Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2010/01

The Method of Simulated Scores for Estimating Multinormal Regression Models with Missing Values

Giorgio Calzolari,
Laura Neri

Università degli Studi
di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

The Method of Simulated Scores for Estimating Multinormal Regression Models with Missing Values

Giorgio Calzolari[a] and **Laura Neri**[b]

[a]Università di Firenze, Dipartimento di Statistica "G.Parenti", Viale Morgagni 59, I-50134 Firenze. <calzolar@ds.unifi.it>
[b]Università di Siena, Dipartimento di Metodi Quantitativi, Piazza San Francesco 8, I-53100 Siena. <neri@unisi.it>

Summary

Given a set of continuous variables with missing data, we prove in this paper that the iterative application of a simple "least-squares estimation/multivariate normal simulation" procedure produces an efficient parameters estimator. There are two main assumptions behind our proof: (1) the missing data mechanism is ignorable; (2) the data generating process is a multivariate normal linear regression. Disentangling the iterative procedure and its convergence conditions, 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 thus provide a non-Bayesian re-interpretation of the estimation/simulation problem. The computational procedure is obtained introducing a simple modification into existing algorithms. Its software implementation is straightforward (few simple statements in any programming language) and easily applicable to datasets with large number of variables. [1]

Keywords. Simulated scores, missing data, multivariate normal regression model, estimation/simulation, general pattern of missingness, simultaneous equations, structural form, reduced form

JEL classifications. C13, C15, C30, C81.

[1]Corresponding author: Giorgio Calzolari <calzolar@ds.unifi.it>. Tel: +39-055-4237217; fax: +39-055-4223560.
This paper is a revised and improved version of an earlier work of the authors, whose results were presented at *Compstat 2002* (Berlin) and *European Meeting of the Econometric Society 2002* (Venezia). We gratefully acknowledge suggestions and comments from Erich Battistin, Monica Billio, Fabio Corradi, Emanuela Dreassi, Giampiero M. Gallo, James M. Lepkowski, Fabrizia Mealli, Federico M. Stefanini, Vijay K. Verma. Of course, the usual caveat applies.

1 Introduction

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

We consider in this paper a particular but important case: the data generating process is a multivariate normal linear regression, where the missing data mechanism is *ignorable*, and data can be missing in any possible non-monotone pattern (general pattern).

We discuss in detail a parametric estimation procedure, that can be viewed as a *simulated scores estimator* (Hajivassiliou and McFadden, 1990, a particular case of the *method of simulated moments* proposed in McFadden, 1989, or Pakes and Pollard, 1989). Its properties are well defined and well known in the literature, in particular its equivalence to ML when the number of replications is conveniently large. In this way we solve (or re-interpret) the estimation problem following a non-Bayesian paradigm, different from the MCMC, which is often applied to this type of problems.

The procedure we propose is straightforward, and can be obtained modifying existing algorithms (Raghunathan et al., 2001). Otherwise it can easily be implemented with few simple statements in any programming language. It is based on iterated "least-squares estimation/multivariate normal simulation" of a system of simultaneous equations in structural form. The procedure estimates parameters (OLS) given previously simulated values, and simulates missing values, given previously estimated parameters.

After describing the problem (section 2), we discuss the evaluation of the likelihood. From a practical/technical viewpoint, when the number of variables with potentially missing data is moderately large, the number of possible patterns of missingness to be explicitly considered becomes too large; the classical likelihood-based procedures become too complex in practice (section 3).

Estimation based on *simulated scores* would be much simpler (section 4), but it requires a simulation step to complete the data. If simulation is performed in the most "intuitive" way (from the reduced form equations, section 5.1) an eplicit expression of the simulated score can be derived (section 5.2).

It is easy to prove that iterated *least squares estimation/simulation* of the reduced form equations sets to zero the simulated score, upon convergence (section 5.3), thus it is a *simulated scores estimator*. This technique, however, does not escape the computational complexity due to large number of possible patterns of missingness, and is therefore practically not feasible, analogously to maximum likelihood (section 5.4).

In section 6 we solve the practical complexity with a *least squares estimation/simulation* procedure based on linear simultaneous equations in *structural form*, that gives exactly the same results as the simulated scores estimator of the previous sections: same results means same "numbers", not just same asymptotic properties. This is explained simply recalling that the simulation procedure is nothing but the Gauss-Seidel technique, well known and used by econometricians specially for large scale equation systems in the sixties and seventies (see, for instance, Evans, Klein and Schink, 1968). In our context, "convergence" means *exact numerical solution* of an iterative procedure, thus easily verifiable in any practical application: no risk of those ambiguities that, not many years ago, led Horton and Lipsitz (2001, p.246) to confess that convergence "remains more of an art form than a science".

Detailed analytical proofs are given in the appendices for the "bivariate normal" case.

i	X_1	.....	X_k	Y_1	.....	Y_p
1	•	•	•	•	•	?
2	•	•	•	?	•	•
3	•	•	•	?	•	?
.	•	•	•	•	?	•
.	•	•	•	•	?	?
.	•	•	•	?	?	?
n-1	•	•	•	•	•	•
n	•	•	•	?	?	•

Table 1: Dataset with missing values.

2 Statement of the problem

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

We distinguish variables with missing data, called Y (a $n \times p$ matrix) from variables with complete data, called X (a $n \times k$ matrix), 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 expected value (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. If no variable is complete, we still use the above notation (in that case X will be a constant column and Π will be the vector of expected values).

For instance, the case of a bivariate normal distribution $Y = (Y_1, Y_2)$, with a single column matrix X of complete data ($p = 2$, $k = 1$), is one of the smallest possible cases (in terms of dimensions). We shall use it often in this paper, as it helps to simplify the study without substantial loss of generality. We specify the multivariate normal data generating process (often called *reduced form* in the rest of this paper) as follows

$$Y = X\Pi' + E = X\Pi' + U\Sigma^{\frac{1}{2}} \tag{2.1}$$

where matrix X $(n \times 1)$ is completely observed, $\Pi = [\Pi_{1,}\Pi_2]'$ denotes the matrix of unknown coefficients (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.

Missing data may affect Y_1 and Y_2 according to a general pattern. Thus, 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.

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

$$[X,\ Y_1,\ Y_2] = \begin{array}{|c|c|c|}\hline X_A & Y_{A1} & Y_{A2} \\ \hline X_B & Y_{B1} & ? \\ \hline X_C & ? & Y_{C2} \\ \hline X_D & ? & ? \\ \hline \end{array} \begin{array}{l} A-block \\ B-block \\ C-block \\ D-block \end{array}$$

Table 2: Bivariate dataset with missing values.

3 Maximum likelihood estimation

We may wish to estimate parameters (Π and Σ) directly by maximum likelihood (e.g. Schafer, 1997, p. 16). Indexing the unique missingness pattern that actually appears in the sample by δ ($\delta = 1, 2, ..., \Delta$), and denoting with $D(\delta)$ the subset of the rows $i = 1, 2, ..., n$ that exhibit pattern δ, the likelihood of the parameters $\theta =$ (elements of Π and Σ) is

$$P(\theta|Y_{obs}) = \prod_{\delta=1}^{\Delta} \prod_{i \in D(\delta)} |\Sigma_\delta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(y_{i,obs} - \Pi_\delta x_i\right)' \Sigma_\delta^{-1} \left(y_{i,obs} - \Pi_\delta x_i\right)\right\} \tag{3.2}$$

where $y_{i,obs}$ denotes the observed variables of unit i, $\Pi_\delta x_i$ and Σ_δ denote the corresponding mean subvector and covariance submatrix. In the above equation, unit i belongs to a block of units whose missing data pattern is δ; there are Δ different patterns in the sample, each contributing to the likelihood in a different way; the value of Δ can be up to 2^p.

If any rows of the data matrix are completely missing, those rows drop out of the likelihood function $P(\theta|Y_{obs})$, under the ignorability assumption. In general, in order to obtain the ML estimator of Π and Σ, assuming ignorability of the missing data mechanism and a general pattern of missing values for the $p-$variate Y variables, we should take into account explicitly up to 2^p blocks, each contributing to the likelihood function in a different way.

Maximization of (3.2) is practically intractable. Not only it requires an iterative alghoritm, because the estimator has no closed form, but specially because the specification of the likelihood itself involves many different terms (one for each block).[2] Of course, algebraic tools like the "sweep operator" would be helpful, as for instance in the case of EM estimator (e.g. Little and Rubin, 1987, ch. 6; see also Kofman and Sharpe, 2003, for a recent application to financial data); the computational complexity of the method, however, would discourage most practitioners.

4 Score and simulated score

Hajivassiliou and McFadden (1990) introduced a simulation-based estimator called *method of simulated scores* (see also Hajivassiliou, 1993, Hajivassiliou and Ruud, 1994, Hajivassiliou and McFadden, 1998, Stern, 2000). In a general context, the intuition behind the method of simulated scores is the following. Suppose that, at least in principle, there are no difficulties in writing explicitly the log-likelihood (based on observed data only) and its first derivatives (the score). 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

[2] As well known, the problem would be much simpler for monotone missing data patterns, where explicit estimates of Π and Σ can be derived using Anderson's (1957) method of *factored likelihoods.* The parameters are transformed in such a way that the likelihood factorizes into distinct factors corresponding to complete data problems (Little and Rubin, 1987, Ch. 6).

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.

Although it is not always easy to construct in practice an unbiased simulator of the score (Stern, 2000, p. 25), it is quite simple in our case.

The simulated score is obtained differentiating the "simulated loglikelihood", that is the loglikelihood of the "completed" data $\widetilde{Y}$. Being related to completed data, this simulated score has the usual expression for the multivariate normal, simple and manageable

$$\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Pi} = \Sigma^{-1}\left(\widetilde{Y} - X\Pi'\right)' X \tag{4.3}$$

$$\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Sigma^{-1}} = \frac{n}{2}\Sigma - \frac{1}{2}\left(\widetilde{Y} - X\Pi'\right)'\left(\widetilde{Y} - X\Pi'\right) \tag{4.4}$$

When only a subset of the data is observable, the score is obtained differentiating the loglikelihood of the observed data (Y_{obs}). Unlike the previous case, due to the missing data, this score has a rather complex expression. For instance, in the simple bivariate case, the three blocks A, B and C contribute to the score with different formulas, while block D does not contribute at all, since it contains no Y_{obs}

$$\frac{\partial \log f(Y_{obs}|X;\Pi,\Sigma)}{\partial \Pi} = \Sigma^{-1}\begin{bmatrix} (Y_{A1} - X_A\Pi_1) & (Y_{A2} - X_A\Pi_2) \\ (Y_{B1} - X_B\Pi_1) & \frac{\sigma_{12}}{\sigma_{11}}(Y_{B1} - X_B\Pi_1) \\ \frac{\sigma_{12}}{\sigma_{22}}(Y_{C2} - X_C\Pi_2) & (Y_{C2} - X_C\Pi_2) \\ 0 & 0 \end{bmatrix}' X \tag{4.5}$$

$$\begin{aligned} \frac{\partial \log f(Y_{obs}|X,\Pi,\Sigma)}{\partial \Sigma^{-1}} = {} & \frac{n_A}{2}\Sigma + \frac{n_B}{2}\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \frac{\sigma_{12}^2}{\sigma_{11}} \end{pmatrix} + \frac{n_C}{2}\begin{pmatrix} \frac{\sigma_{12}^2}{\sigma_{22}} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \\ & -\frac{1}{2}\sum_{i=A}(y_i - \Pi x_i)(y_i - \Pi x_i)' \\ & -\frac{1}{2}\begin{pmatrix} 1 & \frac{\sigma_{12}}{\sigma_{11}} \\ \frac{\sigma_{12}}{\sigma_{11}} & \frac{\sigma_{12}^2}{\sigma_{11}^2} \end{pmatrix}\sum_{i=B}(y_{i1} - \Pi_1 x_i)^2 \\ & -\frac{1}{2}\begin{pmatrix} \frac{\sigma_{12}^2}{\sigma_{22}^2} & \frac{\sigma_{12}}{\sigma_{22}} \\ \frac{\sigma_{12}}{\sigma_{22}} & 1 \end{pmatrix}\sum_{i=C}(y_{i2} - \Pi_2 x_i)^2 \end{aligned} \tag{4.6}$$

(the rule of matrix differentiation has been used, 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).

It is known that, given the parameter values θ = (elements of Π and Σ), *the conditional expectation of the simulated score (eq. 4.3, 4.4), given the observed variables, is equal to the score (eq. 4.5, 4.6).* In our context it means that

$$E_\theta\left[\frac{\partial \log f(\widetilde{Y}|X;\theta)}{\partial \theta}|Y_{obs}, X\right] = \frac{\partial \log f(Y_{obs}|X;\theta)}{\partial \theta} \tag{4.7}$$

The above equality holds in general for latent variable models, when observations are related to the latent variables through a known function (see Gouriéroux and Monfort, 1996, pp. 35-36). In any case, a detailed proof is given in Appendix 1 for the bivariate normal case.

5 Reduced form estimation/simulation

5.1 Simulation

In order to complete the data set we build a simulation step. We first build this step "as if" the parameters of the data generating process were known; later, we shall consider the estimation phase, necessary to get parameter values.

Simulation step: *For fixed values of the reduced form parameters (Π and Σ), missing values (elements of $y_{i,mis}$) are filled in with their conditional expectations (given all observed variables x_i and $y_{i,obs}$) plus simulated pseudo-random errors with appropriate conditional variances and covariances (given the same observations).*

To make things more explicit with the bivariate example, the simulation step produces a completed data matrix, expressed as a block matrix as follows

$$\widetilde{Y} = \begin{bmatrix} Y_{A1} & Y_{A2} \\ Y_{B1} & \widetilde{Y}_{B2} \\ \widetilde{Y}_{C1} & Y_{C2} \\ \widetilde{Y}_{D1} & \widetilde{Y}_{D2} \end{bmatrix} \tag{5.8}$$

Indicating $u_i = (u_{i1}, u_{i2})'$ as random variates from a bivariate standard normal distribution, the completed data ($\widetilde{Y}$) are as follows

$$\begin{cases} y_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}}\ u_{i1} \\ y_{i2} = \Pi_2 x_i + \frac{\sigma_{12}}{\sigma_{11}}(y_{i1} - \Pi_1 x_i) + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\ u_{i2} \end{cases} , i \in A \tag{5.9}$$

$$\begin{cases} y_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}}\ u_{i1} \\ \widetilde{y}_{i2} = \Pi_2 x_i + \frac{\sigma_{12}}{\sigma_{11}}(y_{i1} - \Pi_1 x_i) + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{i2} \end{cases} , i \in B \tag{5.10}$$

$$\begin{cases} y_{i2} = \Pi_2 x_i + \sqrt{\sigma_{22}}\ u_{i2} \\ \widetilde{y}_{i1} = \Pi_1 x_i + \frac{\sigma_{12}}{\sigma_{22}}(y_{i2} - \Pi_2 x_i) + \sqrt{\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}}\ \widetilde{u}_{i1} \end{cases} , i \in C \tag{5.11}$$

$$\begin{cases} \widetilde{y}_{i1} = \Pi_1 x_i + \sqrt{\sigma_{11}}\ \widetilde{u}_{i1} \\ \widetilde{y}_{i2} = \Pi_2 x_i + \frac{\sigma_{12}}{\sigma_{11}}\sqrt{\sigma_{11}}\ \widetilde{u}_{i1} + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{i2} \end{cases} , i \in D \tag{5.12}$$

The simulation step defined is rather straightforward and obvious. When the variables are both observed (A-block) they are obviously left unchanged, and equation (5.9) writes their explicit expression as given by the data generating process (2.1). It must be noticed that the representation in equation (5.9) 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 (2.1). 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 2.1), while the other is replaced by its conditional mean (given the observed variable) plus a zero mean pseudo-random error with the appropriate conditional variance (equations 5.10 and 5.11, where $\widetilde{u}_{i1}$ and $\widetilde{u}_{i2}$ are independent standard normal deviates produced by a pseudo-random number generator; note the difference between the random errors u produced by the data generating process, and the pseudo-random errors $\widetilde{u}$ produced by the pseudo-random number generator). Finally, when both variables are missing (D-block), simulation is performed using the unconditional mean plus a bivariate zero mean pseudo-random error with covariance matrix Σ. Equivalently, as explicitly written in equation (5.12), we first simulate the value $\widetilde{y}_{i1}$, with the appropriate unconditional mean and variance, then simulate $\widetilde{y}_{i2}$ with the appropriate conditional mean and variance (given $\widetilde{y}_{i1}$).

Summarizing, the equations above describe in a unified way observed variables and variables *simulated conditionally on the observed data.*

5.2 Simulated score of data completed from simulation of the reduced form equations

Having completed the data with the above simulation step, the simulated score (4.3), expliciting all simulated values, is (see Appendix 2 for the proof)

$$\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Pi}$$
$$= \Sigma^{-1} \begin{bmatrix} (Y_{A1} - X_A\Pi_1) & (Y_{A2} - X_A\Pi_2) \\ (Y_{B1} - X_B\Pi_1) & \frac{\sigma_{12}}{\sigma_{11}}(Y_{B1} - X_B\Pi_1) + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{B2} \\ \frac{\sigma_{12}}{\sigma_{22}}(Y_{C2} - X_C\Pi_2) + \sqrt{\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}}\ \widetilde{u}_{C1} & (Y_{C2} - X_C\Pi_2) \\ \sqrt{\sigma_{11}}\ \widetilde{u}_{D1} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}}\ \widetilde{u}_{D1} + \sqrt{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{D2} \end{bmatrix}' X \quad (5.13)$$

(analogously for eq. 4.4).

5.3 The estimation/simulation procedure

Of course, to apply the simulation technique in practice, an estimate of the the parameters Π and Σ must be available.

Estimation step: *Estimation of the reduced form parameters (*Π *and* Σ*), using complete data, is performed by Ordinary Least Squares (*Ols*).*

The joint use of the simulation and estimation steps allows to implement an iterative estimation/simulation procedure. We may start from values of Π and Σ estimated by Ols from the fully observed data only, then use the estimated parameters to simulate missing values. This can be called *iteration 0*, and provides a first set of completed data. We can now use the completed data to estimate again by Ols the reduced form parameters, and use the new estimates for a new simulation of the missing data (*iteration 1*). And so on, iteratively. Appendix 3 "explicitly" derives parameters estimated at the k-th iteration using data completed after $k-1$ iterations (therefore based on the $k-1$-th iteration estimates). Still for the bivariate example, for coefficients Π the expression is

$$\widehat{\Pi}'^{(k)} = \widehat{\Pi}'^{(k-1)}$$
$$+ (X'X)^{-1} X' \begin{bmatrix} \widehat{e}_{A1}^{(k-1)} & \widehat{e}_{A2}^{(k-1)} \\ \widehat{e}_{B1}^{(k-1)} & \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}}\ \widehat{e}_{B1}^{(k-1)} + \sqrt{\widehat{\sigma}_{22}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{11}^{(k-1)}}}\ \widetilde{u}_{B2} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}}\ \widehat{e}_{C2}^{(k-1)} + \sqrt{\widehat{\sigma}_{11}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{22}^{(k-1)}}}\ \widetilde{u}_{C1} & \widehat{e}_{C2}^{(k-1)} \\ \sqrt{\widehat{\sigma}_{11}^{(k-1)}}\ \widetilde{u}_{D1} & \frac{\widehat{\sigma}_{12}^{(k-1)}}{\sqrt{\widehat{\sigma}_{11}^{(k-1)}}}\ \widetilde{u}_{D1} + \sqrt{\widehat{\sigma}_{22}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{11}^{(k-1)}}}\ \widetilde{u}_{D2} \end{bmatrix} \quad (5.14)$$

From the above equation it is clear that the procedure converges (for Π) when X is orthogonal to the $k-1$-th iteration vectors of residuals (the two columns in brackets). Σ is estimated from residuals upon convergence. As a conclusion, if the estimation/simulation steps are applied iteratively to the reduced form equations, till numerical convergence of the parameter estimates, then an explicit condition for numerical convergence of the procedure exists, and it can be derived from equation (5.14).

It is quite evident the relationship of the simulated score (5.13) with the convergence condition implied in equation (5.14). The two columns in brackets are the residuals of the completed $\widetilde{Y}$ corresponding to the Π and Σ used for simulation. When computed at the convergence values $\widehat{\Pi}^{(k)} = \widehat{\Pi}^{(k-1)}$ and $\widehat{\Sigma}^{(k)} = \widehat{\Sigma}^{(k-1)}$, residuals are orthogonal to X, thus the simulated score is zero. We can therefore conclude that the OLS estimator of the reduced form parameters, at convergence of the estimation/simulation procedure, sets to zero the simulated score.

Summarizing, we can state that

the iterated estimation/simulation procedure based on the reduced form equations produces a simulated scores estimator with "one" replication.

Finally we observe that in the above statement we refer to "one" replication, because only "one" set of pseudo-random error terms are generated and used for the simulation. It is shown in the literature (Gouriéroux and Monfort, 1996, pp. 35-36) that the asymptotic covariance matrix of the simulated score estimator is larger than for the maximum likelihood estimator, the difference being proportional to the inverse of the number of replications, and thus negligible when the number of replications (S) is very large

$$V_{as}\left[\sqrt{n}\left(\widehat{\theta}_S - \theta\right)\right] = I^{-1} + \frac{1}{S} I^{-1}\left(I^* - I\right) I^{-1} \tag{5.15}$$

If the simulated scores estimator is based on one replication and sample size n, its asymptotic variance-covariance matrix is

$$V_{as}\left[\sqrt{n}\left(\widehat{\theta} - \theta\right)\right] = I^{-1} + I^{-1}\left(I^* - I\right) I^{-1} \tag{5.16}$$

It is clear from (5.16) 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 an efficiency price that must be paid to 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; of course, there would be no simulation). Summarizing, the well known result is that

the method of simulated scores is (asymptotically) inefficient if it is based on "one" replication of simulated pseudo-random errors, but its efficiency increases with the number of replications and becomes equivalent to maximum likelihood when the number of replications tends to infinity.

In Appendix 4 an explicit expression for the variance of the estimator is derived, and compared with the (inverse of) the information matrix.

5.4 Computational complexity

The computational complexity of the previous procedure increases if considering more than two variables ($p > 2$) with missing data on a non-monotone pattern (general pattern). In such a case the rows of the data matrix can be grouped according to their missingness pattern as blocks. We may have *up to* 2^p

possible missingness patterns, and consequently up to 2^p different blocks. There would be no difficulty in the estimation phase (simple OLS of a set of reduced form equations). On the contrary, the simulation phase would be very complex in practice. For each missing data pattern, in fact, we should specify the appropriate simulation step, with the pseudo-random errors conditional on the Y-observed in that data pattern. So we should build, in practice, up to 2^p different simulation steps (one for each block). Thus we face a computational complexity analogous to direct maximum likelihood (section 3).

Concluding, the estimation/simulation procedure based on the reduced form, as well as the ML estimation, do not solve our practical/technical problem.

6 A simple structural form estimation/simulation procedure

The topic dealt with in this section is how to estimate the model parameters in any real case, without considering explicitly the likelihood function stated in terms of the possible 2^p different blocks, and without facing up the technical problem of the simulation in 2^p different blocks.

The method is prompted to the "sequential regression multivariate imputation" approach (SRMI and related software *IVE-ware*, Raghunathan *et al.*, 2001), with a simple modification that considers the multivariate distribution of the error terms also in the structural form equations.

To begin with, we estimate coefficients and variance of the linear regression model related to the variable with fewest missing values (let be Y_1) by OLS, using the observed part of Y_1 ($Y_{obs,1}$). The imputed values are $\widetilde{Y_1} = X_{mis,1}\widehat{\Pi}_1 + \sqrt{\widehat{\sigma}_{11}}\ \widetilde{u}_1$. This step only aims at providing initial values of the iterative procedure, and should have no influence on the final results after convergence (analogous considerations are, for instance in Pollock, 2003).

Having completed Y_1, 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 a simulation step that completes Y_2. Going on, the first iteration 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 simulating the missing values depends on the order in which we select the variables. Thus, the simulated 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 re-estimate and overwrite the simulated values iteratively.

The system of regression equations has, as dependent variable for each equation, the variable to be "simulated if missing" and has on the right hand side all the others variables

$$
\begin{aligned}
Y_1 &= X\gamma_{11} + Y_2\gamma_{12} + Y_3\gamma_{13} + ... + Y_p\gamma_{1p} + \varepsilon_1 \\
Y_2 &= X\gamma_{21} + Y_1\gamma_{22} + Y_3\gamma_{23} + ... + Y_p\gamma_{2p} + \varepsilon_2 \\
& ... \\
Y_p &= X\gamma_{p1} + Y_1\gamma_{p2} + Y_2\gamma_{p3} + ... + Y_{p-1}\gamma_{pp} + \varepsilon_p
\end{aligned} \tag{6.17}
$$

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

Equations (6.17) 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 textbook-type representation of the system in structural form is

$$BY' + \Gamma X' = \in', \tag{6.18}$$

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

$$By_i + \Gamma x_i = \varepsilon_i \tag{6.19}$$

The matrices of the structural form coefficients are

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

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

We remember that we can easily derive the reduced form (or "restricted" reduced form, e.g. Greene, 2008, Ch. 16) "solving" the structural form system (6.19) for y_i

$$y_i = -B^{-1}\Gamma x_i + B^{-1}\varepsilon_i = \Pi x_i + e_i \tag{6.20}$$

Remark 1: *A structural form model like (6.17) is underidentified, as it violates the order condition for identification (eg. Greene, 2008, Ch. 16): 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, without expecting any "good" properties for the $\widehat{\gamma}$, we can estimate each equation by* OLS.

After coefficients have been estimated by OLS, the variance covariance matrix is computed from residuals $\widehat{\Psi} = (1/n)\left(\widehat{B}Y' + \widehat{\Gamma}X'\right)(\widehat{B}Y' + \widehat{\Gamma}X')'$ (this is nothing but the usual way of estimating variances and covariances from residuals; for instance, see Foschi et al. 2003). Differently from the SRMI method, that considers only variances (Raghunathan et al., 2001, p. 94), we use the Cholesky decomposition of the matrix $\widehat{\Psi}$ to produce vectors of pseudo-random numbers for simulation, thus considering also covariances besides variances.

When a value of Y_1 is missing, we simulate the value obtained from the right hand side of the first equation in (6.17), where the γ-s are at the previous iteration estimated value, the value(s) of X is (are) observed, and the values of the Y on the right hand side are, in any case, complete (some of them are observed, the others have been produced by simulation in the previous iteration). The same is done for the second equation of the system, filling in values of Y_2, and so on. The values of ε_1, ε_2, ..., ε_p are jointly produced by the pseudo-random generator with a cross equation variance-covariance matrix equal to the last estimated $\widehat{\Psi}$.

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

A question naturally arises when dealing with an iterative simulation-based method: why and when does the iterative estimation/simulation procedure converge? The transformation between structural form and reduced form helps to answer this question. It is in fact not at all obvious how convergence is achieved if we only consider the procedure as it has just been described (note the strong similarity of the procedure with the well known Gibbs sampling, e.g. Gelfand and Smith, 1990: equation (6.17) can be viewed as producing Y_1 from its conditional distribution, given $X, Y_2, ..., Y_p$, etc.). But we may think at the sequence of iterations in a different order, as if iterations were "grouped": studying the convergence of this "reordered" sequence of estimation and simulation phases becomes conceptually manageable. In fact, if we keep parameter values fixed (the γ-s and the Cholesky decomposition of the matrix $\widehat{\Psi}$), and iterate substitutions of simulated values on the right hand side of equations (6.17), then these iterated substitutions (e.g. Thisted, 1988, sec. 3.11.2) are "exactly" the steps of the Gauss-Seidel method for the "simultaneous solution" of the (6.17) system (also called "stochastic simulation" of the system, because of the presence of the ε terms; see, for example, Bianchi et al, 1978). For several decades, Gauss-Seidel was the main algorithm for solving large scale nonlinear systems of simultaneous equations, and was very well known to all econometricians dealing with macro-models (e.g. Evans et al, 1968, Fair and Taylor, 1983). We get, therefore, the following result.

Convergence: *For fixed values of the structural form parameters, the iterated substitution of simulated values converges to the reduced form derived from the structural form (or "restricted" reduced form, equation 6.20).*

Now we can re-estimate parameters (with OLS on the structural form) and start again a new cycle of iterated substitutions in (6.17), and so on.

The strictly tighted sequence of estimations and simulations for each structural equation is thus disentangled and converted into a reordered sequence of iterations. In each iteration, an OLS estimation of "all" the structural form equations (6.17), using observed and previously simulated values, is followed by the simultaneous solution of the equation system (or derivation of the reduced form 6.20) that produces "all" the values of the variables Y.

We are now ready to show the connection between the structural form approach of this section and the reduced form approach of section 5.3.

Proposition: *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.*

The proof of this proposition can be found in Appendix 5.

The discussion on convergence of the iterated simulations with fixed parameters and the proposition stated above ensures that we can get *exactly* the same results either if we perform estimation and simulation directly on the reduced form, or if we use the structural form.

Remark 2: *It is preferable to work, in practice, with the structural form; it is computationally simpler and more intuitive. Nevertheless, passing to the reduced form is necessary for our proofs, because the reduced form is much more manageable from the analytic point of view.*

To conclude, we underline that our method is based on a different paradigm from the SRMI method; the latter, in fact, follows a Bayesian paradigm, and aims at convergence in distribution. Moreover we underline that, while SRMI draws the pseudo-random deviates for each equation "independently", the method we propose considers multivariate stochastic terms. This simple modification allowed us to prove the "good" properties of the estimator, obtained at convergence achieved.

7 Conclusion

In this paper we have introduced an estimation method for multivariate normal linear regression models with missing values, assuming arbitray patterns for the missing values and ignorable missing data mechanisms. The method is essentially based on an iterative "least-squares estimation/multivariate normal simulation" procedure, with some simple changes with respect to the SRMI. It can be viewed as a non-Bayesian version of incomplete Gibbs sampling. The method seems to be friendly for the data analyst, as it computes the parameters by estimating independently each equation of the system by OLS, and it performs the simulation by introducing pseudo-random errors and solving simultaneously the same system. Besides its technical simplicity and feasibility, peculiar of the method are the properties of the estimator. Being a simulated scores estimator, it is consistent, asymptotically normal, and its efficiency can be improved by increasing the number of replications; it becomes as efficient as maximum likelihood if the iterative procedure is replicated a sufficiently large number of times, each time iterating to convergence.

Appendix 1

The identity (4.7) holds in general for latent variable models, when observations are related to the latent variables through a known function (see Gouriéroux and Monfort, 1996, pp. 35-36). For convenience of the reader, in this appendix we give the detailed proof for the bivariate normal.

We start expliciting step by step equation (4.3). Expliciting derivatives with respect to Π as a block matrix, we have

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

Computing expectation conditional on Y_{obs} and X, we have

$$
E\left[\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Pi}|Y_{obs},X\right] = \Sigma^{-1}\begin{bmatrix} (Y_{A1}-X_A\Pi_1) & (Y_{A2}-X_A\Pi_2) \\ (Y_{B1}-X_B\Pi_1) & \frac{\sigma_{12}}{\sigma_{11}}(Y_{B1}-X_B\Pi_1) \\ \frac{\sigma_{12}}{\sigma_{22}}(Y_{C2}-X_C\Pi_2) & (Y_{C2}-X_C\Pi_2) \\ 0 & 0 \end{bmatrix}' X \qquad \text{(A1.21)}
$$

In the same way, we explicit the derivatives of equation (4.4) with respect to Σ^{-1}

$$
\begin{aligned}
&\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Sigma^{-1}} \\
&= \quad \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} \\
&= \quad \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i\in A}\left\{\begin{bmatrix} y_{i1}-\Pi_1x_i \\ y_{i2}-\Pi_2x_i \end{bmatrix}\begin{bmatrix} y_{i1}-\Pi_1x_i \\ y_{i2}-\Pi_2x_i \end{bmatrix}'\right\} \\
&- \quad \frac{1}{2}\sum_{i\in B}\left\{\begin{bmatrix} y_{i1}-\Pi_1x_i \\ \frac{\sigma_{12}}{\sigma 11}(y_{i1}-\Pi_1x_i)+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{i2} \end{bmatrix}\begin{bmatrix} y_{i1}-\Pi_1x_i \\ \frac{\sigma_{12}}{\sigma 11}(y_{i1}-\Pi_1x_i)+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{i2} \end{bmatrix}'\right\} \\
&- \quad \frac{1}{2}\sum_{i\in C}\left\{\begin{bmatrix} \frac{\sigma_{12}}{\sigma_{22}}(y_{i2}-\Pi_2x_i)+\sqrt{\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}}\ \widetilde{u}_{i1} \\ y_{i2}-\Pi_2x_i \end{bmatrix}\begin{bmatrix} \frac{\sigma_{12}}{\sigma_{22}}(y_{i2}-\Pi_2x_i)+\sqrt{\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}}\ \widetilde{u}_{i1} \\ y_{i2}-\Pi_2x_i \end{bmatrix}'\right\} \\
&- \quad \frac{1}{2}\sum_{i\in D}\left\{\begin{bmatrix} \sqrt{\sigma_{11}}\ \widetilde{u}_{i1} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}}\ \widetilde{u}_{i1}+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{i2} \end{bmatrix}\begin{bmatrix} \sqrt{\sigma_{11}}\ \widetilde{u}_{i1} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}}\ \widetilde{u}_{i1}+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\ \widetilde{u}_{i2} \end{bmatrix}'\right\}
\end{aligned}
$$

Computing conditional expectation, we obtain

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

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

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

Remarking that

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

we have

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

$$\begin{aligned}
= \quad & \frac{n_A}{2}\Sigma+\frac{n_B}{2}\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \frac{\sigma_{12}^2}{\sigma_{11}}\end{pmatrix}+\frac{n_C}{2}\begin{pmatrix} \frac{\sigma_{12}^2}{\sigma_{22}} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}\end{pmatrix}-\frac{1}{2}\sum_{i=A}\left[\left(y_i-\Pi x_i\right)\left(y_i-\Pi x_i\right)'\right] \\
- \quad & \frac{1}{2}\begin{pmatrix} 1 & \frac{\sigma_{12}}{\sigma_{11}} \\ \frac{\sigma_{12}}{\sigma_{11}} & \frac{\sigma_{12}^2}{\sigma_{11}^2}\end{pmatrix}\sum_{i=B}\left(y_{i1}-\Pi_1 x_i\right)^2-\frac{1}{2}\begin{pmatrix} \frac{\sigma_{12}^2}{\sigma_{22}^2} & \frac{\sigma_{12}}{\sigma_{22}} \\ \frac{\sigma_{12}}{\sigma_{22}} & 1\end{pmatrix}\sum_{i=C}\left(y_{i2}-\Pi_2 x_i\right)^2
\end{aligned} \tag{A1.22}$$

This proves the identity (4.7) for the bivariate normal regression. Infact, the explicit form of the right hand side of (4.7) are the expressions (4.5) and (4.6) and the left hand side of (4.7) are the (A1.21) and (A1.22); we reach the result observing that (4.5) is equal to (A1.21) and (4.6) is equal to (A1.22).

Appendix 2: Simulated score

The simulated scores estimator of $\theta =$ (elements of Π and Σ) is obtained solving

$$\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Pi}=0 \tag{A2.23}$$

and

$$\frac{\partial \log f(\widetilde{Y}|X;\Pi,\Sigma)}{\partial \Sigma^{-1}}=0 \tag{A2.24}$$

where $\widetilde{Y}$ is the completed matrix. Expliciting (A2.23) we obtain

$$\begin{bmatrix} Y_{A1}-X_A\Pi_1 \\ Y_{B1}-X_B\Pi_1 \\ \frac{\sigma_{12}}{\sigma_{22}}\ e_{C2}+\sqrt{\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}}}\widetilde{u}_{C1} \\ \sqrt{\sigma_{11}}\widetilde{u}_{D1}\end{bmatrix}' X=0 \tag{A2.25}$$

and

$$\begin{bmatrix} Y_{A2}-X_A\Pi_2 \\ \frac{\sigma_{12}}{\sigma_{11}}\ e_{B1}+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\widetilde{u}_{B2} \\ Y_{C2}-X_C\Pi_C \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}}\widetilde{u}_{D1}+\sqrt{\sigma_{22}-\frac{\sigma_{12}^2}{\sigma_{11}}}\widetilde{u}_{D2}\end{bmatrix}' X=0 \tag{A2.26}$$

Expliciting (A2.24) we obtain

$$n\ \Sigma - \begin{bmatrix} Y_{A1}-X_A\Pi_1 & Y_{A2}-X_A\Pi_2 \\ Y_{B1}-X_B\Pi_1 & \widetilde{Y}_{B2}-X_B\Pi_2 \\ \widetilde{Y}_{C1}-X_C\Pi_1 & Y_{C2}-X_C\Pi_2 \\ \widetilde{Y}_{D1}-X_D\Pi_1 & \widetilde{Y}_{D2}-X_D\Pi_2 \end{bmatrix}' \begin{bmatrix} Y_{A1}-X_A\Pi_1 & Y_{A2}-X_A\Pi_2 \\ Y_{B1}-X_B\Pi_1 & \widetilde{Y}_{B2}-X_B\Pi_2 \\ \widetilde{Y}_{C1}-X_C\Pi_1 & Y_{C2}-X_C\Pi_2 \\ \widetilde{Y}_{D1}-X_D\Pi_1 & \widetilde{Y}_{D2}-X_D\Pi_2 \end{bmatrix} = 0 \qquad \text{(A2.27)}$$

In order to obtain the simulated scores estimator for Σ we have to solve (A2.27) or equivalently the following sistem

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

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

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

Appendix 3: Numerical convergence

The estimation procedure directly considers the data generating process (2.1) and estimates the unknown parameters by ordinary least squares (OLS) method. The procedure starts (**Iteration 0**) considering only the observed part of the data matrix (A-block); the reduced form coefficients (Π) are estimated by OLS and Σ (variance-covariance matrix) is estimated from OLS residuals, 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 $\widehat{E}_A^{(0)} = Y_A - X_A\widehat{\Pi}'^{(0)}$.

After $k-1$ iterations have been performed, let us consider iteration k.

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 simulated as in the previous iterations, so the completed data are

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

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

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

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

or equivalently

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

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 simulating 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

$$
\left(X'X\right)^{-1} X' \begin{bmatrix} \widehat{e}_{A1}^{(k-1)} \\ \widehat{e}_{B1}^{(k-1)} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}}\widehat{e}_{C2}^{(k-1)} + \sqrt{\widehat{\sigma}_{11}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{22}^{(k-1)}}}\ \widetilde{u}_{C1} \\ \sqrt{\widehat{\sigma}_{11}^{(k-1)}}\ \widetilde{u}_{D1} \end{bmatrix} = 0 \tag{A3.28}
$$

$$
\left(X'X\right)^{-1} X' \begin{bmatrix} \widehat{e}_{A2}^{(k-1)} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}}\widehat{e}_{B1}^{(k-1)} + \sqrt{\widehat{\sigma}_{22}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{11}^{(k-1)}}}\ \widetilde{u}_{B2} \\ \widehat{e}_{C2}^{(k-1)} \\ \frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}}\sqrt{\widehat{\sigma}_{11}^{(k-1)}}\ \widetilde{u}_{D1} + \sqrt{\widehat{\sigma}_{22}^{(k-1)} - \frac{\widehat{\sigma}_{12}^{(k-1)2}}{\widehat{\sigma}_{11}^{(k-1)}}}\ \widetilde{u}_{D2} \end{bmatrix} = 0 \tag{A3.29}
$$

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

$$
\begin{aligned}
n\ \widehat{\sigma}_{11}^{(k)} = {} & \widehat{e}_{A1}^{(k)\prime}\widehat{e}_{A1}^{(k)} + \widehat{e}_{B1}^{(k)\prime}\widehat{e}_{B1}^{(k)} + \left(\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}}\right)^2 \widehat{e}_{C2}^{(k-1)\prime}\widehat{e}_{C2}^{(k-1)} + \left(\widehat{\sigma}_{11}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{22}^{(k-1)}}\right) \widetilde{u}_{C1}'\ \widetilde{u}_{C1} \\
& +2\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}}\sqrt{\widehat{\sigma}_{11}^{(k-1)} - \frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^2}{\widehat{\sigma}_{22}^{(k-1)}}}\widehat{e}_{C2}^{\prime(k-1)}\ \widetilde{u}_{C1} + \widehat{\sigma}_{11}^{(k-1)}\ \widetilde{u}_{D1}'\ \widetilde{u}_{D1}
\end{aligned} \tag{A3.30}
$$

$$n\ \widehat{\sigma}_{12}^{(k)}=\widehat{e}_{A1}^{(k)\prime}\widehat{e}_{A2}^{(k)}+\left(\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}}\right)^{2}\widehat{e}_{B1}^{(k-1)\prime}\widehat{e}_{B1}^{(k-1)}+\left(\widehat{\sigma}_{22}^{(k-1)}-\frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^{2}}{\widehat{\sigma}_{11}^{(k-1)}}\right)\widehat{e}_{B1}^{(k-1)\prime}\ \widetilde{u}_{B2}$$

$$+\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{22}^{(k-1)}}\widehat{e}_{C2}^{(k-1)\prime}\widehat{e}_{C2}^{(k-1)}+\sqrt{\widehat{\sigma}_{11}^{(k-1)}-\frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^{2}}{\widehat{\sigma}_{22}^{(k-1)}}}\widehat{e'}_{C2}^{(k-1)}\ \widetilde{u}_{C1}$$

$$+\widehat{\sigma}_{12}^{(k-1)}\ \widetilde{u}_{D1}'\ \widetilde{u}_{D1}+.\sqrt{\widehat{\sigma}_{22}^{(k-1)}\widehat{\sigma}_{11}^{(k-1)}-\left(\widehat{\sigma}_{12}^{(k-1)}\right)^{2}}\ \widetilde{u}_{D1}'\ \widetilde{u}_{D2} \qquad \text{(A3.31)}$$

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

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

$$+\left(\widehat{\sigma}_{22}^{(k-1)}-\frac{\left(\widehat{\sigma}_{12}^{(k-1)}\right)^{2}}{\widehat{\sigma}_{11}^{(k-1)}}\right)\ \widetilde{u}_{D2}'\ \widetilde{u}_{D2}+2\frac{\widehat{\sigma}_{12}^{(k-1)}}{\widehat{\sigma}_{11}^{(k-1)}}\sqrt{\widehat{\sigma}_{22}^{(k-1)}\widehat{\sigma}_{11}^{(k-1)}-\left(\widehat{\sigma}_{12}^{(k-1)}\right)^{2}}\ \widetilde{u}_{D1}'\ \widetilde{u}_{D2} \qquad \text{(A3.32)}$$

We may regard (A3.30-A3.32) as the convergence condition for the iterative estimation of Σ. The estimation/simulation procedure achieves convergence when equations (A3.28-A3.32) are jointly solved. These equations are the same that we solve to obtain the simulated scores estimator (A2.25, A2.26 and A2.27). Thus, the OLS estimator of the reduced form system with completed data (at convergence achieved) is a simulated scores estimator.

Appendix 4: 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 (Gouriéroux, 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 equations (4.5) and (4.6), 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

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

$$= E\begin{bmatrix} \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vec\Pi)\partial(vec\Pi)'}\right] & \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vec\Pi)\partial(vech\Sigma^{-1})'}\right] \\ \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vec\Pi)'}\right] & \left[-\frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] \end{bmatrix} \tag{A4.33}$$

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

$$= E\begin{bmatrix} \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vec\Pi)\partial(vec\Pi)'}\right] & \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vec\Pi)\partial(vech\Sigma^{-1})'}\right] \\ \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vec\Pi)'}\right] & \left[-\frac{\partial^2 \log f(\widetilde{Y}|X_{obs};\Pi,\Sigma)}{\partial(vech\Sigma^{-1})\partial(vech\Sigma^{-1})'}\right] \end{bmatrix} \tag{A4.34}$$

Each expression inside small square brackets 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 off-diagonal blocks, for both (A4.33) and (A4.34), are identically zero (the proof is straightforward).

Expliciting the block $(1,1)$ of the matrix on the right hand side of the (A4.33), we have

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

which remains unchanged when applying the expected value

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

Explicitly, the block $(2,2)$ of the matrix on the right hand side of the (A4.33) is the following (3×3) matrix

$$A = \begin{bmatrix} \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{11}\partial\sigma^{11}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{11}\partial\sigma^{12}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{11}\partial\sigma^{22}} \\ \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{12}\partial\sigma^{11}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{12}\partial\sigma^{12}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{12}\partial\sigma^{22}} \\ \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{22}\partial\sigma^{11}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{22}\partial\sigma^{12}} & \frac{\partial^2 \log f(Y_{obs}|X_{obs};\Pi,\Sigma)}{\partial\sigma^{22}\partial\sigma^{22}} \end{bmatrix} \tag{A4.35}$$

whose elements will be indicated as A_{lm}, $l,m = 1,2,3$.

To compute (A4.35), we first differentiate $\log f(Y_{obs}|X_{obs};\Pi,\Sigma)$ with respect to the elements of $vech(\Sigma^{-1})$ obtaining a (3×1) vector whose elements are labelled with A_l $(l = 1,2,3)$. Reminding that $\sigma_{12}/\sigma_{11} = -\sigma^{12}/\ \sigma^{22}$ and that $\sigma_{12}/\sigma_{22} = -\sigma^{12}/\sigma^{11}$, we can write the A_l elements as follows

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

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

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

With further differentiation, we obtain the second derivatives, which are the elements of the (3×3) matrix (A4.35) (being symmetric, we explicit only the lower triangle)

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

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

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

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

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

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

In order to explicit the lower block of the information matrix I, according to the (A4.33), we have to apply the expected value to the matrix (A4.35), or equivalently to each A_{lm} previously defined, so we have

$$I_{11} = E\left[-A_{11}\right] = \left(\frac{n_A+n_B}{2}\right)\sigma_{11}^2 + \left(\frac{n_C}{2}\right)\frac{\sigma_{12}^4}{\sigma_{22}^2}$$
$$I_{21} = E\left[-A_{21}\right] = (n_A+n_B)\,\sigma_{11}\sigma_{12} + n_C\frac{\sigma_{12}^3}{\sigma_{22}}$$
$$I_{22} = E\left[-A_{22}\right] = n_A\left(\sigma_{12}^2 + \sigma_{11}\sigma_{22}\right) + 2(n_B+n_C)\sigma_{12}^2$$
$$I_{31} = E\left[-A_{31}\right]\frac{n_A+n_B+n_C}{2}\sigma_{12}^2$$

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

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

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 have been computed, and we follow the same procedure to compute the elements of I^*. Let us consider the (A4.34). Expliciting the element $(1,1)$ of the right hand side matrix, we obtain

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

applying the expected value we obtain

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

We explicit now the elements of the block $(2,2)$ of the matrix (A4.34). Labelling with A_l^* the first derivative and reminding (4.4) we easily obtain

$$A_1^* = \frac{n}{2}\sigma_{11} - \frac{1}{2}\sum_i (y_{i1} - \Pi_1 x_i)^2 \tag{A4.36}$$

$$A_2^* = n\sigma_{12} - \sum_i (y_{i1} - \Pi_1 x_i)(y_{i2} - \Pi_2 x_i) \tag{A4.37}$$

$$A_3^* = \frac{n}{2}\sigma_{22} - \frac{1}{2}\sum_i (y_{i2} - \Pi_2 x_i)^2 \tag{A4.38}$$

Now, differentiating (A4.36-A4.38) with respect to $[vech\left(\Sigma^{-1}\right)]'$ we obtain

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

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

The 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

$$V_{as}\left[\sqrt{n}\left(\widehat{\theta}_S - \theta\right)\right] = I^{-1} + \frac{1}{S}I^{-1}\left(I^* - I\right)I^{-1} \tag{A4.39}$$

What we have described in detail in this paper is a simulated scores estimator based on one replication and sample size n, so its asymptotic variance-covariance matrix is

$$V_{as}\left[\sqrt{n}\left(\widehat{\theta}-\theta\right)\right]=I^{-1}+I^{-1}\left(I^{*}-I\right)I^{-1} \tag{A4.40}$$

It is clear from (A4.40) 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; of course, there would be no simulation). The difference $(I^{*}-I)$ should be a positive semidefinite matrix, and should be "quite small" when the latent and the observable model are similar. To show this result we write explicitly the difference

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

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

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

$$\begin{aligned}
I^{*}_{\Sigma\Sigma} \quad - \quad I_{\Sigma\Sigma} &= n\begin{bmatrix} \frac{\sigma_{11}^2}{2} & \sigma_{11}\sigma_{12} & \frac{\sigma_{12}^2}{2} \\ \sigma_{11}\sigma_{12} & \sigma_{12}^2+\sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{22} \\ \frac{\sigma_{12}^2}{2} & \sigma_{12}\sigma_{22} & \frac{\sigma_{22}^2}{2} \end{bmatrix}-\begin{bmatrix} I_{11} & I_{21} & I_{31} \\ I_{21} & I_{22} & I_{32} \\ I_{31} & I_{32} & I_{33} \end{bmatrix} \\
&= \begin{bmatrix} \left[\frac{n_C}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{22}^2}+\frac{n_D}{2}\sigma_{11}^2\right], & -, & - \\ \begin{bmatrix} n_C\frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}}\sigma_{12} \\ +n_D\sigma_{11}\sigma_{12}) \end{bmatrix}, & \begin{bmatrix} (n_B+n_C)\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right) \\ +n_D(\sigma_{12}^2+\sigma_{11}\sigma_{22}) \end{bmatrix}, & - \\ \left[\frac{n_D}{2}\sigma_{12}^2\right], & \begin{bmatrix} n_B(\sigma_{11}\sigma_{22}-\sigma_{12}^2)\frac{\sigma_{12}}{\sigma_{11}} \\ +n_D\sigma_{12}\sigma_{22} \end{bmatrix}, & \begin{bmatrix} \frac{n_B}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{11}^2} \\ +\frac{n_D}{2}\sigma_{22}^2 \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} \frac{n_C}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{22}^2}, & -, & - \\ n_C\frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}}\sigma_{12}, & (n_B+n_C)\left(\sigma_{11}\sigma_{22}-\sigma_{12}^2\right), & - \\ 0 & n_B(\sigma_{11}\sigma_{22}-\sigma_{12}^2)\frac{\sigma_{12}}{\sigma_{11}}, & \frac{n_B}{2}\frac{\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^4}{\sigma_{11}^2} \end{bmatrix} \\
&+ \begin{bmatrix} \frac{n_D}{2}\sigma_{11}^2, & -, & - \\ n_D\sigma_{11}\sigma_{12}, & n_D(\sigma_{12}^2+\sigma_{11}\sigma_{22}), & - \\ \frac{n_D}{2}\sigma_{12}^2, & n_D\sigma_{12}\sigma_{22}, & \frac{n_D}{2}\sigma_{22}^2 \end{bmatrix}=T+R
\end{aligned}$$

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

$$A=QQ' \tag{A4.41}$$

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 (A4.41). The matrix $I^{*}_{\Pi\Pi}-I_{\Pi\Pi}$ can be rearranged as follows

$$
\begin{aligned}
I^*_{\Pi\Pi} &- I_{\Pi\Pi} = \Sigma^{-1}\begin{bmatrix} \sum\limits_{i\in C,D} x_i^2, & -\frac{\sigma_{12}}{\sigma_{22}}\sum\limits_{i\in C} x_i^2 \\ -\frac{\sigma_{12}}{\sigma_{11}}\sum\limits_{i\in B} x_i^2, & \sum\limits_{i\in B,D} x_i^2 \end{bmatrix} \\
&= \frac{1}{\det(\Sigma)}\begin{bmatrix} \sigma_{22}\sum\limits_{i\in C,D} x_i^2 + \frac{\sigma_{12}^2}{\sigma_{11}}\sum\limits_{i\in B} x_i^2, & -\sigma_{12}\sum\limits_{i\in B,C,D} x_i^2 \\ -\sigma_{12}\sum\limits_{i\in B,C,D} x_i^2 & \frac{\sigma_{12}^2}{\sigma_{22}}\sum\limits_{i\in C} x_i^2 + \sigma_{11}\sum\limits_{i\in B,D} x_i^2 \end{bmatrix} \\
&= \frac{1}{\det(\Sigma)}\begin{bmatrix} \sqrt{\frac{\sigma_{12}^2}{\sigma_{11}}}X'_B, & \sqrt{\sigma_{22}}X'_C \\ -\sqrt{\sigma_{11}}X'_B, & \sqrt{\frac{\sigma_{12}^2}{\sigma_{22}}}X'_C \end{bmatrix}\begin{bmatrix} \sqrt{\frac{\sigma_{12}^2}{\sigma_{11}}}X'_B, & \sqrt{\sigma_{22}}X'_C \\ -\sqrt{\sigma_{11}}X'_B, & \sqrt{\frac{\sigma_{12}^2}{\sigma_{22}}}X'_C \end{bmatrix}' \\
&+ \frac{1}{\det(\Sigma)}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}\sum_{i\in,D} x_i^2 = \frac{1}{\det(\Sigma)}Q_{\Pi\Pi}Q'_{\Pi\Pi} + \Sigma^{-1}\sum_{i\in,D} x_i^2 \qquad \text{(A4.42)}
\end{aligned}
$$

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

Considering the matrix $I^*_{\Sigma\Sigma} - I_{\Sigma\Sigma}$

$$
\begin{aligned}
I^*_{\Sigma\Sigma} - I_{\Sigma\Sigma} &= T + R \\
&= \begin{bmatrix} \frac{n_C}{2}\frac{\sigma_{11}^2\sigma_{22}^2 - \sigma_{12}^4}{\sigma_{22}^2}, & -, & - \\ n_C\frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}}\sigma_{12}, & (n_B + n_C)(\sigma_{11}\sigma_{22} - \sigma_{12}^2), & - \\ 0 & n_B(\sigma_{11}\sigma_{22} - \sigma_{12}^2)\frac{\sigma_{12}}{\sigma_{11}}, & \frac{n_B}{2}\frac{\sigma_{11}^2\sigma_{22}^2 - \sigma_{12}^4}{\sigma_{11}^2} \end{bmatrix} \\
&+ \begin{bmatrix} \frac{n_D}{2}\sigma_{11}^2, & -, & - \\ n_D\sigma_{11}\sigma_{12}, & n_D(\sigma_{12}^2 + \sigma_{11}\sigma_{22}), & - \\ \frac{n_D}{2}\sigma_{12}^2, & n_D\sigma_{12}\sigma_{22}, & \frac{n_D}{2}\sigma_{22}^2 \end{bmatrix} \qquad \text{(A4.43)}
\end{aligned}
$$

We show that it can be rewritten specifying both T and R as in (A4.41). Let us first consider T; indicating $t = \sigma_{12}^2 + \sigma_{11}\sigma_{22}$, we have

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

Being T written as (A4.41) 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 (A4.41), 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. The larger the difference $(I^* - I)$, the larger is 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.

What has been discussed till now is *one replication*, till convergence, of the iterative estimation/simulation procedure. 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 replicate the procedure S times independently, each time till convergence. Then we average the S estimates of $\widehat{\theta}$, obtaining the estimate $\widehat{\theta}_S$. 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.

Appendix 5: Proof of the proposition (section 6)

The structural (underidentified) equation system to be estimated is

$$\left\{ \begin{array}{lcll} Y_1 & = & X\gamma_{11} & + \quad Y_2\gamma_{12} + \varepsilon_1 \\ Y_2 & = & X\gamma_{21} & + \quad Y_1\gamma_{22} + \varepsilon_2 \end{array} \right. \qquad \text{with} \quad \Psi = E\left\{ \left[\begin{array}{c} \varepsilon_{i1} \\ \varepsilon_{i2} \end{array} \right] \left[\begin{array}{c} \varepsilon_{i1} \\ \varepsilon_{i2} \end{array} \right]' \right\} \tag{A5.44}$$

Note that each equation is not identified, as it violates the necessary order condition for identification (e.g. Greene, 2008, Ch. 16); 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

$$\left\{ \begin{array}{l} Y_1 = X\Pi_1 + e_1 \\ Y_2 = X\Pi_2 + e_2 \end{array} \right.$$

The OLS direct estimator of the reduced form coefficients

$$\widehat{\Pi}' = (X'X)^{-1}X'Y \tag{A5.45}$$

and the variance-covariance estimator based on the OLS residual ($\widehat{E} = Y - X\widehat{\Pi}'$)

$$\widehat{\Sigma} = \frac{1}{n}\widehat{E}'\widehat{E} \tag{A5.46}$$

are consistent, and we can write

$$\left\{ \begin{array}{l} Y_1 = X\widehat{\Pi}_1 + \widehat{e}_1 \\ Y_2 = X\widehat{\Pi}_2 + \widehat{e}_2 \end{array} \right. \tag{A5.47}$$

The OLS estimates of the structural form coefficients of the equation having Y_1 as dependent variable (A5.44) are

$$\left[\begin{array}{c} \widetilde{\gamma}_{11} \\ \widetilde{\gamma}_{12} \end{array} \right] = \left[\begin{array}{cc} \sum x_i^2 & \sum x_i y_{i2} \\ \sum x_i y_{i2} & \sum y_{i2}^2 \end{array} \right]^{-1} \left[\begin{array}{c} \sum x_i y_{i1} \\ \sum y_{i2} y_{i1} \end{array} \right]$$

We can replace y_{i1}, y_{i2} with the expressions (A5.47), obtaining

$$\begin{aligned} \left[\begin{array}{c} \widetilde{\gamma}_{11} \\ \widetilde{\gamma}_{12} \end{array} \right] &= \left[\begin{array}{cc} \sum x_i^2 & \widehat{\Pi}_2 \sum x_i^2 + \sum x_i \widehat{e}_{i2} \\ \widehat{\Pi}_2 \sum x_i^2 + \sum x_i \widehat{e}_{i2} & \widehat{\Pi}_2{}^2 \sum x_i^2 + 2\widehat{\Pi}_2 \sum x_i \widehat{e}_{i2} + \sum \widehat{e}_{i2}^2 \end{array} \right]^{-1} \\ & \cdot \left[\begin{array}{l} \widehat{\Pi}_1 \sum x_i^2 + \sum x_i \widehat{e}_{i1} \\ \widehat{\Pi}_1 \widehat{\Pi}_2 \sum x_i^2 + \widehat{\Pi}_1 \sum x_i \widehat{e}_{i2} + \widehat{\Pi}_2 \sum x_i \widehat{e}_{i1} + \sum \widehat{e}_{i1} \widehat{e}_{i2} \end{array} \right] \end{aligned}$$

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

$$\begin{aligned}
\begin{bmatrix} \widetilde{\gamma}_{11} \\ \widetilde{\gamma}_{12} \end{bmatrix} &= \begin{bmatrix} q^2 & \widehat{\Pi}_2 q^2 \\ \widehat{\Pi}_2 q^2 & \widehat{\Pi}_2^2 q^2 + \sum \widehat{e}_{i2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\Pi}_1 q^2 \\ \widehat{\Pi}_1 \widehat{\Pi}_2 q^2 + \sum \widehat{e}_{i1}\widehat{e}_{i2} \end{bmatrix} \\
&= \frac{1}{q^2 \sum \widehat{e}_{i2}^2} \begin{bmatrix} \widehat{\Pi}_2^2 q^2 + \sum \widehat{e}_{i2}^2 & -\widehat{\Pi}_2 q^2 \\ -\widehat{\Pi}_2 q^2 & q^2 \end{bmatrix} \begin{bmatrix} \widehat{\Pi}_1 q^2 \\ \widehat{\Pi}_1 \widehat{\Pi}_2 q^2 + \sum \widehat{e}_{i1}\widehat{e}_{i2} \end{bmatrix} \\
&= \frac{1}{q^2 \sum \widehat{e}_{i2}^2} \begin{bmatrix} \widehat{\Pi}_1 \widehat{\Pi}_2^2 q^4 + \widehat{\Pi}_1 q^2 \sum \widehat{e}_{i2}^2 - \widehat{\Pi}_1 \widehat{\Pi}_2^2 q^4 - \widehat{\Pi}_2 q^2 \sum \widehat{e}_{i1}\widehat{e}_{i2} \\ -\widehat{\Pi}_1 \widehat{\Pi}_2 q^4 + \widehat{\Pi}_1 \widehat{\Pi}_2 q^4 + q^2 \sum \widehat{e}_{i1}\widehat{e}_{i2} \end{bmatrix} \\
&= \frac{1}{\sum \widehat{e}_{i2}^2} \begin{bmatrix} \widehat{\Pi}_1 \sum \widehat{e}_{i2}^2 - \widehat{\Pi}_2 \sum \widehat{e}_{i1}\widehat{e}_{i2} \\ \sum \widehat{e}_{i1}\widehat{e}_{i2} \end{bmatrix} = \begin{bmatrix} \widehat{\Pi}_1 - \frac{\sum \widehat{e}_{i1}\widehat{e}_{i2}}{\sum \widehat{e}_{i2}^2} \widehat{\Pi}_2 \\ \frac{\sum \widehat{e}_{i1}\widehat{e}_{i2}}{\sum \widehat{e}_{i2}^2} \end{bmatrix}
\end{aligned} \tag{A5.48}$$

and analogously for the second equation of the structural form system

$$\begin{bmatrix} \widetilde{\gamma}_{21} \\ \widetilde{\gamma}_{22} \end{bmatrix} = \begin{bmatrix} \widehat{\Pi}_2 - \frac{\sum \widehat{e}_{i1}\widehat{e}_{i2}}{\sum \widehat{e}_{i1}^2} \widehat{\Pi}_1 \\ \frac{\sum \widehat{e}_{i1}\widehat{e}_{i2}}{\sum \widehat{e}_{i1}^2} \end{bmatrix} \tag{A5.49}$$

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

$$\widetilde{\Pi} = -\widetilde{B}^{-1}\widetilde{\Gamma} \tag{A5.50}$$

where

$$\widetilde{B} = \begin{bmatrix} 1 & -\widetilde{\gamma}_{12} \\ -\widetilde{\gamma}_{22} & 1 \end{bmatrix} \qquad \widetilde{\Gamma} = \begin{bmatrix} -\widetilde{\gamma}_{11} \\ -\widetilde{\gamma}_{21} \end{bmatrix} \tag{A5.51}$$

Substituting (A5.48) and (A5.49) into (A5.51) and then into (A5.50) we obtain exactly the $\widehat{\Pi}$ as in (A5.45)

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

So that, estimating the structural form coefficients $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22})$ of our underidentified system (A5.44) by OLS and trasforming them into reduced form coefficients (by the proper trasformation) gives exactly the same as estimating directly by OLS the reduced form coefficients.

The same result holds 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

$$\begin{aligned}
\widetilde{\Psi} &= \frac{1}{n} \sum \left\{ \begin{bmatrix} \widetilde{\varepsilon}_{i1} \\ \widetilde{\varepsilon}_{i2} \end{bmatrix} \begin{bmatrix} \widetilde{\varepsilon}_{i1} \\ \widetilde{\varepsilon}_{i2} \end{bmatrix}' \right\} \\
&= \frac{1}{n} \sum \left\{ \begin{bmatrix} \widehat{e}_{i1} - \frac{\sum \widehat{e}_{i1}\widehat{e}_{i2}}{\sum \widehat{e}_{i2}^2} \widehat{e}_{i2} \\ \widehat{e}_{i2} - \frac{\sum \widehat{e}_{i1}\widehat{e}_{i2}}{\sum \widehat{e}_{i1}^2} \widehat{e}_{i1} \end{bmatrix} \begin{bmatrix} \widehat{e}_{i1} - \frac{\sum \widehat{e}_{i1}\widehat{e}_{i2}}{\sum \widehat{e}_{i2}^2} \widehat{e}_{i2} \\ \widehat{e}_{i2} - \frac{\sum \widehat{e}_{i1}\widehat{e}_{i2}}{\sum \widehat{e}_{i1}^2} \widehat{e}_{i1} \end{bmatrix}' \right\}
\end{aligned} \tag{A5.52}$$

The corresponding reduced form variance-covariance matrix is obtained, as usual (see eq. 6.20), computing

$$\widetilde{\Sigma} = \widetilde{B}^{-1}\widetilde{\Psi}\left(\widetilde{B}^{-1}\right)' \tag{A5.53}$$

Substituting (A5.52) and (A5.51) into (A5.53) we obtain exactly the same $\widehat{\Sigma}$ as in (A5.46)

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

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.

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.

Belsley D. A., Kontoghiorghes E. J., Eds. (2003): "Special Issue on Computational Econometrics", *Computational Statistics and Data Analysis*, 42.

Bianchi C., Calzolari G., Corsi P. (1978): "A Program for Stochastic Simulation of Econometric Models", *Econometrica*, 46, 235-236.

Evans M. K., Klein L. R., Schink G. R. (1968): *The Wharton Econometric Forecasting Model.* Philadelphia: University of Pennsylvania, Economics Research Unit, Studies in Quantitative Economics No. 2.

Fair R. C., Taylor J. B. (1983): "Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models", *Econometrica*, 51, 1169-1185.

Foschi P., Belsley D. A., Kontoghiorghes E. J. (2003): "A Comparative Study of Algorithms for Solving Seemingly Unrelated Regressions Models", *Computational Statistics & Data Analysis*, 44, 3-35.

Gelfand A. E., Smith A. F. M. (1990): "Sampling-Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, 85, 398-409.

Gouriéroux C., Monfort A. (1981): "On the Problem of Missing Data in Linear Models", *Review of Economic Studies*, 68, 579-586.

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

Greene W. H. (2008): *Econometric Analysis* (sixth edition). Upper Saddle River, NJ: Prentice-Hall, Inc.

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., 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., McFadden D. L. (1998): "The Method of Simulated Scores for the Estimation of LDV Models", *Econometrica*, 66, 863-896.

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., 2384-2441.

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

Kofman P., Sharpe I. G. (2003): "Using Multiple Imputation in the Analysis of Incomplete Observations in Finance", *Journal of Financial Econometrics*, 1, 216-249.

Lauritzen, S. L. (1995): "The EM Algorithm for Graphical Association Models with Missing Data", *Computational Statistics and Data Analysis*, 19, 191-201.

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.

Pollock D. S. G. (2003): "Recursive Estimation in Econometrics", *Computational Statistics & Data Analysis*, 44, 37-75.

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

Raghunathan T. E., Lepkowski J. M., Van Voewyk J., Solenberger P. (2001): "A Multivariate Technique for Imputing Missing Values Using a Sequence of Regression Models", *Survey Methodology*, 27, 85-95.

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.