# Bayesian Simultaneous Equations Analysis using Reduced Rank Structures

Frank Kleibergen* Herman K. van Dijk†

February [illegible], [illegible]

**Abstract**

Diffuse priors lead to pathological posterior behavior when used in Bayesian analyses of Simultaneous Equation Models (SEMs). This results from the local nonidentification of certain parameters in SEMs. When this, a priori known, feature is not captured appropriately, it results in an a posteriori favoring of certain specific parameter values that is not the consequence of strong data information but of local nonidentification. We show that a proper consistent Bayesian analysis of a SEM explicitly has to consider the reduced form of the SEM as a standard linear model on which nonlinear (reduced rank) restrictions are imposed, which result from a singular value decomposition. The priors/posteriors of the parameters of the SEM are therefore proportional to the priors/posteriors of the parameters of the linear model under the condition that the restrictions hold. This leads to a framework for constructing priors and posteriors for the parameters of SEMs. The framework is used to construct priors and posteriors for one, two and three structural equation SEMs. These examples together with a theorem, showing that the reduced forms of SEMs accord with sets of reduced rank restrictions on standard linear models, show how Bayesian analyses of generally specified SEMs can be conducted.

*Corresponding author: Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands, email:kleibergen@few.eur.nl.

†Econometric and Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands. Part of this research was carried out while the first author was hosted at the department of Econometrics at Tilburg University, and during his PPS fellowship of the Netherlands Foundation for Scientific Research (N.W.O.). We like to thank four anonymous referees, Luc Bauwens, Peter Phillips, Dale Poirier, Arnold Zellner, Eric Zivot, and other seminar participants for helpful comments and suggestions. This paper has previously circulated as "Bayesian Simultaneous Equations Analysis: On the Existence of Structural Posterior Moments" and "Bayesian Simultaneous Equations Analysis using Noninformative Priors", see Kleibergen and van Dijk ([illegible]).

## 1 Introduction

Since the early forties a lot of research has focussed on the development of statistical methods for analyzing Simultaneous Equation Models (SEMs), see a.o. Haavelmo (1943) and Anderson and Rubin (1949). It shows that models which are able to generate variables simultaneously are important since this is a stylized fact of many economic time series. The SEM is not only important but also rather complicated due to the problems regarding the identification of its parameters. The identification of the structural parameters is reflected in the rank and order conditions which result from the implied reduced form, see Hausman (1983). The order condition reflects overall identification while the rank condition reflects local (non) identification. This latter phenomenon, local nonidentification, is shown to lead to pathological posterior behavior when flat priors are used in Bayesian analyses of the SEM. This behavior occurs in the traditional Bayesian analyses of SEMs documented in the literature, see a.o. Drèze (1976), Drèze and Morales (1976) and Drèze and Richard (1983). We show its occurrence in a limited information (one equation) analysis of the SEM. Similar behavior can be found in other specifications of the SEM as well since the origin of the pathological posterior behavior, local nonidentification of parameters, is exemplary to SEMs.

In order to obtain a consistent Bayesian analysis of a SEM, which does not suffer from these pathologies, we construct a framework in which the reduced form of a SEM is specified as a multivariate linear model with nonlinear (reduced rank) restrictions on its parameters. Using singular value decompositions we specify the restrictions such that an one-to-one correspondence with a linear model is obtained when the restrictions do not hold; and the reduced form of the SEM is obtained when they hold. The prior and posterior analysis then results when this specification is used in the framework for analyzing nested models as parameter restrictions of embedding models constructed in Kleibergen (1997). It a.o. leads to invariance of the priors and posteriors with respect to the specification of the model. The resulting posteriors of the parameters of the SEM accord with the posterior of the embedding linear model. Our analysis is therefore similar to the construction of the Savage-Dickey density ratio, see Dickey (1971). That is, we construct the priors/posteriors in the points where the hypothesis (restriction) holds. In contrast, the posterior of the parameters of a SEM, derived in the usual way using a diffuse prior, is inconsistent in the sense that its implied posterior of the parameters of the embedding linear model is not a member of the standard class of posteriors of the parameters of linear models, see Kleibergen (1997).

The contents of the paper is organized as follows. In section 2, we show the pathologies arising in the posteriors of the parameters of an incomplete (one structural equation analysis of a) SEM when flat priors are used. Sections 3 and 4 show how an incomplete SEM is rewritten as a multivariate linear model with nonlinear parameter restrictions. We use this specification jointly with the framework for analyzing nested models as parameter restrictions of embedding models to obtain the prior and posterior analysis. Singular value decompositions are also involved which are similar to the canonical correlations used in a limited information maximum likelihood analysis, see Anderson and Rubin (1949). In section 5, posterior simulators are constructed to sample from the posterior of the parameters of an incomplete SEM. Section 6 extends the one structural equation analysis to a full system analysis by showing that a fully specified SEM accords with a set of reduced rank restrictions on a linear model. Different subsections then show the framework for prior and posterior analysis for two and three structural equations and also show that the order condition for a full system analysis of a SEM can differ from the order condition resulting from an one structural equation analysis. Finally, the seventh section contains conclusions.

## 2 Nonidentification and Pathological Posterior Behavior

To show the consequences which local nonidentification of parameters of SEMs has for posterior distributions, we analyze, as an example, the case of one (set of) structural equation(s). This model is also known as INcomplete Simultaneous Equations Model (INSEM). As the results for the posteriors of the INSEM are exemplary for other specifications of the SEM, the importance of a proper treatment of the issue of local nonidentification is shown by the analysis of the INSEM.

We use as specification of the INSEM, see Zellner *et. al.* (1988),

$$\begin{aligned} y_1 &= Y_2\beta + X_1\gamma + \varepsilon_1, \\ Y_2 &= X_1\Pi_{12} + X_2\Pi_{22} + V_2, \end{aligned} \tag{1}$$

where $y_1 : T \times 1$, and $Y_2 : T \times (m-1)$, are endogenous and $X_1 : T \times k_1$, and $X_2 : T \times k_2$, $k = k_1 + k_2$, contain the (weakly) exogenous, see Engle *et. al.* (1983), and lagged dependent variables, $\beta : (m-1) \times 1$, $\gamma : k_1 \times 1$, $\Pi_2 = (\Pi'_{12}\ \Pi'_{22})' : k \times (m-1)$ and we assume that $(\varepsilon_1\ V_2) \sim n(0, \Sigma \otimes I_T)$. The identification problems arise when the parameter $\Pi_{22} = 0$ (or has reduced rank) as (parts of) the structural form parameter $\beta$ is then nonidentified. This is easily seen when we construct the reduced form of the INSEM (1),

$$\begin{aligned} y_1 &= X_1\pi_{11} + X_2\Pi_{22}\beta + \xi_1, \\ Y_2 &= X_1\Pi_{12} + X_2\Pi_{22} + V_2, \end{aligned} \tag{2}$$

where $\pi_{11} = \gamma + \Pi_{12}\beta$, $\xi_1 = \varepsilon_1 + V_2\beta$, $(\xi_1\ V_2) \sim n(0, \Omega)$, $\Sigma = B'\Omega B$, $B = \begin{pmatrix} 1 & 0 \\ -\beta & I_{m-1} \end{pmatrix}$. When $\Pi_{22} = 0$, $\beta$ is not identified in (2) and the disturbances $\xi_1$ are not affected by the value of $\beta$. So, the likelihood is flat and nonzero in the

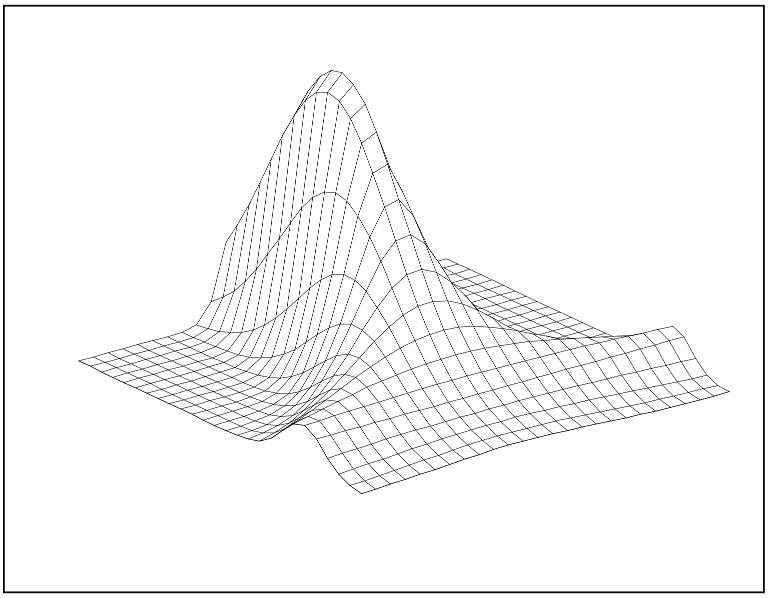

Figure 1: Bivariate Posterior $(\beta, \pi_{22})$ demand equation Tintner model

direction of $\beta$ when $\Pi_{22} = 0$. If we use flat (diffuse) priors in a Bayesian analysis of the INSEM, such that the joint posterior is proportional to the likelihood, also the joint posterior of the different parameters will be flat and nonzero in the direction of $\beta$ for zero values of $\Pi_{22}$. This property is passed on to the marginal posteriors, which are the integrals of the joint posterior over the different parameters. To show the consequences for the marginal posteriors in practice, we calculated the marginal posteriors of the parameters of the demand equation of the "Tintner meat market" model, see Tintner (1952). In this exact identified model, $y_1$ reflects quantity of meat consumed, $y_2$ is the price of meat, $x_1$ is national income per capita, $x_2$ is the cost of processing meat (all variables are in deviation from their mean), $m = 2$, $k_1 = k_2 = 1$.

In figure 1, the joint posterior of $\beta$ and $\Pi_{22}$ is drawn for the Tintner meat market dataset and figure 2 contains the contourlines of this bivariate posterior. The functional form of this posterior is obtained by using a flat prior ($\propto 1$) and integrating out $(\Sigma, \pi_{11}, \Pi_{12})$, and reads,

$$p(\beta, \Pi_{22}|Y, X) \propto |(y_1 - X_2\Pi_{22}\beta)' M_{(X_1\, x_2)} (y_1 - X_2\Pi_{22}\beta)|^{-\frac{1}{2}(T-k_1-m-1)} \quad (3)$$
$$|(\hat{Y}_2 - X_2\Pi_{22})' M_{X_1} (\hat{Y}_2 - X_2\Pi_{22})|^{-\frac{1}{2}(T-k_1-m-1)},$$

as $\hat{Y}_2 = X_1\Pi_{12} + X_2\Pi_{22} + v_2$ and $M_W = I_T - W(W'W)^{-1}W'$, $W = X_1$, $W = (X_1\, x_2)$. Both figures 1, 2, and the functional form of the posterior in (3) show that the marginal posterior does not depend on $\beta$ when $\Pi_{22} = 0$ as it is flat and nonzero in the direction of $\beta$ for zero values of $\Pi_{22}$. This implies that the marginal posterior

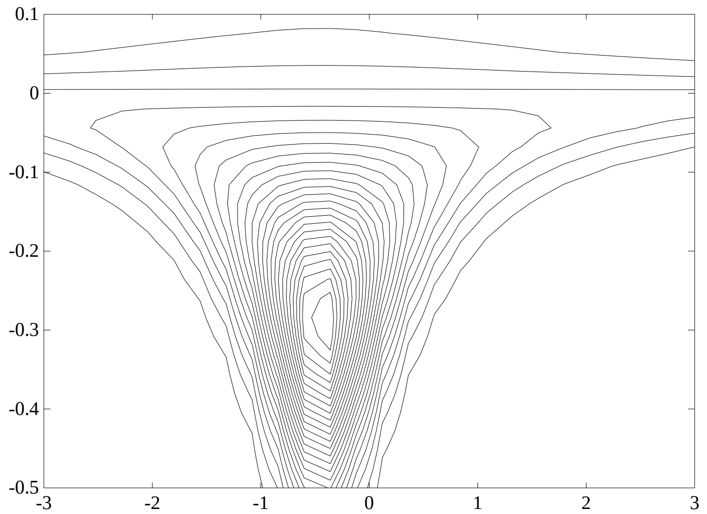


Figure 3: Contourlines marginal posterior $(\beta, \Pi_{22})$ demand equation

of $\Pi_{22}$, which is the integral of the posterior (3) over $\beta$, will be infinite at $\Pi_{22} = 0$ as at this particular value of $\Pi_{22}$, we construct an integral of a function over an infinite parameter region while the function itself does not depend on the parameter $\beta$ over which we integrate. So, the integral will be proportional to the size of the parameter region, *i.e.* infinity. Both the functional form of the marginal posterior of $\Pi_{22}$,

$$p(\Pi_{22}|Y,X) \propto |\Pi_{22}'X_2'M_{(X_1\ y_1)}X_2\Pi_{22}|^{-\frac{1}{2}}\left[\frac{|\Pi_{22}'X_2'M_{(X_1\ Y_2)}X_2\Pi_{22}|}{|\Pi_{22}'X_2'M_{(X_1\ y_1\ Y_2)}X_2\Pi_{22}|}\right]^{-\frac{1}{2}(T-k_1-2(m-1))}$$
$$|(Y_2 - X_2\Pi_{22})'M_{X_1}(Y_2 - X_2\Pi_{22})|^{-\frac{1}{2}(T-k_1-m-1)}, \quad (4)$$

and the marginal posterior of $\Pi_{22}$ for the Tintner dataset from figure 4 show this phenomenon and consequently the value of the posterior of $\Pi_{22}$ is infinite at $\Pi_{22} = 0$.

The nonidentification of $\beta$ has also consequences for its own marginal posterior, which belongs to the class of 1-1 poly $t$ densities, see Bauwens and van Dijk (1989), Drèze (1976), Drèze (1977), Drèze and Richard (1983), and Richard and Tompa (1980) for an efficient algorithm to calculate the moments of this class of densities. This posterior reads

$$p(\beta|Y,X) \propto |(y_1 - Y_2\beta)'M_{(X_1\ X_2)}(y_1 - Y_2\beta)|^{\frac{1}{2}(T-k_1-k_2-m-1)} \quad (5)$$
$$|(y_1 - Y_2\beta)'M_{X_1}(y_1 - Y_2\beta)|^{-\frac{1}{2}(T-k_1-m-1)},$$

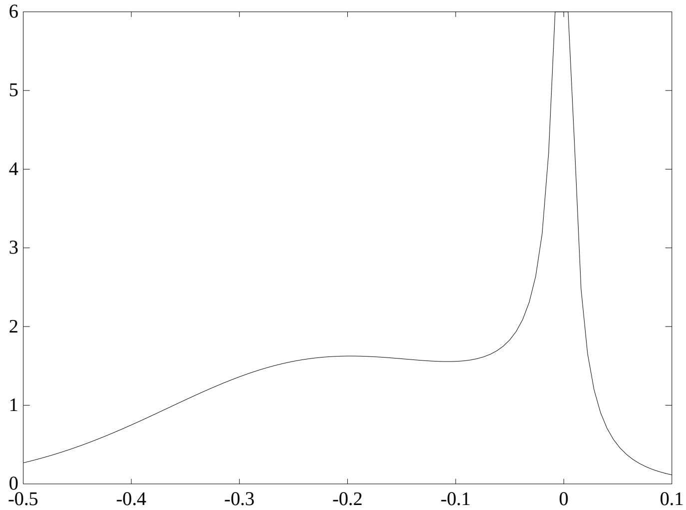


Figure 5: Marginal posterior $\sigma_{22}$ demand equation Tintner model

and it has fat tails resulting from the flat nonzero conditional posterior of $\beta$ given $\Phi_{22} = 0$. For the case of the Tintner model, the marginal posterior is even nonintegrable which is plausible given the fat tails of the marginal posterior of $\beta$ shown in figure 4. In general, the moments of the posterior in (5) exist up to/including the degree of overidentification minus 1 implying that exact identified models lead to nonintegrable posteriors when flat (diffuse) priors are used.

A popular method for numerical calculation of posterior densities is to construct the conditional posteriors and use them to perform Gibbs Sampling, see Gelfand and Smith (1990) and Smith and Roberts (1993). When this Markov Chain Monte Carlo ($MC^2$) algorithm is used to compute the marginal posteriors of the parameters of the INSEM, as in Geweke (1996), the local nonidentification problems lead to a reducible Markov Chain since when a locally nonidentified parameter value is drawn, the sampler continues drawing nonidentified parameter values. Stated differently, the region of locally nonidentified parameter values is an absorbing state in the Markov chain. The posterior, therefore, violates the convergence conditions for Gibbs Samplers as outlined in Roberts and Smith (1994). A solution to this problem is to use informative priors but this approach is questionable when priors are used which are not in accordance with the likelihood, see Kleibergen (1997).

The integrability problems of the posteriors discussed previously result from the dependence of the structural form parameter $\beta$ on $\Phi_{22}$. In classical econometric analysis, see Anderson (1982), Phillips (1983) and Bekker (1994), the pa-

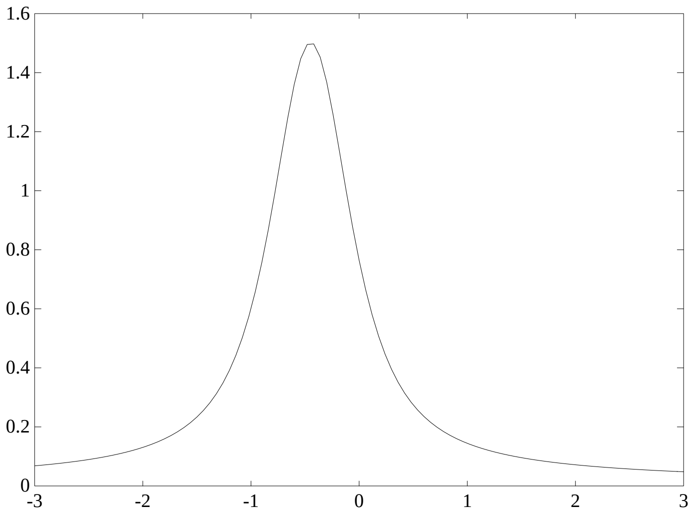

Figure 4: Marginal posterior $\beta$ demand equation Tintner model

rameter $\beta$ is analyzed conditional on a so-called concentration parameter. This is essentially a statistic to test the hypothesis $H_0 : \pi_{22} = 0$ and it shows whether the information in the likelihood is concentrated around $\pi_{22} = 0$. When this concentration parameter tends to infinity when the sample size becomes large, normal asymptotic theory can be applied, see Anderson (1982) and Phillips (1983). When $\pi_{22} = 0$, however, estimators of $\beta$, like 2SLS, converge to random variables, see Phillips (1989). The integrability problems outlined above show that also in a Bayesian analysis $\beta$ has to be analyzed given $\pi_{22}$, which is natural given that the identification problems in the likelihood result from model properties, i.e. the nonidentification of $\beta$ at $\pi_{22} = 0$, and are not the result of inferior data. Since we know *a priori* that these integration problems arise, a framework is needed which formalizes the way the parameters are analyzed conditional on one another and which leads to nonpathological posteriors. This framework is constructed in Kleibergen (1997) and is used in the following sections.

# 3 Priors for the INSEM parameters

In the previous section, we showed that the parameters which suffer from local nonidentification problems should be analyzed conditional on the value of their identifying parameters. This is one of the main properties obtained through the priors constructed in this section. In previous versions of this paper, see Kleibergen and van Dijk (1992,1994a), and also Kleibergen and van Dijk (1994b) and Chao and Phillips (1998), Jeffreys' priors are used to obtain this property. The resulting posterior, when this prior is used, is, however, not nested within the assumed posterior of the parameters of the embedding unrestricted linear model. This is a key property of the priors constructed in this section. The prior we construct in this section results from Kleibergen (1997), where it is shown that a whole range of models can be considered as nonlinear restrictions on the parameters of standard linear models. This gives a general framework for the analysis of a large class of models, see also Kleibergen and Hoek (1996) and Kleibergen and Paap (1997).

## 3.1 SEMs as linear models with nonlinear parameter restrictions

Overidentified SEMs can be considered as a nonlinear restriction on the parameters of a multivariate linear model. It is well known how diffuse and conjugate priors and their resulting posteriors are constructed for the parameters of linear models, see Zellner (1971). When we explicitly consider the SEM as a nonlinear restriction on the parameters of a linear model, the priors and posteriors of the parameters of the SEM result, straightforwardly, as proportional to the priors and posteriors of the parameters of the linear model under the condition that the restrictions on these parameters hold, see Kleibergen (1997).

To analyze the restrictions imposed by a SEM on the parameters of a linear model consider the INSEM (1) and its implied reduced form (2). To show the imposed restrictions, we add a parameter $\lambda$ to this model which is such that when it is nonzero, (*i*) there is an one-to-one correspondence with a standard linear model and when it equals zero both (*ii*) the reduced form of the INSEM results and (*iii*) it is locally uncorrelated with specific other parameters. This latter property is needed to obtain priors and posteriors of the parameters of the INSEM which are invariant with respect to the specification of the model, see Kleibergen (1997) for an exact specification of the conditions the restrictions have to satisfy. Several restrictions imposed on the linear model namely lead to the reduced form of the INSEM but only one restriction leads to priors and posteriors which are invariant with respect to parameter transformations. This invariance property is needed in order to avoid the Borel-Kolmogorov Paradox, see Billingsley (1986) and Press and Richard (1988), and for more details on this posterior invariance, see Kleibergen (1997). The resulting model, which we call the unrestricted SEM, reads,

$$\begin{aligned} \begin{pmatrix} y_1 & Y_2 \end{pmatrix} &= X_1 \begin{pmatrix} \phi_{11} & \Pi_{12} \end{pmatrix} + X_2 \Pi_{22} \begin{pmatrix} \beta & I_{m-1} \end{pmatrix} \\ &\quad + X_2 \Pi_{22\perp} \lambda \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_{\perp} + \begin{pmatrix} \varepsilon_1 & V_2 \end{pmatrix}, \end{aligned} \tag{6}$$

where $\lambda : (k_2 - m + 1) \times 1$ and $\Pi_{22\perp}$, $\begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_{\perp}$ are the orthogonal complements of $A_{22}$, $\begin{pmatrix} \beta & I_{m-1} \end{pmatrix}$ resp., such that $A_{22}'A_{22\perp} \equiv 0$, $\begin{pmatrix} \beta & I_{m-1} \end{pmatrix}\begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_\perp' \equiv 0$, and $A_{22\perp}'A_{22\perp} \equiv I_{k_2-m+1}$, $\begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_\perp \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_\perp' \equiv 1$ (*i.e.* $A_{22\perp} = \begin{pmatrix} -A_{222}A_{221}^{-1} & I_{k_2-m+1} \end{pmatrix}'(I_{k_2-m+1}+A_{222}A_{221}^{-1}A_{221}^{-1\prime}A_{222}')^{-\frac{1}{2}}$, when $A_{22} = \begin{pmatrix} A_{221}' & A_{222}' \end{pmatrix}'$, $A_{221} : (m-1)\times(m-1)$, $A_{222} : (k_2-m+1)\times(m-1)$, and $\begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_\perp = (1+\beta'\beta)^{-\frac{1}{2}}\begin{pmatrix} 1 & -\beta' \end{pmatrix}'$). We note that the orthogonal complements used in other parts of the paper are defined identical to the ones stated above.

It is clear that when $\lambda = 0$, (6) is identical to (2) and since $\lambda$ is multiplied by the orthogonal complements of the matrices containing $\beta$ and $A_{22}$, the information matrix is block diagonal at $\lambda = 0$. We therefore say that $\lambda$ is locally uncorrelated with $\beta$ and $A_{22}$ at $\lambda = 0$. The one-to-one correspondence between the parameters of (6) and a multivariate linear model,

$$\begin{pmatrix} y_1 & Y_2 \end{pmatrix} = \begin{pmatrix} Z_1 & Z_2 \end{pmatrix}\begin{pmatrix} \pi_{11} & \Pi_{12} \\ \pi_1 & \Pi_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 & V_2 \end{pmatrix}, \tag{7}$$

where $\pi_1 : k_2 \times 1$, $\Pi_2 : k_2 \times (m-1)$, can be shown using a Singular Value Decomposition (SVD) of $\Pi = \begin{pmatrix} \pi_1 & \Pi_2 \end{pmatrix}$, see Golub and van Loan (1989) and Magnus and Neudecker (1988) for definitions of a SVD. The equality of (7) and (6) is shown in appendix A and uses the SVD of $\Pi$,

$$\Pi = USV', \tag{8}$$

where $U : k_2 \times k_2$, $U'U = I_{k_2}$; $V : m \times m$, $V'V = I_m$; and $S : k_2 \times m$ is a rectangular matrix containing the (nonnegative) singular values (in decreasing order) on its main diagonal ($= (s_{11}...s_{mm})$). If we now write,

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix},\ S = \begin{pmatrix} S_1 & 0 \\ 0 & s_2 \end{pmatrix} \text{ and } V = \begin{pmatrix} v_{11} & v_{12} \\ V_{21} & v_{22} \end{pmatrix}, \tag{9}$$

where $U_{11}$, $S_1$, $V_{21}$ : $(m-1)\times(m-1)$; $v_{12}$ : $1\times 1$; $v_{11}'$, $v_{22}$ : $(m-1)\times 1$; $U_{12} : (m-1)\times(k_2-m+1)$; $U_{21} : (k_2-m+1)\times(m-1)$; $U_{22} : (k_2-m+1)\times(k_2-m+1)$; $s_2 : (k_2-m+1)\times 1$, then the following relationship between $(A_{22}, \beta, \lambda)$ and $(U, S, V)$ results,

$$A_{22} = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V_{21}',\ \beta = V_{21}'^{-1}v_{11}',\ \text{and } \lambda = (U_{22}U_{22}')^{-\frac{1}{2}}U_{22}s_2v_{12}'(v_{12}v_{12}')^{-\frac{1}{2}}. \tag{10}$$

Furthermore, the SVD shows that $\lambda$ is identified by the smallest singular value of $\Pi$ contained in $s_2$ and is essentially a rotation of $s_2$ since $s_2$ is pre and post-multiplied by orthogonal matrices to obtain $\lambda$. Because the singular value $s_2$ is invariant with respect to the ordering of the variables contained in $Y$ ($= (y_1\ Y_2)$) and $Z_2$, the length of $\lambda$, which is equal to the length of $s_2$ since it is a rotation of $s_2$, is identical for all orderings of the variables contained in $Y$ and $Z_2$. This property is needed to obtain a prior/posterior of the parameters of the INSEM which is invariant with respect to the ordering of the variables in $Y$ and $Z_2$.

When we use the least squares estimator of $\Phi$ in (8), $\hat{\Phi} = (Z_2'M_{Z_1}Z_2)^{-1}Z_2'M_{Z_1}(y_1\ Y_2)$, the estimators for $\beta$ and $\Omega_{22}$ resulting from (10) are identical to the limited information maximum likelihood estimators, see Anderson and Rubin (1949) and Hausman (1983), when the instruments are reasonable, see Kleibergen and Zivot (1998) for a proof of this. The hypothesis $H_0 : \lambda = 0$ can also be tested in that setting to check the validity of the imposed overidentification.

The above shows that the INSEM can be considered as a nonlinear (reduced rank) restriction, $\lambda = 0$, on the parameters of the linear model (7). We therefore construct the priors and posteriors of the parameters of the INSEM (1) as proportional to the priors and posteriors of the parameters of the linear model (7) evaluated in $\lambda = 0$. This framework for constructing priors and posteriors results from Kleibergen (1997) and we discuss its results for the INSEM in the following (sub)section. The framework can also be used in a full system analysis in which SVDs have to be applied recursively. As this becomes notationally more complicated we discuss it in a later section. Note also that the analysis for exact identified SEMs directly results from the standard linear model since in that case there is an one-to-one correspondence between the parameters of the structural form and the linear model.

### 3.2 Prior Framework for SEMs

As shown previously, the INSEM can be considered as a nonlinear restriction on the parameters of a multivariate linear model. It is, however, not possible to analytically construct the conditional posterior of the parameters, $\Pi$, $\sigma_{11}$, $\beta$, $\Omega_{12}$ and $\Omega_{22}$, given the parameter reflecting the restrictions, $\lambda$, see Kleibergen (1997). To show this let $\theta = (\sigma_{11}, \beta, \Omega_{12}, \Omega_{22})$ and $\psi = (\Phi, \sigma_{11}, \Omega_{12})$, then

$$p_{unres}(\theta, \lambda|\Pi, Y, Z) \propto p_{lin}(\psi(\theta, \lambda)|\Pi, Y, Z)\left|\frac{\partial\psi}{\partial(\theta, \lambda)}\right|, \tag{11}$$

where $\psi$ is a function of $\theta$ and $\lambda$, *unres* stands for unrestricted SEM and *lin* for linear model. Assume that the posterior of $\psi$ is well behaved, which is typically the case for the posterior of the parameters of a multivariate linear model, then we cannot give an exact expression of the conditional posterior of $\theta$ given $\lambda$, $p_{unres}(\theta|\lambda, \Pi, Y, Z)$, including its normalizing constants because we cannot construct the marginal posterior of $\lambda$, $p_{unres}(\lambda|Y, Z)$, analytically. This results as $\lambda$ is multiplied by $\Omega_{22}$ and $\begin{pmatrix}\beta & I_{m-1}\end{pmatrix}_{\perp}$ in (6). $\lambda$ is therefore partly a nonlinear function of $\beta$ and $\Omega_{22}$ such that we cannot construct its marginal posterior analytically. So, to obtain a consistent analysis, in the sense that the INSEM has to accord with its embedding linear model, we cannot ignore that the INSEM is a linear model with nonlinear restrictions on its parameters and just proceed by constructing the posterior like in section 2. In that section we namely implicitly assumed that the involved posterior is proportional to $p_{ansem}(\theta,\lambda|X,Y,Z)|_{\lambda=0}$. This implies a posterior for the parameters of the linear model in $\lambda=0$,

$$p_{lin}(\pi|X,Y,Z)|_{\lambda=0} \propto p_{ansem}(\pi(\theta,\lambda)|X,Y,Z)|_{\lambda=0}|(\frac{\partial(\theta,\lambda)}{\partial\pi})|_{\lambda=0}|. \quad (12)$$

As shown in section 2 the posterior $p_{ansem}(\theta,\lambda|X,Y,Z)|_{\lambda=0}$ is badly behaved and the resulting $p_{lin}(\pi|X,Y,Z)|_{\lambda=0}$ is thus also badly behaved. This is, however, a posterior of the parameters of a linear model which is normally well-behaved and well understood. It therefore does not belong to (or is nested within) the standard class of posteriors of parameters of linear models. For more details we refer to Kleibergen (1997). Also slight modifications of the INSEM, so for example an INSEM which is nested in the original INSEM, lead to a different implied posterior of the parameters of the embedding linear model. We therefore use the priors/posteriors of the parameters of the linear model as a base to construct the priors/posteriors of the parameters of the INSEM. So, we specify a prior for the parameters of the linear model, for example a diffuse or natural-conjugate prior, see Zellner (1971), and we evaluate this prior in $\lambda=0$ to obtain the prior for the INSEM, see Kleibergen (1997) and Kleibergen and Paap (1997),

$$\begin{aligned} p_{insem}(\theta,\Omega) &\propto p_{ansem}(\theta,\lambda,\Omega)|_{\lambda=0} \\ &\propto p_{lin}(\pi(\theta,\lambda),\Omega)|_{\lambda=0}|(\frac{\partial\pi}{\partial(\theta,\lambda)})|_{\lambda=0}|. \end{aligned} \quad (13)$$

where *insem* stands for INSEM. We note that we can also perform the construction of the prior vice versa by constructing a prior on the structural form parameters and check whether the implied prior on the parameters of the embedding linear model is plausible, see Kleibergen (1997) and Kleibergen and Zivot (1998).

### 3.2.1 Diffuse Prior

Using the framework resulting from (13), a diffuse (Jeffreys') prior for the parameters of the linear model, $(\pi_{11},\Pi_{12},\Phi,\Omega)$,

$$p_{lin}(\pi_{11},\Pi_{12},\Phi,\Omega) \propto |\Omega|^{-\frac{1}{2}(k+m+1)} \propto |\Omega|^{-\frac{1}{2}(m+1)}|\Omega^{-1}\otimes Z'Z|^{\frac{1}{2}}, \quad (14)$$

implies the prior for the parameters of the INSEM, $(\beta,\pi_{11},\Pi_{12},\Pi_{22},\Omega)$,

$$\begin{aligned} & p_{insem}(\beta,\pi_{11},\Pi_{12},\Pi_{22},\Omega) \quad (15)\\ \propto\ & |\Omega|^{-\frac{1}{2}(m+1)}|\Omega^{-1}\otimes Z'Z|^{\frac{1}{2}}|J(\Phi,(\Pi_{22},\beta,\lambda))|_{\lambda=0}| \\ \propto\ & |\Omega|^{-\frac{1}{2}(k_1+m+1)}|Z_1'Z_1|^{\frac{1}{2}m}|B\Omega B'|^{-\frac{1}{2}(k_2-m+1)}|\Pi_{22}'(Z_2'M_{Z_1}Z_2)^{-1}\Pi_{22}|^{-\frac{1}{2}} \\ & |\begin{pmatrix} B\Omega^{-1}B'\otimes Z_2'M_{Z_1}Z_2 & B\Omega^{-1}e_1\otimes Z_2'M_{Z_1}Z_2\Pi_{22} \\ e_1'\Omega^{-1}B'\otimes\Pi_{22}'Z_2'M_{Z_1}Z_2 & e_1'\Omega^{-1}e_1\otimes\Pi_{22}'Z_2'M_{Z_1}Z_2\Pi_{22} \end{pmatrix}|^{\frac{1}{2}} \end{aligned}$$

where $e_1 : m \times 1$ is the first $m$ dimensional unity vector, $B = \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}$, $|J(\Phi, (\Pi_{22}, \beta, \lambda))| = |\frac{\partial \pi}{\partial(\theta, \lambda)}|$ and is constructed in appendix A.

The prior (15) shows that $\beta$ is analyzed conditional on the value of $\Pi_{22}$, as it should be according to the local nonidentification of $\beta$ for lower rank values of $\Pi_{22}$. Furthermore, the prior shows the functional form of a diffuse prior for the parameters of the INSEM. This accords with our conclusions from the previous section that diffuseness for models like the INSEM has to be defined in a different way than the usual one for parameters of linear models.

We note that the prior (15) is the Jeffreys' prior of the unrestricted reduced form of the INSEM (6) evaluated in $\lambda = 0$. In Kleibergen and van Dijk (1994a), the Jeffreys' prior of the reduced form of the INSEM (2) is used to obtain well-behaved posteriors, see also Chao and Phillips (1996). This prior is apart from $|B_{\perp} \Omega B_{\perp}'|^{-\frac{1}{2}(k_2-m+1)} |\Pi_{22}'(Z_2' M_{Z_1} Z_2)^{-1} \Pi_{22}|^{-\frac{1}{2}}$ identical to (15). We use (15) instead of that prior for three reasons. First, (15) results in a generic manner from the linear model (7) instead of being a tool specially designed to solve the integrability problem of the marginal posterior. Second, the concept for constructing (15) can also be applied in the full system analysis while the Jeffreys' priors of the reduced forms of full system SEMs are intractable. Third, although we use data matrices in (15) to obtain a more interpretable expression of the prior, it is not data dependent as no data matrices appear in the jacobian $J(\Phi, (\Pi_{22}, \beta, \lambda))$ and $|Z'Z|$ can just be left out. A Jeffreys' prior on the reduced form (2) is data dependent however.

The prior (15) is identical to the Jeffreys' prior for the reduced form of the orthonormal SEM, see Phillips (1983) and Chao and Phillips (1996), where $\Omega = I_m$ and $Z_2' M_{Z_1} Z_2 = I_{k_2}$, as $B_{\perp} B_{\perp}' = 1$, $\Pi_{22}' \Pi_{22} = I_{k_2-m+1}$. Using Rayleigh quotients it can also be shown that the ratio of the prior (15) and a Jeffreys' prior on the reduced form (2) is bounded between finite nonzero constants.

### 3.2.2 Natural Conjugate Prior

In case of a natural conjugate prior for the parameters of the linear model, we specify an inverted-Wishart prior for $\Omega$ and a matrix normal prior for $(\pi_{11}, \Pi_{12}, \Phi)$ given $\Omega$,

$$
\begin{aligned}
p_{Lin}(\Omega) &\propto |S|^{\frac{1}{2}h} |\Omega|^{-\frac{1}{2}(h+m+1)} \exp[-\tfrac{1}{2} tr(\Omega^{-1} S)] \qquad (16) \\
p_{Lin}(\pi_{11}, \Pi_{12}, \Phi|\Omega) &\propto |\Omega|^{-\frac{1}{2}m} |A|^{\frac{1}{2}k} \exp[-\tfrac{1}{2} tr(\Omega^{-1}(\begin{pmatrix} \pi_{11} & \Pi_{12} \\ \pi_{21} & \Phi_2 \end{pmatrix} - P)' \\
& \quad A(\begin{pmatrix} \pi_{11} & \Pi_{12} \\ \pi_{21} & \Phi_2 \end{pmatrix} - P))],
\end{aligned}
$$

where $S : m \times m$, $A : k \times k$, $S$ and $A$ are positive definite symmetric (pds) matrices, $P : k \times m$, and $h$ is the prior degrees of freedom parameter. The matrix

$A$ can be decomposed as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{17}$$

where $A_{11} : k_1 \times k_1$, $A_{12} = A_{21}' : k_2 \times k_1$, $A_{22} : k_2 \times k_2$. The prior of the parameters of the INSEM resulting from $p_{lin}(\pi_{11}, \Pi_{12}, \Phi, \Omega)$ can again be constructed using (16),

$$\begin{aligned} & p_{insem}(\beta, \pi_{11}, \Pi_{12}, \Pi_{22}, \Omega) \qquad\qquad (18) \\ \propto\ & p_{lin}(\pi_{11}, \Pi_{12}, \Phi(\beta, \Pi_{22}, \lambda), \Omega)|_{\lambda=0} |J(\Phi, (\Pi_{22}, \beta, \lambda))|_{\lambda=0}| \\ \propto\ & |A|^{\frac{1}{2}h} |\Omega|^{-\frac{1}{2}(h+k_1+m+1)} |A_{11}|^{\frac{1}{2}m} |B_{\perp}' \Omega B_{\perp}|^{-\frac{1}{2}(k_2-m-1)} |\Pi_{22}' A_{22.1} \Pi_{22}|^{-\frac{1}{2}} \\ & \left| \begin{pmatrix} B\Omega^{-1}B' \otimes A_{22.1} & B\Omega^{-1}e_1 \otimes A_{22.1}\Pi_{22} \\ e_1'\Omega^{-1}B' \otimes \Pi_{22}'A_{22.1} & e_1'\Omega^{-1}e_1 \otimes \Pi_{22}'A_{22.1}\Pi_{22} \end{pmatrix} \right|^{\frac{1}{2}} \\ & \exp[-\frac{1}{2}tr(\Omega^{-1}(S + (\begin{pmatrix} \pi_{11} & \Pi_{12} \\ \Pi_{22}\beta & \Pi_{22} \end{pmatrix} - P)'A(\begin{pmatrix} \pi_{11} & \Pi_{12} \\ \Pi_{22}\beta & \Pi_{22} \end{pmatrix} - P)))], \end{aligned}$$

where $A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ and the specification of (18) is not unique in the sense that certain scaling factors are used in order to obtain a more interpretable expression.

It may be that we have more knowledge about possible values of the parameters of the INSEM than about the parameters of the linear model. This knowledge can be used in the construction of the prior of the parameters of the linear model though as these parameters are an exact function of the parameters of the INSEM when the restriction $\lambda = 0$ holds. We can also directly specify a prior on the parameters of the INSEM and check whether the implied prior on the parameters of the embedding linear model is plausible, see Kleibergen and Zivot (1998).

The prior (18) does not belong to a known class of probability density functions and we do not know analytical expressions of its moments (which even only exist up to its first order (but not including)) or normalizing constant. These properties can be calculated using Monte-Carlo simulation and in the fifth section we construct a simulation algorithm to obtain drawings from (18).

# 4 Posteriors of the INSEM parameters

The framework for constructing the priors of the parameters of the INSEM can directly be applied to construct the posteriors of the parameters of the INSEM. This results since the likelihood of the INSEM is a continuous function of the parameters such that the posterior, which is proportional to the product of the prior and the likelihood, can be evaluated in the same way as the prior,

$$p_{insem}(\beta, \pi_{11}, \Pi_{12}, \Pi_{22}, \Omega | Y, Z) \tag{19}$$

$$\propto p_{insem}(\beta, \phi_{11}, \Pi_{12}, \Pi_{22}, \Omega) L_{insem}(Y|\beta, \phi_{11}, \Pi_{12}, \Pi_{22}, \Omega, X)$$
$$\propto p_{ansem}(\beta, \lambda, \phi_{11}, \Pi_{12}, \Pi_{22}, \Omega)|_{\lambda=0} L_{ansem}(Y|\beta, \lambda, \phi_{11}, \Pi_{12}, \Pi_{22}, \Omega, X)|_{\lambda=0}$$
$$\propto p_{lin}(\phi_{11}, \Pi_{12}, \Phi(\beta, \lambda, \Pi_{22}), \Omega)|_{\lambda=0} |J(\Phi, (\Pi_{22}, \beta, \lambda))|_{\lambda=0}| \; L_{lin}(Y|\phi_{11}, \Pi_{12}, \Phi(\beta, \lambda, \Pi_{22}), \Omega, X)|_{\lambda=0}.$$

In the following two subsections, we construct the posteriors for different specifications of the prior, *i.e.* a diffuse and natural conjugate prior.

## 4.1 Posterior INSEM using Diffuse Prior

Using the diffuse prior (15), the joint posterior of the parameters of the INSEM can directly be constructed from this prior and the likelihood using (13),

$$\begin{aligned}
& p_{insem}(\beta, \phi_{11}, \Pi_{12}, \Pi_{22}, \Omega|Y, X) \qquad (20)\\
\propto\; & p_{insem}(\beta, \phi_{11}, \Pi_{12}, \Pi_{22}, \Omega) L(Y|\beta, \phi_{11}, \Pi_{12}, \Pi_{22}, \Omega, X)\\
\propto\; & |\Omega|^{-\frac{1}{2}(T+k_1+m+1)} |X_1'X_1|^{\frac{1}{2}m} |B_\perp' \Omega B_\perp|^{-\frac{1}{2}(k_2-m+1)} |\Pi_{22\perp}'(X_2'M_{X_1}X_2)^{-1}\Pi_{22\perp}|^{-\frac{1}{2}}\\
& \left|\begin{pmatrix} B\Omega^{-1}B' \otimes X_2'M_{X_1}X_2 & B\Omega^{-1}e_1 \otimes X_2'M_{X_1}X_2\Pi_{22} \\ e_1'\Omega^{-1}B' \otimes \Pi_{22}'X_2'M_{X_1}X_2 & e_1'\Omega^{-1}e_1 \otimes \Pi_{22}'X_2'M_{X_1}X_2\Pi_{22} \end{pmatrix}\right|^{\frac{1}{2}}\\
& \exp[-\tfrac{1}{2} tr(\Omega^{-1}(\begin{pmatrix} y_1 & Y_2 \end{pmatrix} - \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} \phi_{11} & \Pi_{12} \\ \Pi_{22}\beta & \Pi_{22} \end{pmatrix})'\\
& (\begin{pmatrix} y_1 & Y_2 \end{pmatrix} - \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} \phi_{11} & \Pi_{12} \\ \Pi_{22}\beta & \Pi_{22} \end{pmatrix}))].
\end{aligned}$$

The posterior (20) does not belong to a known class of probability density functions nor do any of the conditional posteriors, apart from the conditional posterior of $(\phi_{11}, \Pi_{12})$ given $(\beta, \Pi_{22}, \Omega)$, which is matrix-normal, belong to a known class of probability density functions. So, we can only analytically integrate out $(\phi_{11}, \Pi_{12})$ to obtain the marginal posterior of $(\beta, \Pi_{22}, \Omega)$,

$$\begin{aligned}
& p_{insem}(\beta, \Pi_{22}, \Omega|Y, X) \propto |\Omega|^{-\frac{1}{2}(T+m+1)} |B_\perp' \Omega B_\perp|^{-\frac{1}{2}(k_2-m+1)}\\
& \left|\begin{pmatrix} B\Omega^{-1}B' \otimes X_2'M_{X_1}X_2 & B\Omega^{-1}e_1 \otimes X_2'M_{X_1}X_2\Pi_{22} \\ e_1'\Omega^{-1}B' \otimes \Pi_{22}'X_2'M_{X_1}X_2 & e_1'\Omega^{-1}e_1 \otimes \Pi_{22}'X_2'M_{X_1}X_2\Pi_{22} \end{pmatrix}\right|^{\frac{1}{2}}\\
& |\Pi_{22\perp}'(X_2'M_{X_1}X_2)^{-1}\Pi_{22\perp}|^{-\frac{1}{2}} \exp[-\tfrac{1}{2} tr(\Omega^{-1}(\begin{pmatrix} y_1 & Y_2 \end{pmatrix} - X_2\Pi_{22} \begin{pmatrix} \beta & I_{m-1} \end{pmatrix})'\\
& M_{X_1}(\begin{pmatrix} y_1 & Y_2 \end{pmatrix} - X_2\Pi_{22} \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}))]. \qquad (21)
\end{aligned}$$

which shows the functional form of the kernel of the density of a matrix normal distributed random matrix with reduced rank, see Kleibergen (1997). The posterior (21) is proportional to the product of the marginal posterior of $(\Phi, \Omega)$ and the jacobian of the transformation evaluated in $\lambda = 0$,

$$p_{insem}(\beta, \Pi_{22}, \Omega|Y, X) \propto p_{lin}(\Phi(\beta, \lambda, \Pi_{22}), \Omega|Y, X)|_{\lambda=0} |J(\Phi, (\Pi_{22}, \beta, \lambda))|_{\lambda=0}|. \qquad (22)$$

In section 5, we construct Importance and Metropolis-Hastings samplers for calculating the marginal posteriors of (21) which use (22).

## 4.2 Posterior INSEM using Natural Conjugate Prior

Identical to the posterior of the parameters of the INSEM using a diffuse prior (20), we can construct the posterior of the parameters of the INSEM when we use the natural conjugate prior (18),

$$
\begin{aligned}
& p_{insem}(\beta, \phi_{11}, \Pi_{12}, \Pi_{22}, \Omega | Y, X) \qquad (23) \\
\propto \ & |\Omega|^{-\frac{1}{2}(T+k_1+m+1)} |(A + X'X)_{11}|^{\frac{1}{2}m} |B_{\perp}\Omega B_{\perp}'|^{-\frac{1}{2}(k_2-m+1)} \\
& \left| \begin{pmatrix} B\Omega^{-1}B' \otimes (A + X'X)_{22.1} & B\Omega^{-1}e_1 \otimes (A + X'X)_{22.1}\Pi_{22} \\ e_1'\Omega^{-1}B' \otimes \Pi_{22}'(A + X'X)_{22.1} & e_1'\Omega^{-1}e_1 \otimes \Pi_{22}'(A + X'X)_{22.1}\Pi_{22} \end{pmatrix} \right|^{\frac{1}{2}} \\
& |\Pi_{22\perp}'(A + X'X)_{22.1}^{-1}\Pi_{22\perp}|^{-\frac{1}{2}} \exp[-\frac{1}{2}tr(\Omega^{-1}(S + (\begin{pmatrix} \phi_{11} & \Pi_{12} \\ \Pi_{22}\beta & \Pi_{22} \end{pmatrix} - \hat{\Pi})' \\
& (A + X'X)(\begin{pmatrix} \phi_{11} & \Pi_{12} \\ \Pi_{22}\beta & \Pi_{22} \end{pmatrix} - \hat{\Pi})))].
\end{aligned}
$$

where $\hat{\Pi} = (A + X'X)^{-1}(X'Y + A'P)$, $S = G + Y'Y - \hat{\Pi}'(A + X'X)\hat{\Pi}$, $Y = \begin{pmatrix} y_1 & Y_2 \end{pmatrix}$. Again similar to the posterior using a diffuse prior (20), only the conditional posterior of $(\phi_{11}, \Pi_{12})$ given $(\beta, \Pi_{22}, \Omega)$ belongs to a known class of probability density functions and $(\phi_{11}, \Pi_{12})$ are the only parameters which can be integrated out analytically to obtain the marginal posterior of $(\beta, \Pi_{22}, \Omega)$,

$$
\begin{aligned}
& p_{insem}(\beta, \Pi_{22}, \Omega | Y, X) \propto |\Omega|^{-\frac{1}{2}(T+m+1)} |B_{\perp}\Omega B_{\perp}'|^{-\frac{1}{2}(k_2-m+1)} \\
& \left| \begin{pmatrix} B\Omega^{-1}B' \otimes (A + X'X)_{22.1} & B\Omega^{-1}e_1 \otimes (A + X'X)_{22.1}\Pi_{22} \\ e_1'\Omega^{-1}B' \otimes \Pi_{22}'(A + X'X)_{22.1} & e_1'\Omega^{-1}e_1 \otimes \Pi_{22}'(A + X'X)_{22.1}\Pi_{22} \end{pmatrix} \right|^{\frac{1}{2}} \\
& |\Pi_{22\perp}'(A + X'X)_{22.1}^{-1}\Pi_{22\perp}|^{-\frac{1}{2}} \exp[-\frac{1}{2}tr(\Omega^{-1}(S + (\Pi_{22}\begin{pmatrix} \beta & I_{m-1} \end{pmatrix} - \hat{\Pi}_2)' \\
& (A + X'X)_{22.1}(\Pi_{22}\begin{pmatrix} \beta & I_{m-1} \end{pmatrix} - \hat{\Pi}_2)))]. \qquad (24)
\end{aligned}
$$

where $\hat{\Pi} = \begin{pmatrix} \hat{\Pi}_1' & \hat{\Pi}_2' \end{pmatrix}'$, $\hat{\Pi}_1 : k_1 \times m$, $\hat{\Pi}_2 : k_2 \times m$.

Again (22) applies to this posterior and we use it in the following section to construct a posterior simulator.

# 5 Simulating Posteriors

As mentioned before the posteriors (21) and (24) do not belong to a standard class of probability density functions nor do their conditional posteriors. We can therefore not perform Gibbs sampling as the conditional posteriors are nonstandard. The simulation algorithms constructed in this section therefore generate drawings from a probability density function which approximates the true posterior. To correct for not drawing from the true posterior, weights are attached to each drawing of the parameters proportional to the ratio of the posterior and the approximating density in the generated parameter points. These weights can be used both in Importance, see Kloek and van Dijk (1978) and Geweke (1989), and Metropolis-Hastings, see Metropolis *et. al.* (1953) and Hastings (1970), algorithms to draw from the posterior. We first discuss the construction of the weights and the approximating density and hereafter we briefly discuss the two different simulation algorithms.

We use the posterior of the unrestricted SEM, $p_{unsem}(\beta,\lambda,\Pi_{22},\Omega|Y,Z)$, as approximating density of the posterior of the INSEM, $p_{insem}(\beta,\Pi_{22},\Omega|Y,Z)$. The posterior of the unrestricted SEM contains the parameter $\lambda$, however, which is not present in the posterior of the INSEM. In order to obtain a density which both accords with the posterior of the INSEM and contains $\lambda$, we assume that $\lambda$ is generated given $(\beta,\Pi_{22},\Omega)$ from a proper conditional density $g(\lambda|\beta,\Pi_{22},\Omega)$, which we specify ourselves, see Chen (1994), Verdinelli and Wasserman (1995), Kleibergen (1997) and Kleibergen and Paap (1997). Furthermore, we assume that $\beta$, $\Pi_{22}$ and $\Omega$ are generated from $p_{insem}(\beta,\Pi_{22},\Omega|Y,Z)$. So, as density function to be approximated by $p_{unsem}(\beta,\lambda,\Pi_{22},\Omega|Y,Z)$ we have,

$$g(\lambda|\beta,\Pi_{22},\Omega)p_{insem}(\beta,\Pi_{22},\Omega|Y,Z) \propto g(\lambda|\beta,\Pi_{22},\Omega)(p_{unsem}(\beta,\lambda,\Pi_{22},\Omega|Y,Z)|_{\lambda=0}). \tag{25}$$

The weight function thus becomes,

$$w(\beta,\lambda,\Pi_{22},\Omega) = \frac{g(\lambda|\beta,\Pi_{22},\Omega)(p_{unsem}(\beta,\lambda,\Pi_{22},\Omega|Y,Z)|_{\lambda=0})}{p_{unsem}(\beta,\lambda,\Pi_{22},\Omega|Y,Z)}. \tag{26}$$

The quality of the approximating density $p_{unsem}(\beta,\lambda,\Pi_{22},\Omega|Y,Z)$ crucially depends on the chosen specification of $g(\lambda|\beta,\Pi_{22},\Omega)$. In case we use the diffuse prior for the parameters of the INSEM (15), a natural choice of $g(\lambda|\beta,\Pi_{22},\Omega)$ is,

$$\begin{aligned} g(\lambda|\beta,\Pi_{22},\Omega) &= (2\pi)^{-\frac{1}{2}(k_2-m+1)}|B_{\perp}\Omega^{-1}B_{\perp}'|^{\frac{1}{2}(k_2-m+1)}|\Pi_{22\perp}'Z_2'M_{Z_1}Z_2\Pi_{22\perp}|^{\frac{1}{2}} \\ &\exp[-\frac{1}{2}tr(B_{\perp}\Omega^{-1}B_{\perp}'(\lambda-\hat{\lambda})'\Pi_{22\perp}'Z_2'M_{Z_1}Z_2\Pi_{22\perp}(\lambda-\hat{\lambda}))], \end{aligned} \tag{27}$$

where $\hat{\lambda} = (\Pi_{22\perp}'Z_2'M_{Z_1}Z_2\Pi_{22\perp})^{-1}\Pi_{22\perp}'Z_2'M_{Z_1}(Y-Z_2\Pi_{22}B)\Omega^{-1}B_{\perp}'(B_{\perp}\Omega^{-1}B_{\perp}')^{-1}$, while

$$\begin{aligned} g(\lambda|\beta,\Pi_{22},\Omega) &= (2\pi)^{-\frac{1}{2}(k_2-m+1)}|B_{\perp}\Omega^{-1}B_{\perp}'|^{\frac{1}{2}(k_2-m+1)}|\Pi_{22\perp}'(N+Z'Z)_{22.1}\Pi_{22\perp}|^{\frac{1}{2}} \\ &\exp[-\frac{1}{2}tr(B_{\perp}\Omega^{-1}B_{\perp}'(\lambda-\hat{\lambda})'\Pi_{22\perp}'(N+Z'Z)_{22.1}\Pi_{22\perp}(\lambda-\hat{\lambda}))], \end{aligned} \tag{28}$$

where $\hat{\lambda} = (\Pi_{22\perp}'(N+Z'Z)_{22.1}\Pi_{22\perp})^{-1}\Pi_{22\perp}'((NP+Z'Y)_2-(N+Z'Z)_{22.1}\Pi_{22}B)\Omega^{-1}B_{\perp}'(B_{\perp}\Omega^{-1}B_{\perp}')^{-1}$, $NP+Z'Y = ((NP+Z'Y)_1'\ (NP+Z'Y)_2')'$, $(NP+Z'Y)_1 : k_1 \times m$,

$(X'P = S'Y)_2 : k_2 \times m$, is a natural choice of $g(\lambda|\beta, \Pi_{22}, \Omega)$ when we use the natural conjugate prior (16).

The weight function resulting from these choices of $g$ read in both cases,

$$w(\beta, \lambda, \Pi_{22}, \Omega) = \frac{|J(\Pi, (\beta, \lambda, \Pi_{22}))|_{\lambda=0}|}{|J(\Pi, (\beta, \lambda, \Pi_{22}))|} g(\lambda|\beta, \Pi_{22}, \Omega)|_{\lambda=0}, \qquad (29)$$

where $g(\lambda|\beta, \Pi_{22}, \Omega)$ should be chosen from (27) and (28) according to the involved prior. In appendix B, we show that $|J(\Pi, (\beta, \lambda, \Pi_{22}))| \geq |J(\Pi, (\beta, \lambda, \Pi_{22}))|_{\lambda=0}|$ such that the ratio of the jacobians in (29) is always finite. Furthermore as $g(\lambda|\beta, \Pi_{22}, \Omega)|$ is a proper conditional density, it is also finite and the weight function is consequently always finite.

When $\lambda = 0$, the ratio of jacobians in (29) is equal to one and the weight function then simplifies to the proposed conditional density of $\lambda$ evaluated in $\lambda = 0$. The weight function is therefore always finite and nonzero when $\lambda = 0$. All drawings of $(\beta, \lambda, \Pi_{22})$ for which $\lambda = 0$ thus get a finite nonzero weight. This has consequences for the existence of moments of the posterior $p_{invar}(\beta, \Pi_{22}, \Omega|Y, X)$ since it implies that the degree of finite moments is determined by the transformation of $\Pi$ to $(\beta, \lambda, \Pi_{22})$. According to (10), $\beta = V_{21}'^{-1} v_{11}'$. As no restrictions are imposed on the rank of $V_{21}$, this implies that the posterior of $\beta$ has Cauchy type tails and no finite mean and variance, see Kleibergen and Zivot (1998). Note also that the sampling distributions of limited information maximum likelihood estimators have Cauchy type tails, see Anderson (1982) and Phillips (1983).

We summarize the different steps involved in obtaining the weight function, attached to the $i-$th drawing, $i = 1, \ldots, N$, in a simulation algorithm as follows, see also Kleibergen (1997) and Kleibergen and Paap (1997),

- Draw $\Omega^i$ from $p_{lin}(\Omega|Y, X)$
  Draw $\Pi^i$ from $p_{lin}(\Pi|\Omega^i, Y, X)$.
- Perform a singular value decomposition of $\Pi^i = U^i S^i V^{i\prime}$
- Compute $\beta$, $\lambda$, $\Pi_{22}$ according to (8)-(10)
- Compute $w(\beta^i, \lambda^i, \Pi_{22}^i, \Omega^i)$ according to (29)
- Draw $\pi_{11}^i$, $\Pi_{12}^i$ from $p_{lin}(\pi_{11}, \Pi_{12}|\Pi(\beta^i, \lambda, \Pi_{22}^i), \Omega^i, Y, X)|_{\lambda=0}$

The posteriors of the linear model parameters, $\Omega$ and $\Pi$, used in the first step, are standard density functions, *i.e.* inverted-Wishart and matrix normal respectively, in case of diffuse or natural conjugate priors. The exact functional specification of these densities depends on the specification of the involved priors and is straightforward to construct, *i.e.*,

$$p_{lin}(\Omega|Y, X) \propto |\Omega|^{-\frac{1}{2}(T+k+m+1)} \exp[-\frac{1}{2} tr(\Omega^{-1} S)], \qquad (30)$$

where $l = 0$, $S = Y'M_{X}Y$ in case of the diffuse prior, and $l = k$ and $S = \tilde{S}$ in case of the natural conjugate prior, and

$$p_{lim}(\Phi|\Omega, Y, Z) \propto |\Omega|^{-\frac{1}{2}k_2} \exp[-\frac{1}{2}tr(\Omega^{-1}(\Phi - \hat{\Phi})'M(\Phi - \hat{\Phi}))], \tag{31}$$

where $\hat{\Phi} = (Z_2'M_{Z_1}Z_2)^{-1}Z_2'M_{Z_1}Y$, $M = Z_2'M_{Z_1}Z_2$, in case of the diffuse prior, and $\hat{\Phi} = \tilde{P}_2$, $M = (A + Z'Z)_{22.1}$ in case of the natural conjugate prior. In Kleibergen and Zivot (1998) another simulation algorithm to generate drawings from the posterior $p_{inexact}(\beta, \Pi_{22}, \Omega|Y, Z)$ is constructed which is sometimes more efficient but is more difficult to generalize to the full information case.

The weight function can either be used in an Importance or Metropolis-Hastings Sampling algorithm to calculate the marginal posteriors or moments of these. Using the Importance Sampling algorithm, see Kloek and van Dijk (1978) and Geweke (1989), we approximate the moment $E(g(\pi_{11}, \Pi_{12}, \beta, \Pi_{22}, \Omega))$ by

$$\hat{E}(g(\pi_{11}, \Pi_{12}, \beta, \Pi_{22}, \Omega)) = \frac{\sum_{i=1}^{N} w(\beta^i, \lambda^i, \Pi_{22}^i, \Omega^i)g(\pi_{11}^i, \Pi_{12}^i, \beta^i, \Pi_{22}^i, \Omega^i)}{\sum_{i=1}^{N} w(\beta^i, \lambda^i, \Pi_{22}^i, \Omega^i)}, \tag{32}$$

where we use $\hat{E}$ to indicate that it is an estimator of the true expectation $E$. In Geweke (1989), it is shown that under quite general conditions central limit theorems can be used to prove the convergence of the approximation (32) to its true value. As the weights are always finite, they satisfy the conditions for the central limit theorems to apply and statistics can be calculated which show the numerical accuracy of the approximation (32).

The weights (26) can also be used in a Metropolis-Hastings (M-H) algorithm, see Metropolis *et. al.* (1953) and Hastings (1970), known as the independence sampler, see Tierney (1994). This algorithm constructs a Markov Chain from the drawn $(\pi_{11}^i, \Pi_{12}^i, \beta^i, \Pi_{22}^i, \Omega^i)$'s. The $(\pi_{11}^i, \Pi_{12}^i, \beta^i, \Pi_{22}^i, \Omega^i)$'s in this Markov Chain are accepted as drawings from the posterior. This is achieved using the following steps,

0. $i = 1$

1. Draw $(\pi_{11}^{i+1}, \Pi_{12}^{i+1}, \beta^{i+1}, \Pi_{22}^{i+1}, \Omega^{i+1})$ using the simulation scheme stated previously. Given that $(\pi_{11}^i, \Pi_{12}^i, \beta^i, \Pi_{22}^i, \Omega^i)$ is accepted as drawing from the posterior, $(\pi_{11}^{i+1}, \Pi_{12}^{i+1}, \beta^{i+1}, \Pi_{22}^{i+1}, \Omega^{i+1})$ is accepted as the $(i+1)$−th drawing from the posterior with probability, $\min(\frac{w(\beta^i, \lambda^i, \Pi_{22}^i, \Omega^i)}{w(\beta^{i+1}, \lambda^{i+1}, \Pi_{22}^{i+1}, \Omega^{i+1})}, 1)$, otherwise $(\pi_{11}^{i+1}, \Pi_{12}^{i+1}, \beta^{i+1}, \Pi_{22}^{i+1}, \Omega^{i+1}) = (\pi_{11}^i, \Pi_{12}^i, \beta^i, \Pi_{22}^i, \Omega^i)$.

2. $i = i + 1$. Go to 1.

When the resulting Markov Chain, $(\pi_{11}^i, \Pi_{12}^i, \beta^i, \Pi_{22}^i, \Omega^i)$, $i = 1, ...$; has converged to its equilibrium distribution, say after $N$ drawings, we can record

$(\pi_{11}^i, \Pi_{12}^i, \beta^i, \Pi_{22}^i, \Omega^i)$, $i = H+1, \ldots;$ as simulated values of the parameters from the posterior.

The simulation algorithms can also be used to calculate other properties of the posterior, like Bayes Factors and Bayesian Lagrange Multiplier Statistics, see Kleibergen and Paap (1997), and to obtain drawings from the natural conjugate prior (18). In that case, the natural choice of the involved $g(\lambda|\beta, \Pi_{22}, \Omega)$ reads,

$$g(\lambda|\beta, \Pi_{22}, \Omega) = (2\pi)^{-\frac{1}{2}(k_2-m+1)} |B_{\perp}\Omega^{-1}B_{\perp}'|^{\frac{1}{2}(k_2-m+1)} |\Pi_{22\perp}' X_{22.1} \Pi_{22\perp}|^{\frac{1}{2}}$$
$$\exp[-\frac{1}{2}tr(B_{\perp}\Omega^{-1}B_{\perp}'(\lambda-\hat{\lambda})'\Pi_{22\perp}'X_{22.1}\Pi_{22\perp}(\lambda-\hat{\lambda}))], \qquad (33)$$

where $\hat{\lambda} = (\Pi_{22\perp}'X_{22.1}\Pi_{22\perp})^{-1}\Pi_{22\perp}'X_{22.1}(P_2 - \Pi_{22}B)\Omega^{-1}B_{\perp}'(B_{\perp}\Omega^{-1}B_{\perp}')^{-1}$, $P = (P_1' \; P_2')'$, $P_1 : k_1 \times m$, $P_2 : k_2 \times m$, and $p_{lim}(\Omega|Y, X)$, $p_{lim}(\Phi|\Omega, Y, X)$, both result from (16). This also shows the conjugateness of this prior as it equals the posterior using a diffuse prior of some arbitrary set of observations which doesnot hold for the extended natural conjugate priors, which are also specifieds for SEMs, used by Drèze and Morales (1976) and Drèze and Richard (1983). We note that the simulation algorithms do not calculate $\gamma$, as $\gamma = \pi_{11} - \Pi_{12}\beta$, we can easily incorporate $\gamma$ into these algorithms.

# 6 Full System Analysis

The INSEM is a reduced rank restriction on a parameter matrix of a linear model. A full system analysis of a SEM can also be specified as a linear model with nonlinear restrictions on its parameters. Again these restrictions are reduced rank restrictions but the difference with the INSEM is that they can depend on one another in a recursive way. Theorem 1 states that the reduced form of a SEM is a linear model with reduced rank restrictions on its parameter matrices.

**Theorem 1** *Assume that a SEM has the following specification,*

$$\begin{pmatrix} Y_{\bar{m}} & Y_m \end{pmatrix} \begin{pmatrix} B_{\bar{m}\bar{m}} & B_{\bar{m}m} \\ B_{m\bar{m}} & B_{mm} \end{pmatrix} \qquad (34)$$
$$= \begin{pmatrix} X_{\bar{m}} & X_{lm} & X_m \end{pmatrix} \begin{pmatrix} \Gamma_{\bar{m}\bar{m}} & 0 \\ \Gamma_{m\bar{m}} & \Gamma_{lm} \\ 0 & \Gamma_{mm} \end{pmatrix} + \begin{pmatrix} \varepsilon_{\bar{m}} & \varepsilon_m \end{pmatrix}$$

*where the number of variables contained in $Y_m$ is chosen such that $\Gamma_{mm} : l_m \times j_m$ $(l_m \geq j_m)$ and $\Gamma_{lm} : l_m \times j_m$ are unrestricted, the parameter matrices, $\Gamma_{\bar{m}\bar{m}} : l_{\bar{m}} \times j_{\bar{m}}$, $\Gamma_{m\bar{m}} : l_m \times j_{\bar{m}}$, $B_{\bar{m}\bar{m}} : j_{\bar{m}} \times j_{\bar{m}}$, $B_{\bar{m}m} : j_{\bar{m}} \times j_m$, $B_{m\bar{m}} : j_m \times j_{\bar{m}}$, $B_{\bar{m}\bar{m}} : j_{\bar{m}} \times j_{\bar{m}}$, contain (some) parameters which are restricted to zero except for $B_{\bar{m}\bar{m}}$, which has all diagonal elements equal to one and some offdiagonal elements* *equal to zero, and* $B_{nnn} = I_{m}$; *then the reduced form of the SEM from equation (??) is equal to a set of reduced rank restrictions on the standard linear model,*

$$\left( \Pi_{21} \quad \Pi_{22} \right) = \left( \Phi_{21} \quad \Phi_{222} \quad \Phi_{22} \right) \Phi = 0,$$

*where* $\Phi : (k_2 - k_{22} - m_{22}) \times (m_1 - m_{21})$.

**Proof:** see appendix B.

Theorem 1 shows that we can use the framework for prior/posterior analysis used in the previous sections, which results from Kleibergen (1997), in a Bayesian full system analysis of a SEM. An important difference with the analysis from the previous sections is, however, the dependence of the different reduced rank restrictions on one another. For the INSEM, we can either analyze $\Phi$ conditional on $(\pi_{11}, \Pi_{12})$ or vice versa. So, the conditionalization of these parameters on one another does not matter. This does not hold for the full system analysis which we can conclude from the proof of theorem 1. It results in a strict ordering in which the reduced rank restrictions have to be imposed and hence how the parameters have to be analyzed conditional on one another. The reduced form of the SEM constructed in appendix B shows already some important conditionalization rules for the parameters of the SEM. For example, the structural form parameter $\beta_{nn1}$ is analyzed conditional on the structural form parameter $\beta_{nnn}$, which are both defined in appendix B. More of these conditionalization rules will appear when the reduced form is constructed further.

The conditionalization rules also imply rank and order conditions which can differ from the INSEM based conditions used in general. This is part of the point made in Hausman (1976). Regarding the conditionalization rules, the reduced form, constructed in appendix B, shows that $\beta_{nnn}$ is identified when $\Pi_{nn1}$ has full rank (or when that part of $\Pi_{nn1}$ which is multiplied by the nonzero parts of $\beta_{nnn}$ has full rank), where the elements of $\Pi$ are defined in appendix B. When the INSEM based conditions are used, it is assumed that no restrictions are imposed on $\Pi_{nn1}$. If restrictions are imposed, however, the resulting rank and order conditions can become different. In the following, an example of this will be discussed. It can also be seen in $\beta_{nn1}$, which is identified jointly by $\Pi_{nnn}$, $\Pi_{nn1}\beta_{nnn}$ and $\Pi_{nnn}$ and its rank and order conditions therefore depend on the specification of the SEM.

As mentioned before, the framework for prior and posterior analysis used in the previous sections can also be used to construct the priors and posteriors of the parameters in a full system analysis of a SEM. When we apply this framework we have to give an exact specification of the reduced form and its (hyper) parameters reflecting the restrictions which obey the three conditions, that (*i.*) when these (hyper) parameters are nonzero, the model is observationally equivalent with a standard linear model and when these (hyper) parameters are zero, (*ii.*) both the reduced form of the SEM results and (*iii.*) these (hyper) parameters are locally uncorrelated with specific other parameters such that the resulting posterior is invariant with respect to the ordering of those variables for which also the likelihood is invariant, see Kleibergen (1997) for an exact specification of the conditions the restrictions have to satisfy. This enables us to construct the prior/posterior of the parameters of the SEM as proportional to the prior/posterior of the parameters of the linear model under the restriction that the (hyper) parameters reflecting the restrictions are zero which is identical to the construction of priors/posteriors of the parameters of the INSEM. Since there are still some differences compared to the analysis of the INSEM, because the reduced form has a more complicated structure and the number of additional parameters we have to simulate in the posterior simulator increases, see (25), we give two detailed examples, a two and three (sets of) equation(s) model, to indicate all these differences. These examples jointly with theorem 1 show how a Bayesian full system analysis of any kind of SEM is conducted.

## 6.1 Two (sets of) equations

We specify the structural form of the two (sets of) equation(s) model by,

$$\begin{aligned} Y_1 &= Y_2 B_1 + X_1 \Gamma_{11} + X_2 \Gamma_{21} + \varepsilon_1, \\ Y_2 &= Y_1 B_2 + X_1 \Gamma_{12} + X_3 \Gamma_{32} + \varepsilon_2, \end{aligned} \tag{55}$$

where $Y_1 : T \times m_1$, $Y_2 : T \times m_2$; contain the endogenous variables, $X_1 : T \times k_1$, $X_2 : T \times k_2$, $X_3 : T \times k_3$; contain (weakly) exogenous and lagged dependent variables; $k_2 \geq m_1$, $k_3 \geq m_2$, $m = m_1 + m_2$, $(\varepsilon_1\ \varepsilon_2) \sim n(0, \Sigma \otimes I_T)$, $B_1 : m_2 \times m_1$, $B_2 : m_1 \times m_2$, $\Gamma_{11} : k_1 \times m_1$, $\Gamma_{12} : k_1 \times m_2$, $\Gamma_{21} : k_2 \times m_1$, $\Gamma_{32} : k_3 \times m_2$. The reduced form of (55), which can be constructed using the proof of theorem 1, reads

$$\begin{aligned} Y_1 &= X_1 \Pi_{11} + X_2 \Pi_{21} + X_3 \Pi_{32} B_1 + \xi_1, \\ Y_2 &= X_1 \Pi_{12} + X_2 \Pi_{21} B_2 + X_3 \Pi_{32} + \xi_2, \end{aligned} \tag{56}$$

where $\Pi_{11} = (\Gamma_{11} + \Gamma_{12} B_1)(I_{m_1} - B_2 B_1)^{-1}$, $\Pi_{21} = \Gamma_{21}(I_{m_1} - B_2 B_1)^{-1}$, $\Pi_{12} = (\Gamma_{12} + \Gamma_{11} B_2)(I_{m_2} - B_1 B_2)^{-1}$, $\Pi_{32} = \Gamma_{32}(I_{m_2} - B_1 B_2)^{-1}$, $\xi_1 = (\varepsilon_1 + \varepsilon_2 B_1)(I_{m_1} - B_2 B_1)^{-1}$, $\xi_2 = (\varepsilon_2 + \varepsilon_1 B_2)(I_{m_2} - B_1 B_2)^{-1}$, $(\xi_1\ \xi_2) \sim n(0, \Omega \otimes I_T)$, $\Sigma = B' \Omega B$, $B = \begin{pmatrix} I_{m_1} & -B_2 \\ -B_1 & I_{m_2} \end{pmatrix}$.

Similar to the reduced form of the INSEM (2) and as indicated in the proof of theorem 1, we add parameters to the reduced form to obtain a model, which we call unrestricted SEM (UNSEM), which is observationally equivalent with a linear model and when these added parameters are zero both the reduced form (56) results and the added parameters are locally uncorrelated with specific other parameters,

$$\begin{pmatrix} Y_1 & Y_2 \end{pmatrix} = X_1 \begin{pmatrix} \Pi_{11} & \Pi_{12} \end{pmatrix} + X_2 \Pi_{21} \begin{pmatrix} I_{m_1} & B_2 \end{pmatrix} + X_3 \Pi_{32} \begin{pmatrix} B_1 & I_{m_2} \end{pmatrix} \tag{57}$$

$$= X_2 \Pi_{21\perp} \lambda_2 \begin{pmatrix} I_{m_1} & \beta_2 \end{pmatrix}_{\perp} = X_3 \Pi_{32\perp} \lambda_1 \begin{pmatrix} \beta_1 & I_{m_2} \end{pmatrix}_{\perp} = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 \end{pmatrix},$$

where $\lambda_2 : (k_2 - m_1) \times m_2$, $\lambda_3 : (k_3 - m_2) \times m_1$, and the orthogonal complements $\Pi_{21\perp}$, $\Pi_{32\perp}$, $\begin{pmatrix} I_{m_1} & \beta_2 \end{pmatrix}_{\perp}$ and $\begin{pmatrix} \beta_1 & I_{m_2} \end{pmatrix}_{\perp}$ are defined similar to the ones used in (6), see appendix C. It is clear that when $\lambda_2 = 0$, $\lambda_3 = 0$, the reduced form (26) results and that $\lambda_2$ and $\lambda_3$ are locally uncorrelated, when they are equal to zero, with $(\Pi_{21}, \beta_2)$ and $(\Pi_{32}, \beta_1)$ respectively. When $\lambda_2 \neq 0$, $\lambda_3 \neq 0$, again similar to (6), (27) is observationally equivalent with the linear model,

$$\begin{pmatrix} Y_1 & Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix} \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 & \varepsilon_2 \end{pmatrix}, \tag{28}$$

where $\Phi_1 = \begin{pmatrix} \Pi_{11} & \Pi_{12} \end{pmatrix}$, $\Phi_2 : k_2 \times m$, $\Phi_3 : k_3 \times m$. Using a SVD, the equality of (27) and (28) can be shown. SVDs are also used to obtain $(\beta_2, \lambda_2, \Pi_{21})$ from $\Phi_2$ and $(\beta_1, \lambda_3, \Pi_{32})$ from $\Phi_3$, see appendix C. The resulting relationships are similar to (8)-(10) and straightforward to derive given (8)-(10). The SEM (25) is consequently a linear model with nonlinear restrictions on its parameters, $\lambda_2 = 0$, $\lambda_3 = 0$. The framework for prior/posterior analysis of the INSEM used in sections 2-4 can, therefore, directly be extended to the two equation SEM (25). So, we specify a prior for the parameters of the linear model $(\Phi_1, \Phi_2, \Phi_3, \Omega)$, for example a diffuse or natural conjugate prior, and this implies a prior for the parameters of the SEM (26) as this SEM equals the linear model evaluated in $\lambda_2 = 0$, $\lambda_3 = 0$ (Note that we use the reduced form (26) but this model is observationally equivalent with the SEM (25)),

$$\begin{aligned} & p_{sem}(\Pi_{11}, \Pi_{12}, \beta_1, \beta_2, \Pi_{21}, \Pi_{32}, \Omega) \qquad (29) \\ \propto\ & p_{insem}(\Pi_{11}, \Pi_{12}, \beta_1, \beta_2, \lambda_2, \lambda_3, \Pi_{21}, \Pi_{32}, \Omega)|_{\lambda_3=0, \lambda_2=0} \\ \propto\ & p_{lin}(\Phi_1, \Phi_2(\beta_2, \lambda_2, \Pi_{21}), \Phi_3(\beta_1, \lambda_3, \Pi_{32}), \Omega)|_{\lambda_3=0, \lambda_2=0} \\ & |J(\Phi_2, (\beta_2, \lambda_2, \Pi_{21}))|_{\lambda_3=0}||J(\Phi_3, (\beta_1, \lambda_3, \Pi_{32}))|_{\lambda_2=0}| \end{aligned}$$

where $sem$ stands for SEM, $insem$ for INSEM, and $lin$ for linear model and the jacobians $J(\Phi_2, (\beta_2, \lambda_2, \Pi_{21}))$, $J(\Phi_3, (\beta_1, \lambda_3, \Pi_{32}))$ are straightforward to derive given the derivation of the jacobian of the transformation in case of the INSEM and are stated in appendix C. Using (29) and the expressions of diffuse and natural conjugate priors for the linear model, (14) and (16), we can again construct the functional expressions of diffuse and natural conjugate priors for SEMs like (26). For reasons of compactness and similarity with section 2 we do not give the exact functional expressions.

For the posterior exactly the same reasoning as for the prior applies, *i.e.* the posterior of the parameters of the SEM (26) is proportional to the posterior of the parameters of the linear model under the imposed restriction. We can decompose the posterior of the linear model into a product of marginal and conditional posteriors which belong to a standard class of density functions, i.e. normal or inverted-Wishart, see e.g. Zellner (1971). This property can directly be used to decompose the posterior of the SEM,

$$
\begin{aligned}
& p_{sem}(\Gamma_{11}, \Gamma_{12}, \Pi_1, \Pi_2, \Gamma_{21}, \Gamma_{32}, \Omega|Y, X) \qquad (40)\\
\propto\ & p_{nasem}(\Gamma_{11}, \Gamma_{12}, \Pi_1, \Pi_2, \lambda_2, \lambda_3, \Gamma_{21}, \Gamma_{32}, \Omega|Y, X)|_{\lambda_2=0,\lambda_3=0}\\
\propto\ & p_{lin}(\Phi_1, \Phi_2(\Pi_2, \lambda_2, \Gamma_{21}), \Phi_3(\Pi_1, \lambda_3, \Gamma_{32}), \Omega|Y, X)|_{\lambda_2=0,\lambda_3=0}\\
& |J(\Phi_2, (\Pi_2, \lambda_2, \Gamma_{21}))|_{\lambda_2=0}||J(\Phi_3, (\Pi_1, \lambda_3, \Gamma_{32}))|_{\lambda_3=0}|\\
\propto\ & p_{lin}(\Phi_1|\Phi_2(\Pi_2, \lambda_2, \Gamma_{21}), \Phi_3(\Pi_1, \lambda_3, \Gamma_{32}), \Omega, Y, X)|_{\lambda_2=0,\lambda_3=0}\\
& p_{lin}(\Phi_2(\Pi_2, \lambda_2, \Gamma_{21})|\Phi_3(\Pi_1, \lambda_3, \Gamma_{32}), \Omega, Y, X)|_{\lambda_2=0,\lambda_3=0}\\
& |J(\Phi_2, (\Pi_2, \lambda_2, \Gamma_{21}))|_{\lambda_2=0}|\\
& p_{lin}(\Phi_3(\Pi_1, \lambda_3, \Gamma_{32})|\Omega, Y, X)|_{\lambda_3=0}|J(\Phi_3, (\Pi_1, \lambda_3, \Gamma_{32}))|_{\lambda_3=0}|\\
& p_{lin}(\Omega|Y, X).
\end{aligned}
$$

Note that we can also use other orderings in this decomposition. To simulate parameters from the posterior of the SEM (36), we use the decomposition of the posterior of the SEM (40). This allows us to perform the simulation in two different steps. Furthermore, we add, in each of the two different steps, parameters to the model which we, similar to section 3, assume to be generated from some conditional density $g$, which we specify ourselves. In case of diffuse priors, the following choices of these functions are natural,

$$
\begin{aligned}
& g_1(\lambda_3|\Pi_1, \Gamma_{32}, \Omega) = (2\pi)^{-\frac{1}{2}l_3}|B_{1\perp}\Omega^{-1}B_{1\perp}'|^{\frac{1}{2}l_3}|\Gamma_{32\perp}' X_3' M_{(X_1\ X_2)} X_3 \Gamma_{32\perp}|^{\frac{1}{2}m_1}\\
& \exp[-\tfrac{1}{2}tr(B_{1\perp}\Omega^{-1}B_{1\perp}'(\lambda_3 - \hat{\lambda}_3)'\Gamma_{32\perp}' X_3' M_{(X_1\ X_2)} X_3 \Gamma_{32\perp}(\lambda_3 - \hat{\lambda}_3))], \qquad (41)
\end{aligned}
$$

$$
\begin{aligned}
& g_2(\lambda_2|\Pi_2, \Gamma_{21}, \Phi_3, \Omega) = (2\pi)^{-\frac{1}{2}l_2}|B_{2\perp}\Omega^{-1}B_{2\perp}'|^{\frac{1}{2}l_2}|\Gamma_{21\perp}' X_2' M_{X_1} X_2 \Gamma_{21\perp}|^{\frac{1}{2}m_2}\\
& \exp[-\tfrac{1}{2}tr(B_{2\perp}\Omega^{-1}B_{2\perp}'(\lambda_2 - \hat{\lambda}_2)'\Gamma_{32\perp}' X_2' M_{X_1} X_3 \Gamma_{32\perp}(\lambda_2 - \hat{\lambda}_2))], \qquad (42)
\end{aligned}
$$

where $l_2 = k_2 - m_1$, $l_3 = k_3 - m_2$, $B_1 = \begin{pmatrix} \Pi_1 & I_{m_2} \end{pmatrix}$, $B_2 = \begin{pmatrix} I_{m_1} & \Pi_2 \end{pmatrix}$,

$$
\begin{aligned}
\hat{\lambda}_3 &= (\Gamma_{32\perp}' X_3' M_{(X_1\ X_2)} X_3 \Gamma_{32\perp})^{-1}\Gamma_{32\perp}' X_3' M_{(X_1\ X_2)}\\
&\quad (Y - X_3\Gamma_{32}B_1)\Omega^{-1}B_{1\perp}'(B_{1\perp}\Omega^{-1}B_{1\perp}')^{-1},\\
\hat{\lambda}_2 &= (\Gamma_{21\perp}' X_2' M_{X_1} X_2 \Gamma_{21\perp})^{-1}\Gamma_{21\perp}' X_2' M_{X_1}\\
&\quad (Y - X_3\Phi_3 - X_2\Gamma_{21}B_2)\Omega^{-1}B_{2\perp}'(B_{2\perp}\Omega^{-1}B_{2\perp}')^{-1}.
\end{aligned}
$$

The weight functions of the two different steps of the simulation algorithm, involving both (41) and (42), then become,

$$
\begin{aligned}
w_1(\Pi_1, \lambda_3, \Gamma_{32}, \Omega) &= \frac{|J(\Phi_3, (\Pi_1, \lambda_3, \Gamma_{32}))|_{\lambda_3=0}|}{|J(\Phi_3, (\Pi_1, \lambda_3, \Gamma_{32})|}g_1(\lambda_3|\Pi_1, \Gamma_{32}, \Omega)|_{\lambda_3=0}, \qquad (43)\\
w_2(\Pi_2, \lambda_2, \Gamma_{21}, \Omega|\Phi_3) &= \frac{|J(\Phi_2, (\Pi_2, \lambda_2, \Gamma_{21}))|_{\lambda_2=0}|}{|J(\Phi_2, (\Pi_2, \lambda_2, \Gamma_{21})|}g_2(\lambda_2|\Pi_2, \Gamma_{21}, \Phi_3, \Omega)|_{\lambda_2=0} \qquad (44)
\end{aligned}
$$

Again these weights functions are always finite. The different steps involved in obtaining the weight attached to a certain drawing $i$, $i = 1, ..., N$, of the parameters of the SEM, can then be summarized as follows,

1. Draw $\Omega^i$ from $g_{Im}(\Omega|Y, X)$
   Draw $\Phi_3^i$ from $g_{Im}(\Phi_3|\Omega^i, Y, X)$.
2. Compute $\Pi_1^i$, $\lambda_3^i$, $\beta_{32}^i$ from $\Phi_3^i$ using a SVD
3. Compute $w_1(\Pi_1^i, \lambda_3^i, \beta_{32}^i, \Omega^i)$ according to (43)
4. Draw $\Phi_2^i$ from $g_{Im}(\Phi_2|\Phi_3(\Pi_1^i, \lambda_3, \beta_{32}^i), \Omega^i, Y, X)|_{\lambda_2=0}$
5. Compute $\Pi_2^i$, $\lambda_2^i$, $\beta_{21}^i$ from $\Phi_2^i$ using a SVD
6. Compute $w_2(\Pi_2^i, \lambda_2^i, \beta_{21}^i, \Omega^i|\Phi_3(\Pi_1^i, \lambda_3, \beta_{32}^i))|_{\lambda_2=0}$ according to (44)
7. Compute total weight $i-$th drawing :
   $w(\Pi_1^i, \lambda_3^i, \beta_{32}^i, \Pi_2^i, \lambda_2^i, \beta_{21}^i, \Omega^i) = w_1 \times w_2$
8. Draw $\Phi_1^i$ from $g_{Im}(\Phi_1|\Phi_2(\Pi_2^i, \lambda_2, \beta_{21}^i), \Phi_3(\Pi_1^i, \lambda_3, \beta_{32}^i), \Omega^i, Y, X)|_{\lambda_3=0,\lambda_2=0}$

The posteriors from which we simulate are all standard, in case of diffuse or natural conjugate priors, and are similar to the ones used in the algorithm in section 4. The values of other structural form parameters can directly be calculated using the equations used in (36) and the drawings from the above algorithm. The resulting total weights, $w$, can be used in an Importance or M-H sampler as discussed in section 4 to obtain a posterior simulator of the posterior of the parameters of the SEM (36).

## 6.2 Three (sets of) Equations

As an example of a three (sets of) equation(s) model, we use (Note that contrary to the two equation model, the specification of a three equation model is not unique since the model is not invariant with respect to the ordering of the variables),

$$
\begin{aligned}
Y_1 &= Y_2\beta_{21} + X_1\Gamma_{11} + \varepsilon_1, \\
Y_2 &= Y_3\beta_{32} + X_1\Gamma_{12} + X_2\Gamma_{22} + \varepsilon_2, \\
Y_3 &= Y_1\beta_{13} + Y_2\beta_{23} + X_2\Gamma_{23} + X_3\Gamma_{33} + \varepsilon_3,
\end{aligned}
\tag{45}
$$

where $Y_1 : T \times m_1$, $Y_2 : T \times m_2$, and $Y_3 : T \times m_3$, contain the endogenous variables and $X_1 : T \times k_1$, $X_2 : T \times k_2$, and $X_3 : T \times k_3$, contain lagged endogenous and weakly exogenous variables, $\beta_{21} : m_2 \times m_1$, $\beta_{32} : m_3 \times m_2$, $\beta_{13} : m_1 \times m_3$, $\beta_{23} : m_2 \times m_3$, $\Gamma_{11} : k_1 \times m_1$, $\Gamma_{12} : k_1 \times m_2$, $\Gamma_{22} : k_2 \times m_2$, $\Gamma_{23} : k_2 \times m_3$,

$\Pi_{33} : k_3 \times m_3$, $m = m_1 + m_2 + m_3$. $(\varepsilon_1\ \varepsilon_2\ \varepsilon_3) \sim n(0, \Sigma \otimes I_T)$. Since the SEM (45) has to be properly identified, the following (INSEM) order conditions need to be fulfilled, $k_2 + k_3 \geq m_2$, $k_3 \geq m_3$, $k_1 \geq m_1 + m_2$. Using the proof of theorem 1, the reduced form of the model in equation (45) is constructed and reads,

$$\begin{aligned} y_1 &= X_1\Pi_{11} + \begin{pmatrix} X_1 & X_3\Pi_{33} \end{pmatrix} \begin{pmatrix} \Pi_{22} \\ \beta_{32} \end{pmatrix} \beta_{21} + \xi_1, \\ y_2 &= X_1\Pi_{12} + X_2\Pi_{22} + X_3\Pi_{33}\beta_{32} + \xi_2, \\ y_3 &= X_1(\Pi_{11}\ \Pi_{12}) \begin{pmatrix} \beta_{13} \\ \beta_{23} \end{pmatrix} + X_2\Pi_{23} + X_3\Pi_{33} + \xi_3, \end{aligned} \tag{46}$$

where $(\Lambda_{11}\ \Lambda_{12}) = (\Pi_{11}\ \Pi_{12}) \begin{pmatrix} I_{m_1} & -\beta_{13}\beta_{23} \\ -\beta_{21} & I_{m_2} - \beta_{23}\beta_{13} \end{pmatrix}$, $\Lambda_{33} = \Pi_{33}(I_{m_3} - \beta_{32}(\beta_{21}\beta_{13} - \beta_{23}))$, $(\Lambda_{22}\ \Lambda_{23}) = (\Pi_{22}\ \Pi_{23}) \begin{pmatrix} I_{m_2} & -(\beta_{23} + \beta_{21}\beta_{13}) \\ -\beta_{32} & I_{m_3} \end{pmatrix}$, $(\xi_1\ \xi_2\ \xi_3)B = (\varepsilon_1\ \varepsilon_2\ \varepsilon_3)$, $\Sigma = B'\Omega B$, $B = \begin{pmatrix} I_{m_1} & 0 & -\beta_{13} \\ -\beta_{21} & I_{m_2} & -\beta_{23} \\ 0 & -\beta_{32} & I_{m_3} \end{pmatrix}$. The reduced form (46) is again a system of reduced rank matrices like the reduced forms of the one equation (8) and two equation (36) models. An important difference with these models is that its reduced rank matrices depend on another which is a.o. reflected in the identification of $\beta_{21}$ which depends on one of the other structural form parameters, $\beta_{32}$. This difference also leads to a change in the order condition compared to the INSEM. According to the INSEM order condition, $\beta_{21}$ is identified when $k_2 + k_3 \geq m_2$, *i.e.* the number of excluded exogenous variables is at least equal to the number of included endogenous variables, see Hausman (1983). The model (45) shows, however, that $\beta_{21}$ is identified when $\begin{pmatrix} \Pi_{22} \\ \Pi_{33}\beta_{32} \end{pmatrix}$ has full rank. Although this matrix has $k_2 + k_3$ rows, which accords with the standard order condition, it's row rank can never exceed $k_2 + m_3$ $(\leq k_2 + k_3)$ as it can be specified as $\begin{pmatrix} I_{k_2} & 0 \\ 0 & \Pi_{33} \end{pmatrix} \begin{pmatrix} \Pi_{22} \\ \beta_{32} \end{pmatrix}$ and the last matrix in this product has $k_2 + m_3$ rows. It is therefore important that the identification of the different parameters of a SEM in a full system analysis is conducted using the restricted reduced form parameter matrix instead of the unrestricted one as this can lead to different rank and order conditions, see also Maddala (1976). This different order condition results from the dependence of the, by the SEM (46) imposed, reduced rank structures on one another, see also proof of theorem 1. The reduced rank structures appearing in the two equation model do not depend on one another, as can be concluded from (37), and therefore the INSEM order conditions still apply there.

As a consequence of the sequential dependence between the reduced rank structures, not only the order conditions of the INSEM and the SEM (45) differ, as indicated above, but also the parameters which we add to the model (46), to make it observationally equivalent to a linear model, are different from the ones we used before, see also the proof of theorem 1. In the cases of the INSEM (6) and the two equation SEM (57), the parameters added to the reduced form, to make it observationally equivalent to a linear model, do not depend on one another in a sequential way. The parameters added to (46) do, however, depend on each other sequentially. To show this, consider the linear model,

$$\begin{pmatrix} Y_1 & Y_2 & Y_3 \end{pmatrix} = \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix} \begin{pmatrix} \Pi_1 \\ \Pi_2 \\ \Pi_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 \end{pmatrix}. \tag{47}$$

The reduced form model (46) can be obtained by using a, what we call, unrestricted SEM specification of the parameters of (47),

$$\begin{aligned} \Pi_1 &= \begin{pmatrix} \beta_{11} & \beta_{12} \end{pmatrix} \begin{pmatrix} I_{m_1} & 0 & \gamma_{13} \\ 0 & I_{m_2} & \gamma_{23} \end{pmatrix} \\ &\quad + \begin{pmatrix} \beta_{11} & \beta_{12} \end{pmatrix}_{\perp} \lambda_1 \begin{pmatrix} I_{m_1} & 0 & \gamma_{13} \\ 0 & I_{m_2} & \gamma_{23} \end{pmatrix}_{\perp}, \end{aligned} \tag{48}$$

$$\begin{pmatrix} \Pi_2 \\ \Pi_3 \end{pmatrix} = B \begin{pmatrix} \gamma_{21} & I_{m_2} & 0 \\ 0 & 0 & I_{m_3} \end{pmatrix} + B_{\perp} \lambda_2 \begin{pmatrix} \gamma_{21} & I_{m_2} & 0 \\ 0 & 0 & I_{m_3} \end{pmatrix}_{\perp}, \tag{49}$$

$$B = \begin{pmatrix} \beta_{22} & \beta_{23} \\ & B_3 \end{pmatrix}, \tag{50}$$

$$B_3 = \beta_{33} \begin{pmatrix} \gamma_{32} & I_{m_3} \end{pmatrix} + \beta_{33,\perp} \lambda_3 \begin{pmatrix} \gamma_{32} & I_{m_3} \end{pmatrix}_{\perp} \tag{51}$$

where the orthogonal complements are defined similar to the ones used in (6), see also appendix B, $\lambda_1 : (k_1 - m_1 - m_2) \times m_3$, $\lambda_2 : (k_2 + k_3 - m_2 - m_3) \times m_1$, $\lambda_3 : (k_3 - m_3) \times m_2$. To analyze the implications of the different orthogonality conditions in (48)-(51), we substitute the expression of $B$ in $(\Pi_2' \; \Pi_3')'$,

$$\begin{aligned} \begin{pmatrix} \Pi_2 \\ \Pi_3 \end{pmatrix} &= \begin{pmatrix} \beta_{22} & \beta_{23} \\ \beta_{33} \begin{pmatrix} \gamma_{32} & I_{m_3} \end{pmatrix} + \beta_{33,\perp} \lambda_3 \begin{pmatrix} \gamma_{32} & I_{m_3} \end{pmatrix}_{\perp} & \end{pmatrix} \begin{pmatrix} \gamma_{21} & I_{m_2} & 0 \\ 0 & 0 & I_{m_3} \end{pmatrix} \\ &\quad + \begin{pmatrix} \beta_{22} & \beta_{23} \\ \beta_{33} \begin{pmatrix} \gamma_{32} & I_{m_3} \end{pmatrix} + \beta_{33,\perp} \lambda_3 \begin{pmatrix} \gamma_{32} & I_{m_3} \end{pmatrix}_{\perp} & \end{pmatrix}_{\perp} \lambda_2 \begin{pmatrix} \gamma_{21} & I_{m_2} & 0 \\ 0 & 0 & I_{m_3} \end{pmatrix}_{\perp} \end{aligned} \tag{52}$$

It is clear from (48)-(51) that when $\lambda_1 = 0$, $\lambda_2 = 0$, $\lambda_3 = 0$, the model (46) results. Furthermore, when $\lambda_1 = 0$, $\lambda_2 = 0$, $\lambda_3 = 0$, $\lambda_1$ is locally uncorrelated with $(\beta_{11}, \beta_{12}, \gamma_{13}, \gamma_{23})$, $\lambda_3$ with $(\beta_{33}, \gamma_{32})$, and $\lambda_2$ with $\gamma_{21}$ and all parameters contained in $B$, *i.e.* $\beta_{22}, \beta_{23}, \beta_{33}, \lambda_3, \gamma_{32}$. SNCs are needed to obtain $(\beta_{11}, \beta_{12}, \lambda_1, \gamma_{13}, \gamma_{23})$ from $\Pi_1$, $(B, \lambda_2, \gamma_{21})$ from $(\Pi_2, \Pi_3)$ and $(\beta_{22}, \beta_{23}, \beta_{33}, \lambda_3, \gamma_{32})$ from $B$, and to show the observational equivalence between the model imposed by (48)-(51) and (47) when $\lambda_1 \neq 0$, $\lambda_2 \neq 0$, $\lambda_3 \neq 0$. These SNCs are stated in appendix B. The sequential dependence between the structural form parameters now reflects itself in SEMs which have to be applied recursively, e.g. of $\Theta$ which already results from a SEM as it is the reduced form of an INSEM,

$$\begin{aligned} \tilde{y}_1 &= \tilde{y}_2\beta_{32} + \tilde{x}_1\Gamma_{11} + \varepsilon_1, \\ \tilde{y}_2 &= \tilde{x}_1\Gamma_{21} + \tilde{x}_2\Gamma_{22} + \varepsilon_2, \end{aligned} \tag{53}$$

where $\tilde{y}_1$, $\tilde{y}_2$, $\tilde{x}_1$ and $\tilde{x}_2$ are data matrices, $\Gamma_{21} = \Phi_{23}$, $\Gamma_{22} = \Phi_{33}$, $\Gamma_{11} = \Phi_{22} - \Gamma_{21}\beta_{32}$. $\Theta$ is therefore similar to the $(\Phi'_{21}\ \Phi'_{22})'$ parameter matrix used in the proof of theorem 1.

So, the SEM (46) is again a linear model with restrictions on its parameters. We can, therefore, again apply the framework for prior/posterior analysis used in the previous sections, *i.e.* we specify the prior/posterior of the parameters of (46) as proportional to the prior/posterior of the parameters of the linear model under the condition that the restrictions hold,

$$\begin{aligned} & p_{sem}(\beta_{21},\beta_{32},\beta_{13},\beta_{23},\Phi_{11},\Phi_{12},\Phi_{22},\Phi_{23},\Phi_{33},\Omega) \\ \propto\ & p_{unsem}(\beta_{21},\beta_{32},\beta_{13},\beta_{23},\lambda_1,\lambda_2,\lambda_3,\Phi_{11},\Phi_{12},\Phi_{22},\Phi_{23},\Phi_{33},\Omega)|_{(\lambda_1,\lambda_2,\lambda_3)=0} \\ \propto\ & p_{lin}(\Pi_1(\beta_{13},\beta_{23},\lambda_1,\Phi_{11},\Phi_{12}), \\ & (\Pi_2,\Pi_3)(\beta_{21},\lambda_2,\Theta(\Phi_{22},\Phi_{23},\Phi_{33},\lambda_3,\beta_{32}),\Omega)|_{(\lambda_1,\lambda_2,\lambda_3)=0} \\ & |J(\Pi_1,(\beta_{13},\beta_{23},\lambda_1,\Phi_{11},\Phi_{12}))|_{\lambda_1=0}||J(\Theta,(\Phi_{22},\Phi_{23},\Phi_{33},\lambda_3,\beta_{32}))|_{\lambda_3=0}| \\ & |J((\Pi_2,\Pi_3),(\beta_{21},\lambda_2,\Theta(\Phi_{22},\Phi_{23},\Phi_{33},\lambda_3,\beta_{32})|_{\lambda_3=0}))|_{\lambda_2=0}|, \end{aligned} \tag{54}$$

where $J(\Pi_1,(\beta_{13},\beta_{23},\lambda_1,\Phi_{11},\Phi_{12}))$, $J(\Theta,(\Phi_{22},\Phi_{23},\Phi_{33},\lambda_3,\beta_{32}))$, $J((\Pi_2,\Pi_3),(\beta_{21},\lambda_2,\Theta))$ are the jacobians of the transformation from $\Pi_1$ to $(\beta_{13},\beta_{23},\lambda_1,\Phi_{11},\Phi_{12})$, $(\Pi_2,\Pi_3)$ to $(\beta_{21},\lambda_2,\Theta)$ and $\Theta$ to $(\Phi_{22},\Phi_{23},\Phi_{33},\lambda_3,\beta_{32})$ and these jacobians are stated in appendix B. When we specify a diffuse (14) or natural conjugate prior (16) for the parameters of the linear model, (54) shows the implied prior for the parameters of the SEM. We do not give the exact functional expressions as they can be constructed along the lines of section 3.

Also for the posterior, we use the framework from Kleibergen (1997). Furthermore, we use the decomposition of the posterior of the linear model into a product of conditional and marginal densities,

$$\begin{aligned} & p_{sem}(\beta_{21},\beta_{32},\beta_{13},\beta_{23},\Phi_{11},\Phi_{12},\Phi_{22},\Phi_{23},\Phi_{33},\Omega|Y,X) \\ \propto\ & p_{unsem}(\beta_{21},\beta_{32},\beta_{13},\beta_{23},\lambda_1,\lambda_2,\lambda_3,\Phi_{11},\Phi_{12},\Phi_{22},\Phi_{23},\Phi_{33},\Omega|Y,X)|_{(\lambda_1,\lambda_2,\lambda_3)=0} \\ \propto\ & p_{lin}(\Pi_1(\beta_{13},\beta_{23},\lambda_1,\Phi_{11},\Phi_{12}), \\ & (\Pi_2,\Pi_3)(\beta_{21},\lambda_2,\Theta(\Phi_{22},\Phi_{23},\Phi_{33},\lambda_3,\beta_{32}),\Omega|Y,X)|_{(\lambda_1,\lambda_2,\lambda_3)=0} \\ & |J(\Pi_1,(\beta_{13},\beta_{23},\lambda_1,\Phi_{11},\Phi_{12}))|_{\lambda_1=0}||J(\Theta,(\Phi_{22},\Phi_{23},\Phi_{33},\lambda_3,\beta_{32}))|_{\lambda_3=0}| \\ & |J((\Pi_2,\Pi_3),(\beta_{21},\lambda_2,\Theta(\Phi_{22},\Phi_{23},\Phi_{33},\lambda_3,\beta_{32})|_{\lambda_3=0}))|_{\lambda_2=0}|, \\ \propto\ & p_{lin}(\Pi_1(\beta_{13},\beta_{23},\lambda_1,\Phi_{11},\Phi_{12})|(\Pi_2,\Pi_3)(\beta_{21},\lambda_2,\Theta),\Omega|Y,X)|_{(\lambda_1,\lambda_2,\lambda_3)=0} \\ & |J(\Pi_1,(\beta_{13},\beta_{23},\lambda_1,\Phi_{11},\Phi_{12}))|_{\lambda_1=0}| \end{aligned} \tag{55}$$

[illegible equation]

We note that for this model only a few decompositions of the posterior into conditional and marginal posteriors are allowed for, i.e. [illegible] given [illegible] and vice versa, because of the reduced rank structure imposed by the [illegible]. We cannot for example analyze [illegible] given [illegible] or vice versa. We use the decomposition of the posterior (55) to construct a posterior simulator. Again, similar to previous sections, to simulate from the posterior of (46) we add parameters to the model, i.e. [illegible], which we assume to be drawn from a specific conditional density, which we specify ourselves, see (25). In case of a diffuse prior for the linear model (14), natural choices for these conditional densities are,

[illegible equation] (56)

[illegible equation] (57)

[illegible equation] (58)

where [illegible equation]

[illegible equation]

Since we simulate from a density which approximates the posterior of (46), weight functions are involved in the different stage of the posterior simulation. As we simulate three different parameters, i.e. [illegible], which are not present in the original posterior we want to simulate from, three weight functions are involved,

$$w_1(\Phi_{13}, \Phi_{23}, \lambda_1, \Sigma_{11}, \Sigma_{12}, \Omega|\theta_2, \theta_3) \quad (59)$$
$$= \frac{|J(\theta_1, (\Phi_{13}, \Phi_{23}, \lambda_1, \Sigma_{11}, \Sigma_{12}))|_{\lambda_1=0}|}{|J(\theta_1, (\Phi_{13}, \Phi_{23}, \lambda_1, \Sigma_{11}, \Sigma_{12}))|} g_1(\lambda_1|\Phi_{13}, \Phi_{23}, \Sigma_{11}, \Sigma_{12}, \theta_2, \theta_3, \Omega)|_{\lambda_1=0}$$

$$w_2(\Phi_{21}, \lambda_2, \Pi, \Omega) \quad (60)$$
$$= \frac{|J((\theta_2, \theta_3), (\Phi_{21}, \lambda_2, \Pi))|_{\lambda_3=0,\lambda_2=0}|}{|J((\theta_2, \theta_3), (\Phi_{21}, \lambda_2, \Pi))|_{\lambda_2\neq 0}|} g_2(\lambda_2|\Phi_{21}, \Pi, \Omega)|_{\lambda_3=0,\lambda_2=0}$$

$$w_3(\Phi_{32}, \Phi_{21}, \lambda_3, \Sigma_{33}, \Omega) \quad (61)$$
$$= \frac{|J(\Pi_2, (\Phi_{32}, \lambda_3, \Sigma_{33}))|_{\lambda_3=0}|}{|J(\Pi_2, (\Phi_{32}, \lambda_3, \Sigma_{33}))|} g_3(\lambda_3|\Phi_{32}, \Phi_{21}, \Sigma_{33}, \Omega)|_{\lambda_2=0},$$

where $J(\theta_1, (\Phi_{13}, \Phi_{23}, \lambda_1, \Sigma_{11}, \Sigma_{12}))$, $J((\theta_2, \theta_3), (\Phi_{21}, \lambda_2, \Pi))$ and $J(\Pi_2, (\Phi_{32}, \lambda_3, \Sigma_{33}))$ are the jacobians of the different parameter transformations and each of the weight functions is always finite, see appendix B.

The different steps involved in obtaining the weight attached to a certain drawing $i$, $i = 1, ..., N$, of the parameters of the SEM (46), can then be summarized as follows,

1. Draw $\Omega^i$ from $p_{Lim}(\Omega|Y, X)$
   Draw $(\theta_2^i, \theta_3^i)$ from $p_{Lim}(\theta_2, \theta_3|\Omega^i, Y, X)$
2. Compute $\Phi_{21}^i, \lambda_2^i, \Pi^i$ from $(\theta_2^i, \theta_3^i)$ using SVD
3. Compute $\Phi_{32}^i, \lambda_3^i, \Sigma_{33}^i$ from $\Pi_2^i$ using SVD
4. Compute $w_3(\Phi_{32}^i, \Phi_{21}^i, \lambda_3^i, \Sigma_{33}^i, \Omega^i)$
5. Compute $w_2(\Phi_{21}^i, \lambda_2^i, \Pi^i, \Omega^i)$
6. Draw $\theta_1^i$ from $p_{Lim}(\theta_1|\theta_2^i, \theta_3^i, \Omega^i, Y, X)|_{\lambda_3=0,\lambda_2=0}$
7. Compute $\Phi_{13}^i, \Phi_{23}^i, \lambda_1^i, \Sigma_{11}^i, \Sigma_{12}^i$ from $\theta_1^i$
8. Compute $w_1(\Phi_{13}^i, \Phi_{23}^i, \lambda_1^i, \Sigma_{11}^i, \Sigma_{12}^i, \Omega^i|\theta_2^i, \theta_3^i)|_{\lambda_3=0,\lambda_2=0}$
9. Compute total weight $i$-th drawing: $w = w_1 \times w_2 \times w_3$

The total weights can be used in an Importance or M-H sampler, as indicated in section 5, to obtain a posterior simulator of the posterior of the parameters of (46).

The means of the conditional posteriors of $\theta_1$ given $(\theta_2, \theta_3)$ and $(\theta_2, \theta_3)$ given $\theta_1$ can also be used in an iterative scheme to obtain the full information maximum likelihood estimator of $(\Phi_{21}, \Phi_{32}, \Phi_{13}, \Phi_{23}, \Sigma_{11}, \Sigma_{12}, \Sigma_{22}, \Sigma_{23}, \Sigma_{33})$, see Hausman (1983). This is similar to the INSEM where evaluating the posterior of $\pi$ at its posterior mean using a diffuse prior gives the analytical expression of the limited information maximum likelihood estimator of $\beta$ and $\Omega_{22}$ using the involved SNE. The iterative scheme for obtaining the full information maximum likelihood estimator proceeds as follows,

(0.) Initialize $\pi_1 = \hat{\pi}_1$,

(i.) Construct $(\beta_{21}, \beta_{32}, \Omega_{22}, \Omega_{23}, \Omega_{33})$ from $(\tilde{\pi}_2, \tilde{\pi}_3)$ using SNEs from stage 2 and 3 from the simulation scheme,

(ii.) Compute value of $(\pi_2, \pi_3)$ implied by $(\beta_{21}, \beta_{32}, \Omega_{22}, \Omega_{23}, \Omega_{33})$,

(iii.) Construct $(\beta_{13}, \beta_{23}, \Omega_{11}, \Omega_{12})$ from $\tilde{\pi}_1$ using SNE from step 7,

(iv.) Compute value of $\pi_1$ implied by $(\beta_{13}, \beta_{23}, \Omega_{11}, \Omega_{12})$,

(v.) Unless $(\pi_1, \pi_2, \pi_3)$ have converged goto (i.).

where $(\tilde{\pi}_2' \; \tilde{\pi}_3')' = ((X_2 \; X_3)'(X_2 \; X_3))^{-1}(X_2 \; X_3)'(y - X_1\pi_1)$, $\tilde{\pi}_1 = (X_1'X_1)^{-1}X_1'(y - (X_2 \; X_3)(\pi_2' \; \pi_3')')$. Using theorem 1, iterative schemes similar to the one above can be constructed to obtain the full information maximum likelihood estimators of the parameters of generally specified SEMs. Jointly with the examples of the two and three structural equations SEMs, theorem 1 shows how Bayesian analyses of generally specified SEMs are conducted.

# 7 Conclusions

The traditional Bayesian analyses of SEMs using diffuse priors, as proposed by e.g. Drèze (1976), Drèze and Morales (1976) and Drèze and Richard (1983), suffer from local nonidentification problems which lead to an a posteriori favor for certain parameter values while it is not the result of information in the prior or data. We therefore use a framework constructed in Kleibergen (1997) in which the priors/posteriors of the parameters of the SEM are proportional to the priors/posteriors of the parameters of a linear model under the condition that the restrictions, imposed by the SEM on the parameters of the linear model, hold. We applied this framework to examples of one, two and three structural equation SEMs, for which expressions of the priors and posteriors are derived jointly with posterior simulators. Using a theorem, which states that the reduced form of any kind of SEM accords with a linear model with reduced rank restrictions of its parameters, the analysis of the examples can be generalized to other specifications of SEMs in a straightforward way. This theorem also shows how full information maximum likelihood estimators can be constructed.

Using results from Kleibergen and Paap (1997), we can also construct tools for model comparison like Bayes Factors, Posterior Odds Ratios and Bayesian Lagrange Multiplier statistics. In future work we will construct and apply these procedures to analyze the support for (multiple structural equations) SEMs in practice. It is also interesting to analyze the theoretical properties of the derived posteriors, as for example in Chao and Phillips (1996) where functional expressions are constructed for the marginal posterior of the structural form parameters of the INSEM using a Jeffreys' prior, to investigate the similarities/differences between small sample distributions of classical statistical estimators and the marginal posteriors of the structural form parameters, see Kleibergen and Zivot (1998). Both limited information maximum likelihood (LIML) estimators, see Anderson and Rubin (1949), and the posteriors of the parameters of the INSEM are namely constructed using SVDs, see Kleibergen and Zivot (1998), which correspond with canonical correlations in case of the LIML estimator. So, it is interesting to investigate to what extent these similarities hold further.

# Appendix

## A. Jacobian of transformation from linear model to INSEM

For the derivation of the Jacobian of the transformation from the linear model parameters to the parameters of the INSEM, it is notationally convenient to conduct this transformation in two steps, $(i.)$ from $\Pi$ to $(\Sigma_{221}, A_2, \beta, \lambda)$ where $A_2 = \Sigma_{222}\Sigma_{221}^{-1}$, and $(ii.)$ from $(\Sigma_{221}, A_2, \beta, \lambda)$ to $(\Sigma_{221}, \Sigma_{222}, \beta, \lambda)$. In the following we construct the jacobians of the two transformations.

We can denote $\Pi$ as,

$$
\begin{aligned}
\Pi &= \begin{pmatrix} A & A_\perp \end{pmatrix} \begin{pmatrix} \Sigma_{221} & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} B \\ B_\perp \end{pmatrix} \\
&= A\Sigma_{221}B + A_\perp \lambda B_\perp,
\end{aligned}
$$

where $A = (I_{m-1}\ A_2')'$, $A_\perp = (-A_2\ I_{k_2-m+1})'(I_{k_2-m+1} + A_2A_2')^{-\frac{1}{2}}$, $B = (\beta\ I_{m-1})$, $B_\perp = (1 + \beta'\beta)^{-\frac{1}{2}}(1\ -\beta')$. The jacobians of $\Pi$ with respect to $\Sigma_{221}$, $A_2$, $\beta$ and $\lambda$ then read,

$$
\begin{aligned}
J_1 &= \frac{\partial vec(\Pi)}{\partial vec(\Sigma_{221})'} = (B' \otimes A) \\
J_2 &= \frac{\partial vec(\Pi)}{\partial vec(A_2)'} = (B'\Sigma_{221}' \otimes I_{k_2})\frac{\partial vec(A)}{\partial vec(A_2)'} + (B_\perp'\lambda' \otimes I_{k_2})\frac{\partial vec(A_\perp)}{\partial vec(A_2)'} \\
J_3 &= \frac{\partial vec(\Pi)}{\partial vec(\beta)'} = (I_m \otimes A\Sigma_{221})\ \frac{\partial vec(B)}{\partial vec(\beta)'} + (I_m \otimes A_\perp \lambda)\ \frac{\partial vec(B_\perp)}{\partial vec(\beta)'} \\
J_4 &= \frac{\partial vec(\Pi)}{\partial vec(\lambda)'} = (B_\perp' \otimes A_\perp)
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial vec(A)}{\partial vec(A_2)'} &= \left(I_{m-1} \otimes \begin{pmatrix} 0 \\ I_{k_2-m+1} \end{pmatrix}\right), \\
\frac{\partial vec(A_\perp)}{\partial vec(A_2)'} &= -(W^{-\frac{1}{2}\prime} \otimes \begin{pmatrix} I_{m-1} \\ 0 \end{pmatrix})K_{k_2-m+1,m-1} + \\
&\quad (I_{k_2-m+1} \otimes \begin{pmatrix} -A_2' \\ I_{k_2-m+1} \end{pmatrix})\frac{\partial vec((W^{\frac{1}{2}})^{-1})}{\partial vec(W^{\frac{1}{2}})'}\frac{\partial vec(W^{\frac{1}{2}})}{\partial vec(W)'}\frac{\partial vec(W)}{\partial vec(A_2)'}, \\
\frac{\partial vec((W^{\frac{1}{2}})^{-1})}{\partial vec(W^{\frac{1}{2}})'} &= -(W^{-\frac{1}{2}\prime} \otimes W^{-\frac{1}{2}}), \\
\frac{\partial vec(W^{\frac{1}{2}})}{\partial vec(W)'} &= ((I_{k_2-m+1} \otimes W^{\frac{1}{2}}) + (W^{\frac{1}{2}\prime} \otimes I_{k_2-m+1}))^{-1}, \\
\frac{\partial vec(W)}{\partial vec(A_2)'} &= (A_2 \otimes I_{k_2-m+1}) + (I_{k_2-m+1} \otimes A_2)K_{k_2-m+1,m-1},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial vec(B)}{\partial vec(\lambda)'} &= (e_1 \otimes I_{m-1}), \\
\frac{\partial vec(B_2)}{\partial vec(\lambda)'} &= -(\begin{pmatrix} 0 \\ I_{m-1} \end{pmatrix} \otimes S^{-\frac{1}{2}})K_{m-1,1} + \\
& ((\begin{pmatrix} 1 & -\lambda' \end{pmatrix})' \otimes 1)\frac{\partial vec(S^{-\frac{1}{2}})}{\partial vec(S^{\frac{1}{2}})'}\frac{\partial vec(S^{\frac{1}{2}})}{\partial vec(S)'}\frac{\partial vec(S)}{\partial vec(\lambda)'}, \\
\frac{\partial vec(S^{-\frac{1}{2}})}{\partial vec(S^{\frac{1}{2}})'} &= -(S^{-\frac{1}{2}\prime} \otimes S^{-\frac{1}{2}}) = -S^{-1}, \\
\frac{\partial vec(S^{\frac{1}{2}})}{\partial vec(S)'} &= ((1 \otimes S^{\frac{1}{2}}) + (S^{\frac{1}{2}\prime} \otimes 1))^{-1} = \frac{1}{2}S^{-\frac{1}{2}}, \\
\frac{\partial vec(S^{\frac{1}{2}})}{\partial vec(\lambda)'} &= (\lambda' \otimes 1)K_{m-1,1} + (1 \otimes \lambda') = 2\lambda',
\end{aligned}
$$

and $N = I_{k-m-1} + \theta_2\theta_2'$, $N^{\frac{1}{2}}N^{\frac{1}{2}} = N$, $S = (1 + \lambda'\lambda)$, $S^{\frac{1}{2}}S^{\frac{1}{2}} = S$, $e_1$ is the first $m$ dimensional unity vector, $K_{i,j} : ij \times ij$, are so called commutation matrices such that for any $A : i \times j$, $vec(A') = K_{i,j}vec(A)$, $vec(A) = K_{j,i}vec(A')$, $K_{j,i} = K_{i,j}'$, see Magnus and Neudecker (1988). Note that when $X$ is symmetric, $X = P\Lambda P'$, where $P$ are orthogonal eigenvectors and $\Lambda$ is a diagonal matrix containing the eigenvalues, that $X^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P'$ is also symmetric.

The jacobian of the transformation from $\Theta$ to $(\Pi_{21}, \theta_2, \lambda, \Omega)$ then reads,

$$
\frac{\partial vec(\Theta)}{\partial(vec(\Pi_{21})'\ vec(\theta_2)'\ vec(\lambda)'\ vec(\Omega)')} = \begin{pmatrix} J_1 & J_2 & J_3 & J_4 \end{pmatrix}.
$$

Since $\theta_2 = \Pi_{22}\Pi_{21}^{-1}$, the jacobians of the transformations from $(\Pi_{21}, \theta_2, \lambda, \Omega)$ to $\Pi_{21}$, $\Pi_{22}$, $\lambda$, and $\Omega$ read,

$$
\begin{aligned}
J_1 &= \frac{\partial(vec(\Pi_{21})'\ vec(\theta_2)'\ vec(\lambda)'\ vec(\Omega)')'}{\partial vec(\Pi_{21})'} = \begin{pmatrix} I_{m-1} \otimes I_{m-1} \\ -\Pi_{21}^{-1\prime} \otimes \Pi_{22}\Pi_{21}^{-1} \\ 0 \\ 0 \end{pmatrix} \\
J_2 &= \frac{\partial(vec(\Pi_{21})'\ vec(\theta_2)'\ vec(\lambda)'\ vec(\Omega)')'}{\partial vec(\Pi_{22})'} = \begin{pmatrix} 0 \\ \Pi_{21}^{-1\prime} \otimes I_{k-m-1} \\ 0 \\ 0 \end{pmatrix} \\
J_3 &= \frac{\partial(vec(\Pi_{21})'\ vec(\theta_2)'\ vec(\lambda)'\ vec(\Omega)')'}{\partial vec(\lambda)'} = \begin{pmatrix} 0 \\ 0 \\ 1 \otimes I_{m-1} \\ 0 \end{pmatrix}
\end{aligned}
$$

$$M_3 = \frac{\partial (\text{vec}(B_{221})' \; \text{vec}(\theta_2)' \; \text{vec}(\Phi)' \; \text{vec}(\Lambda)')'}{\partial \text{vec}(\Lambda)'} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ I \otimes I_{k_2 - m - 1} \end{pmatrix}$$

The jacobian of the transformation from $\theta$ to $(B_{22}, \Phi, \Lambda)$ then becomes,

$$\begin{aligned}
& |J(\theta, (B_{22}, \Phi, \Lambda))| \\
= \; & \left| \frac{\partial \text{vec}(\theta)}{\partial (\text{vec}(B_{22})' \; \text{vec}(\Phi)' \; \text{vec}(\Lambda)')} \right| \\
= \; & \left| \frac{\partial \text{vec}(\theta)}{\partial (\text{vec}(B_{221})' \; \text{vec}(\theta_2)' \; \text{vec}(\Phi)' \; \text{vec}(\Lambda)')} \right| \\
& \left| \frac{\partial (\text{vec}(B_{221})' \; \text{vec}(\theta_2)' \; \text{vec}(\Phi)' \; \text{vec}(\Lambda)')'}{\partial (\text{vec}(B_{22})' \; \text{vec}(\Phi)' \; \text{vec}(\Lambda)')} \right| \\
= \; & \left| \begin{pmatrix} J_1 & J_2 & J_3 & J_4 \end{pmatrix} \right| \left| \begin{pmatrix} M_1 & M_2 & M_3 & M_4 \end{pmatrix} \right|.
\end{aligned}$$

So,

$$J(\theta, (B_{22}, \Phi, \Lambda))|_{\lambda=0} = \begin{pmatrix} \beta' \otimes I_{k_2} & \alpha_1 \otimes B_{22} & \beta'_{\perp} \otimes B_{22\perp} \end{pmatrix}.$$

To prove that $|(J(\theta, (B_{221}, \theta_2, \Lambda, \Phi)))| \geq |(J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))|_{\lambda=0})|$, we use that

$$\begin{aligned}
& J(\theta, (B_{22}, \Lambda, \Phi)) \\
= \; & J(\theta, (B_{221}, \theta_2, \Lambda, \Phi)) J((B_{221}, \theta_2, \Lambda, \Phi), (B_{22}, \Lambda, \Phi)).
\end{aligned}$$

As shown previously

$$J((B_{221}, \theta_2, \Lambda, \Phi), (B_{22}, \Lambda, \Phi))|_{\lambda=0} = J((B_{221}, \theta_2, \Lambda, \Phi), (B_{22}, \Lambda, \Phi)).$$

It also holds that

$$J(\theta, (B_{221}, \theta_2, \Lambda, \Phi)) = J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))|_{\lambda=0} + R,$$

where

$$R = \begin{pmatrix} 0 & (\beta'_{\perp} \Lambda' \otimes I_{k_2}) \frac{\partial \text{vec}(\theta_{\perp})}{\partial \text{vec}(\theta_2)'} & (I_m \otimes \theta_{\perp} \Lambda) \frac{\partial \text{vec}(\theta_{\perp})}{\partial \text{vec}(\beta)'} & 0 \end{pmatrix}$$

such that

$$(J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))|_{\lambda=0}) R' = 0.$$

This implies that

$$\begin{aligned}
& |J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))| \\
= \; & |J(\theta, (B_{221}, \theta_2, \Lambda, \Phi)) J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))'|^{\frac{1}{2}} \\
= \; & |(J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))|_{\lambda=0})(J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))|_{\lambda=0})' + RR'|^{\frac{1}{2}} \\
\geq \; & |(J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))|_{\lambda=0})(J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))|_{\lambda=0})'|^{\frac{1}{2}} \\
\geq \; & |(J(\theta, (B_{221}, \theta_2, \Lambda, \Phi))|_{\lambda=0})|
\end{aligned}$$

and consequently

$$|\Gamma(\Theta, (\Pi_{22}, \lambda, \psi))| \geq |(\Gamma(\Theta, (\Pi_{22}, \lambda, \psi))|_{\lambda=0})|.$$

## B. Proof of theorem 1.

Assume that the reduced form of the SEM,

$$Y_1 B_{11} = X_1 \Gamma_{11} + X_2 \Gamma_{21} + U_1,$$

reads,

$$Y_1 = X_1 \Pi_{11} + X_2 \Pi_{21} + V_1,$$

where $\Pi_{11} = \Gamma_{11} B_{11}^{-1}$, $\Pi_{21} = \Gamma_{21} B_{11}^{-1}$, and this reduced form is equivalent to a set of nonlinear (reduced rank) restrictions on the parameters of a linear model and the (larger) parameters of this linear model, which are restricted to zero to obtain the reduced form, are locally uncorrelated with specific other parameters.

The parameter matrix of the reduced form of the SEM from theorem 1 reads,

$$\begin{aligned}
& \begin{pmatrix} \Gamma_{11} & 0 \\ \Gamma_{21} & \Gamma_{22} \\ 0 & \Gamma_{32} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}^{-1} \\
= & \begin{pmatrix} \Gamma_{11} & 0 \\ \Gamma_{21} & \Gamma_{22} \\ 0 & \Gamma_{32} \end{pmatrix} \begin{pmatrix} B_{11}^{-1} + B_{11}^{-1} B_{12} B_{22.1}^{-1} B_{21} B_{11}^{-1} & -B_{11}^{-1} B_{12} B_{22.1}^{-1} \\ -B_{22.1}^{-1} B_{21} B_{11}^{-1} & B_{22.1}^{-1} \end{pmatrix} \\
= & \begin{pmatrix} \Pi_{11}(I_{g_1} + \Psi_{12}\Psi_{21}) & -\Pi_{11}\Psi_{12} \\ \Pi_{21}(I_{g_1} + \Psi_{12}\Psi_{21}) - \Pi_{22}\Psi_{21} & \Pi_{22} - \Pi_{21}\Psi_{12} \\ -\Pi_{32}\Psi_{21} & \Pi_{32} \end{pmatrix},
\end{aligned}$$

where $\Pi_{11} = \Gamma_{11} B_{11}^{-1}$, $\Pi_{21} = \Gamma_{21} B_{11}^{-1}$, $\Pi_{32} = \Gamma_{32} B_{22.1}^{-1}$, $\Pi_{22} = \Gamma_{22} B_{22.1}^{-1}$, $B_{11.2} = B_{11} - B_{12} B_{22}^{-1} B_{21} = B_{11} - B_{12} B_{21}$, $B_{22.1} = B_{22} - B_{21} B_{11}^{-1} B_{12}$, $\Psi_{21} = B_{21} B_{11}^{-1}$, $\Psi_{12} = B_{12} B_{22.1}^{-1}$. This implies, as both $\Gamma_{32}$ and $\Gamma_{22}$ are unrestricted, that no restrictions are imposed on $\Pi_{32}$ and $\Pi_{22}$. The linear model of which the reduced form is a nonlinear restriction reads,

$$\begin{pmatrix} Y_1 & Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & X_2 & X_3 \end{pmatrix} \Theta + V,$$

where $\Theta : (k_1 + k_2 + k_3) \times (g_1 + g_2)$ and can be specified as,

$$\Theta = \begin{pmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \\ \Theta_{31} & \Theta_{32} \end{pmatrix},$$

$\Pi_{11} : l_{\bar{m}} \times j_{\bar{m}}$, $\Pi_{21} : l_{m} \times j_{\bar{m}}$, $\Pi_{31} : i_{m} \times j_{\bar{m}}$, $\Pi_{12} : l_{\bar{m}} \times j_{m}$, $\Pi_{22} : l_{m} \times j_{m}$, $\Pi_{32} : i_{m} \times j_{m}$. To obtain the restrictions on the linear model parameters which result in the reduced form, we specify $\Pi$ as,

$$\begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \\ \Pi_{31} & \Pi_{32} \end{pmatrix} = \begin{pmatrix} \tilde{\Pi}_{11} \\ \tilde{\Pi}_{21} \\ 0 \end{pmatrix} \begin{pmatrix} I_{j_m} & 0 \end{pmatrix} + \begin{pmatrix} \tilde{\Pi}_{12} \\ \tilde{\Pi}_{22} \\ A_{mm} \end{pmatrix} \begin{pmatrix} -B_{m\bar{m}} & I_{j_m} \end{pmatrix}$$

$$= \begin{pmatrix} & 0 & \\ & 0 & \\ A_{mm} \Lambda_{mm} \begin{pmatrix} -B_{m\bar{m}} & I_{j_m} \end{pmatrix} \end{pmatrix},$$

$$\begin{pmatrix} \tilde{\Pi}_{11} & \tilde{\Pi}_{12} \\ \tilde{\Pi}_{21} & \tilde{\Pi}_{22} \end{pmatrix} = \begin{pmatrix} A_{\bar{m}\bar{m}} \\ A_{m\bar{m}} \end{pmatrix} \begin{pmatrix} I_{j_m} & -B_{\bar{m}m} \end{pmatrix} + \begin{pmatrix} 0 \\ I_{j_m} \end{pmatrix} A_{\bar{m}m} \begin{pmatrix} 0 & I_{j_m} \end{pmatrix}$$

$$= \begin{pmatrix} A_{\bar{m}\bar{m}} \Lambda_{\bar{m}\bar{m}} \begin{pmatrix} I_{j_m} & -B_{\bar{m}m} \end{pmatrix} \\ 0 \end{pmatrix},$$

where $\tilde{\Pi}_{11} : l_{\bar{m}} \times j_{\bar{m}}$, $\tilde{\Pi}_{21} : l_{m} \times j_{\bar{m}}$, $\tilde{\Pi}_{12} : l_{\bar{m}} \times j_{m}$, $\tilde{\Pi}_{22} : l_{m} \times j_{m}$. It is clear from the chosen specification that when $\Lambda_{mm} = 0$, $\Lambda_{\bar{m}\bar{m}} = 0$, the reduced form results and that $\Lambda_{mm}$ is locally uncorrelated (when it is zero) with the parameters contained in $A_{mm}$ and $B_{m\bar{m}}$, and $\Lambda_{\bar{m}\bar{m}}$ is locally uncorrelated (when it is zero) with the parameters contained in $A_{\bar{m}\bar{m}}$ and $B_{\bar{m}m}$. As we can apply the same kind of decomposition on $A_{mm}$ and $A_{m\bar{m}}$, which we assumed to be possible, and since $A_{\bar{m}m}$ and $A_{mm}$ are unrestricted, such that there is no need to decompose them further, we can recursively apply the above decomposition and thereby the theorem is proved.

### C. Singular Value Decomposition and Jacobians two equation model

For the two equation model, reduced rank restrictions are imposed on the parameter matrices $\Pi_2$ and $\Pi_3$. In the following we state the SVDs and the jacobians involved with these two parameter matrices. We start with $\Pi_2$.

$$\Pi_2 = \begin{pmatrix} a & a_\perp \end{pmatrix} \begin{pmatrix} A_{211} & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} b_2 \\ b_{2\perp} \end{pmatrix}$$

$$= a A_{211} b_2 + a_\perp \lambda_2 b_{2\perp},$$

where $A_{21} = (A_{211}' \; A_{212}')'$, $A_{211} : m_1 \times m_1$, $A_{212} : (k_2 - m_1) \times m_1$, $a_2 = A_{212} A_{211}^{-1}$, $a = (I_{m_1} \; a_2')'$, $a_\perp = (-a_2 \; I_{k_2 - m_1})'(I_{k_2 - m_1} + a_2 a_2')^{-\frac{1}{2}}$, $b_2 = (I_{m_1} \; B_2)$, $b_{2\perp} = (I_{m_2} + B_2' B_2)^{-\frac{1}{2}}(-B_2' \; I_{m_2})$. A SVD can be used to obtain these parameters from $\Pi_2$,

$$\Pi_2 = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}',$$

where $U'U = I_{k_2}$; $V'V = I_m$; $U_{11}$, $S_1$, $V_{11} : m_1 \times m_1$; $S_2 : (k_2 - m_1) \times m_2$; $V_{22} : m_2 \times m_2$; $U_{21} : (k_2 - m_1) \times m_1$; $U_{12} : m_1 \times (k_2 - m_1)$; $U_{22} : (k_2 - m_1) \times (k_2 - m_1)$;

$U_{21}, U_{12}^t : m_2 \times m_1$; and $S_2$ contains the smallest $m_2$ singular values of $X_2$. This leads to the relations,

$$B_{211} = U_{11} S_1 V_{11}^t, \; A_2 = U_{21} U_{11}^{-1},$$
$$Y_2 = (V_{21} V_{11}^{-1})^t, \; Z_2 = (U_{22} U_{22}^t)^{-\frac{1}{2}} U_{22} S_2 V_{22}^t (V_{22} V_{22}^t)^{-\frac{1}{2}}.$$

The jacobians of $X_2$ with respect to $B_{211}$, $A_2$, $Y_2$ and $Z_2$ read,

$$J_1 = \frac{\partial vec(X_2)}{\partial vec(B_{211})^t} = (L_2^t \otimes K)$$
$$J_2 = \frac{\partial vec(X_2)}{\partial vec(A_2)^t} = (L_2^t B_{211}^t \otimes I_k) \frac{\partial vec(K)}{\partial vec(A_2)^t} + (L_{2\perp}^t Z_2^t \otimes I_k) \frac{\partial vec(K_\perp)}{\partial vec(A_2)^t}$$
$$J_3 = \frac{\partial vec(X_2)}{\partial vec(Y_2)^t} = (I_n \otimes K B_{211}) \frac{\partial vec(L_2)}{\partial vec(Y_2)^t} + (I_n \otimes K_\perp Z_2) \frac{\partial vec(L_{2\perp})}{\partial vec(Y_2)^t}$$
$$J_4 = \frac{\partial vec(X_2)}{\partial vec(Z_2)^t} = (L_{2\perp}^t \otimes K_\perp)$$

where

$$\frac{\partial vec(K)}{\partial vec(A_2)^t} = \begin{pmatrix} I_{m_1} \otimes \begin{pmatrix} 0 \\ I_{k-m_1} \end{pmatrix} \end{pmatrix},$$
$$\frac{\partial vec(K_\perp)}{\partial vec(A_2)^t} = -\left(H^{-\frac{1}{2}t} \otimes \begin{pmatrix} I_{m_1} \\ 0 \end{pmatrix}\right) P_{k-m_1,m_1} +$$
$$\left(I_{k-m_1} \otimes \begin{pmatrix} -A_2^t \\ I_{k-m_1} \end{pmatrix}\right) \frac{\partial vec(H^{-\frac{1}{2}})}{\partial vec(H^{\frac{1}{2}})^t} \frac{\partial vec(H^{\frac{1}{2}})}{\partial vec(H)^t} \frac{\partial vec(H)}{\partial vec(A_2)^t},$$
$$\frac{\partial vec(H^{-\frac{1}{2}})}{\partial vec(H^{\frac{1}{2}})^t} = -(H^{-\frac{1}{2}t} \otimes H^{-\frac{1}{2}}),$$
$$\frac{\partial vec(H^{\frac{1}{2}})}{\partial vec(H)^t} = ((I_{k-m_1} \otimes H^{\frac{1}{2}}) + (H^{\frac{1}{2}t} \otimes I_{k-m_1}))^{-1},$$
$$\frac{\partial vec(H)}{\partial vec(A_2)^t} = (A_2 \otimes I_{k-m_1}) + (I_{k-m_1} \otimes A_2) P_{k-m_1,m_1},$$
$$\frac{\partial vec(L_2)}{\partial vec(Y_2)^t} = \left(\begin{pmatrix} 0 \\ I_{n_2} \end{pmatrix} \otimes I_{m_1}\right),$$
$$\frac{\partial vec(L_{2\perp})}{\partial vec(Y_2)^t} = -\left(\begin{pmatrix} I_{m_1} \\ 0 \end{pmatrix} \otimes G^{-\frac{1}{2}}\right) P_{m_1,n_2} +$$
$$\left(\begin{pmatrix} -Y_2^t & I_{n_2} \end{pmatrix}^t \otimes I_{n_2}\right) \frac{\partial vec(G^{-\frac{1}{2}})}{\partial vec(G^{\frac{1}{2}})^t} \frac{\partial vec(G^{\frac{1}{2}})}{\partial vec(G)^t} \frac{\partial vec(G)}{\partial vec(Y_2)^t},$$
$$\frac{\partial vec(G^{-\frac{1}{2}})}{\partial vec(G^{\frac{1}{2}})^t} = -(G^{-\frac{1}{2}t} \otimes G^{-\frac{1}{2}}),$$

$$\frac{\partial vec(S^{\frac{1}{2}})}{\partial vec(S)'} = ((I_{m_2} \otimes S^{\frac{1}{2}}) + (S^{\frac{1}{2}\prime} \otimes I_{m_2}))^{-1},$$

$$\frac{\partial vec(S^{\frac{1}{2}})}{\partial vec(\Phi_2)'} = (\Phi_2' \otimes I_{m_2})K_{m_1,m_2} + (I_{m_2} \otimes \Phi_2'),$$

and $W = I_{k_2-m_1} + \phi_2\phi_2'$, $W^{\frac{1}{2}}W^{\frac{1}{2}} = W$, $S = (I_{m_2} + \Phi_2'\Phi_2)$, $S^{\frac{1}{2}}S^{\frac{1}{2}} = S$. The jacobian of the transformation from $\Theta_2$ to $(\Pi_{211}, \phi_2, \Phi_2, \lambda_2)$ then reads,

$$\frac{\partial vec(\Theta_2)}{\partial(vec(\Pi_{211})'\ vec(\phi_2)'\ vec(\Phi_2)'\ vec(\lambda_2)')} = \begin{pmatrix} J_1 & J_2 & J_3 & J_4 \end{pmatrix}.$$

Since $\phi_2 = \Pi_{212}\Pi_{211}^{-1}$, the jacobians of the transformations from $(\Pi_{211}, \phi_2, \Phi_2, \lambda_2)$ to $\Pi_{211}$, $\Pi_{212}$, $\Phi_2$, and $\lambda_2$ read,

$$L_1 = \frac{\partial(vec(\Pi_{211})'\ vec(\phi_2)'\ vec(\Phi_2)'\ vec(\lambda_2)')'}{\partial vec(\Pi_{211})'} = \begin{pmatrix} I_{m_1} \otimes I_{m_1} \\ -\Pi_{211}^{-1\prime} \otimes \Pi_{212}\Pi_{211}^{-1} \\ 0 \\ 0 \end{pmatrix}$$

$$L_2 = \frac{\partial(vec(\Pi_{211})'\ vec(\phi_2)'\ vec(\Phi_2)'\ vec(\lambda_2)')'}{\partial vec(\Pi_{212})'} = \begin{pmatrix} 0 \\ \Pi_{211}^{-1\prime} \otimes I_{k_2-m_1} \\ 0 \\ 0 \end{pmatrix}$$

$$L_3 = \frac{\partial(vec(\Pi_{211})'\ vec(\phi_2)'\ vec(\Phi_2)'\ vec(\lambda_2)')'}{\partial vec(\Phi_2)'} = \begin{pmatrix} 0 \\ 0 \\ I_{m_2} \otimes I_{m_1} \\ 0 \end{pmatrix}$$

$$L_4 = \frac{\partial(vec(\Pi_{211})'\ vec(\phi_2)'\ vec(\Phi_2)'\ vec(\lambda_2)')'}{\partial vec(\lambda_2)'} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ I_{m_2} \otimes I_{k_2-m_1} \end{pmatrix}$$

The jacobian of the transformation from $\Theta_2$ to $(\Pi_{21}, \Phi_2, \lambda_2)$ then becomes,

$$\begin{aligned}
& |J(\Theta_2, (\Pi_{21}, \Phi_2, \lambda_2))| \\
= & \left|\frac{\partial vec(\Theta_2)}{\partial(vec(\Pi_{21})'\ vec(\Phi_2)'\ vec(\lambda_2)')}\right| \\
= & \left|\frac{\partial vec(\Theta_2)}{\partial(vec(\Pi_{211})'\ vec(\phi_2)'\ vec(\Phi_2)'\ vec(\lambda_2)')}\right| \\
& \left|\frac{\partial(vec(\Pi_{211})'\ vec(\phi_2)'\ vec(\Phi_2)'\ vec(\lambda_2)')'}{\partial(vec(\Pi_{21})'\ vec(\Phi_2)'\ vec(\lambda_2)')}\right| \\
= & \left|\begin{pmatrix} J_1 & J_2 & J_3 & J_4 \end{pmatrix}\right| \left|\begin{pmatrix} L_1 & L_2 & L_3 & L_4 \end{pmatrix}\right|.
\end{aligned}$$

The specification of $\Pi_3$ reads,

$$\Pi_3 = \begin{pmatrix} A & A_{\perp} \end{pmatrix} \begin{pmatrix} S_{321} & 0 \\ 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} B_1 \\ B_{1\perp} \end{pmatrix},$$

where $A = (I_{m_2}\ A_2')'$, $B_1 = (\Psi_1\ I_{m_2})$, $S_{32} = (S_{321}'\ S_{322}')'$, $S_{321} : m_2 \times m_2$, $S_{322} : (k_3 - m_2) \times m_2$, $A_2 = S_{322}S_{321}^{-1}$. So, the specification of $\Pi_3$ is identical to the specification of $\Pi$ for the INSEM. The parameters $(S_{32}, \Psi_1, \lambda_3)$ can therefore be obtained using the SVDs (8)-(10) and changing the sizes of the involved matrices, *i.e.* $k_2$ to $k_3$, $m-1$ to $m_2$, 1 to $m_1$. Also the jacobian involved in the parameter transformation of the INSEM is identical to the jacobian in case of $\Pi_3$ when we change the sizes of the involved matrices in the outlined manner.

## B. Singular Value Decomposition and Jacobians three equation model

For the three equation model, reduced rank restrictions are imposed on the parameter matrices $(\Pi_2'\ \Pi_3')'$, $\Phi$ and $\Pi_1$. The important difference with the INSEM and the two equation model lies in $\Phi$ which itself already results from a reduced rank restriction. As we have to analyze $\Phi$ given $(\Pi_2'\ \Pi_3')'$, we start with the SVD and jacobian involved with $(\Pi_2'\ \Pi_3')'$. The specification of $(\Pi_2'\ \Pi_3')'$ reads,

$$\begin{pmatrix} \Pi_2 \\ \Pi_3 \end{pmatrix} = \Phi \begin{pmatrix} \Psi_{21} & I_{m_1} & 0 \\ 0 & 0 & I_{m_3} \end{pmatrix} + \Phi_{\perp}\lambda_2 \begin{pmatrix} \Psi_{21} & I_{m_1} & 0 \\ 0 & 0 & I_{m_3} \end{pmatrix}_{\perp}.$$

This implies that when $\Pi_2 = (\Pi_{21}\ \Pi_{22})$, $\Pi_{21} : k_2 \times (m_1 + m_2)$, $\Pi_{22} : k_2 \times m_3$; $\Pi_3 = (\Pi_{31}\ \Pi_{32})$, $\Pi_{31} : k_3 \times (m_1 + m_2)$, $\Pi_{32} : k_3 \times m_3$; $\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix}$, $\Phi_{11} : k_2 \times (m_1 + m_2)$, $\Phi_{12} : k_2 \times m_3$, $\Phi_{21} : k_3 \times (m_1 + m_2)$, $\Phi_{22} : k_3 \times m_3$; that the following equality holds,

$$\begin{pmatrix} \Phi_{12} \\ \Phi_{22} \end{pmatrix} = \begin{pmatrix} \Pi_{22} \\ \Pi_{32} \end{pmatrix}.$$

and we are left with,

$$\begin{pmatrix} \Pi_{21} \\ \Pi_{31} \end{pmatrix} = \begin{pmatrix} \Phi_{11} \\ \Phi_{21} \end{pmatrix} \begin{pmatrix} \Psi_{21} & I_{m_1} \end{pmatrix} + \begin{pmatrix} \Phi_{11} \\ \Phi_{21} \end{pmatrix}_{\perp} \lambda_2 \begin{pmatrix} \Psi_{21} & I_{m_1} \end{pmatrix}_{\perp},$$

which is again identical to the specification of $\Pi$ for the INSEM such that when we change the sizes of the matrices in the appropriate manner, *i.e.* $k_2$ to $k_2 + k_3$, $m-1$ to $m_2$ and 1 to $m_1$, we can directly use the SVDs and jacobians for $\Pi$ of the INSEM.

The SVDs and jacobians for $\Phi_2$ are constructed using (50) and (51),

$$\begin{pmatrix} \Phi_{21} & \Phi_{22} \end{pmatrix} = S_{33} \begin{pmatrix} \Psi_{32} & I_{m_2} \end{pmatrix} + S_{33\perp}\lambda_3 \begin{pmatrix} \Psi_{32} & I_{m_2} \end{pmatrix}_{\perp}.$$

Again this specification is identical to the specification of $\Phi$ for the INSEM such that we can use the SVD and jacobians specified for the INSEM when we change the sizes of the matrices in the appropriate manner, i.e. $k_2$ to $k_3$, $m-1$ to $m_3$ and 1 to $m_2$.

The specification of $\Phi_1$ reads,

$$\begin{aligned}\Phi_1 &= \begin{pmatrix} \Pi_{11} & \Pi_{12} \end{pmatrix} \begin{pmatrix} I_{m_1+m_2} & \begin{pmatrix} \beta_{13} \\ \beta_{23} \end{pmatrix} \end{pmatrix} + \\ & \begin{pmatrix} \Pi_{11} & \Pi_{12} \end{pmatrix}_{\perp} \lambda_1 \begin{pmatrix} I_{m_1+m_2} & \begin{pmatrix} \beta_{13} \\ \beta_{23} \end{pmatrix} \end{pmatrix}_{\perp}.\end{aligned}$$

This specification is identical to the specification of $\Phi_2$ in the two equation model such that we can use the jacobians and the SVD listed there when we change the sizes of the matrices in the appropriate manner, i.e. $k_2$ to $k_1$, $m_1$ to $m_1+m_2$ and $m_2$ to $m_3$.

## B. Obtaining SEMs from Linear Models using SVDs

The specification of $\Phi$ reads,

$$\begin{aligned}\Phi &= \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & s_2 \end{pmatrix} \begin{pmatrix} V'_{11} & V'_{21} \\ V'_{12} & V'_{22} \end{pmatrix} \\ &= \Pi_2 \begin{pmatrix} \beta & I_{m-1} \end{pmatrix} + \Pi_{2\perp} \lambda \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_{\perp},\end{aligned}$$

such that

$$\Pi_2 \begin{pmatrix} \beta & I_{m-1} \end{pmatrix} = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 \begin{pmatrix} V'_{11} & V'_{21} \end{pmatrix}$$

and

$$\Pi_{2\perp} \lambda \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_{\perp} = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} s_2 \begin{pmatrix} V'_{12} & V'_{22} \end{pmatrix}.$$

Consequently,

$$\Pi_2 = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V'_{21} \text{ and } \beta = V'^{-1}_{21} V'_{11}.$$

Substituting these expressions in the specification of $\Pi_{2\perp}$ and $\begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_{\perp}$ gives,

$$\begin{aligned}\Pi_{2\perp} &= \begin{pmatrix} -\Pi^{-1\prime}_{21}\Pi'_{22} \\ I_{k_2-m+1} \end{pmatrix} (I_{k_2-m+1} + \Pi_{22}\Pi^{-1}_{21}\Pi^{-1\prime}_{21}\Pi'_{22})^{-\frac{1}{2}} \\ &= \begin{pmatrix} -U'^{-1}_{11} S^{-1}_1 V^{-1}_{21} V_{21} S_1 U'_{21} \\ I_{k_2-m+1} \end{pmatrix} (I_{k_2-m+1} + U_{21}U^{-1}_{11}U'^{-1}_{11}U'_{21})^{-\frac{1}{2}} \\ &= \begin{pmatrix} -U'^{-1}_{11} U'_{21} \\ I_{k_2-m+1} \end{pmatrix} (I_{k_2-m+1} + U_{21}U^{-1}_{11}U'^{-1}_{11}U'_{21})^{-\frac{1}{2}},\end{aligned}$$

as $V'_{11}V_{12} + V'_{21}V_{22} = 0$ (because of the orthogonality of $V$), $V_{12}V_{22}^{-1} = -V'^{-1}_{11}V'_{21}$, and $V'_{12}V_{12} + V'_{22}V_{22} = I_{n-m-1}$, such that

$$\begin{aligned}
\mathbf{B}_{2m} &= \begin{pmatrix} V_{12}V_{22}^{-1} \\ I_{n-m-1} \end{pmatrix} (I_{n-m-1} + V_{22}^{-1\prime}V'_{12}V_{12}V_{22}^{-1})^{-\frac{1}{2}} \\
&= \begin{pmatrix} V_{12} \\ V_{22} \end{pmatrix} V_{22}^{-1}(V_{22}'^{-1}(V'_{22}V_{22} + V'_{12}V_{12})V_{22}^{-1})^{-\frac{1}{2}} \\
&= \begin{pmatrix} V_{12} \\ V_{22} \end{pmatrix} V_{22}^{-1}((V_{22}'^{-1}V_{22}^{-1})^{-1})^{\frac{1}{2}} \\
&= \begin{pmatrix} V_{12} \\ V_{22} \end{pmatrix} V_{22}^{-1}(V_{22}V'_{22})^{\frac{1}{2}}.
\end{aligned}$$

Similarly for $\begin{pmatrix} Y & I_{m-1} \end{pmatrix}_m$,

$$\begin{aligned}
\begin{pmatrix} Y & I_{m-1} \end{pmatrix}_m &= (1 + Y'Y)^{-\frac{1}{2}} \begin{pmatrix} 1 & -Y' \end{pmatrix} \\
&= (1 + u_{11}U_{21}^{-1}U_{21}'^{-1}u'_{11})^{-\frac{1}{2}} \begin{pmatrix} 1 & -u_{11}U_{21}^{-1} \end{pmatrix} \\
&= (1 + u_{12}^{-1\prime}u'_{22}u_{22}u_{12}^{-1})^{-\frac{1}{2}} \begin{pmatrix} 1 & u_{12}^{-1\prime}u'_{22} \end{pmatrix} \\
&= (u_{12}^{-1\prime}(u'_{12}u_{12} + u'_{22}u_{22})u_{12}^{-1})^{-\frac{1}{2}}u_{12}'^{-1} \begin{pmatrix} u'_{12} & u'_{22} \end{pmatrix} \\
&= (u_{12}^{-1\prime}u_{12}^{-1})^{-\frac{1}{2}}u_{12}'^{-1} \begin{pmatrix} u'_{12} & u'_{22} \end{pmatrix} \\
&= (u_{12}u'_{12})^{\frac{1}{2}}u_{12}'^{-1} \begin{pmatrix} u'_{12} & u'_{22} \end{pmatrix},
\end{aligned}$$

since $u'_{11}u_{12} + U'_{21}u_{22} = 0$, such that $-U_{21}'^{-1}u'_{11} = u_{22}u_{12}^{-1}$, and $u'_{12}u_{12} + u'_{22}u_{22} = 1$. Consequently in order to have equivalence,

$$\begin{aligned}
\lambda &= (V_{22}V'_{22})^{-\frac{1}{2}}V_{22}s_2u'_{12}(u_{12}u'_{12})^{-\frac{1}{2}} \\
&= ks_2a,
\end{aligned}$$

where $k = (V_{22}V'_{22})^{-\frac{1}{2}}V_{22}$, and $a = u'_{12}(u_{12}u'_{12})^{-\frac{1}{2}}$. Both $k$ and $a$ are orthogonal matrices (scalars) which results from the singular value decomposition because when $X = USV'$, where both $U$ and $V$ are orthogonal that

$$(XX')^{\frac{1}{2}} = (USV'VSU')^{\frac{1}{2}} = (US^2U')^{\frac{1}{2}} = USU',$$

such that

$$(XX')^{-\frac{1}{2}}X = US^{-1}U'USV' = UV',$$

which is an orthogonal matrix, such that $\lambda$ equals the smallest singular value pre and postmultiplied by orthogonal vectors/matrices or stated differently, $\lambda$ is a rotation of the singular values.

# References


[1] Anderson, T.W., 1982, Some Recent Developments on the Distributions of Single Equation Estimators, in W. Hildebrand, ed., *Advances in Econometrics*, North-Holland Publishing Co., Amsterdam

[2] Anderson, T.W., and H. Rubin, 1949, Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations, *Annals of Mathematical Statistics*, 21, 570-582

[3] Bauwens, L., and H.K. van Dijk, 1989, Bayesian Limited Information Analysis Revisited, in J.J. Gabszewicz et. al., eds., *Economic Decision-Making : Games, Econometrics and Optimisation*, North-Holland Publishing Co., Amsterdam

[4] Billingsley, P., 1986, *Probability and Measure*, Wiley, New York

[5] Chao, J.C., and P.C.B. Phillips, 1996, *Bayesian Posterior Distributions in Limited Information Analysis of the Simultaneous Equation Model using the Jeffreys' Prior*, Unpublished Manuscript, Cowles Foundation, Yale University

[6] Chen, M-H., 1994, Importance-Weighted Marginal Bayesian Posterior Density Estimation, *Journal of the American Statistical Association*, 89, 818-824

[7] Dickey, J., 1971, The Weighted Likelihood Ratio, Linear Hypothesis on Normal Location Parameters, *The Annals of Mathematical Statistics*, 42, 204-223

[8] Drèze, J.H., 1976, Bayesian Limited Information Analysis of the Simultaneous Equations Model, *Econometrica*, 44, 1045-1075

[9] Drèze, J.H., 1977, Bayesian Regression using poly-t densities, *Journal of Econometrics*, 6, 329-354

[10] Drèze, J.H., and J.A. Morales, 1976, Bayesian Full Information Analysis of Simultaneous Equations, *Journal of the American Statistical Association*, 71, 329-354

[11] Drèze, J.H., and J.F. Richard, 1983, Bayesian Analysis of Simultaneous Equations Systems, in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics*, Vol.I., North-Holland Publishing Co., Amsterdam

[12] Engle, R.F., D.F. Hendry and J.F. Richard, 1983, Exogeneity, *Econometrica*, 51, 277-304

[13] Gelfand, A.E., and A.F.M. Smith, 1990, Sampling based Approaches to calculating Marginal Densities, *Journal of the American Statistical Association*, 85, 398-409

[14] Golub, G.H., and C.F. van Loan, 1989, *Matrix Computations*, The John Hopkins University Press, Baltimore

[15] Geweke, J., 1989, Bayesian Inference in Econometric Models using Monte-Carlo Integration, *Econometrica*, 57, 1317-1339

[16] Geweke, J., 1996, Bayesian Reduced Rank Regression in Econometrics, *Journal of Econometrics*, 75, 63-96

[17] Haavelmo, T., 1943, The Statistical Implications of a System of Simultaneous Equations, *Econometrica*, 11, 1-12

[18] Hastings, W.K., 1970, Monte Carlo Sampling using Markov Chains and their applications, *Biometrika*, 57, 97-109

[19] Hausman, J.A., 1983, Specification and Estimation of Simultaneous Equations Systems, in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics*, Vol.I., North-Holland Publishing Co., Amsterdam

[20] Kleibergen, F., 1997, Equality Restricted Random Variables: Densities and Sampling Algorithms, *Econometric Institute Report 9662/A*, Erasmus University Rotterdam, available at http://www.eur.nl/few/ei/faculty/kleibergen.

[21] Kleibergen, F. and H. Hoek, 1996, Bayesian Analysis of ARMA models using Noninformative Priors, *Econometric Institute Report 9553/A*, Erasmus University Rotterdam, available at http://www.eur.nl/few/ei/faculty/kleibergen.

[22] Kleibergen, F. and R. Paap, 1997, Priors, Posterior Odds and Lagrange Multiplier Statistics in Bayesian Analyses of Cointegration, *Econometric Institute Report 9668/A*, Erasmus University Rotterdam, available at http://www.eur.nl/few/ei/faculty/kleibergen.

[23] Kleibergen, F.R., and H.K. van Dijk, 1992, Bayesian Simultaneous Equation Model analysis: On the Existence of Structural Posterior Moments, *Econometric Institute Reports*, 9269.

[24] Kleibergen, F.R., and H.K. van Dijk, 1994a, Bayesian Analysis of Simultaneous Equation Models using Noninformative Priors, *Tinbergen Institute Discussion Papers*, TI 94-134.

[25] Kleibergen, F.R., and H.K. van Dijk, 1994b, On the Shape of the Likelihood/Posterior in Cointegration Models, *Econometric Theory*, 10, 514-551

[26] Kleibergen, F.R., and E. Zivot, 1998, *Single Equation Analysis: How do Classical and Bayesian Procedures Relate?*, Unpublished Manuscript, available at http://www.eur.nl/few/ei/faculty/kleibergen.

[27] Kloek, T., and H.K. van Dijk, 1978, Bayesian Estimates of Equation System Parameters : An Application of Integration by Monte-Carlo, *Econometrica*, 46, 1-19

[28] Maddala, G.S., 1976, Weak Priors and Sharp Posteriors in Simultaneous Equation Models, *Econometrica*, 44, 345-351

[29] Magnus, J.R. and H. Neudecker, 1988, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley & Sons, Chichester

[30] Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, 1953, Equations of State Calculations by fast computing Machines, *Journal of Chemical Physics*, 21, 1087-1092

[31] Phillips, P.C.B., 1983, Exact Small Sample Theory in the Simultaneous Equations Model, in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics*, Vol.1., North-Holland Publishing Co., Amsterdam

[32] Phillips, P.C.B., 1989, Partially Identified Econometric Models, *Econometric Theory*, 5, 181-240

[33] Poirier, D., 1994, *Prior Beliefs about Fit*, Unpublished Manuscript, University of Toronto

[34] Richard, J.-F. and H. Tompa, 1980, On the Evaluation of 1-1 Poly t density functions, *Journal of Econometrics*, 12, 335-351

[35] Roberts, G.O. and A.F.M. Smith, 1994, Simple Conditions for the convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms, *Stochastic Processes and their Applications*, 49, 207-216

[36] Smith, A.F.M., and G.O. Roberts, 1993, Bayesian Computation via the Gibbs Sampler and related Markov Chain Methods, *Journal of the Royal Statistical Society B*, 55, 3-23

[37] Tierney, L., 1994, Markov Chains for exploring Posterior Distributions, *Annals of Statistics*, 22, 1701-1762

[38] Tintner, G., 1952, *Econometrics*, Wiley, New York

[39] Verdinelli, I. and L. Wasserman, 1995, Computing Bayes Factors using a Generalization of the Savage-Dickey Density Ratio, *Journal of the American Statistical Association*, 90, 614-618

[40] Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York

[41] Zellner, A., L. Bauwens, and H.K. van Dijk, 1988, Bayesian Specification Analysis and Estimation of Simultaneous Equation Models using Monte-Carlo Integration, *Journal of Econometrics*, 38, 39-72