Equality Restricted Random Variables: Densities and Sampling Algorithms

Frank Kleibergen*

Econometric Institute Reports, no. 9662/A
November 1996

Abstract

Many common statistical models can be specified as linear models with restrictions imposed on the parameters. A large amount of these models impose restrictions which do not allow for the analytical construction of the probability density function (pdf) of the parameters given the restrictions. This is often implicitly assumed which leads to an inconsistency as the pdf of the parameters of the linear specification under the imposed restrictions is then not nested within the assumed pdf of the unrestricted linear specification. The paper shows how these restrictions need to be incorporated by constructing the pdfs incorparating the restrictions and algorithms to sample from these pdfs. We show how these methods are applied to some common statistical models, i.e. ARMA, cointegration and simultaneous equation models.

1 Introduction

Many statistical models impose restrictions on the parameters of their implied linear specification. Consider, for example an ARMA(1,1) time series model. This model has an implied linear specification which is an AR(∞) model with restrictions on its parameters. In a Bayesian setting, the probability density functions (pdf) of the parameters of these models have typically

---

*Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, email:kleibergen@few.eur.nl. The author thanks the Netherlands Organisation for Scientific Research (N.W.O.) for their financial support.

been constructed by directly constructing the pdfs (posteriors) of the parameters conditional on the imposed restrictions. This was seen as the Bayesian analog of the classical statistical manner of incorporating parameter restrictions, analyze the likelihood conditional on the restrictions imposed on the parameters. A transformation of parameters, from the parameters of the unrestricted linear model to the parameters spanning the restrictions and the remaining parameter space, shows that the restrictions imposed by many common statistical models do not allow for an analytical construction of the pdfs of the parameters conditional on the imposed restrictions. If mistakenly the pdf of the parameters is constructed conditional on the imposed restrictions, an inconsistent analysis results as the pdf under the imposed restrictions is then not nested within the pdf with unrestricted parameters. This phenomenon especially appears in case the restricted model is compared with the unrestricted model as the pdf of the parameters of the first is not nested through parameter transformations within the assumed pdf of the latter.

In this paper we discuss how the pdfs of parameters of statistical models, which impose restrictions on implied linear specifications, are analyzed. Section 2 of the paper discusses the general theory to derive the pdfs of the parameters and how to sample from these pdfs using specific simulation algorithms. Section 3 discusses some examples of restrictions imposed by common statistical models. The first example discusses linear restrictions where the pdf of the parameters conditional on the imposed restrictions can be constructed analytically such that the standard kind of analysis is allowed for. In the other examples of restrictions imposed by ARMA, cointegration and simultaneous equations models, it is not possible to construct the pdfs of the parameters conditional on the restrictions. In the Bayesian analyzes of these models in the literature, for ARMA models a.o. Monahan (1983) and Chib and Greenberg (1994), for cointegration models a.o. Geweke (1996) and Kleibergen and van Dijk (1994b) and for simultaneous equation models a.o. Drèze (1976), Drèze and Morales (1976) and Drèze and Richard (1983), it is implicitly assumed that the pdfs of the parameters conditional on the restrictions can be constructed. The resulting analysis then leads to pdfs which are, after parameter transformation, not nested within the assumed pdf of the parameters of the implied linear model without restrictions, which is an inconsistency. The fourth section gives a summary of the involved simulation algorithms and the fifth section concludes.

2 Random variables with imposed equality restrictions

Without loss of generality, consider the k dimensional continuous random variable $\pi \in \mathbf{R}^k$ (Note that π can also be a scalar or matrix) with known probability density function (pdf) $p(\pi)$ which is defined on the $\mathbf{R}^k$ (Also other nonempty probability spaces can be considered). Suppose that m $(k > m)$ equality restrictions are imposed on π,

$$g(\pi) = 0, \tag{1}$$

where $g(\pi) = (g_1(\pi) ... g_m(\pi))'$ and all elements g_i, $i = 1, ..., m$, are continuous differentiable and $\frac{\partial g}{\partial \pi'}$ is only allowed to have a lower rank value on probability spaces with measure zero, i.e. lower dimensional probability subspaces. The purpose of this section is to construct the pdfs of the parameters which span the, through the equality restrictions (1), restricted probability space and a sampling algorithm to generate drawings of the random variable π which accord with (1). Examples of models where these kind of random variables appear are discussed in section 3. In the construction of the pdfs and the algorithm we use the following definition.

Definition 1 *Let* $\theta \in \mathbf{R}^{k-m}$ *and* $\lambda \in \mathbf{R}^m$ *be such that*

$$\lambda = g(\pi) \tag{2}$$

and

$$\theta = f(\pi) \tag{3}$$

where g *is defined in (1) and the* $k-m$ *dimensional continuous differentiable function* f *is constructed such that* $\begin{pmatrix} g \\ f \end{pmatrix}$ *is an invertable function.*

As there is an invertable relationship between π and (θ, λ) we can construct the pdf of (θ, λ) implied by $p(\pi)$,

$$p(\theta, \lambda) = p(h(\lambda, \theta)) \left| \frac{\partial h}{\partial(\lambda, \theta)} \right|, \tag{4}$$

where $h = \begin{pmatrix} g \\ f \end{pmatrix}^{-1}$. The pdf of (θ, λ) evaluated in $\lambda = 0$ is proportional to the pdf of the random variables which span the, through the equality restrictions (1), restricted probability space,

$$p_{res}(\theta) \propto p(\theta, \lambda = 0), \tag{5}$$

where res indicates that the pdf is defined on the probability space restricted by (1). Depending on the possible analytical construction of the conditional and marginal pdfs in (4), we can distinguish four different cases for the construction of $p_{res}(\theta)$ and an algorithm to obtain random drawings from (5):

1. The conditional pdf of θ given λ and the marginal pdf of λ can be constructed analytically, $p(\theta,\lambda) = p(\theta|\lambda)p(\lambda)$. So, $p_{res}(\theta) = p(\theta|\lambda = 0)$ and we can directly simulate from $p_{res}(\theta)$. The drawings of θ, θ^i, can then be transformed to drawings of π, π^i, using $\pi^i = h(0,\theta^i)$.

2. Only the conditional pdf of λ given θ and the marginal pdf of θ can be constructed analytically, $p(\theta,\lambda) = p(\lambda|\theta)p(\theta)$. So, $p_{res}(\theta) \propto p(\lambda = 0|\theta)p(\theta)$. Drawings from $p_{res}(\theta)$ can now be constructed by drawing θ from $p(\theta)$ and attaching a weight to each drawing θ^i equal to $w(\theta^i) = p(\lambda = 0|\theta^i)$. These weights can then be used in an Importance, Acception-Rejection (A-R) or Metropolis-Hastings (M-H) sampling algorithm to generate drawings from $p_{res}(\theta)$, see section 4.

3. Combination of cases 1 and 2 where we can analytically construct the conditional pdf of some elements of θ, say θ_2, given the other elements of θ, say θ_1, and λ, $p(\theta,\lambda) = p(\theta_2|\lambda,\theta_1)p(\theta_1,\lambda)$. The pdf of θ on the restricted probability space now becomes, $p_{res}(\theta) \propto p(\theta_2|\lambda = 0,\theta_1)p(\theta_1,\lambda = 0)$ such that when we integrate out θ_2 given $\lambda = 0$, θ_1, we again have a pdf for θ_1 on the restricted probability space identical to (5), as $p_{res}(\theta_1) \propto p(\theta_1,\lambda = 0)$. This density can then again belong to one of the four cases.

4. Combination of cases 1 and 2 where we can analytically construct the conditional pdf of some elements of λ, say λ_2, given θ and the other elements of λ, say λ_1, $p(\theta,\lambda) = p(\lambda_2|\lambda_1,\theta)p(\theta,\lambda_1)$. The pdf of θ on the restricted probability space now becomes, $p_{res}(\theta) \propto p(\lambda_2 = 0|\lambda_1 = 0,\theta)p(\theta,\lambda_1 = 0)$. Given a drawing of θ from $p(\theta,\lambda_1 = 0)$, θ^i, we construct a drawing from $p_{res}(\theta)$ by attaching a weight to it equal to $w(\theta^i) = p(\lambda_2 = 0|\lambda_1 = 0,\theta^i)$ in an Importance, A-R or M-H sampling algorithm.

Cases 3 and 4 can be applied iteratively to obtain a sampling algorithm which can be applied to quite complicated models. In the next section we show some examples of models which lead to pdfs of the parameters which belong to one of the four cases outlined above. Note that the cases 2-4 show an important difference with the classical statistical likelihood analysis where only case 1 occurs as the likelihood is directly specified in terms of the, in this paradigm fixed, parameters. Many Bayesian analyses of models with restricted parameters, like of the models discussed in the next section, have neglected the sometimes implicit restrictions on the parameters and the resulting decompositions mentioned above. They therefore implicitly assume that these restrictions imply decompositions of the pdfs which belong to case 1. This assumption is often incorrect, as shown in the next examples, which then leads to an inconsistent analysis as the pdf of the parameters under the restriction is not nested within the pdf in case of unrestricted parameters. The cases 2-4 happen in important and widely used models, some of which are used for expository purposes in the next section.

3 Examples

All examples use a normal density for π which results as the conditional posterior of π given σ^2 in a standard linear model with diffuse prior, see a.o. Zellner (1971) (note that natural conjugate priors lead to similar posteriors and can be used as well in the analysis),

$$y = X\pi + \varepsilon, \tag{6}$$

where $y : T \times 1$, $X : T \times k$, and $\varepsilon : T \times 1$ contains the T disturbances, $\varepsilon_1, ..., \varepsilon_T$, which are i.i.d. normal with mean 0 and variance σ^2. This conditional posterior is then proportional to a normal pdf,

$$p(\pi|\sigma^2, X) \propto \sigma^{-k}|X'X|^{\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(\pi - \hat{\pi})'X'X(\pi - \hat{\pi})], \tag{7}$$

where $\hat{\pi} = (X'X)^{-1}X'y$. In the following subsections we consider different restrictions imposed, by common statistical models, on π. Note that other densities for π can be considered as well.

3.1 Linear Restrictions

Let π be distributed according to (7) and let a set of m $(< k)$ linear restrictions be imposed on π,

$$g(\pi) = G\pi - b = 0, \tag{8}$$

where $G : m \times k$ with full rank m and $b : m \times 1$. We now construct,

$$\begin{aligned} \lambda &= g(\pi) = G\pi - b, \\ \theta &= f(\pi) = F\pi, \end{aligned} \tag{9}$$

where $\lambda : m \times 1$, $\theta : (k-m) \times 1$, $F : (k-m) \times k$ and $\begin{pmatrix} G \\ F \end{pmatrix}$ is invertable, $H = \begin{pmatrix} G \\ F \end{pmatrix}^{-1}$. The density of (θ, λ) then reads,

$$\begin{aligned} p(\theta, \lambda|\sigma^2, X) \;\propto\; & \sigma^{-k}|H'X'XH|^{\frac{1}{2}} \\ & \exp[-\frac{1}{2\sigma^2}(\begin{pmatrix} \lambda \\ \theta \end{pmatrix} - \begin{pmatrix} \hat{\lambda} \\ \hat{\theta} \end{pmatrix})'H'X'XH(\begin{pmatrix} \lambda \\ \theta \end{pmatrix} - \begin{pmatrix} \hat{\lambda} \\ \hat{\theta} \end{pmatrix})], \end{aligned} \tag{10}$$

where $\hat{\lambda} = G\hat{\pi}$, $\hat{\theta} = F\hat{\pi}$. As (10) is a standard normal pdf, we can construct the conditional pdf of θ given λ,

$$\begin{aligned} p(\theta|\lambda, \sigma^2, X) \;\propto\; & \sigma^{-(k-m)}|H_2'X'XH_2|^{\frac{1}{2}} \\ & \exp[-\frac{1}{2\sigma^2}(\theta - \hat{\theta} - (H_2'X'XH_2)^{-1}H_2'X'XH_1(\lambda - \hat{\lambda}))' \\ & H_2'X'XH_2(\theta - \hat{\theta} - (H_2'X'XH_2)^{-1}H_2'X'XH_1(\lambda - \hat{\lambda}))], \end{aligned} \tag{11}$$

where $H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}$, $H_1 : m \times k$, $H_2 : (k-m) \times k$. The pdf of θ on the restricted probability space now equals the conditional pdf (11) evaluated in $\lambda = 0$. This shows that the linear restriction example belongs to case 1 discussed in the previous section. The degenerate pdf of π which incorporates the equality restrictions can also be constructed from (11) using $\pi = H\begin{pmatrix} 0 \\ \theta \end{pmatrix}$.

3.2 Nonlinear Restrictions

Many widely used models impose nonlinear restrictions on the parameters of the implied linear specification, for example ARMA time series models, fractionally integrated time series models, factor models, cointegration models and simultaneous equation models. These restrictions are often such that the pdf of the parameters in a Bayesian setting does not belong to case 1 from the previous section. So, if the restrictions are neglected and the pdf is treated as if belonging to case 1, an inconsistent analysis results as the pdf of the parameters under the restriction is not nested within the pdf of the parameters in the unrestricted case. In the following subsections some examples of models, i.e. ARMA, cointegration and simultaneous equations, are discussed were this occurs.

3.2.1 ARMA(1,1)

AutoRegressive Moving Average (ARMA) models imply equality restrictions for the parameters of a linear AR(∞) model. Consider, for example, the ARMA(1,1) model,

$$(1-\rho L)y_t = (1-\alpha L)\varepsilon_t, \qquad t=1,...,T, \tag{12}$$

where L is a lag operator and we condition on y_0, y_{-1}. By recurrently substituting ε_{t-i}, $i=1,...,T$, the implied AR(∞) specification of this model is obtained,

$$y_t = \sum_{i=1}^{T} \pi_i y_{t-i} + \varepsilon_t, \tag{13}$$

where $\pi_i = \pi_2 \left(\frac{\pi_2}{\pi_1}\right)^{i-2} = 0$, $i=3,...,T+1$, $\pi_1 = \rho-\alpha$, $\pi_2 = \alpha(\rho-\alpha)$, $y_{-i}=0$, $i>1$. As (13) is a linear model, the pdf of π is in the unrestricted case similar to (7) with $k=T$,

$$p(\pi|\sigma^2, X) \propto \sigma^{-T}|X'X|^{\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(\pi-\hat{\pi})'X'X(\pi-\hat{\pi})], \tag{14}$$

where $X: T\times T$, $X=(x_1...x_T)$, $x_i=(y_{1-i}...y_{T-i})'$, $\pi=(\pi_1...\pi_T)'$, $y=(x_1...x_T)', \hat{\pi}=(X'X)^{-1}X'y=(\hat{\pi}_1...\hat{\pi}_T)'$. Note that (14) is a degenerate density as $X'X$ is not invertable but by reducing the number of elements of π (T) an invertable $X'X$ is obtained. The restrictions specified on π read,

$$g_i(\pi) = \pi_{i+2} - \pi_2\left(\frac{\pi_2}{\pi_1}\right)^i = 0, \quad i=1,...,T-2. \tag{15}$$

We now construct,

$$\begin{aligned} \lambda_i &= g_i(\pi) = \pi_{i+2} - \pi_2\left(\frac{\pi_2}{\pi_1}\right)^i & i&=1,...,T-2, \\ \theta_i &= f_i(\pi) = \pi_i & i&=1,2. \end{aligned} \tag{16}$$

The inverse relation of this is,

$$\begin{aligned} \pi_i &= h_i(\theta,\lambda) = \theta_i & i&=1,2, \\ \pi_i &= h_i(\theta,\lambda) = \lambda_{i-2} + \tilde{\pi}_i & i&=3,...,T, \end{aligned} \tag{17}$$

where $\tilde{\pi}_i = \theta_2\left(\frac{\theta_2}{\theta_1}\right)^{i-2}$, $i=3,...,T$. The jacobian of the transformation from π to (θ,λ) is equal to 1. Note that we didnot explicitly use the parameters of the ARMA(1,1), (α,ρ), to construct θ. We do this as the pdfs are easier to express in θ than in (α,ρ) and there is a one-to-one relationship between

θ and (α, ρ). The pdfs for (α, ρ) can, therefore, be obtained by transforming θ to (α, ρ) (jacobian is $|\rho - \alpha|$). The joint pdf of (θ, λ) reads,

$$p(\theta, \lambda|\sigma^2, X) \propto \sigma^{-T}|X'X|^{\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(\pi(\theta, \lambda) - \hat{\pi})'X'X(\pi(\theta, \lambda) - \hat{\pi})], \quad (18)$$

which can be decomposed into a marginal density for θ and a conditional density for λ given θ, $p(\theta, \lambda|\sigma^2, X) = p(\lambda|\theta, \sigma^2, X)p(\theta|\sigma^2, X)$, with

$$\begin{aligned} p(\theta|\sigma^2, X) &\propto \sigma^{-2}|(x_1\ x_2)'M_{X_2}(x_1\ x_2)|^{\frac{1}{2}} \quad (19) \\ &\quad \exp[-\frac{1}{2\sigma^2}(\theta - \hat{\theta})'(x_1\ x_2)'M_{X_2}(x_1\ x_2)(\theta - \hat{\theta})], \\ p(\lambda|\theta, \sigma^2, X) &\propto \sigma^{-(T-2)}|X_2'X_2|^{\frac{1}{2}} \exp[-\frac{1}{2\sigma^2}(\lambda - \hat{\lambda})'X_2'X_2(\lambda - \hat{\lambda})], \quad (20) \end{aligned}$$

where $X = (x_1\ x_2\ X_2)$, $\hat{\theta} = (\tilde{\pi}_1\ \tilde{\pi}_2)' = ((x_1\ x_2)'M_{X_2}\ (x_1\ x_2))^{-1}(x_1\ x_2)'M_{X_2}y$, $M_V = I_T - V(V'V)^{-1}V'$, $\hat{\lambda} = (X_2'X_2)^{-1}X_2'(y - (x_1\ x_2)(\theta - \hat{\theta})) - (\tilde{\pi}_3...\tilde{\pi}_T)'$. The pdfs (19) and (20) show that the ARMA(1,1) puts restrictions on the parameters which make the pdfs belong to case 2 discussed in the previous section. The pdf for θ, implied by the pdf for π and the restrictions imposed by the ARMA(1,1) model, then becomes,

$$p_{res}(\theta) \propto p(\lambda = 0|\theta, \sigma^2, X)p(\theta|\sigma^2, X). \quad (21)$$

The pdf of θ, (19), does allow to generate drawings from but it has a near singular covariance matrix as $X_2'X_2$ is near singular. By reducing the number of components in X_2 in (19) the near singularity will disappear. If we simultaneously reduce the number of components of X_2 in $\hat{\lambda}$ in an identical manner, there is no effect on $p_{res}(\theta)$. Note that the pdf of θ (21) can easily be transformed into the pdf of α, ρ. The conditional pdf of λ given θ evaluated in $\lambda = 0$ equals the weight function, $w(\theta^i) = p(\lambda = 0|\theta^i, \sigma^2, X)$, where θ^i is a drawing from $p(\theta|\sigma^2, X)$, to be used in an Importance, A-R or M-H sampling algorithm to draw the parameters which span the ARMA(1,1) parameter space. For more details on these algorithms, see section 4 and Kleibergen and Hoek (1996). From the drawn θ^i's, we can solve for α^i and ρ^i. Note that this approach importantly differs from the analysis of the parameters of ARMA models discussed in Chib and Greenberg (1994) and Monahan (1983) where ARMA models are not analyzed as restricted linear models, for more comments on this see Kleibergen and Hoek (1996).

3.2.2 ARMA(p,q)

The analysis discussed in the previous subsection for the ARMA(1,1) model (12) can be extended to an ARMA(p, q) model,

$$\rho(L)y_t = \alpha(L)\varepsilon_t, \quad (22)$$

where $\rho(L) = (1 - \sum_{i=1}^{p} \rho_i L^i)$, $\alpha(L) = (1 - \sum_{i=1}^{q} \alpha_i L^i)$. Again this model implies a restricted AR(∞) model like (13) with restrictions, see Fuller (1976) and Kleibergen and Hoek (1996),

$$g_i(\pi) = \pi_{p+q+i} - (\pi_{p+1} ... \pi_{p+q})\Pi_p^{-1} \begin{pmatrix} \pi_{p+q+i-1} \\ . \\ \pi_{p+i} \end{pmatrix} = 0, \qquad i = 1, ..., T-(p+q), \tag{23}$$

where

$$\begin{aligned}
\Pi_p &= \begin{pmatrix} \pi_p & . & . & . & \pi_{p+q-1} \\ . & & . & & . \\ . & & & . & . \\ . & & & . & . \\ \pi_{p-q+1} & . & . & . & \pi_p \end{pmatrix}, \\
\alpha' \Pi_p &= \begin{pmatrix} \pi_{p+1} & ... & \pi_{p+q} \end{pmatrix}, \\
\rho' &= \begin{pmatrix} \pi_1 & ... & \pi_p \end{pmatrix} - \alpha' \begin{pmatrix} 0 & \pi_1 & ... & \pi_p \\ . & . & . & . \\ . & & . & \pi_1 \\ 0 & . & . & 0 \end{pmatrix}, \\
\alpha &= \begin{pmatrix} \alpha_1 & ... & \alpha_q \end{pmatrix}', \ \rho = \begin{pmatrix} \rho_1 & ... & \rho_p \end{pmatrix}'.
\end{aligned}$$

Similar to the ARMA(1,1) example, we specify λ parameters to reflect the restrictions (23) and θ parameters to span the space in which the parameters can vary in a unrestricted way. Note again that for notational purposes we do not use the ARMA parameters explicitly to specify θ. The jacobian of the transformation from θ to (α, ρ) is $|\Pi_p|$.

$$\lambda_i = g_i(\pi), \qquad i = 1, ..., T - (p+q), \tag{24}$$

$$\theta_i = f_i(\pi) = \pi_i, \qquad i = 1, ..., p+q, \tag{25}$$

where $g_i(\pi)$ is defined in (23). The resulting pdf of (λ, θ) can be decomposed into a marginal pdf of θ and a conditional pdf of λ given θ,

$$\begin{aligned}
p(\theta|\sigma^2, X) &\propto \sigma^{-(p+q)} |(x_1 ... x_{p+q})' M_{X_{p+q}} (x_1 ... x_{p+q})|^{\frac{1}{2}} \qquad (26) \\
&\quad \exp[-\frac{1}{2\sigma^2}(\theta - \hat{\theta})'(x_1 ... x_{p+q})' M_{X_{p+q}} (x_1 ... x_{p+q})(\theta - \hat{\theta})], \\
p(\lambda|\theta, \sigma^2, X) &\propto \sigma^{-(T-p-q)} |X_{p+q}' X_{p+q}|^{\frac{1}{2}} \\
&\quad \exp[-\frac{1}{2\sigma^2}(\lambda - \hat{\lambda})' X_{p+q}' X_{p+q} (\lambda - \hat{\lambda})], \qquad (27)
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\pi}_i &= \theta_i, && i = 1, ..., p+q, \\
&= (\tilde{\pi}_{p+1} ... \tilde{\pi}_{p+q}) \tilde{\Pi}_p^{-1} \begin{pmatrix} \tilde{\pi}_{p+q+i-1} \\ . \\ \tilde{\pi}_{p+i} \end{pmatrix} && i = p+q+1, ..., T,
\end{aligned}$$

$X = (x_1 ... x_{p+q} \; X_{p+q})$, $\hat{\theta} = ((x_1 ... x_{p+q})' M_{X_{p+q}} (x_1 ... x_{p+q}))^{-1} (x_1 ... x_{p+q})' M_{X_{p+q}} y$, $\hat{\lambda} = (X'_{p+q} X_{p+q})^{-1} X'_{p+q} (y - (x_1 ... x_{p+q})(\theta - \hat{\theta})) - (\tilde{\pi}_{p+q+1} ... \tilde{\pi}_T)'$.

Similar arguments with respect to the dimensionality of X_{p+q} for (26) apply as in the ARMA(1,1) case. Also the ARMA(p, q) model belongs to case 2 discussed in the previous section such that the pdf for θ in the ARMA(p, q) model is proportional to the product of (26) and (27) evaluated in $\lambda = 0$. This conditional pdf of λ given θ evaluated in $\lambda = 0$ can then again be used as weight function in a sampling algorithm, see section 4. The α and ρ parameters are constructed from the resulting drawings of θ. Note that the matrix $X'_{p+q} X_{p+q}$ may be close to singularity. When we again reduce the number of elements in X_{p+q} in (26) and $\hat{\lambda}$ in an identical manner, the pdf $p_{res}(\theta)$ remains unaffected. In Kleibergen and Hoek (1996), we essentially removed all columns of X_{t+q} where it is allowed for, such that X_{t+q} does not appear in (26) and $\hat{\lambda}$.

Also other (nonlinear) restrictions can be imposed on the π parameters in (13) such as restrictions imposed by for example fractionally integrated models. A specification of θ and λ can then again be constructed to obtain the pdf of θ in the restricted probability space. Also mixtures of ARMA and fractionally integrated models or ARMA models with exogenous variables are possible. Essentially all these models belong to cases 2 or 3 discussed in the previous section.

3.2.3 Error Correction Cointegration Model

Cointegration imposes restrictions on the long run multiplier of Vector AutoRegressive (VAR) models, see Engle and Granger (1987) and Johansen (1991). This long run multiplier is more conveniently analyzed using an Error Correction Model (ECM) specification of a VAR model. Consider the ECM specification of a VAR(1),

$$\Delta x_t = \Pi' x_{t-1} + \varepsilon_t, \qquad t = 1, ..., T, \tag{28}$$

where $x_t, \varepsilon_t : k \times 1$ and $\Pi : k \times k$ is the long run multiplier. This model can be specified in matrix notation by,

$$\Delta X = X_{-1} \Pi + \varepsilon, \tag{29}$$

where $X = (x_1 ... x_T)'$, $\varepsilon = (\varepsilon_1 ... \varepsilon_t)'$. When we assume that the disturbances ε_t are i.i.d. normal with covariance matrix Ω, we can construct the conditional posterior of Π given Ω, which is a normal pdf, see Zellner (1971) (again also natural conjugate priors can be used as they lead to similar results),

$$p(\Pi|\Omega, X) \propto |\Omega|^{-\frac{1}{2}k} |X_{-1}'X_{-1}|^{\frac{1}{2}k} \exp[-\frac{1}{2} tr(\Omega^{-1}(\Pi - \hat{\Pi})' X_{-1}'X_{-1}(\Pi - \hat{\Pi}))], \tag{30}$$

where $\hat{\Pi} = (X_{-1}'X_{-1})^{-1}X_{-1}'\Delta X$.

Cointegration with r cointegrating vectors restricts the long run multiplier to have rank r, see Engle and Granger (1987),

$$g(\Pi) = rank(\Pi) - r = 0. \tag{31}$$

This rank condition can also be formulated by specifying the $k \times k$ long run multiplier as the product of two $k \times r$ matrices, see Engle and Granger (1987), Johansen (1991),

$$\begin{aligned} g(\Pi) &= \Pi - \beta\alpha \qquad (32) \\ &= \begin{pmatrix} \Pi_{11} - \alpha_1 & \Pi_{12} - \alpha_2 \\ \Pi_{21} + \beta_2\alpha_1 & \Pi_{22} + \beta_2\alpha_2 \end{pmatrix} = 0, \end{aligned}$$

where $\beta, \alpha' : k \times r$, $\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix}$, $\Pi = \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{pmatrix}$; $\Pi_{11}, \alpha_1 : r \times r$, $\Pi_{21}, \beta_2 : (k-r) \times r$, $\Pi_{12}, \alpha_2 : r \times (k-r)$, $\Pi_{22} : (k-r) \times (k-r)$. The cointegration vectors β are specified by $\beta = (I_r \; -\beta_2')'$ to obtain properly identified parameters. The dimension of the different matrices is such that we can specify the λ parameters spanning the restrictions (32) using the lower right block of $g(\Pi)$ in (32) and the θ parameters spanning the free parameter space as equal to α, β, see Kleibergen and van Dijk (1994a,b),

$$\begin{aligned} \lambda &= g(\Pi) = \Pi_{22} + \beta_2\alpha_2, \qquad (33) \\ \theta &= \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_2 \end{pmatrix} = f(\Pi) = \begin{pmatrix} \Pi_{11} \\ \Pi_{12} \\ -\Pi_{21}\Pi_{11}^{-1} \end{pmatrix}. \qquad (34) \end{aligned}$$

The jacobian of the transformation of Π to (λ, θ), see Kleibergen and van Dijk (1994a),

$$J(\lambda, \theta) = \begin{pmatrix} I_k \otimes I_r & 0 \\ -I_k \otimes \beta_2 & \begin{pmatrix} \alpha' & \begin{matrix} 0 \\ I_{k-r} \end{matrix} \end{pmatrix} \otimes I_{k-r} \end{pmatrix}, \tag{35}$$

can be combined with the normalizing constants of the normal pdf (30), $|\Omega|^{-\frac{1}{2}k}|X_{-1}'X_{-1}|^{\frac{1}{2}k}$, see Kleibergen and Paap (1996),

$$\begin{aligned} & |J(\lambda,\theta)||\Omega|^{-\frac{1}{2}k}|X_{-1}'X_{-1}|^{\frac{1}{2}k} \\ = & |J(\lambda,\theta)'(\Omega^{-1}\otimes X_{-1}'X_{-1})J(\lambda,\theta)|^{\frac{1}{2}} \\ = & |\Omega|^{-\frac{1}{2}r}|X_{1,-1}'M_{X_{2,-1}}X_{1,-1}|^{\frac{1}{2}k}|\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)}|X_{2,-1}'X_{2,-1}|^{\frac{1}{2}r} \\ & |\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}'|^{-\frac{1}{2}(k-r)}|X_{2,-1}'X_{2,-1}|^{\frac{1}{2}(k-r)}, \end{aligned} \tag{36}$$

where $X=(X_1\ X_2)$, $X_1: T\times r$, $X_2: T\times(k-r)$. The decomposition (36) is proved in the appendix. The different elements of (36) again form the normalizing constants of the conditional and marginal pdfs of (λ,θ) implied by (30). The only permissible decomposition of the joint pdf of (λ,θ) implied by (30) is, see Kleibergen and Paap (1996),

$$p(\lambda,\theta|\Omega,X)\propto p(\beta_2|\lambda,\alpha,\Omega,X)p(\lambda|\alpha,\Omega,X)p(\alpha|\Omega,X), \tag{37}$$

where $\theta=(\alpha,\beta_2)$, and

$$\begin{aligned} p(\alpha|\Omega,X) \propto & |\Omega|^{-\frac{1}{2}r}|X_{1,-1}'M_{X_{2,-1}}X_{1,-1}|^{\frac{1}{2}k} \\ & \exp[-\frac{1}{2}tr(\Omega^{-1}(\alpha-\hat{\alpha})'X_{1,-1}'M_{X_{2,-1}}X_{1,-1}(\alpha-\hat{\alpha}))], \quad (38) \\ p(\lambda|\alpha,\Omega,X) \propto & |\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}'|^{-\frac{1}{2}(k-r)} \\ & |X_{2,-1}'X_{2,-1}|^{\frac{1}{2}(k-r)}\exp[-\frac{1}{2}tr((\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix} \quad (39) \\ & \Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}')^{-1}(\lambda-\hat{\lambda})'X_{2,-1}'X_{2,-1}(\lambda-\hat{\lambda}))], \\ p(\beta_2|\lambda,\alpha,\Omega,X) \propto & |\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)}|X_{2,-1}'X_{2,-1}|^{\frac{1}{2}r} \quad (40) \\ & \exp[-\frac{1}{2}tr(X_{2,-1}'X_{2,-1}((\beta_2-\hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2-\hat{\beta}_2)')], \end{aligned}$$

and

$$\begin{aligned} \hat{\alpha} &= (X_{1,-1}'M_{X_{2,-1}}X_{1,-1})^{-1}X_{1,-1}'M_{X_{2,-1}}\Delta X, \\ \hat{\lambda} &= (X_{2,-1}'X_{2,-1})^{-1}X_{2,-1}'\Delta X\begin{pmatrix} -\alpha_1^{-1}\alpha_2 \\ I_{k-r} \end{pmatrix}, \\ \hat{\beta}_2 &= (\begin{pmatrix} 0 & \lambda \end{pmatrix}-(X_{2,-1}'X_{2,-1})^{-1}X_{2,-1}'(\Delta X-X_{1,-1}\alpha))\Omega^{-1}\alpha'\ (\alpha\Omega^{-1}\alpha')^{-1}. \end{aligned}$$

For a proof of the above decomposition, we refer to the appendix. The joint pdf (37) shows that the cointegration restrictions are such that the pdf of the parameters belongs to case 3 discussed in section 2 as we can integrate out part of θ, β_2, as these parameters are analyzed conditional on the parameters spanning the restrictions, λ. The λ parameters are again analyzed conditional on another part of θ, α. Note that the pdf $p(\theta, \lambda = 0)$ gives the matrix normal density for a random matrix with fixed reduced rank. The pdf of α on the space restricted by the cointegration restrictions (32),

$$p_{res}(\alpha|\Omega, X) \propto p(\lambda = 0|\alpha, \Omega, X)p(\alpha|\Omega, X), \tag{41}$$

shows that we can generate drawings from this density by generating α^i, $i = 1, ..., n$, from $p(\alpha|\Omega, X)$ and attaching a weight equal to $w(\alpha^i) = p(\lambda = 0|\alpha^i, \Omega, X)$ to each α^i. This weight is also attached to subsequent draws β_2^i from $p(\beta_2|\alpha^i, \lambda = 0, \Omega, X)$ (40). These weights can then be used in an Importance, A-R or M-H sampling algorithm. Note when we directly substitute $\Pi = \beta\alpha$, (32), in the pdf (30), as for example in Geweke (1996), the pdf under cointegration is not nested in the unrestricted pdf (30) since the pdf of the parameter spanning the cointegration restrictions, λ, is neglected. So, from a Bayesian perspective, the posterior of the parameters of the ECM under cointegration incorporate part of the conditional posterior of λ, a parameter which is not even present in the ECM under cointegration. As shown partly by (36), diffuse priors for the ECM under cointegration importantly differ from standard kind of diffuse priors, see also Kleibergen and Paap (1996). In Kleibergen and van Dijk (1994b), we used Jeffreys' priors on the parameters of the cointegration model which were derived using the information matrix of the cointegration model. The above shows that the Jeffreys' priors should be derived using the information matrix of the unrestricted model.

3.2.4 Simultaneous Equations Models

Another well known class of models, which impose nonlinear restrictions on the parameters of implied linear models, are the Simultaneous Equations Models (SEM). For an overview of these kind of models, see Hausman (1983). In a Bayesian setting, the pdfs of the parameters of these kind of models have been analyzed a.o. in Drèze (1976), Drèze and Morales (1976), and Drèze and Richard (1983). In matrix notation the SEM can be specified by,

$$YB + X\Gamma = \varepsilon, \tag{42}$$

where $Y : T \times m$ contains the endogenous variables, $X : T \times k$ contains the lagged endogenous and (weakly) exogenous variables, $\varepsilon : T \times m$, $B : m \times m$, $\Gamma : k \times m$, $\varepsilon \sim n(0, \Sigma \otimes I_T)$, and restrictions are imposed on B and Γ to make them estimable, see Hausman (1983). The reduced form of the SEM (42) then becomes,

$$Y = X\Pi + U, \tag{43}$$

where $U = \varepsilon B^{-1}$, $\Omega = B^{-1\prime}\Sigma B^{-1}$ and $\Pi = -\Gamma B^{-1}$ shows the restrictions imposed on the linear model (43). When a diffuse prior is used in a Bayesian setting (natural conjugate prior can be used as well), the posterior of Π given Ω is proportional to a normal pdf,

$$p(\Pi|\Omega, Y, X) \propto |\Omega|^{-\frac{1}{2}k}|X'X|^{\frac{1}{2}m}\exp[-\frac{1}{2}tr(\Omega^{-1}(\Pi - \hat{\Pi})'X'X(\Pi - \hat{\Pi}))], \quad (44)$$

where $\hat{\Pi} = (X'X)^{-1}X'Y$. The restrictions specified on Π are,

$$g(\Pi) = \Pi B + \Gamma = 0. \quad (45)$$

Depending on the form of the restrictions (45), a specification for the parameters λ, spanning these restrictions, and the parameters θ, spanning the space in which the parameters can freely vary, can be given. Here we only consider the single equation or limited information analysis, where only the structural equation of a single variable is analyzed and the other equations are in reduced form. For the full information case, we refer to Kleibergen and van Dijk (1996). The resulting analysis can be compared with the Bayesian limited information analysis in Drèze (1976) and Drèze and Richard (1983). We specify the structural equation as,

$$y_1 = Y_2\beta + X_1\gamma + \varepsilon_1, \quad (46)$$

where $Y_2 : T \times (m-1)$, $X_1 : T \times k_1$, $\beta : (m-1) \times 1$, $\gamma : k_1 \times 1$, and the $m-1$ reduced form equations as,

$$Y_2 = (X_1\ X_2\ X_3)\begin{pmatrix} \Pi_{12} \\ \Pi_{22} \\ \Pi_{32} \end{pmatrix} + U_2, \quad (47)$$

where $X_2 : T \times (m-1)$, $X_3 : T \times (k - k_1 - m + 1)$, $U_2 : T \times (m-1)$, $\Pi_{12} : k_1 \times (m-1)$, $\Pi_{22} : (m-1) \times (m-1)$, $\Pi_{32} : (k - k_1 - m + 1) \times (m-1)$, $\Pi_2 = (\Pi_{12}'\ \Pi_{22}'\ \Pi_{32}')'$. The parameters Π_{32} show the degree of overidentification, $k - k_1 - m + 1$, of the structural equation. The restrictions imposed by (46) on the parameters of the linear model (43) then read,

$$g(\Pi) = \begin{pmatrix} \pi_{11} & \Pi_{12} \\ \pi_{21} & \Pi_{22} \\ \pi_{31} & \Pi_{32} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -\beta & I_{m-1} \end{pmatrix} - \begin{pmatrix} \gamma & \Pi_{12} \\ 0 & \Pi_{22} \\ 0 & \Pi_{32} \end{pmatrix} = 0. \quad (48)$$

Similar to the cointegration model equation (33), the restrictions (48) can be spanned by a λ parameter,

$$\lambda = g(\Pi) = \pi_{31} - \Pi_{32}\beta, \quad (49)$$

and a θ parameter can be constructed to reflect the space in which the parameters can freely vary,

$$\theta = \begin{pmatrix} \gamma \\ \beta \\ \Pi_2 \end{pmatrix} = f(\Pi) = \begin{pmatrix} \pi_{11} - \Pi_{12}\beta \\ \Pi_{22}^{-1}\pi_{21} \\ \Pi_2 \end{pmatrix}. \tag{50}$$

The parameter $\lambda \neq 0$ implies an exact identified single equation model,

$$y_1 = Y_2\beta + X_1\gamma + X_3\lambda + \varepsilon_1, \tag{51}$$

which shows that we analyze an exact identified SEM conditional on $\lambda = 0$, which makes it an overidentified SEM. In a classical likelihood setting this implies that we can just condition on $\lambda = 0$ and analyze the likelihood with $\lambda = 0$. In Drèze (1976), a Bayesian limited information analysis is constructed by extending the classical approach of using the likelihood with $\lambda = 0$ and diffuse priors to form the posteriors. This implicitly assumes that the conditional posteriors of θ given λ can be constructed which is not possible for all elements of θ in (50). The resulting approach is inconsistent as the implied pdf of Π in the unrestricted model then no longer corresponds with (44). In the following, we show the Bayesian analysis where we analyze the pdf in $(\lambda = 0, \theta)$.

The jacobian of the transformation of $(\pi_{11}, \pi_{12}, \pi_{13}, \Pi_2)$ to (θ, λ) reads,

$$J(\theta,\lambda) = \begin{pmatrix} I_{k_1} & \Pi_{12} & 0 & \\ 0 & \Pi_{22} & 0 & \beta' \otimes I_{m-1} \\ 0 & \Pi_{32} & I_{k-k_1-m+1} & \\ 0 & 0 & 0 & I_{m-1} \otimes I_k \end{pmatrix}, \tag{52}$$

which can similar to (35) in the cointegration case be combined with the normalizing constants in (44) to obtain the normalizing constants of the conditional pdfs,

$$\begin{aligned} & |J(\theta,\lambda)||\Omega|^{-\frac{1}{2}k}|X'X|^{\frac{1}{2}m} \qquad (53) \\ = \;& |J(\theta,\lambda)'(\Omega^{-1} \otimes X'X)J(\theta,\lambda)|^{\frac{1}{2}} \\ = \;& |\Omega_{22}|^{-\frac{1}{2}k}|X'X|^{\frac{1}{2}(m-1)}|\omega_{11.2}|^{-\frac{1}{2}k} \\ & |\begin{pmatrix} I_{k_1} & \Pi_{12} & 0 \\ 0 & \Pi_{22} & 0 \\ 0 & \Pi_{32} & I_{k-k_1-m+1} \end{pmatrix}' X'X \begin{pmatrix} I_{k_1} & \Pi_{12} & 0 \\ 0 & \Pi_{22} & 0 \\ 0 & \Pi_{32} & I_{k-k_1-m+1} \end{pmatrix}|^{\frac{1}{2}} \\ = \;& |\Omega_{22}|^{-\frac{1}{2}k}|X'X|^{\frac{1}{2}(m-1)}|\omega_{11.2}|^{-\frac{1}{2}k}|X_3'M_{(X_1\,X\Pi_2)}X_3|^{\frac{1}{2}} \\ & |X_1'X_1|^{\frac{1}{2}}|\Pi_2'X'M_{X_1}X\Pi_2|^{\frac{1}{2}}. \end{aligned}$$

where $\omega_{11.2}=\omega_{11}-\omega_{12}\Omega_{22}^{-1}\omega_{21}$, $\Omega=\begin{pmatrix}\omega_{11} & \omega_{12}\\ \omega_{21} & \Omega_{22}\end{pmatrix}$.

The accompanying conditional pdfs of the parameters read,

$$\begin{aligned}
p(\lambda|\Pi_2,\Omega,Y,X) &\propto |\omega_{11.2}|^{-\frac{1}{2}(k-k_1-m+1)}|X_3'M_{(X_1\,X\Pi_2)}X_3|^{\frac{1}{2}} && (54)\\
&\quad \exp[-\frac{1}{2\omega_{11.2}}(\lambda-\hat{\lambda})'X_3'M_{(X_1\,X\Pi_2)}X_3(\lambda-\hat{\lambda})],\\
p(\beta|\lambda,\Pi_2,\Omega,Y,X) &\propto |\omega_{11.2}|^{-\frac{1}{2}(m-1)}|\Pi_2'X'M_{X_1}X\Pi_2|^{\frac{1}{2}} && (55)\\
&\quad \exp[-\frac{1}{2\omega_{11.2}}(\beta-\hat{\beta})'\Pi_2'X'M_{X_1}X\Pi_2(\beta-\hat{\beta})],\\
p(\gamma|\beta,\lambda,\Pi_2,\Omega,Y,X) &\propto |\omega_{11.2}|^{-\frac{1}{2}k_1}|X_1'X_1|^{\frac{1}{2}} && (56)\\
&\quad \exp[-\frac{1}{2\omega_{11.2}}(\gamma-\hat{\gamma})'X_1'X_1(\gamma-\hat{\gamma})],\\
p(\Pi_2|\Omega,Y,X) &\propto |\Omega_{22}|^{-\frac{1}{2}k}|X'X|^{\frac{1}{2}(m-1)}\\
&\quad \exp[-\frac{1}{2}tr(\Omega_{22}^{-1}(\Pi_2-\hat{\Pi}_2)'X'X(\Pi_2-\hat{\Pi}_2))], && (57)
\end{aligned}$$

where

$$\begin{aligned}
\hat{\gamma} &= (X_1'X_1)^{-1}X_1'(y_1-X\Pi_2\beta-X_3\lambda-(Y_2-X\Pi_2)\Omega_{22}^{-1}\omega_{21}),\\
\hat{\beta} &= (\Pi_2'X'M_{X_1}X\Pi_2)^{-1}\Pi_2'M_{X_1}X'(y_1-X_3\lambda-(Y_2-X\Pi_2)\Omega_{22}^{-1}\omega_{21}),\\
\hat{\lambda} &= (X_3'M_{(X_1\,X\Pi_2)}X_3)^{-1}X_3'M_{(X_1\,X\Pi_2)}(y_1-Y_2\Omega_{22}^{-1}\omega_{21}),\\
\hat{\Pi}_2 &= (X'X)^{-1}X'Y_2.
\end{aligned}$$

Note that we can construct $p(\gamma|\lambda,\Pi_2,\Omega,Y,X)$ and $p(\beta|\gamma,\lambda,\Pi_2,\Omega,Y,X)$ as well. Equations (54)-(56) show that the restrictions (49) make the pdf (44) belong to case 3 of section 2 as we can integrate out β and γ. The pdf for Π_2 on the restricted parameter space then becomes,

$$p_{res}(\Pi_2|\Omega,Y,X)\propto p(\lambda=0|\Pi_2,\Omega,Y,X)p(\Pi_2|\Omega,Y,X). \tag{58}$$

We can again simulate drawings of Π_2 from $p_{res}(\Pi_2|\Omega,Y,X)$ by attaching weights equal to $w(\Pi_2^i)=p(\lambda=0|\Pi_2^i,\Omega,Y,X)$ to drawings Π_2^i generated from $p(\Pi_2|\Omega,Y,X)$. β and γ can then be drawn conditional on Π_2^i. A variation of this simulation scheme is also used in the full information analysis, see Kleibergen and van Dijk (1996), where the conditionalization rules for constructing the conditional pdfs are more complicated. Note that the pdf of Π_2 in (58) differs from the pdf of Π_2 resulting from the Bayesian limited information analysis developed by Drèze (1976), which even has a pole at $\Pi_{22}=0$, see Kleibergen and van Dijk (1996).

4 Simulation Algorithms

To generate drawings from the pdfs of the, in the previous (sub)sections, discussed models, we can use both Markov Chain Monte Carlo (MC^2) as direct simulation procedures. Both procedures are briefly discussed in the following two subsections.

4.1 Metropolis-Hastings Sampling

The weights, $w(\theta^i)$, mentioned in the previous subsection can be used in a MC^2 procedure known as a Metropolis-Hastings (M-H) sampler, see Metropolis *et al.* (1953), Hastings (1970) and more recently Smith and Roberts (1993) and Tierney (1994), to generate drawings of the parameters from the associated pdfs. Since not all of the full conditional distributions of the involved parameters are of a standard type, standard Gibbs sampling is not possible.

The simulation scheme is based on the following decomposition of the pdf (5),

$$p_{res}(\theta) \propto w(\theta)p(\theta) \tag{59}$$

where $w(\theta) = p(\lambda = 0|\theta)$ and $p(\theta)$ is a density from which we can directly simulate. Since $w(\theta)$ is a bounded function, we also can use an Acceptance-Rejection (A-R) or Importance sampling algorithm which are discussed in the next subsection. In case A-R leads to large rejection frequencies, Chib and Greenberg (1995) show that a M-H algorithm can speed up the simulation process.

The *candidate-generating* density is $p(\theta)$. The acceptance-rejection step uses a ratio of weight functions $w(\theta)$. Given a drawing for θ, we generate a drawing of other parameters, for example β_2 in the error correction cointegration model where $\theta = \alpha$ or β, γ in the SEM where $\theta = \Pi_2$, conditional on θ from a normal pdf.

The steps to generate from the pdf with parameter restrictions using the M-H algorithm can be summarized as follows,

1. Draw θ^i from $p(\theta)$
2. Accept θ^i with probability
$$\min\left(\frac{p(\theta^i)w(\theta^i)p(\theta^{i-1})}{w(\theta^{i-1})p(\theta^{i-1})p(\theta^i)}, 1\right) = \min\left(\frac{w(\theta^i)}{w(\theta^{i-1})}, 1\right),$$
otherwise $\theta^i = \theta^{i-1}$.

The steps of this iterative scheme generate a Markov Chain. After the chain has converged, say after n iterations, the simulated values $[\theta^i, i > n]$ can be used as a sample from the pdf $p_{res}(\theta)$, see Tierney (1994) for details.

4.2 Importance and Acception-Rejection Sampling

The weight function, $w(\theta)$, from (59) can also be used in a Importance Sampling, see Kloek and van Dijk (1978) and Geweke (1989), and Acception-Rejection (A-R) algorithm.

Given n draws, θ^i, from $p(\theta)$ Importance Sampling approximates $E(g(\theta))$ by $\frac{\sum_{i=1}^{n} g(\theta^i)w(\theta^i)}{\sum_{i=1}^{n} w(\theta^i)}$. Geweke (1989) shows that central limit theorems can be used to calculate the quality of the approximations.

The A-R algorithm also uses drawings, θ^i, from $p(\theta^i)$, and accepts them with probability $\frac{1}{c}w(\theta^i)$ as drawings from $p_{res}(\theta)$, where $c = \max_\theta \, w(\theta)$.

5 Conclusions

The paper shows that many models, which are commonly used in econometrics and statistics, can be seen as linear models with parameter restrictions. As we know how a Bayesian analysis of a linear model proceeds, we perform a parameter transformation, of the parameters of the linear model towards parameters spanning the restrictions and parameters spanning the remaining space, to generalize the analysis of the linear model to allow for restrictions on the parameters and therefore to the large class of models which can be specified as such. So, by conditioning on the knowledge of how to analyze linear models, we construct procedures for analyzing models which can be specified as linear models with restrictions on the parameters. Many of these models impose restrictions which are such that the joint pdf of the parameters spanning the restrictions and the parameters spanning the remaining space, which results after the transformation of parameters, cannot be specified as the product of a conditional pdf of the parameters spanning the remaining space given the parameters spanning the restrictions and a marginal pdf of the latter parameters. This is implicitly assumed in many well known applications as shown in the examples, a.o. ARMA, cointegration and simultaneous equations models. The resulting analysis is in that case shown to be inconsistent as the resulting pdf is not nested within the assumed pdf of the parameters for the linear model in the unrestricted case.

In future work, we will extend the developed theory to allow for a.o. nonlinear models and derive posterior odds ratios/bayes factors for model comparison, for the latter see Kleibergen and Hoek (1996) and Kleibergen and Paap (1996).

Appendix

Proof of decomposition to obtain equation (36).

$$
\begin{aligned}
& |J(\lambda,\theta)||\Omega|^{-\frac{1}{2}k}|X_{-1}'X_{-1}|^{\frac{1}{2}k} \\
= \; & |J(\lambda,\theta)'(\Omega^{-1}\otimes X_{-1}'X_{-1})J(\lambda,\theta)|^{\frac{1}{2}} \\
= \; & |\Omega|^{-\frac{1}{2}r}|X_{1,-1}'M_{X_{2,-1}}X_{1,-1}|^{\frac{1}{2}k} \\
& |\begin{pmatrix} -(\alpha'\otimes I_{k-r}) & (\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}\otimes I_{k-r}) \end{pmatrix}'(\Omega^{-1}\otimes X_{2,-1}'X_{2,-1}) \\
& \begin{pmatrix} -(\alpha'\otimes I_{k-r}) & (\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}\otimes I_{k-r}) \end{pmatrix}|^{\frac{1}{2}} \\
= \; & |\Omega|^{-\frac{1}{2}r}|X_{1,-1}'M_{X_{2,-1}}X_{1,-1}|^{\frac{1}{2}k} \\
& |\begin{pmatrix} \alpha\Omega^{-1}\alpha'\otimes X_{2,-1}'X_{2,-1} & \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}'\Omega^{-1}\alpha'\otimes X_{2,-1}'X_{2,-1} \\ \alpha\Omega^{-1}\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}\otimes X_{2,-1}'X_{2,-1} & \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}'\Omega^{-1}\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}\otimes X_{2,-1}'X_{2,-1} \end{pmatrix}|^{\frac{1}{2}} \\
= \; & |\Omega|^{-\frac{1}{2}r}|X_{1,-1}'M_{X_{2,-1}}X_{1,-1}|^{\frac{1}{2}k}(|\alpha\Omega^{-1}\alpha'\otimes X_{2,-1}'X_{2,-1}| \\
& |(\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}'(\Omega^{-1}-\Omega^{-1}\alpha(\alpha\Omega^{-1}\alpha')^{-1}\alpha'\Omega^{-1})\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix})\otimes X_{2,-1}'X_{2,-1}|)^{\frac{1}{2}} \\
= \; & |\Omega|^{-\frac{1}{2}r}|X_{1,-1}'M_{X_{2,-1}}X_{1,-1}|^{\frac{1}{2}k}|\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)} \\
& |\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}'\alpha_{\perp}(\alpha_{\perp}\Omega\alpha_{\perp}')^{-1}\alpha_{\perp}'\begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}|^{\frac{1}{2}(k-r)}|X_{2,-1}'X_{2,-1}|^{\frac{1}{2}k} \\
= \; & |\Omega|^{-\frac{1}{2}r}|X_{1,-1}'M_{X_{2,-1}}X_{1,-1}|^{\frac{1}{2}k}|\alpha\Omega^{-1}\alpha'|^{\frac{1}{2}(k-r)} \\
& |\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}'|^{-\frac{1}{2}(k-r)}|X_{2,-1}'X_{2,-1}|^{\frac{1}{2}k}
\end{aligned}
$$

Proof of decomposition of the pdf (37).

$$
\begin{aligned}
& tr(\Omega^{-1}(\Pi-\hat{\Pi})'X_{-1}'X_{-1}(\Pi-\hat{\Pi})) \\
= \; & tr(\Omega^{-1}((\alpha-\hat{\alpha})'X_{1,-1}'M_{X_{2,-1}}X_{1,-1}(\alpha-\hat{\alpha}) \\
& +(\beta_2\alpha-(0\ \lambda)-\hat{\Pi}_2)'X_{2,-1}'X_{2,-1}(\beta_2\alpha-(0\ \lambda)-\hat{\Pi}_2)) \\
= \; & tr(\Omega^{-1}((\alpha-\hat{\alpha})'X_{1,-1}'M_{X_{2,-1}}X_{1,-1}(\alpha-\hat{\alpha})) \\
& +tr(X_{2,-1}'X_{2,-1}(\beta_2\alpha-(0\ \lambda)-\hat{\Pi}_2)\Omega^{-1}(\beta_2\alpha-(0\ \lambda)-\hat{\Pi}_2)') \\
= \; & tr(\Omega^{-1}(\alpha-\hat{\alpha})'X_{1,-1}'M_{X_{2,-1}}X_{1,-1}(\alpha-\hat{\alpha})) \\
& +tr(X_{2,-1}'X_{2,-1}((\beta_2-\hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2-\hat{\beta}_2)'+ \\
& +(0\ \lambda)-\hat{\Pi}_2)(\Omega^{-1}-\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1})((0\ \lambda)-\hat{\Pi}_2)')
\end{aligned}
$$

$$\begin{aligned}
&= tr(\Omega^{-1}(\alpha-\hat{\alpha})'X_{1,-1}'M_{X_{2,-1}}X_{1,-1}(\alpha-\hat{\alpha})) \\
&\quad +tr(X_{2,-1}'X_{2,-1}(\beta_2-\hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2-\hat{\beta}_2)') \\
&\quad +tr(X_{2,-1}'X_{2,-1}((0\ \lambda)-\hat{\Pi}_2)\alpha_{\perp}'(\alpha_{\perp}\Omega\alpha_{\perp}')^{-1}\alpha_{\perp}((0\ \lambda)-\hat{\Pi}_2)') \\
&= tr(\Omega^{-1}(\alpha-\hat{\alpha})'X_{1,-1}'M_{X_{2,-1}}X_{1,-1}(\alpha-\hat{\alpha})) \\
&\quad +tr(X_{2,-1}'X_{2,-1}(\beta_2-\hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2-\hat{\beta}_2)') \\
&\quad +tr((\alpha_{\perp}\Omega\alpha_{\perp}')^{-1}\alpha_{\perp}((0\ \lambda)-\hat{\Pi}_2)'X_{2,-1}'X_{2,-1}((0\ \lambda)-\hat{\Pi}_2)\alpha_{\perp}') \\
&= tr(\Omega^{-1}(\alpha-\hat{\alpha})'X_{1,-1}'M_{X_{2,-1}}X_{1,-1}(\alpha-\hat{\alpha})) \\
&\quad +tr(X_{2,-1}'X_{2,-1}(\beta_2-\hat{\beta}_2)\alpha\Omega^{-1}\alpha'(\beta_2-\hat{\beta}_2)') \\
&\quad +tr((\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix}\Omega\begin{pmatrix} -\alpha_2'\alpha_1^{-1\prime} & I_{k-r} \end{pmatrix})^{-1} \\
&\quad (\lambda-\hat{\lambda})'X_{2,-1}'X_{2,-1}(\lambda-\hat{\lambda}))
\end{aligned}$$

where $\hat{\Pi}_2 = (X_{2,-1}'X_{2,-1})^{-1}X_{2,-1}(\Delta X - X_{1,-1}\alpha)$.

References

[1] Chib, S. and E. Greenberg. Bayes Inference in Regression Models with ARMA(p,q) Errors. *Journal of Econometrics*, **64**:183–206, (1994).

[2] Chib, S. and E. Greenberg. Understanding the Metro-Polis Algorithm. *The American Statistician*, **49**:327–335, (1995).

[3] Drèze, J.H. Bayesian Limited Information Analysis of the Simultaneous Equations Model. *Econometrica*, **44**:1045–1075, (1976).

[4] Drèze, J.H. and J.A. Morales. Bayesian Full Iinformation Analysis of Simultaneous Equations. *Journal of the American Statistical Association*, **71**:329–354, (1976).

[5] Drèze, J.H. and R.F. Richard. Bayesian Analysis of Simultaneous Equations systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), (1983).

[6] Engle, R.F. and C.W.J. Granger. Co-integration and error correction : Representation, estimation and testing. *Econometrica*, **55**:251–276, (1987).

[7] Fuller, W.A. *Introduction to Statistical Time Series*. Wiley (New York), (1976).

[8] Geweke, J. Bayesian Inference in Econometric Models using Monte Carlo Integration. *Econometrica*, **57**:1317–1339, (1989).

[9] Geweke, J. Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, **75**:121–146, (1996).

[10] W.K. Hastings. Monte Carlo Sampling using Markov Chains and their Applications. *Biometrika*, **57**:97–10, (1970).

[11] Hausman, J.A. Specification and Estimation of Simultaneous Equations Systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), (1983).

[12] Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, **59**:1551–1580, (1991).

[13] Kleibergen, F. and H. Hoek. Bayesian Analysis of ARMA models using Noninformative Priors. (1996). working paper.

[14] Kleibergen, F. and H.K. van Dijk. Direct Cointegration Testing in Error Correction Models. *Journal of Econometrics*, **63**:61–103, (1994a).

[15] Kleibergen, F. and H.K. van Dijk. On the Shape of the Likelihood/Posterior in Cointegration Models. *Econometric Theory*, **10**:514–551, (1994b).

[16] Kleibergen, F. and H.K. van Dijk. Bayesian Simultaneous Equations Analysis using Reduced Rank Structures. (1996). working paper.

[17] Kleibergen, F. and R. Paap. Priors, Posterior Odds and Lagrange Multiplier Statistics in Bayesian Analyses of Cointegration. (1996). working paper.

[18] Kloek, T. and H.K. van Dijk. Bayesian Estimates of Equation System Parameters : An Application of Iintegration by Monte-Carlo. *Econometrica*, **44**:345–351, (1978).

[19] Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of State Calculations by fast Computing Machines. *Journal of Chemical Physics*, **21**:1087–1092, (1953).

[20] Monahan, J.F. Fully Bayesian Analysis of ARMA Time Series Models. *Journal of Econometrics*, **21**:307–331, (1983).

[21] Smith, A.F.M. and G.O. Roberts. Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society* **B**, **55**:3–23, (1993).

[22] Tierney, L. Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, **22**:1701–1762, (1994).

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