_{i\in C} (y_{i2} - \Pi_2 x_i)^2 \\
&+\sqrt{\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}} \sum_{i\in C} (y_{i2} - \Pi_2 x_i)^2\ \ \tilde{u}_{i1} + \sigma_{12} \sum_{i\in D}\ \tilde{u}_{i1}^2 + \sqrt{\sigma_{22}\sigma_{11} - \sigma_{12}^2} \sum_{i\in D}\ \tilde{u}_{i1}\ \tilde{u}_{i2}
\end{aligned}
$$

$$
\begin{aligned}
n\ \sigma_{22} \ &= \sum_{i\in A,C} (y_{i2} - \Pi_2 x_i)^2 + \left(\frac{\sigma_{12}}{\sigma_{11}}\right)^2 \sum_{i\in B} (y_{i1} - \Pi_1 x_i)^2 \\
&+ \left(\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right) \sum_{i\in B}\ \tilde{u}_{i2}^2 + 2\frac{\sigma_{12}}{\sigma_{11}}\sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}} \sum_{i\in B} (y_{i1} - \Pi_1 x_i)\ \ \tilde{u}_{i2} + \frac{\sigma_{12}^2}{\sigma_{11}} \sum_{i\in D}\ \tilde{u}_{i1}^2 \\
&+ \left(\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right) \sum_{i\in D}\ \tilde{u}_{i2}^2 + 2\frac{\sigma_{12}}{\sigma_{11}}\sqrt{\sigma_{22}\sigma_{11} - \sigma_{12}^2} \sum_{i\in D}\ \tilde{u}_{i1}\ \tilde{u}_{i2}
\end{aligned}
$$

Concluding, the equations that we solved to obtain the simulated scores estimator (6.44, 6.45 and 6.46) are the same conditions we have at convergence achieved of the iterated estimation/ imputation procedure (5.21-5.25). The consequence is that the OLS estimators of the reduced form system with completed data (at convergence achieved) are simulated scores estimators.

The conclusion is the following: the estimation/imputation iterative procedure provides, at convergence achieved, strongly consistent and asymptotically normal estimators (Gourieroux, Monfort, 1996, pp. 35-37). About the efficiency of such estimator it is worth spending some more words.

6.5 Asymptotic (in)efficiency

The potential advantage of the method of simulated scores is to use an estimator with the efficiency properties of the ML and the consistency properties of the method of simulated moments MSM. The MSM estimator is asymptotically efficient if the proper weights are used (those that turn the moment condition into the score statistic) and the simulated scores estimator ensures that such weights are used (Gourieroux, Monfort, 1996, p. 35).

In order to define and make explicit the asymptotic variance-covariance matrix, it is necessary to introduce some convenient notations: the expression of Σ^{-1} in terms of its elements is

$$\Sigma^{-1} = \begin{bmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{12} & \sigma^{22} \end{bmatrix}$$

The vectorization of such a matrix is

$$vec\left(\Sigma^{-1}\right) = \left[\sigma^{11}, \sigma^{12}, \sigma^{12}, \sigma^{22}\right]'$$

but being Σ^{-1} a symmetric matrix, the shorter form

$$vech\left(\Sigma^{-1}\right) = \left[\sigma^{11}, \sigma^{12}, \sigma^{22}\right]'$$

has been used in practice. The information matrix derived from the likelihood (3.11), also called information matrix of the *observable model*, will be indicated as I; the information matrix derived from the multivariate normal with complete variables, also called information matrix of the *latent model*, will be indicated as I^*. They are, respectively

$$I = E\left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\theta)}{\partial\theta\partial\theta'}\right] = \begin{bmatrix} [I_{\Pi\Pi}] & [0] \\ [0] & [I_{\Sigma\Sigma}] \end{bmatrix}$$

$$= E\begin{bmatrix} \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Pi)'}\right] & \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Sigma^{-1})'}\right] \\ \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Pi)'}\right] & \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] \end{bmatrix} \tag{6.47}$$

$$I^* = E\left[-\frac{\partial^2 \log f(\widetilde{Y}|X;\theta)}{\partial\theta\partial\theta'}\right] = \begin{bmatrix} [I^*_{\Pi\Pi}] & [0] \\ [0] & [I^*_{\Sigma\Sigma}] \end{bmatrix}$$

$$= E\begin{bmatrix} \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Pi)'}\right] & \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Sigma^{-1})'}\right] \\ \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Pi)'}\right] & \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] \end{bmatrix} \tag{6.48}$$

Each expression inside small square brackets, in (6.47) and (6.48), represents a block of the matrix inside big brackets. The upper-left block has dimensions (2×2), the lower-right block has dimensions (3×3), and the two off-diagonal blocks have dimensions (2×3) –the upper-right- and (3×2) –the lower-left. The steps performed in order to explicit the expressions (6.47) and (6.48) are in Appendix 3; here we report only the main results.

The off diagonal blocks, for both (6.47) and (6.48), are identically zero (we omit the proof because of its simplicity).

For the observable model we have:

$$I_{\Pi\Pi} = E\left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Pi)'}\right] = \Sigma^{-1}\begin{bmatrix} \sum\limits_{i\in A,B} x_i^2, & \frac{\sigma_{12}}{\sigma_{22}}\sum\limits_{i\in C} x_i^2 \\ \frac{\sigma_{12}}{\sigma_{11}}\sum\limits_{i\in B} x_i^2, & \sum\limits_{i\in A,C} x_i^2 \end{bmatrix}$$

$$I_{\Sigma\Sigma} = E\left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] = \begin{bmatrix} I_{11} & I_{21} & I_{31} \\ I_{21} & I_{22} & I_{32} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}$$

where

$$I_{11} = \left(\frac{n_A + n_B}{2}\right)\sigma_{11}^2 + \left(\frac{n_C}{2}\right)\frac{\sigma_{12}^4}{\sigma_{22}^2}$$

$$I_{21} = (n_A + n_B)\,\sigma_{11}\sigma_{12} + n_C\frac{\sigma_{12}^3}{\sigma_{22}}$$

$$I_{22} = n_A\left(\sigma_{12}^2 + \sigma_{11}\sigma_{22}\right) + 2(n_B + n_C)\sigma_{12}^2$$

$$I_{31} = \frac{n_A + n_B + n_C}{2}\sigma_{12}^2$$

$$I_{32} = (n_A + n_C)\,\sigma_{12}\sigma_{22} + n_B\frac{\sigma_{12}^3}{\sigma_{11}}$$

$$I_{33} = \left(\frac{n_A + n_C}{2}\right)\sigma_{22}^2 + \left(\frac{n_B}{2}\right)\frac{\sigma_{12}^4}{\sigma_{11}^2}$$

For the latent model we have:

$$I_{\Pi\Pi}^* = E\left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Pi)'}\right] = \Sigma^{-1}\begin{bmatrix} \sum_i x_i^2, & 0 \\ 0 & \sum_i x_i^2 \end{bmatrix}$$

$$I_{\Sigma\Sigma}^* = E\left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] = n\begin{bmatrix} \frac{\sigma_{11}^2}{2} & \sigma_{11}\sigma_{12} & \frac{\sigma_{12}^2}{2} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 + \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{22} \\ \frac{\sigma_{12}^2}{2} & \sigma_{12}\sigma_{22} & \frac{\sigma_{22}^2}{2} \end{bmatrix}$$

The asymptotic variance-covariance matrix for a simulated scores estimator, which is discussed in Appendix 2, is

$$V_{as}\left[\sqrt{n}\left(\widehat{\theta} - \theta\right)\right] = I^{-1} + I^{-1}\left(I^* - I\right)I^{-1} \tag{6.49}$$

It is clear from (6.49) that there is a loss of efficiency with respect to maximum likelihood, whose asymptotic variance-covariance matrix would be I^{-1}. Roughly speaking, we can say that there is price that must be paid for the simulation, and it is $I^{-1}\left(I^* - I\right)I^{-1}$. We notice that it is proportional to the difference between the information of the latent and observable model (so, if the latent and the observable model were the same -no missing data- the difference would be zero and obviously there would be no loss of efficiency, but there would be no simulation). The difference $(I^* - I)$ is

$$(I^* - I) = \begin{bmatrix} [I_{\Pi\Pi}^*] & [0] \\ [0] & [I_{\Sigma\Sigma}^*] \end{bmatrix} - \begin{bmatrix} [I_{\Pi\Pi}] & [0] \\ [0] & [I_{\Sigma\Sigma}] \end{bmatrix}$$

and, subtracting the corresponding blocks, we have (formulas are derived in Appendix 3):

$$I^*_{\Pi\Pi} - I_{\Pi\Pi} = \Sigma^{-1} \begin{bmatrix} \sum\limits_{i \in C,D} x_i^2, & -\frac{\sigma_{12}}{\sigma_{22}} \sum\limits_{i \in C} x_i^2 \\ -\frac{\sigma_{12}}{\sigma_{11}} \sum\limits_{i \in B} x_i^2, & \sum\limits_{i \in B,D} x_i^2 \end{bmatrix}$$

$$I^*_{\Sigma\Sigma} - I_{\Sigma\Sigma} = \begin{bmatrix} \frac{n_C}{2} \frac{\sigma_{11}^2 \sigma_{22}^2 - \sigma_{12}^4}{\sigma_{22}^2}, & -, & - \\ n_C \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}} \sigma_{12}, & (n_B + n_C)(\sigma_{11}\sigma_{22} - \sigma_{12}^2), & - \\ 0 & n_B(\sigma_{11}\sigma_{22} - \sigma_{12}^2)\frac{\sigma_{12}}{\sigma_{11}}, & \frac{n_B}{2} \frac{\sigma_{11}^2\sigma_{22}^2 - \sigma_{12}^4}{\sigma_{11}^2} \end{bmatrix}$$

$$+ \begin{bmatrix} \frac{n_D}{2}\sigma_{11}^2, & -, & - \\ n_D \sigma_{11}\sigma_{12}, & n_D(\sigma_{12}^2 + \sigma_{11}\sigma_{22}), & - \\ \frac{n_D}{2}\sigma_{12}^2, & n_D \sigma_{12}\sigma_{22}, & \frac{n_D}{2}\sigma_{22}^2 \end{bmatrix}$$

which are both positive semidefinite matrices, so the difference $(I^* - I)$ is positive semidefinite (proof in Appendix 3): the larger this difference, the larger the loss of efficiency with respect to maximum likelihood. Although appearing in a complicated form, this difference is larger when larger is the contribution of the sections with missing data B, C and D, while section A (complete data) does not contribute at all.

6.6 Improving asymptotic efficiency

What has been discussed till now is *one replication*, till convergence, of the iterative estimation/imputation procedure. We may replicate the procedure S times independently, each time till convergence. Then we average the S estimates of $\widehat{\theta}$, obtaining the estimate $\widehat{\theta}_S$. Its asymptotic variance-covariance matrix would be (see Appendix 2 and Gourieroux, Monfort, 1996, p. 36)

$$V_{as}\left[\sqrt{n}\left(\widehat{\theta}_S - \theta\right)\right] = I^{-1} + \frac{1}{S} I^{-1}(I^* - I) I^{-1} \tag{6.50}$$

which is smaller than the variance-covariance matrix given in (6.49), corresponding to one replication only. As it usually happens when an estimator is computed by simulation, its efficiency increases when the number of replications increases. Thus, the way to improve the efficiency of the estimator is, also in our case, to increase S. It is clear that if $S \to \infty$, the asymptotic variance-covariance matrix is I^{-1}; in other words, the simulated scores estimator would reach the efficiency of the ML estimator.

Only after having computed $\widehat{\theta}_S$ we obtain the "optimum" imputed data performing the imputation by the usual function defined in Section 4. In this way the procedure to get the "optimum" imputed data proceeds through two steps: the first step is repeated S times in order to compute $\widehat{\theta}_S$ (the imputations in this step are only instruments to produce the parameter estimates), the second step needs only to perform imputation with the optimum $\widehat{\theta}_S$.

7 Multiple imputation

This section does not aim at explaining the multiple imputation (MI) technique with its characteristics and properties, the real aim is to explain how to create multiple imputations by using the "least-squares estimation/multivariate normal imputation". Using such a procedure, we estimate the unknown parameters of the encompassing multivariate model for the entire dataset and we consider as definitive completed data those obtained when the convergence of the estimation procedure is achieved. Repeating m times the procedure, we can build a chain of estimates $\left(\widehat{\theta}^1, \widehat{\theta}^2, ..., \widehat{\theta}^m\right)$ and the corresponding chain of simulated values $\left(\widetilde{Y}^1, \widetilde{Y}^2, ..., \widetilde{Y}^m\right)$. In this way, each of the m completed data set is analyzed by standard complete data methods.

The overall procedure for computing multiple imputation proceeds in two steps. First, m simulated versions of the missing data are created under a data model. Second, the m complete datasets are analyzed by complete data statistical techniques, and the results are combined to produce one overall inference. Sometimes the analysis of the second stage involves different models than the one used to produce imputions at the first step; this is not a serious problem, especially if the amount of missing information is not too large. Infact, multiple imputation is robust to departures from the complete-data model. Hence, even if the model examined in this paper may not realistically describe many datasets relatively to real applications, still it is a useful framework for imputing missing data for continuous variables.

8 Conclusion

In this paper we have introduced a method for creating multiple imputations assuming data generated by a multivariate normal linear regression model, assuming an arbitray pattern for the missing values and an ignorable missing data mechanism. The method is essentially based on an iterative "least-squares estimation/multivariate normal imputation" procedure. The method seems to be friendly for the data analyst, as it computes the parameters by estimating independently each equation of a system by OLS , and it performs the imputation by the "reasonably simple" function defined in Section 4. Besides its technical simplicity and feasibility, the peculiarity of the method is in the properties of the estimator. First of all the parameters estimator is consistent and asymptotically normal. Moreover, beeing a simulated scores estimator, its efficiency can be improved by increasing the number of replications. Finally, the estimator becomes as efficient as maximum likelihood if the iterative procedure is replicated a sufficiently large number of times, each time iterating to convergence.

Obviously, the optimum solution when dealing with missing data should be performing multiple imputation building each imputed data set with the optimum estimator $\widehat{\theta}_S$. This might require a large computational effort. But also using the estimator produced by *one replication only* to perform imputation of each final dataset is inferentially superior to the other method easily implementable (complete case, single imputation, etc.). This is surely an advantage of the method discussed in this paper.

We dealt with the missing data problem in the context of a linear normal model in which some observations of some variables (treated as "endogenous" variables) were missing, while other variables (treated as "exogenous" variables) were completely observed. The approach presented could be used in a similar way if some exogenous variables are missing and we can explain them by auxiliary linear relationship involving other exogenous variables.

Of course, in many real cases missing data do not affect only continuous variables. The problem exists also for categorical data, count data, or censored variables: generalization of the method here proposed is left to future research.

Appendix 1: Feasible imputation for bivariate normal data: reduced form approach

Iteration 0.

The two reduced form equations (3.8, or 6.32), considered as two normal linear regression models on the observed part of the data matrix (A), are estimated using OLS

$$\begin{array}{l} y_{i1} = \Pi_1 x_i + e_{i1} \\ y_{i2} = \Pi_2 x_i + e_{i2} \end{array}, i \in A$$

obtaining the initial estimates of the reduced form coefficients:

$$\widehat{\Pi}'^{(0)} = \left[\widehat{\Pi}_1^{(0)}, \widehat{\Pi}_2^{(0)}\right] = \left(X_A' X_A\right)^{-1} X_A' Y_A$$

and (without degrees of freedom correction)

$$\widehat{\Sigma}^{(0)} = \begin{bmatrix} \widehat{\sigma}_{11}^{(0)} & \widehat{\sigma}_{12}^{(0)} \\ \widehat{\sigma}_{12}^{(0)} & \widehat{\sigma}_{22}^{(0)} \end{bmatrix} = \frac{1}{n_A}\left(\widehat{E}_A'^{(0)} \widehat{E}_A^{(0)}\right)$$

as the residual variances, where $\widehat{E}_A^{(0)} = Y_A - X_A \widehat{\Pi}'^{(0)}$.

Missing values in section B and C are imputed as values generated from the conditional distribution of the variable with missing data given the corresponding observed variables, while missing values of the D section are generated from the bivariate distribution. According to the explained procedure, the first values of the completed data are the following:

$$\begin{array}{l} y_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}}\ u_{i1} \\ y_{i2} = \Pi_2 x_i + \frac{\sigma_{12}}{\sigma_{11}}(y_{i1} - \Pi_1 x_i) + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\ u_{i2} \end{array}, i \in A$$

$$
\begin{aligned}
y_{i1} &= \Pi_1 x_i + \sqrt{\sigma_{11}}\ u_{i1} \\
\tilde{y}_{i2}^{(0)} &= \widehat{\Pi}_2^{(0)} x_i + \frac{\widehat{\sigma}_{12}^{(0)}}{\widehat{\sigma}_{11}^{(0)}}(y_{i1} - \widehat{\Pi}_1^{(0)} x_i) + \sqrt{\widehat{\sigma}_{22}^{(0)} - \frac{\widehat{\sigma}_{12}^{(0)2}}{\widehat{\sigma}_{11}^{(0)}}}\ \tilde{u}_{i2}
\end{aligned}
\quad , i \in B
$$

$$
\begin{aligned}
y_{i2} &= \Pi_2 x_i + \sqrt{\sigma_{22}}\ u_{i2} \\
\tilde{y}_{i1}^{(0)} &= \widehat{\Pi}_1^{(0)} x_i + \frac{\widehat{\sigma}_{12}^{(0)}}{\widehat{\sigma}_{22}^{(0)}}(y_{i2} - \widehat{\Pi}_2^{(0)} x_i) + \sqrt{\widehat{\sigma}_{11}^{(0)} - \frac{\widehat{\sigma}_{12}^{(0)2}}{\widehat{\sigma}_{22}^{(0)}}}\ \tilde{u}_{i1}
\end{aligned}
\quad , i \in C
$$

$$
\begin{aligned}
\tilde{y}_{i1}^{(0)} &= \widehat{\Pi}_1^{(0)} x_i + \sqrt{\widehat{\sigma}_{11}^{(0)}}\ \tilde{u}_{i1} \\
\tilde{y}_{i2}^{(0)} &= \widehat{\Pi}_2^{(0)} x_i + \frac{\widehat{\sigma}_{12}^{(0)}}{\widehat{\sigma}_{11}^{(0)}}\sqrt{\widehat{\sigma}_{11}^{(0)}}\ \tilde{u}_{i1} + \sqrt{\widehat{\sigma}_{22}^{(0)} - \frac{\widehat{\sigma}_{12}^{(0)2}}{\widehat{\sigma}_{11}^{(0)}}}\ \tilde{u}_{i2}
\end{aligned}
\quad , i \in D
$$

The dataset is now completed and it can be represented as in the following matrix

$$
\tilde{Y}^{(0)} = \begin{bmatrix} Y_{A1} & Y_{A2} \\ Y_{B1} & \tilde{Y}_{B2}^{(0)} \\ \tilde{Y}_{C1}^{(0)} & Y_{C2} \\ \tilde{Y}_{D1}^{(0)} & \tilde{Y}_{D2}^{(0)} \end{bmatrix}
$$

Also the matrix of residuals can now be completed (so far, residuals were available only in block A) as $\widehat{E}^{(0)} = \tilde{Y}^{(0)} - X\widehat{\Pi}'^{(0)}$.

Iteration 1

The two normal linear regressions on the data (A+B+C+D) completed in iteration 0 are estimated obtaining $\widehat{\Pi}'^{(1)} = \left[\widehat{\Pi}_1^{(1)}, \widehat{\Pi}_2^{(1)}\right]$ and $\widehat{\Sigma}^{(1)}$. Since the completed $\tilde{Y}_1^{(0)}$ can be represented as $\tilde{Y}_1^{(0)} = X\widehat{\Pi}_1'^{(0)} + \widehat{e}_1^{(0)}$, OLS estimation gives

$$
\widehat{\Pi}_1'^{(1)} = (X'X)^{-1} X'\tilde{Y}_1^{(0)} = \widehat{\Pi}_1'^{(0)} + (X'X)^{-1} X' \begin{bmatrix} \widehat{e}_{A1}^{(0)} \\ \widehat{e}_{B1}^{(0)} \\ \frac{\widehat{\sigma}_{12}^{(0)}}{\widehat{\sigma}_{22}^{(0)}}\ \widehat{e}_{C2}^{(0)} + \sqrt{\widehat{\sigma}_{11}^{(0)} - \frac{\widehat{\sigma}_{12}^{(0)2}}{\widehat{\sigma}_{22}^{(0)}}}\ \tilde{u}_{C1} \\ \sqrt{\widehat{\sigma}_{11}^{(0)}}\ \tilde{u}_{D1} \end{bmatrix}
$$

where the vector $\widehat{e}_1^{(0)}$ has been explicitly divided into its four components: $\widehat{e}_{A1}^{(0)}$ and $\widehat{e}_{B1}^{(0)}$ are the residuals of iteration 0 related to the fully observable y_{i1} of sections A and B; in sections C and D the y_{i1} are not observable, so they have been replaced by the values imputed in iteration 0 (notice that, when $i \in C$, $\widehat{e}_{i2}^{(0)} = y_{i2} - \widehat{\Pi}_2^{(0)} x_i$, being y_{i2} fully observable).

Analogously, when estimating Π_2 we have

$$\widehat{\Pi}_2'^{(1)} = \widehat{\Pi}_2'^{(0)} + (X'X)^{-1} X' \begin{bmatrix} \widehat{e}_{A2}^{(0)} \\ \frac{\widehat{\sigma}_{12}^{(0)}}{\widehat{\sigma}_{11}^{(0)}} \widehat{e}_{B1}^{(0)} + \sqrt{\widehat{\sigma}_{22}^{(0)} - \frac{\widehat{\sigma}_{12}^{(0)2}}{\widehat{\sigma}_{11}^{(0)}}} \; \widetilde{u}_{B2} \\ \widehat{e}_{C2}^{(0)} \\ \frac{\widehat{\sigma}_{12}^{(0)}}{\widehat{\sigma}_{11}^{(0)}} \sqrt{\widehat{\sigma}_{11}^{(0)}} \; \widetilde{u}_{D1} + \sqrt{\widehat{\sigma}_{22}^{(0)} - \frac{\widehat{\sigma}_{12}^{(0)2}}{\widehat{\sigma}_{11}^{(0)}}} \; \widetilde{u}_{D2} \end{bmatrix}$$

$$\widehat{\Sigma}^{(1)} = \begin{bmatrix} \widehat{\sigma}_{11}^{(1)} & \widehat{\sigma}_{12}^{(1)} \\ \widehat{\sigma}_{12}^{(1)} & \widehat{\sigma}_{22}^{(1)} \end{bmatrix} = \frac{1}{n} \left[\widetilde{Y}^{(0)} - X\widehat{\Pi}'^{(1)}\right]' \left[\widetilde{Y}^{(0)} - X\widehat{\Pi}'^{(1)}\right]$$

Missing values in section B and C are imputed again as values generated from the conditional distribution of the variable with missing data given the corresponding observed variables, while missing values of the D section are generated from the bivariate distribution. According to the explained procedure values of the completed data are the following:

$$\begin{matrix} y_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}} \; u_{i1} \\ y_{i2} = \Pi_2 x_i + \frac{\sigma_{12}}{\sigma_{11}}(y_{i1} - \Pi_1 x_i) + \sqrt{\sigma_{22} \frac{\sigma_{12}^2}{\sigma_{11}}} \; u_{i2} \end{matrix} \quad , i \in A$$

$$\begin{matrix} y_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}} \; u_{i1} \\ \widetilde{y}_{i2}^{(1)} = \widehat{\Pi}_2^{(1)} x_i + \frac{\widehat{\sigma}_{12}^{(1)}}{\widehat{\sigma}_{11}^{(1)}}(y_{i1} - \widehat{\Pi}_1^{(1)} x_i) + \sqrt{\widehat{\sigma}_{22}^{(1)} - \frac{\widehat{\sigma}_{12}^{(1)2}}{\widehat{\sigma}_{11}^{(1)}}} \; \widetilde{u}_{i2} \end{matrix} \quad , i \in B$$

$$\begin{matrix} y_{i2} = \Pi_2 x_i + \sqrt{\sigma_{22}} \; u_{i2} \\ \widetilde{y}_{i1}^{(1)} = \widehat{\Pi}_1^{(1)} x_i + \frac{\widehat{\sigma}_{12}^{(1)}}{\widehat{\sigma}_{22}^{(1)}}(y_{i2} - \widehat{\Pi}_2^{(1)} x_i) + \sqrt{\widehat{\sigma}_{11}^{(1)} - \frac{\widehat{\sigma}_{12}^{(1)2}}{\widehat{\sigma}_{22}^{(1)}}} \; \widetilde{u}_{i1} \end{matrix} \quad , i \in C$$

$$\begin{matrix} \widetilde{y}_{i1}^{(1)} = \widehat{\Pi}_1^{(1)} x_i + \sqrt{\widehat{\sigma}_{11}^{(1)}} \; \widetilde{u}_{i1} \\ \widetilde{y}_{i2}^{(1)} = \widehat{\Pi}_2^{(1)} x_i + \frac{\widehat{\sigma}_{12}^{(1)}}{\widehat{\sigma}_{11}^{(1)}} \sqrt{\widehat{\sigma}_{11}^{(1)}} \; \widetilde{u}_{i1} + \sqrt{\widehat{\sigma}_{22}^{(1)} - \frac{\widehat{\sigma}_{12}^{(1)2}}{\widehat{\sigma}_{11}^{(1)}}} \; \widetilde{u}_{i2} \end{matrix} \quad , i \in D$$

The dataset has been updated, it can be represented as in the following matrix

$$\widetilde{Y}^{(1)} = \begin{bmatrix} Y_{A1} & Y_{A2} \\ Y_{B1} & \widetilde{Y}_{B2}^{(1)} \\ \widetilde{Y}_{C1}^{(1)} & Y_{C2} \\ \widetilde{Y}_{D1}^{(1)} & \widetilde{Y}_{D2}^{(1)} \end{bmatrix}$$

Iteration k

The two normal linear regressions on the data (A+B+C+D) completed at the (k-1) iteration are estimated obtaining $\widehat{\Pi}'^{(k)} = \left[\widehat{\Pi}_1^{(k)}, \widehat{\Pi}_2^{(k)}\right]$ and $\widehat{\Sigma}_1^{(k)}$, as follows:

$$\widehat{\Pi}_1'^{(k)} = \widehat{\Pi}_1'^{(k-1)} + (X'X)^{-1} X' \begin{bmatrix} \widehat{e}_{A1}^{(k-1)} \\ \widehat{e}_{B1}^{(k-1)} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}} \widehat{e}_{C2}^{(k-1)} + \sqrt{\widehat{\sigma}_{11}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{22}^{(k-1)}}} \widetilde{u}_{C1} \\ \sqrt{\widehat{\sigma}_{11}^{(k-1)}} \widetilde{u}_{D1} \end{bmatrix}$$

$$\widehat{\Pi}_2'^{(k)} = \widehat{\Pi}_2'^{(k-1)} + (X'X)^{-1} X' \begin{bmatrix} \widehat{e}_{A2}^{(k-1)} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}} \widehat{e}_{B1}^{(k-1)} + \sqrt{\widehat{\sigma}_{22}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{11}^{(k-1)}}} \widetilde{u}_{B2} \\ \widehat{e}_{C2}^{(k-1)} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}} \sqrt{\widehat{\sigma}_{11}^{(k-1)}} \widetilde{u}_{D1} + \sqrt{\widehat{\sigma}_{22}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{11}^{(k-1)}}} \widetilde{u}_{D2} \end{bmatrix}$$

$$\begin{aligned} \widehat{\Sigma}^{(k)} &= \begin{bmatrix} \widehat{\sigma}_{11}^{(k)} & \widehat{\sigma}_{12}^{(k)} \\ \widehat{\sigma}_{12}^{(k)} & \widehat{\sigma}_{22}^{(k)} \end{bmatrix} = \frac{1}{n}\left[\widetilde{Y}^{(k-1)} - X\widehat{\Pi}'^{(k)}\right]'\left[\widetilde{Y}^{(k-1)} - X\widehat{\Pi}'^{(k)}\right] = \\ &= \frac{1}{n} \begin{bmatrix} Y_{A1} - X_A\widehat{\Pi}_1^{(k)} & Y_{A2} - X_A\widehat{\Pi}_2^{(k)} \\ Y_{B1} - X_B\widehat{\Pi}_1^{(k)} & \widetilde{Y}_{B2} - X_B\widehat{\Pi}_2^{(k)} \\ \widetilde{Y}_{C1} - X_C\widehat{\Pi}_1^{(k)} & Y_{C2} - X_C\widehat{\Pi}_2^{(k)} \\ \widetilde{Y}_{D1} - X_D\widehat{\Pi}_1^{(k)} & \widetilde{Y}_{D2} - X_D\widehat{\Pi}_2^{(k)} \end{bmatrix}' \begin{bmatrix} Y_{A1} - X_A\widehat{\Pi}_1^{(k)} & Y_{A2} - X_A\widehat{\Pi}_2^{(k)} \\ Y_{B1} - X_B\widehat{\Pi}_1^{(k)} & \widetilde{Y}_{B2} - X_B\widehat{\Pi}_2^{(k)} \\ \widetilde{Y}_{C1} - X_C\widehat{\Pi}_1^{(k)} & Y_{C2} - X_C\widehat{\Pi}_2^{(k)} \\ \widetilde{Y}_{D1} - X_D\widehat{\Pi}_1^{(k)} & \widetilde{Y}_{D2} - X_D\widehat{\Pi}_2^{(k)} \end{bmatrix} \end{aligned}$$

In particular, we display the explicit expression of the element $(1,1)$ (the others would be analogous):

$$\begin{aligned} n\widehat{\sigma}_{11}^{(k)} &= \widehat{e}_{A1}^{(k)\prime}\widehat{e}_{A1}^{(k)} + \widehat{e}_{B1}^{(k)\prime}\widehat{e}_{B1}^{(k)} + \left(\widehat{\Pi}_1^{(k-1)} - \widehat{\Pi}_1^{(k)}\right)^2 \sum_{i\in C} x_i^2 \\ &+ \left(\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}}\right)^2 \widehat{e}_{C2}^{(k-1)\prime}\widehat{e}_{C2}^{(k-1)} + \left(\widehat{\sigma}_{11}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{22}^{(k-1)}}\right) \widetilde{u}_{C1}' \widetilde{u}_{C1} \\ &+2\left(\widehat{\Pi}_1^{(k-1)} - \widehat{\Pi}_1^{(k)}\right)\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}} \sum_{i\in C} x_i \,\widehat{e}_{i2}^{(k-1)} + 2\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}}\sqrt{\widehat{\sigma}_{11}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{22}^{(k-1)}}}\; \widehat{e}_{C2}'^{(k-1)} \widetilde{u}_{C1} \\ &+2\left(\widehat{\Pi}_1^{(k-1)} - \widehat{\Pi}_1^{(k)}\right)\sqrt{\widehat{\sigma}_{11}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{22}^{(k-1)}}} \sum_{i\in C} x_i\widetilde{u}_{i1} \\ &+ \left(\widehat{\Pi}_1^{(k-1)} - \widehat{\Pi}_1^{(k)}\right)^2 \sum_{i\in D} x_i^2 + \widehat{\sigma}_{11}^{(k-1)}\, \widetilde{u}_{D1}'\, \widetilde{u}_{D1} + 2\left(\widehat{\Pi}_1^{(k-1)} - \widehat{\Pi}_1^{(k)}\right)\sqrt{\widehat{\sigma}_{11}^{(k-1)}} \sum_{i\in D} x_i \,\widetilde{u}_{i1} \end{aligned}$$

Missing values are imputed as in the previous iterations, so the completed data are now

$$\begin{matrix} y_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}}\; u_{i1} \\ y_{i2} = \Pi_2 x_i + \frac{\sigma_{12}}{\sigma_{11}}(y_{i1} - \Pi_1 x_i) + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\; u_{i2} \end{matrix} \quad , i \in A$$

$$
\begin{aligned}
y_{i1} &= \Pi_1 x_i + \sqrt{\sigma_{11}}\ u_{i1} \\
\tilde{y}_{i2}^{(k)} &= \widehat{\Pi}_2^{(k)} x_i + \frac{\widehat{\sigma}_{12}^{(k)}}{\widehat{\sigma}_{11}^{(k)}}(y_{i1} - \widehat{\Pi}_1^{(k)} x_i) + \sqrt{\widehat{\sigma}_{22}^{(k)} - \frac{\widehat{\sigma}_{12}^{(k)2}}{\widehat{\sigma}_{11}^{(k)}}}\ \tilde{u}_{i2}
\end{aligned}
\quad , i \in B
$$

$$
\begin{aligned}
y_{i2} &= \Pi_2 x_i + \sqrt{\sigma_{22}}\ u_{i2} \\
\tilde{y}_{i1}^{(k)} &= \widehat{\Pi}_1^{(k)} x_i + \frac{\widehat{\sigma}_{12}^{(k)}}{\widehat{\sigma}_{22}^{(k)}}(y_{i2} - \widehat{\Pi}_2^{(k)} x_i) + \sqrt{\widehat{\sigma}_{11}^{(k)} - \frac{\widehat{\sigma}_{12}^{(k)2}}{\widehat{\sigma}_{22}^{(k)}}}\ \tilde{u}_{i1}
\end{aligned}
\quad , i \in C
$$

$$
\begin{aligned}
\tilde{y}_{i1}^{(k)} &= \widehat{\Pi}_1^{(k)} x_i + \sqrt{\widehat{\sigma}_{11}^{(k)}}\ \tilde{u}_{i1} \\
\tilde{y}_{i2}^{(k)} &= \widehat{\Pi}_2^{(k)} x_i + \frac{\widehat{\sigma}_{12}^{(k)}}{\widehat{\sigma}_{11}^{(k)}}\sqrt{\widehat{\sigma}_{11}^{(k)}}\ \tilde{u}_{i1} + \sqrt{\widehat{\sigma}_{22}^{(k)} - \frac{\widehat{\sigma}_{12}^{(k)2}}{\widehat{\sigma}_{11}^{(k)}}}\ \tilde{u}_{i2}
\end{aligned}
\quad , i \in D
$$

so the whole dataset is now

$$
\widetilde{Y}^{(k)} = \begin{bmatrix} Y_{A1} & Y_{A2} \\ Y_{B1} & \widetilde{Y}_{B2}^{(k)} \\ \widetilde{Y}_{C1}^{(k)} & Y_{C2} \\ \widetilde{Y}_{D1}^{(k)} & \widetilde{Y}_{D2}^{(k)} \end{bmatrix}
$$

The parameters estimated at the current iteration and at the previous iteration are compared. If they are equal, or very close to each other, the estimation/imputation procedure has come to its end, otherwise the procedure continues.

Appendix 2: The method of simulated scores

Let us consider a complete data model, where y_i is the set of the endogenous variables, x_i is the set of the exogenous variables and $f(y_i|x_i, \theta)$ denotes the conditional p.d.f. of y_i given x_i.

The application of the GMM (Generalized Method of Moments, Hansen, 1982) requires a closed form for the specification of the moments. Sometimes such a closed form does not exist and it can be replaced by an approximation, based on observations and simulations, called *simulator* . The derivation of the GMM estimator is based on empirical moments, computed *only* on observations: parameters estimates are obtained matching the empirical moments with the theoretical moments (or minimizing their differences). The MSM estimator (Method of Simulated Moments, McFadden, 1989; Pakes and Pollard, 1989) is derived in the same way, but using the simulator, thus using a different empirical moment that includes not only observations, but also simulated values.

Hajivassiliou and McFadden (1990), Hajivassiliou (1993), Hajivassiliou and Ruud (1994), Stern (2000) consider a particular case. Maximum Likelihood can be viewed as a GMM estimator, where the theoretical moments restriction is that, at the *true* parameter value, the expected value of the score must be zero

$$E\left[\left.\frac{\partial \log f(y_i|x_i;\theta)}{\partial \theta}\right| x_i\right] = 0$$

If the p.d.f (and thus also the score) has an intractable form, and $g\left(y_i, x_i, \widetilde{u}_i; \theta\right)$ is an *unbiased simulator of the score* (Gouriéroux and Monfort, 1996, example 2.3), that is if

$$E\left[g\left(y_i, x_i, \widetilde{u}_i; \theta\right)|y_{i,}x_i\right] = \frac{\partial \log f(y_i|x_i;\theta)}{\partial \theta},$$

then the *method of simulated scores* estimator of θ (MSS) is the MSM estimator based on the particular simulated moment g (the simulated random term $\widetilde{u}_i$ has a known and fixed distribution conditional on y_i, x_i; expectation is with respect to the simulation process).

A natural way to construct an unbiased simulator of the score arises for latent variable models. Hajivassiliou and Ruud (1994, sec. 2.6 and 4.4) show that every score function can be expressed as the expectation of the score of a latent data generating process taken conditional on the observed data. In other words, we may consider a latent model with variables $(y_i^*, x_i), i = 1, ..., n$, and denote by $f^*(y_i^*|x_i;\theta)$ the conditional p.d.f. of y_i^* given x_i (Gouriéroux and Monfort, 1996, example 2.4). If the endogenous observable variables are a known function of the latent variables y_i^*: $y_i = a(y_i^*)$, then it can be proved that the score function is such that:

$$E_\theta\left[\left.\frac{\partial \log f^*(y_i^*|x_i;\theta)}{\partial \theta}\right| y_i, x_i\right] = \frac{\partial \log f(y_i|x_i;\theta)}{\partial \theta}$$

So it would be natural to consider the previous equality as an unbiasedness condition and to propose the unbiased simulator $\partial \log f^*(y_i^{*s}|x_i;\theta)/\partial\theta$ where y_i^{*s} is produced by simulation, drawing from the conditional distribution of y_i^* given y_i and x_i.

However, "natural" does not necessarily imply "easy". (Stern, 2000, p.25) observes that, in practice, it is not always easy to construct an unbiased estimator of the score. Still in example 2.4, Gouriéroux and Monfort (1996) observe that such an unbiased simulator is used, in practice, only when if the above random drawing from the conditional distribution is simple. And this is in fact the case we treat in this paper, where the search for an appropriate and manageable latent model is rather straightforward.

The simulated scores estimator $\widehat{\theta}_S$ based on S replications and sample size n, being a MSM estimator, when n goes to infinity and S is fixed, is consistent and asymptotically normal (Gouriéroux and Monfort, 1996); about the efficiency the same authors (p. 36) show that the asymptotic variance covariance matrix is

$$V_{as}\left[\sqrt{n}\left(\widehat{\theta}_S - \theta\right)\right] = I^{-1} + \frac{1}{S}I^{-1}\left(I^* - I\right)I^{-1} \tag{A2.51}$$

where I and I^* are respectively the information matrix of the observable model,

$$I = E\left[-\frac{\partial^2 \log f(y_i|x_i;\theta)}{\partial\theta\partial\theta'}\right]$$

and the information matrix of the latent model:

$$I^* = E\left[-\frac{\partial^2 \log f^*(y_i^*|x_i;\theta)}{\partial\theta\partial\theta'}\right].$$

The price that must be paid for the simulation is $I^{-1}(I^*-I)\,I^{-1}/S$; as usual for simulation-based estimators, it decreases with S; moreover, it is proportional to the difference between the information of the latent and observable model.

Appendix 3: The information matrices

In this Appendix we report all the steps performed to explicit the information matrix of the observable and of the latent model. We rewrite here the expressions for I (6.47) and I^* (6.48), respectively

$$I = E\left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\theta)}{\partial\theta\partial\theta'}\right] = \begin{bmatrix} [I_{\Pi\Pi}] & [0] \\ [0] & [I_{\Sigma\Sigma}] \end{bmatrix}$$

$$= E\begin{bmatrix} \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Pi)'}\right] & \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Sigma^{-1})'}\right] \\ \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Pi)'}\right] & \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] \end{bmatrix} \tag{A3.52}$$

$$I^* = E\left[-\frac{\partial^2 \log f(\widetilde{Y}|X;\theta)}{\partial\theta\partial\theta'}\right] = \begin{bmatrix} [I^*_{\Pi\Pi}] & [0] \\ [0] & [I^*_{\Sigma\Sigma}] \end{bmatrix}$$

$$= E\begin{bmatrix} \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Pi)'}\right] & \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Sigma^{-1})'}\right] \\ \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Pi)'}\right] & \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] \end{bmatrix} \tag{A3.53}$$

Expliciting the block $(1,1)$ of the matrix on the right hand side of the (A3.52), we have

$$\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Pi)'} = -\Sigma^{-1}\begin{bmatrix} \sum_{i\in A,B} x_i^2, & \frac{\sigma_{12}}{\sigma_{22}}\sum_{i\in C} x_i^2 \\ \frac{\sigma_{12}}{\sigma_{11}}\sum_{i\in B} x_i^2, & \sum_{i\in A,C} x_i^2 \end{bmatrix}$$

which remains unchanged when applying the expected value

$$I_{\Pi\Pi} = E\left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\left(vech\Pi vech\Pi'\right)}\right] = \Sigma^{-1}\begin{bmatrix} \sum_{i\in A,B} x_i^2, & \frac{\sigma_{12}}{\sigma_{22}}\sum_{i\in C} x_i^2 \\ \frac{\sigma_{12}}{\sigma_{11}}\sum_{i\in B} x_i^2, & \sum_{i\in A,C} x_i^2 \end{bmatrix}.$$

We indicate the block $(2,2)$ of the matrix on the right hand side of the (A3.52) as the following (3×3) matrix

$$[A_{lm}] = \begin{bmatrix} \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{11}\partial\sigma^{11}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{11}\partial\sigma^{12}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{11}\partial\sigma^{22}} \\ \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{12}\partial\sigma^{11}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{12}\partial\sigma^{12}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{12}\partial\sigma^{22}} \\ \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{22}\partial\sigma^{11}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{22}\partial\sigma^{12}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{22}\partial\sigma^{22}} \end{bmatrix}$$

$$l,m = 1,2,3 \tag{A3.54}$$

In order to compute (A3.54), we differentiate $\log f(Y_{obs}|X_{obs};\Pi,\Sigma)$ with respect to $vech(\Sigma^{-1})$ obtaining a (3×1) vector whose elements are labelled with A_l $(l = 1,2,3)$. Reminding that $\sigma_{12}/\sigma_{11} = -\sigma^{12}/\sigma^{22}$, $\sigma_{12}/\sigma_{22} = -\sigma^{12}/\sigma^{11}$, we can write the A_l elements as follows

$$\begin{aligned} A_1 &= \frac{n_A}{2}\sigma_{11} - \frac{1}{2}\sum_{i\in A}(y_{i1} - \Pi_1 x_i)^2 + \frac{1}{2}\left[n_B\sigma_{11} - \sum_{i\in B}(y_{i1} - \Pi_1 x_i)^2\right] \\ &+\frac{1}{2}\left(\frac{\sigma^{12}}{\sigma^{11}}\right)^2\left[n_C\sigma_{22} - \sum_{i\in C}(y_{i2} - \Pi_2 x_i)^2\right] \end{aligned}$$

$$\begin{aligned} A_2 &= n_A\sigma_{12} - \sum_{i\in A}(y_{i1} - \Pi_1 x_i)(y_{i2} - \Pi_2 x_i) - \frac{\sigma^{12}}{\sigma^{22}}\left[n_B\sigma_{11} - \sum_{i\in B}(y_{i1} - \Pi_1 x_i)^2\right] \\ &-\frac{\sigma^{12}}{\sigma^{11}}\left[n_C\sigma_{22} - \sum_{i\in C}(y_{i2} - \Pi_2 x_i)^2\right] \end{aligned}$$

$$\begin{aligned} A_3 &= \frac{n_A}{2}\sigma_{22} - \frac{1}{2}\sum_{i\in A}(y_{i1} - \Pi_1 x_i)^2 + \frac{1}{2}\left(\frac{\sigma^{12}}{\sigma^{22}}\right)^2\left[n_B\sigma_{11} - \sum_{i\in B}(y_{i1} - \Pi_1 x_i)^2\right] \\ &+\frac{1}{2}\left[n_C\sigma_{22} - \sum_{i\in C}(y_{i2} - \Pi_2 x_i)^2\right]. \end{aligned}$$

With further differentiation, we obtain the second derivatives, which are the elements of the (3×3) matrix (A3.54) (being symmetric, we explicit only the lower triangle):

$$A_{11} = -\frac{n_A}{2}\sigma_{11}^2 - \frac{1}{2}n_B\sigma_{11}^2 - \frac{(\sigma^{12})^2}{(\sigma^{11})^3}\left[n_C\sigma_{22} - \sum_{i\in C}(y_{i2} - \Pi_2 x_i)^2\right] + \frac{1}{2}\left(\frac{\sigma^{12}}{\sigma^{11}}\right)^2\left(-n_C\sigma_{12}^2\right)$$

$$A_{21} = -n_A\sigma_{11}\sigma_{12} - \frac{\sigma^{12}}{\sigma^{22}}\left(-n_B\sigma_{11}^2\right) + \frac{\sigma^{12}}{(\sigma^{11})^2}\left[n_C\sigma_{22} - \sum_{i\in C}(y_{i2} - \Pi_2 x_i)^2\right] - \frac{\sigma^{12}}{\sigma^{11}}\left(-n_C\sigma_{12}^2\right)$$

$$\begin{aligned} A_{22} &= -n_A\left(\sigma_{12}^2 + \sigma_{11}\sigma_{22}\right) - \frac{1}{\sigma^{22}}\left[n_B\sigma_{11} - \sum_{i\in B}(y_{i1} - \Pi_1 x_i)^2\right] - \frac{\sigma^{12}}{\sigma^{22}}(-2n_B\sigma_{11}\sigma_{12}) \\ &-\frac{1}{\sigma^{11}}\left[n_C\sigma_{22} - \sum_{i\in C}(y_{i2} - \Pi_2 x_i)^2\right] - \frac{\sigma^{12}}{\sigma^{11}}(-2n_C\sigma_{12}\sigma_{22}) \end{aligned}$$

$$A_{31} = -\frac{n_A}{2}\sigma_{12}^2 + \frac{1}{2}\left(\frac{\sigma^{12}}{\sigma^{22}}\right)^2 \left(-n_B\sigma_{11}^2\right) - \frac{1}{2}n_C\sigma_{12}^2$$

$$\begin{aligned} A_{32} &= -\frac{n_A}{2}(2\sigma_{12}\sigma_{22}) + 2\frac{\sigma^{12}}{(\sigma^{22})^2}\left[n_B\sigma_{11} - \sum_{i\in B}(y_{i1} - \Pi_1 x_i)^2\right] \\ &\quad +\frac{1}{2}\left(\frac{\sigma^{12}}{\sigma^{22}}\right)^2(-2n_B\sigma_{11}\sigma_{12}) - \frac{1}{2}(-2n_C\sigma_{12}\sigma_{22}) \end{aligned}$$

$$A_{33} = -\frac{n_A}{2}\sigma_{22}^2 - \frac{(\sigma^{12})^2}{(\sigma^{22})^3}\left[n_B\sigma_{11} - \sum_{i\in B}(y_{i1} - \Pi_1 x_i)^2\right] + \frac{1}{2}\left(\frac{\sigma^{12}}{\sigma^{22}}\right)^2\left(-n_B\sigma_{12}^2\right) - \frac{1}{2}n_C\sigma_{22}^2$$

In order to explicit the lower block of the information matrix I, according to the (A3.52), we have to apply the expected value to each of the (A3.54), or equivalently to each A_{lm} previously defined, so we have:

$$I_{11} = E[-A_{11}] = \left(\frac{n_A + n_B}{2}\right)\sigma_{11}^2 + \left(\frac{n_C}{2}\right)\frac{\sigma_{12}^4}{\sigma_{22}^2}$$

$$I_{21} = E[-A_{21}] = (n_A + n_B)\,\sigma_{11}\sigma_{12} + n_C\frac{\sigma_{12}^3}{\sigma_{22}}$$

$$I_{22} = E[-A_{22}] = n_A\left(\sigma_{12}^2 + \sigma_{11}\sigma_{22}\right) + 2(n_B + n_C)\sigma_{12}^2$$

$$I_{31} = E[-A_{31}]\,\frac{n_A + n_B + n_C}{2}\sigma_{12}^2$$

$$I_{32} = E[-A_{32}] = (n_A + n_C)\,\sigma_{12}\sigma_{22} + n_B\frac{\sigma_{12}^3}{\sigma_{11}}$$

$$I_{33} = E[-A_{33}] = \left(\frac{n_A + n_C}{2}\right)\sigma_{22}^2 + \left(\frac{n_B}{2}\right)\frac{\sigma_{12}^4}{\sigma_{11}^2}$$

So we can indicate

$$I_{\Sigma\Sigma} = E\left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] = \begin{bmatrix} I_{11} & I_{21} & I_{31} \\ I_{21} & I_{22} & I_{32} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}$$

Now, all the elements of the I matrix are computed, and we have to follow the same procedure to compute the elements of I^*.

Let us consider the (A3.53). Expliciting the element $(1,1)$ of the right hand side matrix, we obtain:

$$\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial\,(vec\Pi vec\Pi')} = -\Sigma^{-1}\begin{bmatrix} \sum_i x_i^2, & 0 \\ 0 & \sum_i x_i^2 \end{bmatrix};$$

applying the expected value we obtain:

$$I^*_{\Pi\Pi} = E\left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Pi)\partial(vech\Pi)'}\right] = \Sigma^{-1}\begin{bmatrix} \sum_i x_i^2, & 0 \\ 0 & \sum_i x_i^2 \end{bmatrix}.$$

We explicit now the elements of the block $(2,2)$ of the matrix (A3.53). Labelling with A_l^* the first derivative and reminding (3.10) we easily obtain

$$A_1^* = \frac{n}{2}\sigma_{11} - \frac{1}{2}\sum_i (y_{i1} - \Pi_1 x_i)^2 \tag{A3.55}$$

$$A_2^* = n\sigma_{12} - \sum_i (y_{i1} - \Pi_1 x_i)(y_{i2} - \Pi_2 x_i) \tag{A3.56}$$

$$A_3^* = \frac{n}{2}\sigma_{22} - \frac{1}{2}\sum_i (y_{i2} - \Pi_2 x_i)^2 \tag{A3.57}$$

Now, differentiating (A3.55-A3.57) with respect to $[vech\,(\Sigma^{-1})]'$ we obtain:

$$A_{lm}^* = -n \begin{bmatrix} \frac{\sigma_{11}^2}{2} & \sigma_{11}\sigma_{12} & \frac{\sigma_{12}^2}{2} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 + \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{22} \\ \frac{\sigma_{12}^2}{2} & \sigma_{12}\sigma_{22} & \frac{\sigma_{22}^2}{2} \end{bmatrix}$$

$$I_{\Sigma\Sigma}^* = E\left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] = E\left[A_{lm}^*\right] = n \begin{bmatrix} \frac{\sigma_{11}^2}{2} & \sigma_{11}\sigma_{12} & \frac{\sigma_{12}^2}{2} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 + \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{22} \\ \frac{\sigma_{12}^2}{2} & \sigma_{12}\sigma_{22} & \frac{\sigma_{22}^2}{2} \end{bmatrix}$$

Let us now observe I^* and I more in detail, and particularly the difference $(I^* - I)$ (equations 6.47 and 6.48). This difference should be a positive semidefinite matrix; furthermore such a difference should become quite small when the latent and the observable model are similar. In order to show this result we write the explicit form of the difference $(I^* - I)$:

$$(I^* - I) = \begin{bmatrix} [I_{\Pi\Pi}^*] & [0] \\ [0] & [I_{\Sigma\Sigma}^*] \end{bmatrix} - \begin{bmatrix} [I_{\Pi\Pi}] & [0] \\ [0] & [I_{\Sigma\Sigma}] \end{bmatrix}.$$

The difference between these two matrices can be represented subtracting the corresponding blocks, so that:

$$\begin{aligned} I_{\Pi\Pi}^* - I_{\Pi\Pi} &= \Sigma^{-1} \begin{bmatrix} \sum_i x_i^2, & 0 \\ 0 & \sum_i x_i^2 \end{bmatrix} - \Sigma^{-1} \begin{bmatrix} \sum_{i \in A,B} x_i^2, & \frac{\sigma_{12}}{\sigma_{22}} \sum_{i \in C} x_i^2 \\ \frac{\sigma_{12}}{\sigma_{11}} \sum_{i \in B} x_i^2, & \sum_{i \in A,C} x_i^2 \end{bmatrix} \\ &= \Sigma^{-1} \begin{bmatrix} \sum_{i \in C,D} x_i^2, & -\frac{\sigma_{12}}{\sigma_{22}} \sum_{i \in C} x_i^2 \\ -\frac{\sigma_{12}}{\sigma_{11}} \sum_{i \in B} x_i^2, & \sum_{i \in B,D} x_i^2 \end{bmatrix} \end{aligned}$$

$$I_{\Sigma\Sigma}^* \quad - \quad I_{\Sigma\Sigma} = n \begin{bmatrix} \frac{\sigma_{11}^2}{2} & \sigma_{11}\sigma_{12} & \frac{\sigma_{12}^2}{2} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2 + \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{22} \\ \frac{\sigma_{12}^2}{2} & \sigma_{12}\sigma_{22} & \frac{\sigma_{22}^2}{2} \end{bmatrix} - \begin{bmatrix} I_{11} & I_{21} & I_{31} \\ I_{21} & I_{22} & I_{32} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}$$

$$
= \begin{bmatrix}
\left[\frac{n_C}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{22}^2}+\frac{n_D}{2}\sigma_{11}^2\right], & -, & - \\
\begin{bmatrix} n_C\frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}}\sigma_{12} \\ +n_D\sigma_{11}\sigma_{12}) \end{bmatrix}, & \begin{bmatrix} (n_B+n_C)(\sigma_{11}\sigma_{22}-\sigma_{12}^2) \\ +n_D(\sigma_{12}^2+\sigma_{11}\sigma_{22}) \end{bmatrix}, & - \\
\left[\frac{n_D}{2}\sigma_{12}^2\right], & \begin{bmatrix} n_B(\sigma_{11}\sigma_{22}-\sigma_{12}^2)\frac{\sigma_{12}}{\sigma_{11}} \\ +n_D\sigma_{12}\sigma_{22} \end{bmatrix}, & \begin{bmatrix} \frac{n_B}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{11}^2} \\ +\frac{n_D}{2}\sigma_{22}^2 \end{bmatrix}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\frac{n_C}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{22}^2}, & -, & - \\
n_C\frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}}\sigma_{12}, & (n_B+n_C)(\sigma_{11}\sigma_{22}-\sigma_{12}^2), & - \\
0 & n_B(\sigma_{11}\sigma_{22}-\sigma_{12}^2)\frac{\sigma_{12}}{\sigma_{11}}, & \frac{n_B}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{11}^2}
\end{bmatrix}
$$

$$
+ \begin{bmatrix}
\frac{n_D}{2}\sigma_{11}^2, & -, & - \\
n_D\sigma_{11}\sigma_{12}, & n_D(\sigma_{12}^2+\sigma_{11}\sigma_{22}), & - \\
\frac{n_D}{2}\sigma_{12}^2, & n_D\sigma_{12}\sigma_{22}, & \frac{n_D}{2}\sigma_{22}^2
\end{bmatrix} = T + R
$$

To prove that the difference $(I^* - I)$ is positive semidefinite, we remember that if a symmetric matrix A can be written as

$$A = QQ' \tag{A3.58}$$

then the matrix is positive semidefinite. So, it is sufficient to show that both matrices $I^*_{\Pi\Pi} - I_{\Pi\Pi}$ and $I^*_{\Sigma\Sigma} - I_{\Sigma\Sigma}$ can be written in a form like (A3.58). Let us consider the matrix $I^*_{\Pi\Pi} - I_{\Pi\Pi}$, it can be rearranged as follows

$$
\begin{aligned}
I^*_{\Pi\Pi} \; - \; I_{\Pi\Pi} &= \Sigma^{-1}\begin{bmatrix} \sum_{i\in C,D} x_i^2, & -\frac{\sigma_{12}}{\sigma_{22}}\sum_{i\in C} x_i^2 \\ -\frac{\sigma_{12}}{\sigma_{11}}\sum_{i\in B} x_i^2, & \sum_{i\in B,D} x_i^2 \end{bmatrix} \\
&= \frac{1}{\det(\Sigma)}\begin{bmatrix} \sigma_{22}\sum_{i\in C,D} x_i^2 + \frac{\sigma_{12}^2}{\sigma_{11}}\sum_{i\in B} x_i^2, & -\sigma_{12}\sum_{i\in B,C,D} x_i^2 \\ -\sigma_{12}\sum_{i\in B,C,D} x_i^2 & \frac{\sigma_{12}^2}{\sigma_{22}}\sum_{i\in C} x_i^2 + \sigma_{11}\sum_{i\in B,D} x_i^2 \end{bmatrix} \\
&= \frac{1}{\det(\Sigma)}\begin{bmatrix} \sqrt{\frac{\sigma_{12}^2}{\sigma_{11}}}X_B', & \sqrt{\sigma_{22}}X_C' \\ -\sqrt{\sigma_{11}}X_B', & \sqrt{\frac{\sigma_{12}^2}{\sigma_{22}}}X_C' \end{bmatrix}\begin{bmatrix} \sqrt{\frac{\sigma_{12}^2}{\sigma_{11}}}X_B', & \sqrt{\sigma_{22}}X_C' \\ -\sqrt{\sigma_{11}}X_B', & \sqrt{\frac{\sigma_{12}^2}{\sigma_{22}}}X_C' \end{bmatrix}' \\
&+ \frac{1}{\det(\Sigma)}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}\sum_{i\in,D} x_i^2 = \frac{1}{\det(\Sigma)}Q_{\Pi\Pi}Q'_{\Pi\Pi} + \Sigma^{-1}\sum_{i\in,D} x_i^2 \quad \text{(A3.59)}
\end{aligned}
$$

concluding that $I^*_{\Pi\Pi} - I_{\Pi\Pi}$ is a positive semidefinite matrix being sum of two positive semidefinite matrices.

Let us consider the matrix $I^*_{\Sigma\Sigma} - I_{\Sigma\Sigma}$

$$I^*_{\Sigma\Sigma} - I_{\Sigma\Sigma} \;=\; T + R$$

$$
= \begin{bmatrix} \frac{n_C}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{22}^2}, & -, & - \\ n_C\frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}}\sigma_{12}, & (n_B+n_C)\,(\sigma_{11}\sigma_{22}-\sigma_{12}^2), & - \\ 0 & n_B(\sigma_{11}\sigma_{22}-\sigma_{12}^2)\frac{\sigma_{12}}{\sigma_{11}}, & \frac{n_B}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{11}^2} \end{bmatrix}
$$

$$
+ \begin{bmatrix} \frac{n_D}{2}\sigma_{11}^2, & -, & - \\ n_D\sigma_{11}\sigma_{12}, & n_D(\sigma_{12}^2+\sigma_{11}\sigma_{22}), & - \\ \frac{n_D}{2}\sigma_{12}^2, & n_D\sigma_{12}\sigma_{22}, & \frac{n_D}{2}\sigma_{22}^2 \end{bmatrix} \tag{A3.60}
$$

We show that it can be rewritten specifying both T and R as in (A3.58). Let us first consider T; indicating $t = \sigma_{12}^2 + \sigma_{11}\sigma_{22}$, we have

$$
T = \begin{bmatrix} \frac{1}{\sigma_{22}}\sqrt{\frac{n_c}{2}}t, & 0, & 0 \\ \frac{\sigma_{12}\sqrt{2n_c}}{\sqrt{t}}, & \sqrt{n_B+n_C-\frac{2n_c\sigma_{12}}{t}}, & 0 \\ 0, & \frac{\sigma_{12}}{\sigma_{22}}n_B\sqrt{\frac{t}{t(n_B+n_C)-2n_c\sigma_{12}}} & \sqrt{\frac{n_B}{2}\frac{t}{\sigma_{11}^2}-\frac{\sigma_{12}^2}{\sigma_{11}^2}\frac{t}{t(n_B+n_C)-2n_c\sigma_{12}}n_B^2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sigma_{22}}\sqrt{\frac{n_c}{2}}t, & 0, & 0 \\ \frac{\sigma_{12}\sqrt{2n_c}}{\sqrt{t}}, & \sqrt{n_B+n_C-\frac{2n_c\sigma_{12}}{t}}, & 0 \\ 0, & \frac{\sigma_{12}}{\sigma_{22}}n_B\sqrt{\frac{t}{t(n_B+n_C)-2n_c\sigma_{12}}} & \sqrt{\frac{n_B}{2}\frac{t}{\sigma_{11}^2}-\frac{\sigma_{12}^2}{\sigma_{11}^2}\frac{t}{t(n_B+n_C)-2n_c\sigma_{12}}n_B^2} \end{bmatrix}'
$$

Being T written as (A3.58) we can conclude that T is a positive semidefinite matrix.

Let us consider R; indicating $r = \sigma_{12}^2 - \sigma_{11}\sigma_{22}$, we have

$$
R = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{2}}, & 0, & 0 \\ \sqrt{2}\sigma_{12}, & \sqrt{r}, & 0 \\ \frac{\sigma_{12}^2}{\sqrt{2}\sigma_{11}} & \frac{\sigma_{12}}{\sigma_{11}}\sqrt{r} & \frac{r}{\sigma_{11}\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{2}}, & 0, & 0 \\ \sqrt{2}\sigma_{12}, & \sqrt{r}, & 0 \\ \frac{\sigma_{12}^2}{\sqrt{2}\sigma_{11}} & \frac{\sigma_{12}}{\sigma_{11}}\sqrt{r} & \frac{r}{\sigma_{11}\sqrt{2}} \end{bmatrix}'
$$

The R matrix has been arranged as (A3.58), so we can conclude that also R is a positive semidefinite matrix.

Being $I^*_{\Sigma\Sigma} - I_{\Sigma\Sigma} = T + R$ we can conclude that it is a positive semidefinite matrix; having proved the same for the matrix $I^*_{\Pi\Pi} - I_{\Pi\Pi}$, we can conclude that $(I^* - I)$ is a positive semidefinite matrix.

References

Anderson T. W. (1957): "Maximum Likelihood Estimates for the Multivariate Normal Distribution when Some Observations are Missing", *Journal of the American Statistical Association*, 52, 200-203.

Amemiya T. (1985): *Advanced Econometrics.* Cambridge (Mass.): Harvard University Press.

Dempster A. P., Laird N., Rubin D. B. (1977): "Maximum Likelihood from Incomplete Data Via the EM Algorithm", *Journal of the Royal Statistical Society, Series B*, 39, 1, 1-38, with discussion and reply.

Gourieroux C., Monfort A. (1981): "On the Problem of Missing Data in Linear Models", *Review of Economic Studies*, XLVIII, 579-586.

Gourieroux C., Monfort A. (1996): *Simulation-Based Econometric Methods.* Oxford University Press.

Greene W. H. (2000): *Econometric Analysis* (fourth edition). Upper Saddle River, NJ: Prentice-Hall, Inc.

Hajivassiliou V., McFadden D. (1990): "The Method of Simulated Scores, with Application to Models of External Debt Crises", Cowles Foundation Discussion Paper No. 967, Yale University.

Hajivassiliou V. A. (1993): "Simulation Estimation Methods for Limited Dependent Variable Models", in *Handbook of Statistics, Vol. 11*, ed. by G. S. Maddala, C. R. Rao and H. D. Vinod. Amsterdam: Elsevier Science Publishers B. V.

Hajivassiliou V. A., Ruud P. A. (1994): "Classical Estimation Methods for LDV Models using Simulation", in *Handbook of Econometrics, Vol. IV*, ed. by R. F. Engle and D. L. McFadden. Amsterdam: Elsevier Science B. V.

Hansen L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

Horton N. J. and Lipsitz S. R. (2001): "Multiple Imputation in Practice: Comparison of Software Packages for Regression Models with Missing Variables", *The American Statistician*, 55, 244-254.

Little R. J. A., Rubin D. B. (1987): *Statistical Analysis with Missing Data.* New York : Wiley.

Mc Fadden D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Model without Numerical Integration", *Econometrica* , 57, 995-1026.

Pakes A., Pollard D. (1989): "Simulation of the Asymptotic of Optimization Estimators", *Econometrica*, 57, 1027-1057.

Raghunathan T. E.: www.isr.umich.edu/src/smp/ive.

Raghunathan T. E., Lepkowski J., Van Voewyk J., Solenberger P. (1997): "A Multivariate Technique for Imputing Missing Values Using a Sequence of Regression Models", Technical Report, Survey Methodology Program, Survey Research Center, ISR, University of Michigan.

Rubin D. B. (1974): "Characterizing the Estimation of Parameters in Incomplete Data Problems", *Journal of the American Statistical Association*, 69, 467-474.

Rubin D. B. (1976): "Inference with Missing Data", *Biometrika* , 63, 581-592.

Rubin D. B. (1978): "Multiple Imputations in Sample Surveys-A Phenomeno-Logical Bayesian Approach to Nonresponse", *The Proceeding of the Survey Research Methods Section of the American Statistical Association* , 20-34, with discussion and reply.

Rubin D. B. (1987): *Multiple Imputation for Nonresponse in Survey* . New York: Wiley.

Rubin D. B. (1994): "Comments on: Missing data, Imputation and the Bootstrap, by B. Efron", *Journal of the American Statistical Association*, 89, 485-488.

Rubin D. B. (1996): "Multiple Imputation After 18+ Years", *Journal of the American Statistical Association*, 91, 473-489.

Rubin D. B. (2000): "The Broad Role of Multiple Imputations in Statistical Science", in *Proceeding in Computational Statistics, 14th Symposium, Utrecht- The Netherlands, 2000*, ed. by J. G. Bethlehem and P. G. M. van der Heijden. Vienna: Physica-Verlag, 3-14.

Schafer J. L.: www.stat.psu.edu/˜jls.

Schafer J. L. (1997): *Analysis of Incomplete Multivariate Data.* London: Chapman & Hall.

Stern S. (2000): "Simulation-Based Inference in Econometrics: Motivation and Methods", in *Simulation-Based Inference in Econometrics* , ed. by R. Mariano, T. Schuermann, and M. J. Weeks. Cambridge University Press, 9-37.

Thisted, R. A. (1988): *Elements of Statistical Computing.* New York: Chapman and Hall.