TI 2011-137/4
Tinbergen Institute Discussion Paper

Instrumental Variables, Errors in Variables, and Simultaneous Equations Models: Applicability and Limitations of Direct Monte Carlo

Arnold Zellner[a]
Tomohiro Ando[b]
Nalan Basturk[c;d]
Lennart Hoogerheide[e;f]
Herman K. van Dijk[c;e;f]

[a](posthumous) Booth School of Business, University of Chicago, USA;
[b]Graduate School of Business Administration, Keio University, Japan;
[c]Econometric Institute, Erasmus University Rotterdam, The Netherlands;
[d]The Rimini Centre for Economic Analysis, Rimini, Italy;
[e]VU University Amsterdam, The Netherlands;
[f]Tinbergen Institute, The Netherlands.

Instrumental Variables, Errors in Variables, and Simultaneous Equations Models: Applicability and Limitations of Direct Monte Carlo[1]

Arnold Zellner [a], Tomohiro Ando [b], Nalan Baştürk [c,d], Lennart Hoogerheide [e,f], and Herman K. van Dijk [c,e,f]

[a] *(posthumous) Booth School of Business, University of Chicago, USA.*

[b] *Graduate School of Business Administration, Keio University, Japan.*

[c] *Econometric Institute, Erasmus University Rotterdam, The Netherlands*

[d] *The Rimini Centre for Economic Analysis, Rimini, Italy*

[e] *Vrije Universiteit Amsterdam, The Netherlands*

[f] *Tinbergen Institute, The Netherlands*

September 27, 2011

Abstract

A Direct Monte Carlo (DMC) approach is introduced for posterior simulation in the Instrumental Variables (IV) model with one possibly endogenous regressor, multiple instruments and Gaussian errors under a flat prior. This DMC method can also be applied in an IV model (with one or multiple instruments) under an informative prior for the endogenous regressor's effect. This DMC approach can not be applied to more complex IV models or Simultaneous Equations Models with multiple endogenous regressors. An Approximate DMC (ADMC) approach is introduced that makes use of the proposed Hybrid Mixture Sampling (HMS) method, which facilitates Metropolis-Hastings (MH) or Importance Sampling from a proper marginal posterior density with highly non-elliptical shapes that tend to infinity for a point of singularity. After one has simulated from the irregularly shaped marginal distribution using the HMS method, one easily samples the other parameters from their conditional Student-t and Inverse-Wishart posteriors. An example illustrates the close approximation and high MH acceptance rate. While using a simple candidate distribution such as the Student-t may lead to an infinite variance of Importance Sampling weights. The choice between the IV model and a simple linear model under the restriction of exogeneity may be based on predictive likelihoods, for which the efficient simulation of all model parameters may be quite useful. In future work the ADMC approach may be extended to more extensive IV models such as IV with non-Gaussian errors, panel IV, or probit/logit IV.

[1] This paper started through intense, lively discussions between Arnold Zellner and Herman K. van Dijk in April 2010 when the latter was visiting Chicago.

1 Introduction

In many areas of economics and other sciences, sets of variables are often jointly generated with instantaneous feedback effects present. For instance, a fundamental feature of markets is that prices and quantities are jointly determined. The Simultaneous Equations Model (SEM), that incorporates instantaneous feedback relationships, was systematically analyzed in the nineteen forties and early nineteen fifties and documented in the well known Cowles Commission Monographs (Koopmans, 1950; Hood and Koopmans, 1950) and has been widely employed to analyze the behavior of markets, economies and other multivariate systems. For a survey, see e.g. Aliprantis, Barnett, Cornet, and Durlauf (2007) and the references given therein.

Full system analysis of the SEM is rather involved, see e.g. Bauwens and Van Dijk (1990); Van Dijk (2003). Instead, Zellner, Bauwens, and Van Dijk (1988) proceeded with a single equation analysis of the SEM that can be linked to the Instrumental Variable Regression (IV) analysis. A substantial literature on the issue of endogeneity, another expression for the immediate feedback mechanism, in IV models exists (see e.g. Angrist and Krueger (1991)). In this paper we make a connection between SEMs, the basic IV model and a simple errors-in-variables model (EV). These models focus on, respectively: immediate feedback mechanisms (SEM), on weak and strong instrumental variables (IV) and on correlation between errors in variables (EV). They possess a common statistical structure and they create therefore a common problem for inference: possible strong correlation between a right hand side variable in an equation and the disturbance of that equation. This may create nontrivial problems for simulation based Bayesian inference.

As workhorse model we take the IV model with one possibly endogenous regressor under a flat prior, and we make a distinction between the case of exact identification (a single instrumental variable) and the case of over-identification (more than one instrumental variables). We discuss the theoretical existence conditions for joint, conditional and marginal posterior distributions for the parameters of this model using a flat prior. The most relevant condition for empirical analysis is the well-known condition of non-singularity of the parameter matrix of instrumental variables. We emphasize that in the frequentist literature, parameters are constant and this condition refers to the fixed rank condition of a matrix. In the Bayesian approach the rank of this matrix is a random variable. For the case of exact identification or one instrumental variable and for the case of over-identification or many instruments, we analyze the existence of the joint posterior distribution. For the exactly identified model, application of any MC method is erroneous, because the posterior is improper. However, conditional distributions of each parameter exist and, if one is not aware of the non-existence of the joint posterior, one may apply Gibbs sampling erroneously.

A very attractive Monte Carlo method is Direct Monte Carlo (DMC) where one simulates directly from the posterior distributions. If this is possible, DMC is straightforward to apply and has as attractive property that the generated random drawings are independent, which greatly helps convergence and is convenient in case one aims to compute numerical standard errors or predictive likelihoods. The important issue is to determine whether the posterior or predictive distribution studied allows for DMC. In this paper we discuss that DMC is possible in the IV model with one possibly endogenous regressor, multiple instruments, and Gaussian errors under a flat prior.

In empirical econometrics there exist, however, many situations where the data information is weak in the sense of weak identifiability or weak instrumental variables, and strong endogeneity and to the lack of many available instruments. In these situations, it is common that the parameters have substantial mass of the likelihood, or the posterior under flat priors near the boundary of the parameter region. Examples of such data include nearly non-stationary processes or nearly non-identified processes such as inflation, interest rates, GDP processes or IV regression models with possibly weak instruments (De Pooter, Ravazzolo, Segers, and Van Dijk, 2008)). The important issue is the following: given that much data information may exist at or near the boundary of singularity, an empirical researcher may not want to exclude this information by a strong informative prior that focuses on the center of the parameter space and seriously down-weights or truncates relevant information near the boundary. In such a situation one faces a most important problem for empirical research, that is, the appearance of highly non-elliptical shapes of the posterior and predictive distributions. The Gibbs sampling method may then be very inefficient.

Although we show that a DMC method is possible in the IV model with one possibly endogenous regressor, multiple instruments and Gaussian errors, for more general models with multiple possibly endogenous regressors (such as the general SEM) this is not possible. We also present an Approximate DMC (ADMC) method to simulate from a marginal posterior density that exhibits both a an elliptical part and a singularity where the density tends to infinity. Extended or adapted versions of this ADMH approach may be useful for posterior simulation in IV models with multiple possibly regressors, in cointegration models or factor models.

For illustrative purposes, we present posterior shapes for a simulated data set and for an Incomplete Simultaneous Equations Model for Fulton fish market data.

The remainder of this paper is organized as follows. Section 2 presents the general SEM. Section 3 shows the connection between SEM, EV and IV models. Section 4 summarizes properties of the posterior densities of the IV model under flat priors. Section 5 presents the Direct Monte Carlo (DMC) method, its applicability and its limitations. Section 5 introduces the Approximate DMC (ADMC) method, and illustrates its flexility in a simple example. Section 7 presents a predictive likelihood approach to assess the degree of endogeneity in IV models, an application in which the efficient simulation of all parameters in the IV model may be quite useful. Section 8 presents an illustration using empirical data. Section 9 concludes.

2 Simultaneous Equations Model

We first review the results in Zellner, Bauwens, and Van Dijk (1988). Consider the following m equation SEM:

$$YB = X\Gamma + U, \tag{1}$$

where $Y = (y_1, ..., y_m)$ is a $T \times m$ matrix of observations on m endogenous variables, the $m \times m$ nonsingular matrix B is a matrix coefficient for the endogenous variables, $X = (x_1, \ldots, x_p)$ is a $T \times p$ matrix of observations on the p predetermined variables, the $p \times m$ matrix Γ is the coefficient matrix for the predetermined variables, and $U = (u_1, \ldots, u_m)$ is the $T \times m$ matrix of disturbances. Equation (1) shows the direct feedback mechanism between variables in the model. We assume that enough restrictions on B and Γ are made to have the model in (1) identified. Multiplying both sides of (1) by B^{-1}, the restricted reduced form equations are

$$Y = X\Pi_r + V_r, \tag{2}$$

where $\Pi_r = \Gamma B^{-1}$ is a $p \times m$ restricted reduced form coefficient matrix, and $V_r = UB^{-1} = (v_1, ..., v_m)$ is the restricted reduced form disturbance matrix. The corresponding unrestricted reduced form for the model in (1) is a multivariate regression model of the form $Y = X\Pi + V$ with no restrictions on matrix Π. The T rows of V, v_i $(i = 1, ..., T)$, are assumed to be independently drawn from a multivariate normal distribution with zero mean vector and $m \times m$ pds covariance matrix.

The problem is how to estimate the unknown parameters in the restricted model. Full-information analysis of this model is rather involved (Kleibergen and Van Dijk, 1998) and is outside the scope of this paper. A single identified equation of a SEM (involving possibly endogenous regressor(s) Y_1 and included instruments X_1, a subset of the instruments X) and the unrestricted reduced form equation for Y_1 are:

$$y_1 = Y_1\beta_1 + X_1\delta_1 + u_1, \tag{3}$$
$$Y_1 = X\Pi_1 + V_1, \tag{4}$$

where $\text{vec}(u_1,\ V_1) \sim N(0, \Omega \otimes I_T)$ and $\Omega = \begin{pmatrix} \omega_{11} & \omega_{12}' \\ \omega_{21} & \Omega_{22} \end{pmatrix}$ is a pds matrix, I_T is the identity matrix of size T and parameter Π_1 is unrestricted. Noting that the m-multivariate normal density of $(u_{1i}, v_{1i})'$, the ith row of $(u_1,\ V_1)$, can be expressed as a conditional normal density of u_{1i} given a value of v_{1i} and a marginal multivariate normal density of v_{1i}, Zellner, Bauwens, and Van Dijk (1988) derived $u_{1i}|v_{1i} \sim N(v_{1i}'\eta_1, \omega_{11} - \omega_{12}'\Omega_{22}^{-1}\omega_{21})$ with $\eta_1 = \Omega_{22}^{-1}\omega_{11}$ and $v_{1i} \sim N(0, \Omega_{22})$. One can obtain the orthogonal structural form:

$$y_1 = Y_1\beta_1 + X_1\delta_1 + V_1\eta_1 + \varepsilon_1, \tag{5}$$
$$Y_1 = X\Pi_1 + V_1, \tag{6}$$

where $X = (X_1, X_0)$ are the exogenous variables in (1) and $(\varepsilon_{1i}, v_i')'$ $i = 1, ..., T$ are independent random drawings from a multivariate normal distribution with mean zero and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & 0' \\ 0 & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \omega_{11} - \omega_{12}'\Omega_{22}^{-1}\omega_{21} & 0' \\ 0 & \Omega_{22} \end{pmatrix}.$$

The likelihood function is

$$L(Y|\beta_1, \delta_1, \eta_1, \Pi_1, \sigma_{11}, \Sigma_{22}, X) \propto |\Sigma_{22}|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{\Sigma_1^{-1}V_1'V_1\right\}\right] \times \sigma_{11}^{-\frac{T}{2}} \exp\left\{-\frac{\varepsilon_1'\varepsilon_1}{2\sigma_{11}}\right\}$$

with $Y = (y_1, Y_1)$.

Zellner, Bauwens, and Van Dijk (1988) used the following flat prior for the parameters, namely,

$$p(\beta_1, \delta_1, \Pi_1, \Omega) \propto |\Omega|^{-\frac{m+2+\nu_0}{2}}, \tag{7}$$

and the corresponding flat prior density for the parameters, $\{\beta_1, \delta_1, \Pi_1, \eta_1, \sigma_{11}, \Sigma_{22}\}$ is

$$p(\beta_1, \delta_1, \Pi_1, \eta_1, \sigma_{11}, \Sigma_{22}) \propto |\Sigma_{22}|^{-\frac{m+\nu_0}{2}} \times \sigma_{11}^{-\frac{m+2+\nu_0}{2}}.$$

This flat prior is similar to those employed in Kleibergen and Van Dijk (1998), and Kleibergen and Zivot (2003).

3 Basic EV and IV model structures

Relevant issues in the SEM can be illustrated with less complex models such as a basic IV model or an EV model. For a discussion we refer to Anderson (1976). We explore the issue of identification and non-regular posteriors in the simple model structure of an IV model and an EV model.

Consider a basic IV model with the following structural form:

$$y_i = x_i\beta + u_i, \tag{8}$$
$$x_i = \pi + v_i, \tag{9}$$

for $i = 1, \ldots, T$, with an exact identification and a constant instrument. For convenience, we changed the notation compared to (3) and (4): in (8) and (9) the (possibly) endogenous regressor is x (instead of Y_1). The zero-mean disturbances u_i and v_i are assumed to be independent and to have a bivariate normal distribution with a positive definite symmetric (pds) covariance matrix: $(u_i, v_i)' \sim \mathrm{NID}(0, \Omega)$. Unless Ω is a diagonal matrix, u_i and v_i are correlated and x_i is correlated with u_i.

Inserting (9) into (8), the so-called restricted reduced form (RRF) for the IV model is:

$$y_i = \pi\beta + \tilde{\varepsilon}_i, \tag{10}$$
$$x_i = \pi + v_i, \tag{11}$$

with $(\tilde{\varepsilon}_i, v_i)' \sim NID(0, (\begin{smallmatrix} 1 & \beta \\ 0 & 1 \end{smallmatrix})\Omega(\begin{smallmatrix} 1 & 0 \\ \beta & 1 \end{smallmatrix}))$. From the RRF representation in (10) and (11), it is clear that parameter β is not identified for $\pi = 0$, as β then disappears from the model. The issue of non-identification is the issue of *weak instruments* in IV estimation, where the strength of the instruments is based on the extent to which instruments can explain the endogenous variable. The extreme case, where the instruments are *irrelevant* corresponds to the non-identification, $\pi = 0$. See Van Dijk (2003) for the connection of these two concepts, and Kleibergen and Zivot (2003) for a summary of the problems associated with weak instruments.

The orthogonal structural form (OSF) for the IV model is obtained by decomposing u_i in (8) into two independent components $u_i = v_i\eta + \varepsilon_i$:

$$y_i = x_i\beta + v_i\eta + \varepsilon_i, \tag{12}$$
$$x_i = z_i\pi + v_i, \tag{13}$$

where $\eta = \omega_{12}/\omega_{22}$, $(\varepsilon_i, v_i)' \sim NID(0, \Sigma)$, $\Sigma = \left(\begin{smallmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{smallmatrix}\right)$, $\sigma_{11} = \omega_{11} - \omega_{12}^2/\omega_{22}$ and $\sigma_{22} = \omega_{22}$. By definition, $(\eta, \beta, \pi) \in \mathbb{R}^3$, $(\sigma_{11}, \sigma_{22}) \in \mathbb{R}_+^2$. We note that the OSF for this simplified IV model is similar to the decomposition in general SEM shown in (5) and (6). The issue of nonidentification can be seen from (12). For $\pi = 0$ we have $v_i = x_i$, so that the right hand side of (12) becomes $x_i(\beta + \eta) + \varepsilon_i$, hence parameters β and η are not jointly identified. Note that the IV model considered in (8) and (9) is a simple case of the SEM. The identification issue occurs in the general case of n-equation SEM model as well.

Furthermore, from the IV representation in (10) and (11), we obtain a simplified EV model by defining $\tilde{\eta} = \beta\pi$:

$$y_i = \tilde{\eta} + \varepsilon_i, \quad -\infty < \tilde{\eta} < \infty, \tag{14}$$
$$x_i = \pi + v_i, \quad -\infty < \pi < \infty, \tag{15}$$

where $(\varepsilon_i, v_i)' \sim \mathrm{NID}\left(0, \left(\begin{smallmatrix} 1 & \beta \\ 0 & 1 \end{smallmatrix}\right)\Omega\left(\begin{smallmatrix} 1 & 0 \\ \beta & 1 \end{smallmatrix}\right)\right)$.

Note that a usual EV model is more general than the model in (14) and (15). The restriction $\tilde{\eta} = \beta\pi$ may not necessarily hold in the general EV model, and the unobserved components are allowed to differ across observations, with parameters $\tilde{\eta}$ and π replaced by $\tilde{\eta}_i$ and δ_i, respectively. In this general case, a model has to be specified for these unobserved components. For expository purpose we take constant values for $\tilde{\eta}$ and π.

The EV model with the restriction $\tilde{\eta} = \beta\pi$ can be interpreted as a model that describes the permanent income hypothesis (see e.g. Friedman (1957); Attfield (1976) among others). Let y_i and x_i be measured consumption and income; $\tilde{\eta}$ and π be unobserved permanent components of consumption and income; and the disturbances in (14) and (15) be the *transitionary components* in income and consumption, respectively. Then $\beta = \tilde{\eta}/\pi$ is the ratio of permanent consumption to permanent income.

In the next section, we summarize and illustrate the issue of non-regular posteriors resulting from the identification problem in these models. For illustrative purposes, we consider the basic IV model as the example model.

4 Properties of posterior distributions for the IV model under flat priors

We discuss the local non-identification problem of the IV model under uninformative priors. Suppose a flat prior is proposed for the structural form parameters in (8) and (9):

$$p(\beta, \pi, \Omega) \propto |\Omega|^{-h/2} \text{ with } h > 0, \tag{16}$$

where the choice of the value of h may differ (see e.g. Drèze (1976) and Zellner (1971)). We choose the specification $h = 3$ that leads to a marginal posterior of (β, π) that is equal to the concentrated likelihood function for (β, π) (Bauwens and Van Dijk, 1990).

This flat prior on the structural form coefficients in (16) is not invariant to the change of variables leading to the RRF model in (10) and (11). Jeffreys' principle gives a prior for (β, π, Ω) that is proportional to $|\pi|$. Lancaster (2004) interprets this as a prior that assigns 0 probability density to the troublesome ridge $\pi = 0$, and argues that a possible objection to the use of Jeffreys' prior is that in many econometric applications an instrumental variable that has no regression on the included endogenous variable is all too probable, and to rule it out, dogmatically, a priori, may be unwise. Throughout this paper, we focus on the flat prior. However, we note that alternative priors such as the Jeffreys prior, can also be suitable for IV models, and for SEMs in general.

Define $y = (y_1, \ldots, y_T)'$, $x = (x_1, \ldots, x_T)'$ and ι is a $T \times 1$ vector of ones. The likelihood of the model in (8) and (9) is:

$$p(y, x \mid \beta, \pi, \Omega) \propto |\Omega|^{-T/2} \exp\left\{-\frac{1}{2}\text{tr}\left(\Big(y - x\beta,\, x - \iota\pi\Big)'\Big(y - x\beta,\, x - \iota\pi\Big)\,\Omega^{-1}\right)\right\}. \tag{17}$$

We are interested in the shape of the likelihood in (17) in the parameter space, and the shapes of the posteriors under flat priors.

4.1 Improperness of the posterior densities under flat priors

Combining the prior in (16) and the likelihood in (17) with $h = 3$, a kernel of the joint posterior is:

$$p(\beta, \pi, \Omega \mid y, x) \propto |\Omega|^{-(T+3)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left(y - x\beta,\, x - \iota\pi\right)'\left(y - x\beta,\, x - \iota\pi\right)\Omega^{-1}\right)\right\}. \tag{18}$$

The posterior density in (18) is improper for the IV model with exact identification, that is, with $k = 1$ instrument (see the Appendix for a discussion). For convenience, we illustrate the improperness of this density focusing on the marginal posterior $p(\beta, \pi)$. If $T \geq 2$ and given that $(y - x\beta, \;\; x - \iota\pi)'(y - x\beta, \;\; x - \iota\pi)$ is a pds matrix for all values of (β, π) in the parameter space, a kernel for the marginal posterior density of (β, π) is (Zellner, 1971):

$$p(\beta, \pi \mid y, x) \propto \left|\left(y - x\beta,\, x - \iota\pi\right)'\left(y - x\beta,\, x - \iota\pi\right)\right|^{-T/2}. \tag{19}$$

The marginal posterior density in (19) has a ridge along the line $\pi = 0$, since the right-hand-side of (19) is constant with value $(x'xy'M_x y)^{-T/2}$:

$$p(\beta, \pi \mid y, x, \pi = 0) \propto \left|\left(y - x\beta,\, x\right)'\left(y - x\beta,\, x\right)\right|^{-T/2} = (x'xy'M_x y)^{-T/2}, \tag{20}$$

where we use the determinant decomposition rule, and M_α is the projection matrix outside the span of α.

For the exactly identified IV model summarized in this section, it can be shown that this ridge of the posterior leads to an improper posterior density. For over-identified models, however, the joint posterior is a proper density despite this ridge. This issue will become more clear in the following subsections.

We finally note that also for the exactly identified case with $k = 1$ instrument the conditional densities of β, π, Ω are proper densities for the whole parameter space $(\beta, \pi) \in \mathbb{R}^2$. The improperness of the joint posterior is shown in the Appendix. See also De Pooter, Ravazzolo, Segers, and Van Dijk (2008) for a simple illustration of the improperness of this posterior and Hobert and Casella (1998) for an illustration of how the Gibbs sampler can be employed erroneously on models with proper conditionals and improper joint posterior.

4.2 Marginal posterior densities of β and π

For the basic IV model in (10) and (11) using the prior in (16) with $h = 3$, the marginal density kernel of β is:

$$p(\beta \mid y, x) \propto \left(\frac{(y - x\beta)'(y - x\beta)}{(\tilde{y} - \tilde{x}\beta)'(\tilde{y} - \tilde{x}\beta)} \right)^{-(T-1)/2} \left((\tilde{y} - \tilde{x}\beta)'(\tilde{y} - \tilde{x}\beta) \right)^{-1/2}, \tag{21}$$

where $\tilde{y}$ and $\tilde{x}$ are demeaned data y and x, respectively. This kernel does not correspond to a proper density, since the tails are too fat due to the second factor that decreases at the too slow rate $|\beta|^{-1}$ as $|\beta|$ increases.

For our basic IV model, the marginal density of π is:

$$p(\pi \mid y, x) \propto \left[(x - \iota\pi)'(x - \iota\pi) \right]^{-(T-1)/2} / |\pi|, \tag{22}$$

where ι is the $T \times 1$ vector of ones. Due to the factor $|\pi|^{-1}$, the marginal density in (22) has a non-integrable asymptote at $\pi = 0$, so that the kernel does not correspond to a proper density. These results hold for the IV model with a single instrument, regardless of the strength of the instrument and the level of endogeneity in the data.

For the IV model with k instruments z_i (in $T \times k$ matrix z)

$$y_i = x_i\beta + u_i, \tag{23}$$
$$x_i = z_i\pi + v_i, \tag{24}$$

the marginal posterior density of β is given in Drèze (1976) and Drèze (1977) as

$$p(\beta \mid y, x, z) \propto \left(\frac{(y - x\beta)'(y - x\beta)}{(y - x\beta)' M_z (y - x\beta)} \right)^{-(T-1)/2} \left((y - x\beta)' M_z (y - x\beta) \right)^{-k/2}, \tag{25}$$

also see the Appendix. For $k \geq 2$ this kernel corresponds to a proper density, since the second factor (a kernel of a t-density with $k - 1$ degrees of freedom) decreases at the fast enough rate $|\beta|^{-k}$ as $|\beta|$ increases (whereas the first factor is smaller than 1). That is, the tails of the marginal posterior of β become thinner for larger number k of instruments, regardless of the explanatory power of the instruments.

Kleibergen and Van Dijk (1994, 1998) derive the marginal density for this IV model (see Hoogerheide et al. (2007) for an expository analysis of this issue):

$$p(\pi \mid y, x, z) \propto ((x - z\pi)'(x - z\pi))^{-\frac{T-1}{2}} (\pi' z' M_x z\pi)^{-\frac{1}{2}} \left(\frac{\pi' z' M_x z\pi}{\pi' z' M_{(y\ x)} z\pi} \right)^{\frac{T-1}{2}}. \tag{26}$$

For $k \geq 2$ this kernel corresponds to a proper density. In the appendix it is discussed that the integrability of (26) amounts to the integrability of

$$\int_{\{\pi^* | \pi^{*\prime}\pi^* \leq 1\}} (\pi^{*\prime}\pi^*)^{-1/2} d\pi^*. \tag{27}$$

For $k = 2$ we have

$$\int_{\{\pi^* | \pi^{*\prime}\pi^* \leq 1\}} (\pi^{*\prime}\pi^*)^{-1/2} d\pi^* = \pi + \int_1^\infty \pi \frac{1}{f^2}\, df = \pi + \left[-\pi\frac{1}{f}\right]_1^\infty = \pi + (0 - (-\pi)) = 2\pi, \tag{28}$$

with $\pi = 3.14159\ldots$ (as opposed to the $k \times 1$ vector of coefficients at the instruments throughout the text) on the right hand side. Here we used that the volume under the graph of $f = (\pi^{*\prime}\pi^*)^{-1/2}$ at $\{\pi^* |\, \pi^{*\prime}\pi^* \leq 1\}$ can be computed by integrating the surfaces $\pi\frac{1}{f^2}$ of circles with radius $\frac{1}{f}$ for $1 \leq f < \infty$ and the surfaces π of circles with radius 1 for $0 \leq f < 1$. Figure 1 illustrates this: for each function value $f = (\pi^{*\prime}\pi^*)^{-1/2}$ with $f \geq 1$ the horizontal 'slice' through the graph is a circle with radius $1/f$. For $k \geq 3$ a similar derivation involving an integral over k-dimensional balls yields a different *finite* value.

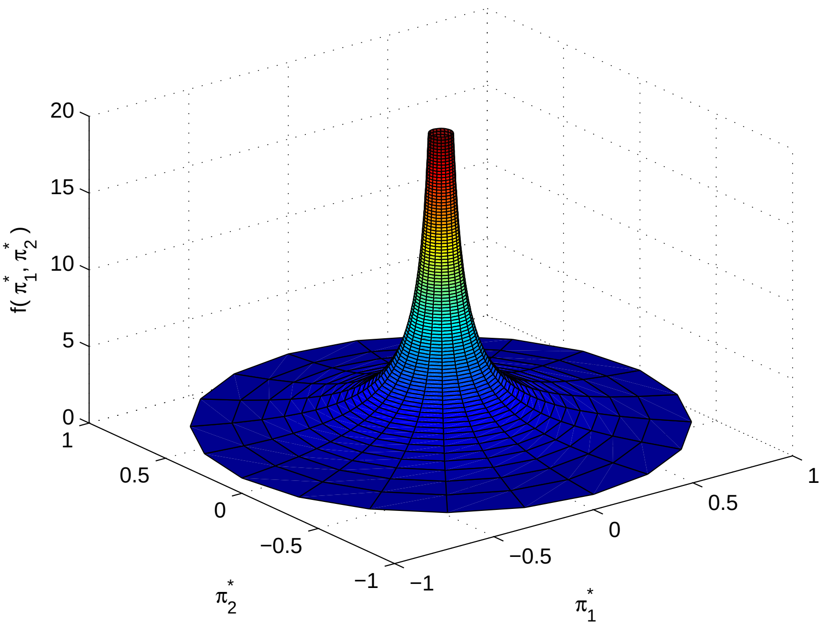


Figure 1. $f(\pi_1^*, \pi_2^*) = \left((\pi_1^*)^2 + (\pi_2^*)^2\right)^{-1/2}$ *at* $\{\pi^* |\, \pi^{*\prime}\pi^* \leq 1\}$.

We note that a special case in the above models is the case of *exogeneity*, that is, when Ω is diagonal. The local non-identification problem for $\pi = 0$ disappears if we a priori impose this exogeneity assumption. The analysis of exogeneity is discussed in Section 7.

5 Direct Monte Carlo: applicability and limitations

In this paper we aim to simulate from our IV model by a Direct Monte Carlo (DMC) method. Obvious advantages of DMC are that the method is straightforward and that the drawings are independent, which helps quick convergence and is convenient in case one desires to compute Numerical Standard Errors (NSEs) or predictive likelihoods. Furthermore, even if one desires to make use of an alternative method such as the Gibbs sampler, the use of both (fundamentally different) methods is arguably one of the best ways to check the results – and thereby the derivations, code and convergence of both simulation methods. First of all, we assume that we have $k \geq 2$ instruments, since for $k = 1$ instrument the improperness of the posterior implies that any simulation method necessarily provides erroneous results (if any).

The orthogonal structural form (OSF) for the IV model is:

$$y_i = x_i\beta + v_i\eta + \varepsilon_i, \tag{29}$$

$$x_i = z_i\pi + v_i, \tag{30}$$

where $\eta = \omega_{12}/\omega_{22}$, $(\varepsilon_i, v_i)' \sim NID(0, \Sigma)$, $\Sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$, $\sigma_{11} = \omega_{11} - \omega_{12}^2/\omega_{22}$ and $\sigma_{22} = \omega_{22}$. From the OSF it may seem as if we can decompose the posterior

$$\begin{aligned} p(\beta, \eta, \pi, \sigma_{11}, \sigma_{22} \mid y, x, z) =& p_1(\beta, \eta \mid \pi, \sigma_{11}, y, x, z) \times p_2(\sigma_{11}|\pi, y, x, z) \\ & \times p_3(\pi|\sigma_{22}, y, x, z) \times p_4(\sigma_{22}|y, x, z), \end{aligned} \tag{31}$$

where $p_1(\beta, \eta|\sigma_{11}, \pi, y, x, z)$ and $p_3(\pi|\sigma_{22}, y, x, z)$ are multivariate normal densities, and $p_2(\sigma_{11}|\pi, y, x, z)$ and $p_4(\sigma_{22}|y, x, z)$ are inverted gamma and Inverse-Wishart densities, respectively. However, one must note that (given the data x,z) the term $v_i = x_i - z_i\pi$ in (29) is a function of π. Therefore, the marginal posterior of π in (29)-(30) is not simply the marginal Student-t posterior (or conditional normal posterior) in the model (30), as already stressed in the previous sections. Therefore it is not possible to obtain posterior drawings using a Direct Monte Carlo (DMC) approach by simulating from 'standard' marginal and conditional distributions p_1 to p_4.

However, in this simple IV model with one possibly endogenous regressor x_i it is possible to obtain posterior drawings by a different DMC approach. For the marginal posterior of β is a 1-dimensional distribution from which one can directly simulate using a (numerical) inverse CDF method. Therefore, one can apply the following approach:

DMC approach in IV model (23)-(24) under flat prior with $k \geq 2$ instruments:

Step 1: Draw β from its marginal posterior, using the numerical inverse CDF method in two sub-steps. First, use the numerical inverse CDF method to simulate $\beta^* = \Psi(\beta)$

where Ψ is the CDF (with pdf ψ) of the Student-t distribution with mode the 2SLS estimator $\hat{\beta}_{2SLS}$, with scale the variance of $\hat{\beta}_{2SLS}$ times a multiplication factor (e.g. 4), and lower degrees of freedom than the marginal posterior of β in (25). For example, $k-2$ degrees of freedom for $k \geq 3$ instruments, and 1/2 degree of freedom for $k=2$ instruments. The pdf of β^* is

$$p(\beta^* \mid y, x, z) \propto \frac{(y - x\Psi^{-1}(\beta^*))' \, (y - x\Psi^{-1}(\beta^*))^{-(T-1)/2}}{(y - x\Psi^{-1}(\beta^*))' \, M_z \, (y - x\Psi^{-1}(\beta^*))^{-(T-k-1)/2}} \times \frac{1}{\psi\left(\Psi^{-1}(\beta^*)\right)}, \tag{32}$$

with β^* in [0,1], so that the use of a very fine grid on [0,1] yields drawings from the distribution of β^*.[1] Second, transform $\beta = \Psi^{-1}(\beta^*)$.

Step 2: Draw π conditionally on β from its conditional posterior, a k-dimensional Student-t distribution with mode $\hat{\pi} = (z'M_u z)^{-1} z'M_u x$, scale matrix $s^2_{\hat{\pi}}\,(z'M_u z)^{-1}$ and $T-k$ degrees of freedom, with $u = y - x\beta$, $(T-k)s^2_{\hat{\pi}} = (x - z\pi)'M_u(x - z\pi)$.

Step 3: Draw Ω conditionally on (β, π) from its conditional posterior, an Inverse-Wishart distribution with parameters $(u\ v)'(u\ v)$ and T degrees of freedom, where $u = y - x\beta$, $v = x - z\pi$. That is, take the inverse of a draw from a Wishart distribution with mean $\left[\frac{1}{T}(u\ v)'(u\ v)\right]^{-1}$ and T degrees of freedom.

If one is only interested in β, then one can obviously merely use step 1, or use a deterministic integration (quadrature) method like like the extrapolated or adaptive Simpson's method or Gaussian quadrature. However, one may often be interested in the strength of the instruments (given by π), the uncertainty on y (or x) for individual observations (given by Ω), or one may wish to investigate whether there is evidence of endogeneity (inspecting Ω, typically $\rho \equiv \omega_{12}/\sqrt{\omega_{11}\,\omega_{22}}$).

It should be noted that the DMC method can also be used if one specifies a different prior of the form

$$p(\beta, \pi, \Omega) \propto p(\beta) \times \Omega^{h/2} \qquad \text{with } h = 3, \tag{33}$$

for example with a normal pdf $p(\beta)$. The only difference is that (32) must be multiplied by the factor $p(\beta) = p(\Psi^{-1}(\beta^*))$. If one specifies an informative normal prior $p(\beta)$ or a uniform prior $p(\beta)$ at a bounded interval, the posterior is also proper for $k=1$, so that DMC is then applicable for any number of instruments $k \geq 1$.

Further, the DMC method can also be applied if there are *included instruments* or

[1] The exact distribution Ψ is not important, it only matters that (i) the range $[0,1]$ of β^* is finite, so that we do not need to truncate the range when choosing a grid for the numerical inverse CDF method; (ii) the pdf of β does not tend to ∞ for β tending to 0 or 1. For the latter it is required that ψ is a more 'wide' distribution with fatter tails than the marginal posterior of β.

control variables w (known as X_1 in the aforementioned INSEM) in both equations. First, x,y and z are transformed to become residuals after regression on w: $M_w x$,$M_w y$ and $M_w z$. This is equivalent with integrating out the coefficients at w under a flat prior. Second, one applies the DMC method. Third, if one is interested in the coefficients at w, then one simulates these by making use of the matricvariate normal conditional posterior of the coefficients in the model with regressands $(M_w y - M_w x\beta)$ and $(M_w x - M_w z\pi)$, regressors w for both regressands, and (known) error covariance matrix Ω.

Finally, note that this DMC method is not applicable in an IV model with multiple possibly endogenous regressors (nor in the general SEM model), since we require a 1-dimensional β for the inverse CDF method.

6 Approximate Direct Monte Carlo: applicability and limitations

The posterior of (β, π, Ω) can be decomposed as

$$p(\beta, \pi, \Omega \mid y, x, z) = p(\pi \mid y, x, z) \times p(\beta \mid \pi, y, x, z) \times p(\Omega \mid \beta, \pi, y, x, z), \tag{34}$$

where $p(\pi \mid y, x, z)$ is given by the non-standard distribution in (26), $p(\beta \mid \pi, y, x, z)$ is a Student-t distribution with mode $\hat{\beta} = (x' M_v x)^{-1}(x' M_v y)$, scale $s^2_{\hat{\beta}}(x' M_v x)^{-1}$ and $(T-1)$ degrees of freedom, where $v = x - z\pi$, $(T-1)s^2_{\hat{\beta}} = (y - x\hat{\beta})' M_v (y - x\hat{\beta})$; $p(\Omega \mid \beta, \pi, y, x, z)$ is an Inverse-Wishart distribution with parameters $(u\ v)'(u\ v)$ and T degrees of freedom, where $u = y - x\beta$, $v = x - z\pi$. Hence, if one can simulate from $p(\pi \mid y, x, z)$, then draws from β and Ω are easily simulated from their conditional posteriors.

One may think that one can simulate from $p(\pi \mid y, x, z)$ by Importance Sampling (IS) – or the independence chain Metropolis-Hastings (MH) algorithm – with candidate density $q(\pi)$ equal to the Student-t posterior of π in the first stage regression (24). However, in this case the variance of the IS weights $W = p(\pi \mid x, y, z)/q(\pi)$ may not be finite. For the case with $k = 2$ instruments we have

$$E[W^2] = \int \frac{(p(\pi \mid x, y, z))^2}{q(\pi)}\, d\pi = \infty,$$

since for $\pi \to 0$ the numerator $(p(\pi \mid x, y, z))^2$ tends to ∞ too quickly (due to the factor $(\pi' z M_x z\pi)^{-1}$), whereas the denominator $q(\pi)$ is bounded from above. This is easily seen from the fact that for $\pi^* \equiv (z M_x z)^{1/2}\pi$ we have

$$\int_{\{\pi^* \mid \pi^{*\prime}\pi^* \leq 1\}} (\pi^{*\prime}\pi^*)^{-1} d\pi^* = \pi + \int_1^\infty \pi\, \frac{1}{f}\, df = \pi + [\pi \log f]_1^\infty = \pi + (\infty - 0) = \infty, \tag{35}$$

with $\pi = 3.14159\ldots$ (as opposed to the $k \times 1$ vector of coefficients at the instruments throughout the text), where we used that the volume under the graph of $f = (\pi^{*\prime}\pi^*)^{-1}$ at $\{\pi^* \mid \pi^{*\prime}\pi^* \leq 1\}$ can be computed by integrating the surfaces $\pi\frac{1}{f}$ of circles with radius $\frac{1}{\sqrt{f}}$ for $1 \leq f < \infty$ and the surfaces π of circles with radius 1 for $0 \leq f < 1$. This means that for any candidate density $q(\pi)$ that does not tend to ∞ for $\pi \to 0$ — also the mixture of Student-t densities of Hoogerheide et al. (2007) or Hoogerheide et al. (2011) — the IS weights have infinite variance.

Therefore we propose the following candidate pdf $q(\pi)$ for approximating the shapes of $p(\pi \mid y, x, z)$, a hybrid mixture of two components, a $t^k_{0,1-k}$ density and a Student-t density:

$$q(\pi) = w_{t^k_{0,1-k}}\, p_{t^k_{0,1-k}}(\pi|A) + (1 - w_{t^k_{0,1-k}})\, p_t(\pi \mid \mu, \Sigma, \nu)$$

with mixing weight $w_{t^k_{0,1-k}}$ in $[0,1]$, Student-t pdf

$$p_t(\theta|\mu, \Sigma, \nu) \propto |\Sigma|^{-1/2} \left(1 + \frac{(\theta-\mu)'\Sigma^{-1}(\theta-\mu)}{\nu}\right)^{-(k+\nu)/2} \tag{36}$$

with positive definite symmetric (pds) Σ, $\nu \geq 1$; and with the density of the k-dimensional $t^k_{0,1-k}$ distribution defined as

$$p_{t^k_{0,1-k}}(\theta|A) \equiv \begin{cases} |A|^{-1/2} \left(\frac{k}{k-1}\, \frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)}\right)^{-1} (\theta' A^{-1} \theta)^{-1/2} & \text{for } (\theta' A^{-1}\theta)^{1/2} \leq 1 \\ \\ 0 & \text{for } (\theta' A^{-1}\theta)^{1/2} > 1 \end{cases}. \tag{37}$$

That is, the $t^k_{0,1-k}$ distribution has one parameter A, a positive definite symmetric (pds) $k \times k$ matrix. We denote this distribution the $t^k_{0,1-k}$ distribution, since it is obtained by letting $\nu \downarrow 0$ in $\frac{(\theta-\mu)'\Sigma^{-1}(\theta-\mu)}{\nu}$ and substituting $\nu = 1-k$ into the exponent $-(k+\nu)/2$ in the pdf (36) of the k-dimensional Student-t distribution (and by taking $\mu = 0$, which can be relaxed to allow for an asymptote around a different value than $\theta = 0$). For $k = 1$ the kernel in (37) would correspond to an (improper) density kernel of a Student-t distribution with 0 degrees of freedom. [2]

Simulating θ from (37) is done by simulating $\tilde{\theta}$ from (37) with $A = I_k$ and taking $\theta = A^{1/2}\tilde{\theta}$. For simulating $\tilde{\theta}$ we first sample $F = (\tilde{\theta}'\tilde{\theta})^{-1/2}$ with cumulative distribution function

$$CDF_F(x) = \Pr[F \leq x] = 1 - \Pr[F \geq x] = 1 - F^{-(k-1)}$$

for $x \geq 1$; 0 else. So, we simulate $U \sim UNIF(0,1)$ and compute

$$F = (1-U)^{-1/(k-1)}.$$

Second, we simulate $\tilde{\theta}$ uniformly from the set $\{\tilde{\theta}|(\tilde{\theta}'\tilde{\theta})^{1/2} = 1/F\}$. This is done by simulating $\theta^* \sim N(0, I_2)$ and taking $\tilde{\theta} = \theta^*(\theta^{*\prime}\theta^*)^{-1/2}\ (1/F)$. For $k = 2$ we have $1/F = (1-U)$ so that the norm of $\tilde{\theta}$ is simulated uniformly between 0 and 1.

[2] Proofs of the scaling constant and moments of the pdf of θ, and of the CDF of $(\theta'\theta)^{-1/2}$, are easily derived making use of the k-volume V_k of the k-dimensional ball with radius 1, $V_k = \frac{\pi^{k/2}}{\Gamma(\frac{k}{2}+1)}$.

The mean and covariance matrix of θ with pdf in (37) are given by:

$$E[\theta] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \text{cov}(\theta) = \frac{k-1}{k(k+1)}\, A.$$

For $k = 2$ and $A = I_2$ the graph of the $t^k_{0,1-k}$ pdf is proportional to the graph of $f(\pi_1^*, \pi_2^*) = \left((\pi_1^*)^2 + (\pi_2^*)^2\right)^{-1/2}$ at $\{\pi^* | \, \pi^{*\prime}\pi^* \leq 1\}$ in Figure 1.

We propose the following approach for posterior simulation from $p(\pi \mid y, x, z)$:

Hybrid Mixture Sampling (HMS) for posterior simulation from $p(\pi \mid y, x, z)$ in IV model (23)-(24) under flat prior with $k \geq 2$ instruments:

Step 1: Our initial choices for the candidate's parameters are as follows:

- $\mu = \hat{\pi}_{OLS} = (z'z)^{-1}z'x$ and $\Sigma = \text{cov}(\hat{\pi}_{OLS}) = s^2(z'z)^{-1}$ with $s^2 = e'e/(T-k)$, $e = x - z\hat{\pi}_{OLS}$. The OLS estimator in (24) and its covariance matrix provide a logical first approximation of the location and scale of the 'regular part' of the posterior distribution (as opposed to the 'asymptote part' for $\pi \approx 0$);

- $\nu = 4$; i.e., low degrees of freedom to ensure that no relevant parts of the parameter space are missed;

- $w_{t^k_{0,1-k}} = 0.1$. As a first approximation we assume that the main part of the posterior is the 'regular part'. Otherwise the instruments may have so little power that it is arguably unwise to use the IV model in the first space;

- $A = c\; z'M_x z$ with scalar $c > 0$ chosen such that the minimum π^*_{min} of $p(\pi \mid y, x, z)$ on the line between $\pi = 0$ (the vertical asymptote) and $\hat{\pi}_{OLS}$ (typically near a regular mode) satisfies $\pi^{*\prime}_{min}\, A^{-1}\, \pi^*_{min} = 1$. Intuitively stated, the $t^k_{0,1-k}$ distribution aims at approximating the asymptote around $\pi = 0$, covering the region with $\pi \approx 0$ where the factor $(\pi' z' M_x z \pi)^{-1/2}$ is 'more important' than the other factors in $p(\pi \mid y, x, z)$.

Step 2: We use this initial candidate distribution in an independence chain Metropolis-Hastings method, which we use to adapt the candidate distribution:

- μ and Σ are the mean and covariance matrix of MH draws π for which $(\pi' A^{-1}\pi)^{1/2} > 1$. ν is chosen to match the maximum kurtosis of the k elements of the π draws for which $(\pi' A^{-1}\pi)^{1/2} > 1$, if this maximum kurtosis is larger than 3. Otherwise ν is set to a rather large value, e.g. 30;

- $w_{t^k_{0,1-k}}$ is the fraction of MH draws for which $(\pi' A^{-1}\pi)^{1/2} \leq 1$. In case of strong instruments this fraction may be 0 (based on a finite number of draws); in that case we set $w_{t^k_{0,1-k}} = 0.01$;

· A is $\frac{k(k+1)}{k-1}$ times the covariance matrix of the MH draws with $(\pi' A^{-1}\pi)^{1/2} \leq 1$.

Step 3: Use the adapted candidate in the independence chain Metropolis-Hastings method or Importance Sampling to simulate efficiently from $p(\pi \mid y, x, z)$.

After one has simulated draws π from $p(\pi \mid y, x, z)$ using the HMS method, one easily samples draws from β and Ω from their conditional Student-t and Inverse-Wishart posteriors. We name this approach the Approximate Direct Monte Carlo (ADMC) method.

The application of (extended or adapted versions of) ADMC to more extensive IV models (e.g., IV with multiple possibly endogenous regressors, IV with non-Gaussian errors, panel IV, probit/logit IV) is outside the scope of this paper, and left as a topic for further research. In any case, a relevant lesson is that one should always be careful to use a candidate distribution that can cope with the shapes of the posterior. Otherwise IS weights may have infinite variance; MH may have an absorbing state or different convergence problems.

Another topic for further research is the inclusion of the $t^k_{0,1-k}$ distribution within the Mixture of t by Importance Sampling and Expectation Maximization (MitISEM) approach of Hoogerheide et al. (2011). This may improve the robustness, flexibility and applicability of MitISEM even further.

6.1 Approximate Direct Monte Carlo (ADMC): an example for simulated data

We simulate $T = 1000$ data from model (23)-(24) with

$$\beta = 0.1, \qquad \pi = (0.025, 0.025)', \qquad \Omega = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

For illustrative purposes, we choose rather weak instruments z. However, they do have a significant effect on x in the frequentist sense. The multiple F-test in (24) has a p-value of 0.0311.

Figure 2 gives the posterior kernel $p(\pi \mid y, x, z)$, which we approximate using the Hybrid Mixture Sampling (HMS) approach. Note that the vertical axis is restricted to the interval $[0, 1]$ (where the values of the posterior kernel $p(\pi \mid y, x, z)$ are scaled to have maximum 1 over the graph's set of grid points), whereas $p(\pi \mid y, x, z)$ obviously tends to ∞ for $\pi \rightarrow 0$. Figure 3 shows the posterior kernel $p(\pi_1, \pi_2 \mid y, x, z)$ on the line between $\pi = 0$ and $\pi = \hat{\pi}_{OLS} = (z'z)^{-1}z'x$. Note that the horizontal axis refers to π_1, but also π_2 varies over the points. The minimum on the line between $\pi = 0$ and $\pi = \hat{\pi}_{OLS}$ is located at $\pi^*_{min} = (0.0101, 0.0106)$. The first hybrid mixture approximation of the posterior in the Hybrid Mixture Sampling (HMS) approach is in Figure 4. The adapted candidate density $q(\pi)$, a mixture with weights $w_{t^k_{0,1-k}} = 0.0463$ and $1 - w_{t^k_{0,1-k}} = 0.9537$, is in Figure 5. Note the close approximation of the posterior shapes.

This whole procedure, yielding 10000 MH draws, takes merely 2.4 s on a Intel Centrino$^{\mathrm{TM}}$ Dual Core processor. The MH acceptance rate is very high: 91.1%. The first order serial correlation of the MH draws is very low: 0.1123 for π_1 and 0.0979 for π_2. The coefficient of variation of the IS weights is also very low: 0.192.

After one has simulated draws π from $p(\pi \mid y, x, z)$ using this HMS method, one can easily sample draws from β and Ω from their conditional Student-t and Inverse-Wishart posteriors. Given the close approximation and high MH acceptance rate, one can truly name this approach the Approximate Direct Monte Carlo (ADMC) method.

Finally, note that we discuss the application of ADMC to posterior simulation in this simple IV model only for illustrative purposes, since here one can simply use our DMC method.

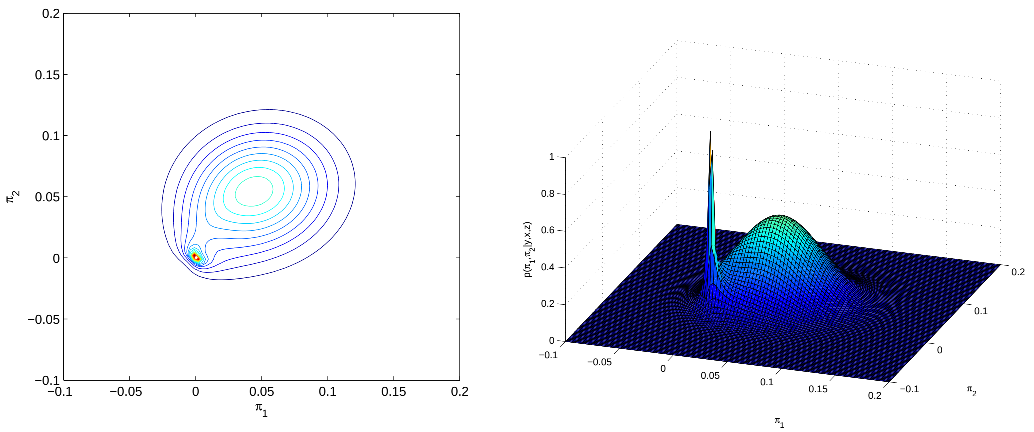


Figure 2. Example of Approximate Direct Monte Carlo (ADMC): contour plot and graph of posterior density kernel $p(\pi \mid y, x, z)$. Note that the vertical axis of the 3d-graph is restricted to the interval $[0, 1]$ (where the values of the posterior kernel $p(\pi \mid y, x, z)$ are scaled to have maximum 1 over the graph's set of grid points), whereas $p(\pi \mid y, x, z)$ obviously tends to ∞ for $\pi \to 0$.

7 Model comparison and testing exogeneity

One of the important aspects of the model structure is the existence of the simultaneous relationship or endogeneity problem in the first place. If the exogeneity

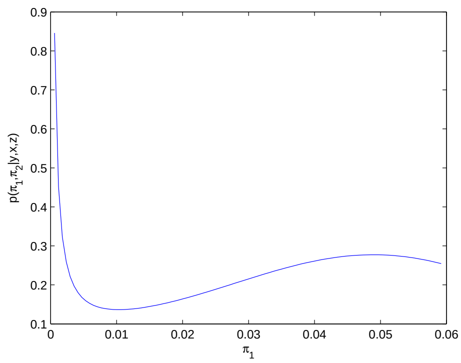


Figure 3. Example of Approximate Direct Monte Carlo (ADMC): posterior density kernel $p(\pi_1, \pi_2 \mid y, x, z)$ on line between $\pi = 0$ and $\pi = \hat{\pi}_{OLS} = (z'z)^{-1}z'x$. Note: the horizontal axis refers to π_1, but also π_2 varies over the points.

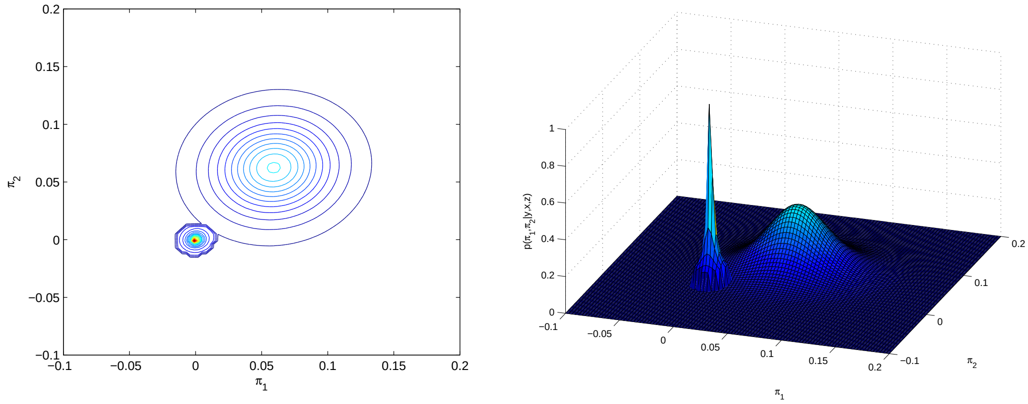


Figure 4. Example of Approximate Direct Monte Carlo (ADMC): contour plot and graph of first hybrid mixture approximation of the posterior $p(\pi \mid y, x, z)$.

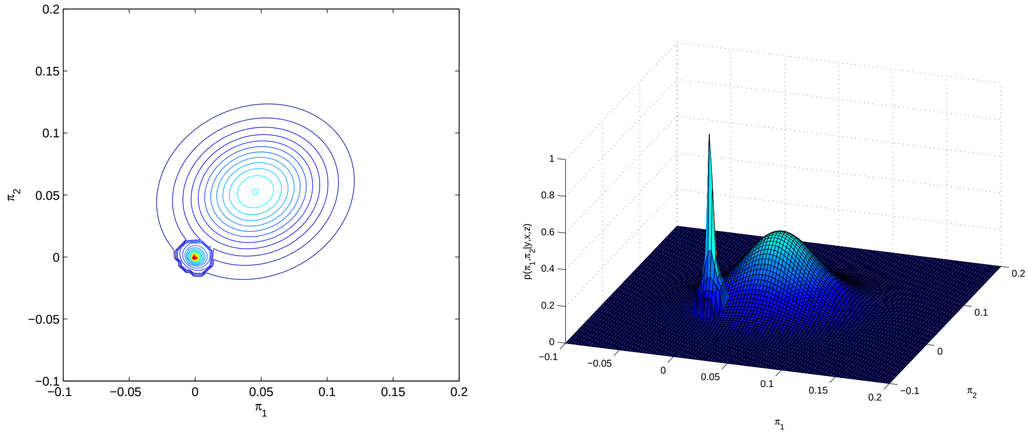


Figure 5. Example of Approximate Direct Monte Carlo (ADMC): contour plot and graph of adapted hybrid mixture approximation $q(\pi)$ of the posterior $p(\pi \mid y, x, z)$.

restriction $\rho = \omega_{12}/(\omega_{11}\omega_{22})^{1/2} = 0$ is set beforehand, we obtain a proper marginal density of π for any $k \geq 1$. In our simple EV model, the posterior density for (β, π) after integrating out the remaining variance terms ω_{11}, ω_{22} is (see Zellner (1971)):

$$p(\beta, \pi \mid y, x) \propto \left[(y - x\beta)'(y - x\beta)\right]^{-T/2} \left[(x - \iota\pi)'(x - \iota\pi)\right]^{-T/2} \tag{38}$$
$$\propto p(\beta \mid y, x)\, p(\pi \mid y, x), \tag{39}$$

i.e. the conditional and marginal distributions β and π are two independent student-t densities.

In the Bayesian context, the exogeneity test corresponds to a simple model comparison. Let M_0 denote the model with the exogeneity restriction for which $\rho = \omega_{12}/(\omega_{11}\omega_{22})^{1/2} = 0$ in (8) and (9), and M_1 denote the unrestricted model. The posterior odds ratio, K_{01} for M_0 is the product of the Bayes factor and the prior odds ratio:

$$K_{01} = \frac{p(y \mid M_0)}{p(y \mid M_1)} \times \frac{p(M_0)}{p(M_1)}, \tag{40}$$

where in this section we disregard the conditioning on x, z for simplicity; that is, y contains all the observed data (previously denoted by $\{y, x, z\}$), and the prior model probabilities are $(p(M_1), p(M_0)) \in (0,1) \times (0,1)$ and $p(M_1) + p(M_0) = 1$.

For the IV model and SEMs in general, calculation of the marginal likelihood is non-trivial. Several methods are proposed to approximate the above integrals (see e.g. Chib (1995); Frühwirth-Schnatter and Wagner (2008); Ardia et al. (2010)). A straightforward method is to use the Savage-Dickey Density Ratio (SDDR) to calculate model probabilities (Dickey, 1971). In this case the Bayes factor can be calculated using a single model if the alternative models are nested and the prior densities satisfy the condition that the prior for $\theta_{-\rho}$ in the restricted model M_0 equals the conditional prior for $\theta_{-\rho}$ given $\rho = 0$ in the model M_1, i.e. $p_1(\theta_{-\rho} \mid \rho = 0) = p_0(\theta_{-\rho})$ [3]. In this case, (40) becomes:

$$K_{01} = \frac{p(\rho = 0 \mid y, M_1)}{p(\rho = 0 \mid M_1)} \times \frac{p(M_0)}{p(M_1)}, \tag{41}$$

where $p(\rho \mid y, M_1) = \int p(\rho, \theta_{-\rho} \mid y, M_1) d\theta_{-\rho}$. [4]

One important consideration in model comparison is the effect of relatively non-informative priors. Choosing a prior $p(\rho, \theta_{-\rho})$ flat enough compared to $p(\theta_{-\rho})$, the posterior odds ratio in (40) becomes larger independent of the data. Hence if we consider non-informative priors, the most restrictive model will typically be favored.

[3] Notice that the condition for SDDR holds if we define the prior for $\theta_{-\rho}$ in the restricted model equal to the conditional prior of $\theta_{-\rho}$ given $\rho = 0$ in the unrestricted model.

[4] As a generalization, Verdinelli and Wasserman (1995) show that K_{01} is equal to the Savage-Dickey density ratio in (41) times a correction factor when the prior condition fails.

This phenomenon is called Bartlett's paradox (Bartlett, 1957). Specifically, the prior $p(\rho \mid \theta_{-\rho})$ must be proper for the Bayes factor to be well defined.

In particular for the flat prior we consider, a model comparison relying on the marginal likelihood under these priors is erroneous. Especially in the time series context, it is shown that model comparison in these cases can be based on predictive likelihoods (Laud and Ibrahim, 1995; Eklund and Karlsson, 2007). Here we summarize a *predictive likelihoods* approach to testing exogeneity.

A predictive likelihood for the model M_1 is computed by splitting the data $y = (y_1, \ldots, y_T)$ into a training sample $y^* = (y_1, \ldots, y_m)$ and a hold-out sample $\tilde{y} = (y_{m+1}, \ldots, y_T)$. Then the predictive likelihood is given by:

$$p(\tilde{y} \mid y^*, M_1) = \int p(\tilde{y} \mid \theta_1, y^*, M_1) p(\theta_1 \mid y^*, M_1) d\theta_1, \tag{42}$$

where θ_1 are the model parameters for model M_1. Notice that equation (42) corresponds to the marginal posterior likelihood for the training sample $\tilde{y}$ and the exact posterior density after observing y^* as the prior. The exact posterior density $p(\theta_1 \mid y^*, M_1)$ is obtained by Bayes' rule:

$$p(\theta_1 \mid y^*, M_1) \propto \frac{p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1)}{p(y^* \mid M_1)} = \frac{p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1)}{\int p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1}. \tag{43}$$

Substituting (43) into (42) leads to:

$$p(\tilde{y} \mid y^*, M_1) = \frac{\int p(\tilde{y} \mid \theta_1, y^*, M_1) p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1}{\int p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1} = \frac{\int p(y \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1}{\int p(y^* \mid \theta_1, M_1) p(\theta_1 \mid M_1) d\theta_1}. \tag{44}$$

In case of predictive likelihoods, model probabilities are again calculated from the posterior odds ratio:

$$\frac{p(M_0 \mid y)}{p(M_1 \mid y)} = \frac{p(\tilde{y} \mid y^*, M_0)}{p(\tilde{y} \mid y^*, M_1)} \frac{p(M_0)}{p(M_1)}. \tag{45}$$

Combining the predictive likelihood formula in (45) and SDDR in (41), the posterior odds ratio becomes:

$$K_{01} = \frac{p(M_0 \mid \tilde{y}, y^*)}{p(M_1 \mid \tilde{y}, y^*)} = \frac{p(\rho = 0 \mid \tilde{y}, y^*, M_1)}{p(\rho = 0 \mid y^*, M_1)} \times \frac{p(M_0)}{p(M_1)},$$

where $p(\rho \mid \tilde{y}, y^*) = \int p(\rho, \theta_{-\rho} \mid \tilde{y}, y^*) d\theta_{-\rho}$ and $p(\rho \mid y^*) = \int p(\rho, \theta_{-\rho} \mid y^*) d\theta_{-\rho}$ are the *exact* marginal posterior densities using the full data, and the training sample, respectively.

A final point concerning the calculation of predictive likelihoods is the size of the training sample. More stable results may be achieved as the training sample size decreases, but the training sample should be large enough to provide a proper density given the originally flat/uninformative prior of parameters. Different training sample sizes have been proposed in the literature (see Gelfand and Dey (1994) for an overview of the forms of predictive likelihood under different training sample choices).

This analysis of the possible validity of the exogeneity restriction is an example of an application in which one is not only interested in the IV model's posterior of β. Efficient simulation from the posterior of Ω may be quite useful here; e.g., when one computes kernel estimates of $p(\rho = 0 \mid \tilde{y}, y^*, M_1)$ and $p(\rho = 0 \mid y^*, M_1)$ based on draws of $\rho = \omega_{12}/(\omega_{11}\omega_{22})^{1/2}$. One may also apply a Rao-Blackwellization step, averaging the conditional posterior of ρ at $\rho = 0$ for each draw of $(\beta, \pi, \omega_{11}, \omega_{12})$, for which again draws of β, π and Ω are required.

8 Empirical example with $k = 2$ instruments: Fulton fish market data

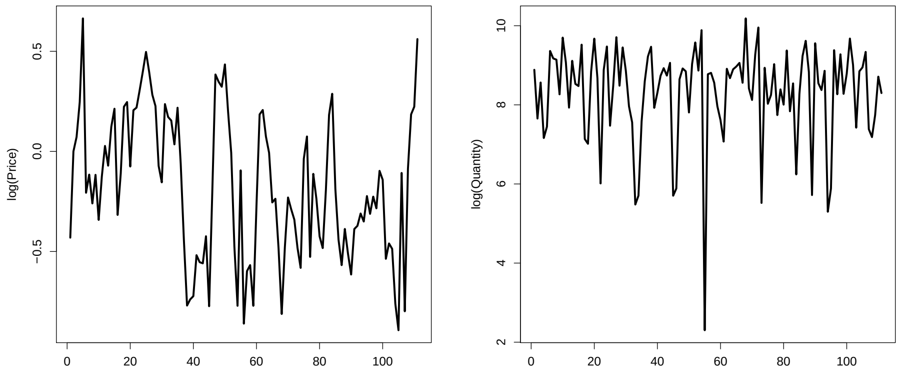
We next illustrate the issue of irregular posterior shapes in a simple analysis of the demand for fish. The data provide the price and quantity of fresh whiting sold in the Fulton fish market over the five month period from December 2, 1991 to May 8, 1992, and are collected from a single dealer (Graddy, 1995; Chernozhukov and Hansen, 2008). The price is measured as the average daily price and the quantity is measured as the total amount of fish sold. The number of observations, namely, the number of days the market was open over the sample period, is $T = 111$. Figure 6 provides a plot of the data.

Following Chernozhukov and Hansen (2008), we consider the following Incomplete Simultaneous Equations Model (INSEM) or overidentified IV model:

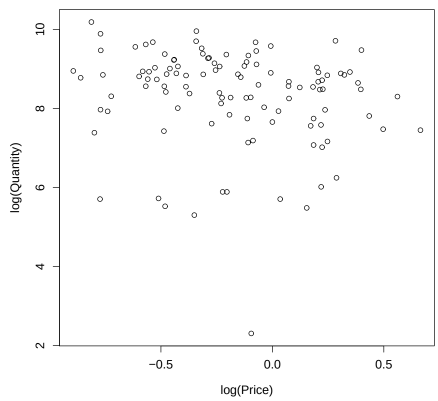
$$\begin{aligned} \log Q_t &= \alpha_q + \beta \log P_t + \varepsilon_t, \\ \log P_t &= \alpha_p + \pi_1 Z_{1t} + \pi_2 Z_{2t} + v_t, \end{aligned} \tag{46}$$

where Z_{1t} and Z_{2t} are two different instruments that capture weather conditions at sea. Z_{1t} is a dummy variable, *Stormy*, which indicates wave height greater than 4.5 ft and wind speed greater than 18 knots. Z_{2t} is also a dummy variable, *Mixed*, indicating wave height greater than 3.8 ft and wind speed greater than 13 knots. Chernozhukov and Hansen (2008) explain that these variables are plausible instruments for price in the demand equation, since weather conditions at sea should influence the amount of fish on the market but should not influence demand for fish.

The constant terms α_q and α_p are simply integrated out under a flat prior by taking all variables in deviation from their sample means. In the model with data in de-



log(Daily Price) log(Daily Quantity)



log(Daily Price) and log(Daily Quantity)

Figure 6. Demand for fish data. The data contain observations on price and quantity of fresh whiting sold in the Fulton fish market in New York City over the five month period from December 2, 1991 to May 8, 1992. The price is measured as the average daily price and the quantity as the total amount of fish sold that day. In total, the sample consists of 111 observations for the days in which the market was open over the sample period. The bottom graph shows the relationship between price and quantity.

viation from their sample means we have $\hat{\pi}_{OLS} = (z'z)^{-1}z'x = (0.437, 0.236)'$ with standard errors $0.078, 0.077$ (implying t-values $5.599, 3.078$). The multiple F-test in the first stage regression has $F = 15.981$ with p-value 0.000. Figure 7 shows the shapes of the marginal posterior of (π_1, π_2). Here the instruments are stronger than in the aforementioned example for simulated data. The volume of the asymptote around $\pi = 0$ (showing up only as a point in the contour plot or a needle in the 3d-graph) is negligible, as compared to the 'regular', bell-shaped part of the posterior near $\hat{\pi}_{OLS} = (z'z)^{-1}z'x$. For a finite set of draws, the Approximate DMC method may work well even if only a Student-t candidate distribution is used to simulate from the marginal posterior of π. However, the fact that the theoretical variance of the Importance Sampling weights is infinite, implies that occasional draws near $\pi = 0$ may cause problems. For this reason, the HMS candidate distribution with a very small weight for the $t^k_{0,1-k}$ distribution around $\pi = 0$ may still be preferred as a 'safer' alternative. Finally, again note that we discuss the application of ADMC to posterior simulation in this simple IV model only for illustrative purposes, since here one can simply use our DMC method.

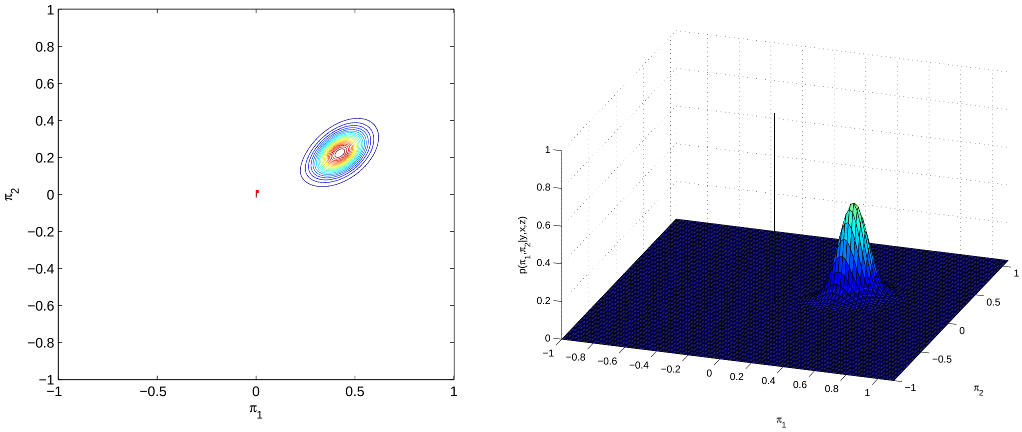


Figure 7. Fulton fish market: posterior density kernel $p(\pi_1, \pi_2 \mid y, x, z)$. Note that the vertical axis of the 3d-graph is restricted to the interval $[0, 1]$ (where the values of the posterior kernel $p(\pi \mid y, x, z)$ are scaled to have maximum 1 over the graph's set of grid points), whereas $p(\pi \mid y, x, z)$ obviously tends to ∞ for $\pi \to 0$. The volume of the asymptote around $\pi = 0$ (showing up only as a point in the contour plot or a needle in the 3d-graph) is negligible, as compared to the 'regular', bell-shaped part of the posterior.

9 Conclusions and Future Work

A Direct Monte Carlo (DMC) approach is introduced for posterior simulation in the Instrumental Variables (IV) model with one possibly endogenous regressor, multiple instruments and Gaussian errors under a flat prior. This DMC method can also be applied in an IV model (with one or multiple instruments) under an informative prior for the possibly endogenous regressor's effect. This DMC approach can not be applied to more complex IV models or Simultaneous Equations Models with multiple possibly endogenous regressors. An Approximate DMC (ADMC) approach is introduced that makes use of the proposed Hybrid Mixture Sampling (HMS) method, which facilitates Metropolis-Hastings (MH) or Importance Sampling from a proper marginal posterior density with highly non-elliptical shapes that tend to infinity for a point of singularity. After one has simulated from the irregularly shaped marginal distribution using the HMS method, one easily samples the other parameters from their conditional Student-t and Inverse-Wishart posteriors. An example illustrates the close approximation and high MH acceptance rate. On the other hand, using a simple candidate distribution such as the Student-t may lead to an infinite variance of Importance Sampling weights. The choice between the IV model and a simple linear model under the restriction of exogeneity (or the model weights in a Bayesian Model Averaging application of these models) may be based on predictive likelihoods, for which the efficient simulation of all model parameters may be quite useful. In future work the ADMC approach may be extended to more extensive IV models such as IV with multiple possibly endogenous regressors, IV with non-Gaussian errors, panel IV, or probit/logit IV. Also for other reduced rank models such as cointegration or factor models, extended or adapted versions of the ADMC method may be useful.

References

Aliprantis, C. D., Barnett, W. A., Cornet, B., Durlauf, S., 2007. Special issue editors' introduction: The interface between econometrics and economic theory. Journal of Econometrics 136 (2), 325–329.

Anderson, T., 1976. Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics. Journal of the Royal Statistical Society. Series B (Methodological) 38 (1), 1–36.

Angrist, J. D., Krueger, A. B., 1991. Does compulsory school attendance affect schooling and earnings? The Quarterly Journal of Economics 106 (4), 979–1014.

Ardia, D., Baştürk, N., Hoogerheide, L., Van Dijk, H. K., 2010. A comparative study of monte carlo methods for efficient evaluation of marginal likelihood. Computational Statistics & Data Analysis XX, forthcoming.

Arnold, B. C., Castillo, E., Sarabia, J. M., 1999. Conditional specification of statistical models. Springer Verlag.

Attfield, C. L. F., 1976. Estimation of the structural parameters in a permanent income model. Economica 43 (171), 247–254.

Bartlett, M. S., 1957. A comment on D. V. Lindley's statistical paradox. Biometrika 44 (3–4), 533.

Bauwens, L., Van Dijk, H. K., 1990. Bayesian limited information analysis revisited. In: Gabszewicz, J. J., Richard, J. F., Wolsey, L. A. (Eds.), Economic Decision-Making: Games, Econometrics and Optimisation: Contributions in Honour of Jacques H. Drèze. North-Holland, Ch. 18, pp. 385–424.

Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: A robust inference approach. Journal of Econometrics 142 (1), 379–398.

Chib, S., 1995. Marginal likelihood from the gibbs output. Journal of the American Statistical Association 90 (432).

De Pooter, M., Ravazzolo, F., Segers, R., Van Dijk, H. K., 2008. Bayesian near-boundary analysis in basic macroeconomic time series models. In: Chib, S., Griffiths, W., Koop, G., Terrell, D. (Eds.), Bayesian Econometrics. Vol. 23 of Advances in Econometrics. Emerald Group, pp. 331–402.

Dickey, J. M., 1971. The weighted likelihood ratio, linear hypotheses on normal location parameters. The Annals of Mathematical Statistics 42 (1), 204–223.

Drèze, J. H., 1976. Bayesian limited information analysis of the simultaneous equations model. Econometrica 44, 1045–1075.

Drèze, J. H., 1977. Bayesian regression analysis using poly-t densities. Journal of Econometrics 6 (3), 329–354.

Eklund, J., Karlsson, S., 2007. Forecast combination and model averaging using predictive measures. Econometric Reviews 26 (2), 329–363.

Friedman, M., 1957. Introduction to A Theory of the Consumption Function. Princeton University Press.

Frühwirth-Schnatter, S., Wagner, H., 2008. Marginal likelihoods for non-gaussian models using auxiliary mixture sampling. Computational Statistics & Data Analysis 52 (10), 4608–4624.

Gelfand, A. E., Dey, D. K., 1994. Bayesian model choice: asymptotics and exact calculations. Journal of the Royal Statistical Society. Series B (Methodological)

56 (3), 501–514.

Graddy, K., 1995. Testing for imperfect competition at the fulton fish market. The RAND Journal of Economics 26 (1), 75–92.

Hobert, J. P., Casella, G., 1998. Functional compatibility, Markov chains, and Gibbs sampling with improper posteriors. Journal of Computational and Graphical Statistics 7 (1), 42–60.

Hood, W. M. C., Koopmans, T. C. (Eds.), 1950. Studies in Econometric Method. No. 14 in Cowles Commision Monographs. Wiley, New York.

Hoogerheide, L., Opschoor, A., Van Dijk, H. K., 2011. A class of Adaptive EM–based Importance Sampling Algorithms for efficient and robust posterior and predictive simulation. Tinbergen Institute Discussion Paper 2011–004/4.

Hoogerheide, L. F., Kaashoek, J. F., Van Dijk, H. K., 2007. On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: an application of flexible sampling methods using neural networks. Journal of Econometrics 139 (1), 154–180.

Kleibergen, F., Van Dijk, H. K., 1994. On the shape of the likelihood/posterior in cointegration models. Econometric Theory 10 (3-4), 514–551.

Kleibergen, F., Van Dijk, H. K., 1998. Bayesian simultaneous equations analysis using reduced rank structures. Econometric Theory 14, 701–743.

Kleibergen, F., Zivot, E., 2003. Bayesian and classical approaches to instrumental variable regression. Journal of Econometrics 114 (1), 29–72.

Koop, G., 2003. Bayesian econometrics. Wiley, New York.

Koopmans, T. C. (Ed.), 1950. Statistical inference in dynamic economic models. No. 10 in Cowles Commision Monographs. Wiley, New York.

Lancaster, T., 2004. An introduction to modern Bayesian econometrics. Blackwell Publishing, Oxford.

Laud, P. W., Ibrahim, J. G., 1995. Predictive model selection. Journal of the Royal Statistical Society. Series B (Methodological) 57 (1), 247–262.

Van Dijk, H. K., 2003. On Bayesian structural inference in a simultaneous equation model. In: Stigum, B. P. (Ed.), Econometrics and the philosophy of economics. Vol. 23. Princeton University Press, pp. 642–682.

Verdinelli, I., Wasserman, L., 1995. Computing bayes factors using a generalization of the savage-dickey density ratio. Journal of the American Statistical Association 90 (430).

Zellner, A., 1971. An introduction to Bayesian inference in econometrics. Wiley, New York.

Zellner, A., Bauwens, L., Van Dijk, H. K., 1988. Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo integration. Journal of Econometrics 38 (1-2), 39–72.

A Derivation of posterior densities for the IV model with one endogenous variable and k instruments

We consider the generalization of the IV model in Section 3, with one possibly endogenous regressor and k instruments. As an introduction we note that for the model $y = x\beta + u$ with $u \sim \text{NID}(0, \omega)$, the posterior density of β under flat priors is a student-t density with posterior mean equal to the Maximum Likelihood or least squares estimator. The scaling factor of this density is also standard, see e.g. Koop (2003) and direct simulation from this posterior is possible. For the IV model however, this posterior density is more complex, in fact it is a student-t density times a polynomial or a rational function. In Figure A.1 we summarize the existence conditions and the derivation steps for the posterior densities in the IV model under flat priors.

Figure A.1 presents the steps for the decomposition of the joint posterior into conditional and marginal posteriors, where we extend the scheme of integration steps in Bauwens and Van Dijk (1990). For the step-by-step derivation of these posterior densities see the Appendix. Under flat priors, conditional posteriors of $\beta|\pi, \Omega,$ data, $\pi|\beta, \Omega,$ data and $\Omega|\beta, \pi,$ data are Normal and Inverted Wishart densities. Moments of these densities exist, and this result does not depend on the number of instruments. However, Gibbs sampling using these conditional densities can only be used if the joint posterior is a proper density, which is not the case for an exactly identified model ($k = 1$). Hence a straightforward application of the Gibbs sampling procedure on these posteriors can be erroneous. See e.g. Arnold, Castillo, and Sarabia (1999); Hobert and Casella (1998) for a discussion and how Markov Chain and Gibbs sampling methods might be employed *erroneously* on models with improper posteriors.

Likelihood function and the joint posterior under a Flat Prior Consider the structural form (SF) representation of the IV model with one endogenous variable and k instruments:

$$y = x\beta + u, \tag{A.1}$$
$$x = z\pi + v, \tag{A.2}$$

where y is the $T \times 1$ vector of data of the dependent variable, x is the $T \times 1$ vector of data on the possibly endogenous explanatory variable, z is the $T \times k$ matrix of data on the instruments, and the disturbances follow an iid normal distribution: $(u', v')' \sim N(0, \Omega \otimes I)$, where I is the identity matrix of size T and $\Omega = \left(\begin{smallmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{smallmatrix}\right)$.

The orthogonal structural form (OSF) for the IV model is obtained by decomposing u in (A.1) into two independent components $u = v\eta + \varepsilon$:

$$y = x\beta + v\eta + \varepsilon, \tag{A.3}$$
$$x = z\pi + v, \tag{A.4}$$

Figure A.1. Scheme of Derivation Steps for Posterior Densities of the IV Model with One Endogenous Variable and k Instruments, under a Flat Prior

Joint posterior

$$p\,(\beta, \pi, \Omega \mid \text{data})$$

Posterior has a ridge at $\pi = 0$,
Density is improper for $k = 1$ and proper for $k \geq 2$.

Conditional posteriors

$$p\,(\beta, \pi, \Omega \mid \text{data})$$

$\downarrow$	$\downarrow$	$\downarrow$
complete sum of squares in β	complete sum of squares in π	use properties of Inverse-Wishart distribution
$\downarrow$	$\downarrow$	$\downarrow$
$\beta \vert \pi, \Omega, \text{data} \sim N(\hat{\beta}, V_\beta)$	$\pi \vert \beta, \Omega, \text{data} \sim N(\hat{\pi}, V_\pi)$	$\Omega \vert \beta, \pi, \text{data} \sim \text{IW}\,(\Xi, T)$
where $V_\beta = \frac{\vert\Omega\vert}{\omega_{22} x'x}$,	where $V_\pi = \frac{\vert\Omega\vert}{\omega_{11}}(z'z)^{-1}$,	where $\Xi = (u,\ v)'\,(u,\ v)$
$\hat{\beta} = \frac{x'y}{x'x} - \frac{\omega_{12}}{\omega_{22}}(1 - \frac{x'z\pi}{x'x})$	$\hat{\pi} = (z'z)^{-1}\left(x'z - \frac{\omega_{12} z'(y - x\beta)}{\omega_{11}}\right)$	for $u = y - x\beta$, $v = x - z\pi$

Moments of $p\,(\beta \vert \pi, \Omega, data)$, $p\,(\pi \vert \beta, \Omega, data)$ and $p\,(\Omega \vert \beta, \pi, data)$ exist for all values of π in their domain and for any number of instruments, $k = 1, 2, \ldots, K$.

Marginal posteriors of β, π ***

$$p\,(\beta, \pi, \Omega \mid \text{data})$$
$$\downarrow$$
Inverse-Wishart step on Ω in $p\,(\beta, \pi, \Omega \mid \text{data})$ (see Zellner (1971); Bauwens and Van Dijk (1990))
$$\downarrow$$
$$p\,(\beta, \pi \mid \text{data}) \propto \left|(u,\ v)'\,(u,\ v)\right|^{-T/2} \text{ for } u = y - x\beta,\ v = x - z\pi$$

$\downarrow$	$\downarrow$
apply determinant decomposition	apply determinant decomposition
$\left\vert(u,\ v)'\,(u,\ v)\right\vert = (u'u)\,(v'M_u v)$	$\left\vert(u,\ v)'\,(u,\ v)\right\vert = (v'v)\,(u'M_v u)$
$\downarrow$	$\downarrow$
complete sum of squares on π	complete sum of squares on β
$\downarrow$	$\downarrow$
$p\,(\pi \vert \beta, \text{data}) \propto$ multivariate t-density*	$p\,(\beta \vert \pi, \text{data}) \propto t$-density
Moments exist for all values of β in its domain.	*The conditional posterior of β given π does not exist for $\pi = 0$.*
$\downarrow$	$\downarrow$
t-density step on π	t-density step on β
$\downarrow$	$\downarrow$
$p\,(\beta \mid y, x, z) \propto \vert z'M_u z\vert^{-\frac{1}{2}}\,(u'u)^{-\frac{T}{2}} \times \left((T-k)\,s^2_{\hat{\pi}}\right)^{-\frac{T-k}{2}}$	$p\,(\pi \mid y, x, z) \propto (x'M_v x)^{-\frac{1}{2}}\,(v'v)^{-\frac{T}{2}} \times \left((T-1)\,s^2_{\hat{\beta}}\right)^{-\frac{T-1}{2}}$
where $s^2_{\hat{\pi}} = (M_u x)'\,M_{M_u z}\,(M_u x)\,/(T-k)$	where $s^2_{\hat{\beta}} = (M_x y)'\,M_{M_v x}\,(M_x y)\,/(T-1)$
$\downarrow$	$\downarrow$
use matrix decomposition and properties of the projection matrix:	use matrix decomposition and properties of the projection matrix:
$\vert z'M_u z\vert \propto (u'M_z u)\,/\,(u'u)$	$x'M_v x \propto v'M_x v/(v'v)$
$(T-k)s^2_{\hat{\pi}} \propto (u'M_z u)^{-1}$	$(T-1)\,s^2_{\hat{\beta}} \propto v'M_{(y\ \ x)}v\,(v'M_x v)^{-1}$
$\downarrow$	$\downarrow$
$p\,(\beta \mid \text{data}) \propto (u'u)^{-\frac{T-1}{2}}\,(u'M_z u)^{\frac{T-k-1}{2}}$	$p\,(\pi \mid \text{data}) \propto (v'v)^{-\frac{T-1}{2}}\,(\pi'z'M_x z\pi)^{-\frac{1}{2}} \times \left(\frac{\pi'z'M_x z\pi}{\pi'z'M_{(y\ \ x)}z\pi}\right)^{\frac{T-1}{2}}$
$p\,(\beta \mid data)$ is a t-density form times a polynomial. It is an improper density for an exactly identified model ($k = 1$); and a proper density for an overidentified model ($k \geq 2$).	*$p\,(\pi \mid data)$ is a t-density form times a rational function**. It is an improper density for an exactly identified model ($k = 1$); and a proper density for an overidentified model ($k \geq 2$). It is not trivial to simulate from this distribution.*

* The conditional posterior simplifies to student-t density for the exactly identified model ($k = 1$).
** See Kleibergen and Van Dijk (1994, 1998) for the derivation of this density.
*** Derivation of the marginal posterior of Ω is left to the reader.

where $\eta = \omega_{12}/\omega_{22}$, $(\varepsilon', v')' \sim NID(0, \Sigma \otimes I)$, $\Sigma = \left(\begin{smallmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{smallmatrix}\right)$, $\sigma_{22} = \omega_{22}$ and $\sigma_{11} = \omega_{11} - \omega_{12}^2/\omega_{22}$. By definition, $(\eta, \beta, \pi) \in \mathbb{R}^{k+2}$, $(\sigma_{11}, \sigma_{22}) \in \mathbb{R}^2_+$ [5].

The likelihood of IV model in terms of the SF representation in (A.1) and (A.2) is equivalent to the following kernels:

$$
\begin{aligned}
p(y, x \mid \beta, \pi, \Omega, z) \propto & |\Omega|^{-T/2} \exp\Big\{-\frac{1}{2} |\Omega|^{-1} \Big(\omega_{22}(y - x\beta)'(y - x\beta) \\
& - 2\omega_{12}(y - x\beta)'(x - z\pi) + \omega_{11}(x - z\pi)'(x - z\pi)\Big)\Big\} \qquad (A.5) \\
= & |\Omega|^{-T/2} \exp\Big\{-\frac{1}{2} \mathrm{tr}\left((y - x\beta,\ x - z\pi)'(y - x\beta,\ x - z\pi)\Omega^{-1}\right)\Big\} \qquad (A.6)
\end{aligned}
$$

A flat prior for the model in (A.1) and (A.2) is:

$$
p(\beta, \pi, \Omega) \propto |\Omega|^{-3/2}. \qquad (A.7)
$$

Combining the prior in (A.7) with the likelihood in (A.5) and (A.6), the posterior density of parameters is:

$$
\begin{aligned}
p(\beta, \pi, \Omega \mid y, x, z) \propto & |\Omega|^{-(T+3)/2} \exp\Big\{-\frac{1}{2} |\Omega|^{-1} \Big(\omega_{22}(y - x\beta)'(y - x\beta) \\
& - 2\omega_{12}(y - x\beta)'(x - z\pi) + \omega_{11}(x - z\pi)'(x - z\pi)\Big)\Big\}, \qquad (A.8) \\
= & |\Omega|^{-(T+3)/2} \exp\Big\{-\frac{1}{2} \mathrm{tr}\left((y - x\beta,\ x - z\pi)'(y - x\beta,\ x - z\pi)\Omega^{-1}\right)\Big\}. \qquad (A.9)
\end{aligned}
$$

Conditional posterior densities for the IV model under a Flat Prior Conditional posterior of $\beta \mid \pi, \Omega, y, x, z$ is derived using (A.8):

$$
\begin{aligned}
p(\beta \mid \pi, \Omega, y, x, z) \propto & \exp\left\{-\frac{1}{2} |\Omega|^{-1} \left(\omega_{22}(y - x\beta)'(y - x\beta) - 2\omega_{12}(y - x\beta)'(x - z\pi)\right)\right\} \\
\propto & \exp\left\{-\frac{1}{2} |\Omega|^{-1} \left(\omega_{22}\beta^2(x'x) - 2\beta\left(\omega_{22}y'x - \omega_{12}x'(x - z\pi)\right)\right)\right\} \\
\Rightarrow & \beta \mid \pi, \Omega, y, x, z \sim N(\hat{\beta}, V_\beta) \qquad (A.10)
\end{aligned}
$$

where $V_\beta = |\Omega|/(\omega_{22}x'x)$ and $\hat{\beta} = y'x/(x'x) - \omega_{12}/\omega_{22}(1 - x'z\pi/(x'x))$.

[5] The support for the variable η is unrestricted: Define $\rho = \omega_{12}/(\omega_{11}\omega_{22})^{1/2}$ where $\rho \in (-1, 1)$ for a pds matrix Ω. Then the transformation for η is: $\eta = \omega_{12}/\omega_{22} = \rho(\omega_{11}/\omega_{22})^{1/2}$. Given that $(\omega_{11}, \omega_{22}) \in \mathbb{R}^2_+$ and $\rho \in (-1, 1)$, $\eta \in \mathbb{R}$.

Conditional posterior $\pi \mid \beta, \Omega, y, x, z$ is derived from (A.8):

$$
\begin{aligned}
p(\pi \mid \beta, \Omega, y, x, z) &\propto \exp\left\{-\frac{1}{2}\left|\Omega\right|^{-1}\left(-2\omega_{12}(y - x\beta)'(x - z\pi) + \omega_{11}(x - z\pi)'(x - z\pi)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left|\Omega\right|^{-1}\left(\omega_{11}\pi' z' z\pi - 2\pi'(\omega_{11}z'x - \omega_{12}z'(y - x\beta))\right)\right\} \\
&\Rightarrow \pi \mid \beta, \Omega, y, x, z \sim N(\hat{\pi}, V_\pi) \qquad \text{(A.11)}
\end{aligned}
$$

where $V_\pi = |\Omega|(\omega_{11}z'z)^{-1}$ and $\hat{\pi} = (z'z)^{-1}z'x - \omega_{12}/\omega_{11}(z'z)^{-1}z'(y - x\beta)$.

Conditional posterior of $\Omega \mid \beta, \pi, y, x, z$ follows from (A.9) and the properties of the Inverse-Wishart distribution. Given that $(y - x\beta,\ x - z\pi)'(y - x\beta,\ x - z\pi)$ is a pds matrix and $T > 1$, conditional posterior of Ω is (see e.g. (Zellner, 1971, pp. 395)):

$$
\begin{aligned}
&p(\Omega \mid \beta, \pi, y, x, z) \propto |\Omega|^{\frac{-(T+3)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left((y - x\beta,\ x - z\pi)'(y - x\beta,\ x - z\pi)\Omega^{-1}\right)\right\} \\
&\Rightarrow \Omega \mid \beta, \pi, y, x, z \sim IW\left((y - x\beta,\ x - z\pi)'(y - x\beta,\ x - z\pi), T\right), \qquad \text{(A.12)}
\end{aligned}
$$

where $IW(\Xi, m)$ denotes the Inverse-Wishart distribution with the inverse scale matrix Ξ and m degrees of freedom.

We conclude that the under flat priors, conditional posteriors in (A.10), (A.11) and (A.12) are conventional Normal and Inverted Wishart densities. Moments of these densities exist for all values of π in their domain and for any number of instruments $k = 1, 2, \ldots, K$. However, Gibbs sampling using these conditional densities can only be used if the joint posterior is a proper density. Hence a straightforward application of the Gibbs sampling procedure on these posteriors can be erroneous.

We next derive the marginal posteriors of β, π for the IV model under flat priors. A graphical illustration of these integration steps to obtain the marginal posteriors under the flat prior is given in Bauwens and Van Dijk (1990).

Marginal posterior density of β for the IV model under a Flat Prior As an intermediary step, consider the marginal posterior of $\beta, \pi \mid y, x, z$, obtained by the Inverse-Wishart step on Ω:

$$
\begin{aligned}
p(\beta, \pi \mid y, x, z) &\propto \int_\Omega p(\beta, \pi, \Omega \mid y, z) d\Omega \\
&\propto \int_\Omega |\Omega|^{-(T+3)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left((y - x\beta,\ x - z\pi)'(y - x\beta,\ x - z\pi)\Omega^{-1}\right)\right\} d\Omega \\
&\propto \left|(y - x\beta,\ x - z\pi)'(y - x\beta,\ x - z\pi)\right|^{-T/2}. \qquad \text{(A.13)}
\end{aligned}
$$

Marginal density $p(\beta \mid y, z)$ is achieved by the following determinant decomposition and by completing the squares on π in (A.13):

$$
\begin{aligned}
p\left(\pi, \beta \mid y, x, z\right) &\propto \left|(y - x\beta,\ x - z\pi)'\left(y - x\beta,\ x - z\pi\right)\right|^{-T/2} \\
&= \left((y - x\beta)'(y - x\beta)\right)^{-T/2}\left((x - z\pi)' M_u (x - z\pi)\right)^{-T/2} \qquad \text{(A.14)}
\end{aligned}
$$

where $u = y - x\beta$ and $M_\alpha = I - \alpha(\alpha'\alpha)\alpha'$ is the projection matrix out of the span of α.

We next rewrite the sum of squares in π in (A.14). Define $\hat{\pi} = (z'M_u z)^{-1} z'M_u x$ and $s_{\hat{\pi}}^2 = ((x - z\hat{\pi})'M_u(x - z\hat{\pi}))/(T-k) = (M_u x)' M_{M_u z}(M_u x)/(T-k)$. Completing the squares in π yields:

$$p(\pi, \beta \mid y, x, z) \propto \left((T-k)\, s_{\hat{\pi}}^2\right)^{-\frac{T}{2}} \left(1 + \frac{(\pi - \hat{\pi})'(z'M_u z)(\pi - \hat{\pi})}{(T-k)\, s_{\hat{\pi}}^2}\right)^{-\frac{T}{2}} \tag{A.15}$$

where it is assumed that the condition $M_u z \neq 0$ holds.

From (A.15), conditional posterior of π is a matric-variate t density:

$$\pi \mid \beta, y, x, z \sim t\left(\hat{\pi}, s_{\hat{\pi}}^2 |z'M_u z|^{-1}, T-k\right). \tag{A.16}$$

From (A.15) and (A.16), marginal density of β is:

$$\begin{aligned} p(\beta \mid y, x, z) &= \int p(\beta, \pi \mid y, z)\, d\pi \\ &\propto ((y - x\beta)'(y - x\beta))^{-\frac{T}{2}} \left((T-k)\, s_{\hat{\pi}}^2\right)^{-\frac{T}{2}} \left(\left|z'M_u z\right| / s_{\hat{\pi}}^2\right)^{-1/2} \\ &\quad \times \int \left(\left|z'M_u z\right| / s_{\hat{\pi}}^2\right)^{1/2} \left(1 + \frac{(\pi - \hat{\pi})'(z'M_u z)(\pi - \hat{\pi})}{(T-k)\, s_{\hat{\pi}}^2}\right)^{-\frac{T}{2}} d\pi \quad \text{(A.17)} \\ &\propto \left|z'M_u z\right|^{-1/2} \left((T-k)\, s_{\hat{\pi}}^2\right)^{-\frac{T-k}{2}} (u'u)^{-\frac{T}{2}}, \quad \text{(A.18)} \end{aligned}$$

where the last equality holds since the integral in (A.17) is a multivariate student-t density apart from the integrating constant.

Simplifying the first term on the right-hand side of (A.18):

$$\left|z'M_u z\right| = \left(u'M_z u\right) |z'z| / (u'u) \propto \left(u'M_z u\right) / (u'u). \tag{A.19}$$

We next simplify the second term in (A.18), using the determinant decomposition:

$$\begin{aligned} (T-k)s_{\hat{\pi}}^2 &= (M_u x)' M_{M_u z}(M_u x) \\ &= (M_u z M_{M_u x} M_u z)(x'M_u x)\,|z'M_u z|^{-1}, \end{aligned} \tag{A.20}$$

where the first term on the right-hand side of (A.20) is the sum of squared residuals (SSR) in a regression of $M_u z$ on $M_u x$, which is equal to the SSR in a regression of z on u and x by the Freisch-Waugh theorem:

$$(M_u z M_{M_u x} M_u z) = (M_x z M_{M_x u} M_x z).$$

Substituting $u = y - x\beta$ in $M_x u$ yields:

$$M_x u = M_x(y - x\beta) = M_x y. \tag{A.21}$$

Hence the first term on the right-hand side of (A.20) is independent of the data, and can be disregarded for the marginal posterior of β:

$$(M_u z M_{M_u x} M_u z) = (M_x z M_{M_x u} M_x z) = (M_x z M_x y M_x z)$$

Therefore the following simplification holds

$$(T-k)s_{\hat{\pi}}^2 \propto (x' M_u x)\, |z' M_u z|^{-1} \propto \frac{u' M_x u x' x}{u' u} \left(\frac{u' M_z u\, |z' z|}{u' u}\right)^{-1} \propto (u' M_z u)^{-1}, \tag{A.22}$$

where we use the determinant decomposition, (A.21), and disregard the factors not depending on π.

Substituting (A.19) and (A.22) in (A.18), we have the simplified marginal posterior for β:

$$p(\beta \mid y, x, z) \propto \left(\frac{u^{'} M_z u}{u' u}\right)^{-1/2} (u' M_z u)^{\frac{T-k}{2}} (u' u)^{-\frac{T}{2}} = (u' M_z u)^{\frac{T-k-1}{2}} (u' u)^{\frac{T-1}{2}} \tag{A.23}$$

which is a polynomial, $(u' M_z u)^{(T-k-1)/2}$, times a t-density form, $(u' u)^{(T-1)/2}$. Marginal posterior of β is a proper density for all parameter values. However, it is not trivial to simulate from this posterior density.

Marginal posterior density of π for the IV model under a Flat Prior Marginal density $p(\pi \mid y, z)$ is achieved by the following determinant decomposition and by completing the squares on β using (A.13):

$$p\,(\pi, \beta \mid y, x, z) \propto ((x - z\pi)'(x - z\pi))^{-T/2} \left((y - x\beta)' M_v (y - x\beta)\right)^{-T/2} \tag{A.24}$$

where $v = x - z\pi$, $M_\alpha = I - \alpha\,(\alpha'\alpha)\,\alpha'$.

The solution to the quadratic form for the last term in (A.24) exists if and only if $M_v x \neq 0$, i.e. $P_v x \neq x$. This condition holds when the model is identified, i.e. $\pi \neq 0$. Assuming this condition, define $\hat{\beta} = \left(x^{'} M_v x\right)^{-1} x^{'} M_v y$ and $s_{\hat{\beta}}^2 = y' M_{M_v x} y/(T-1) = (M_x y)^{'} M_{M_v x} \,(M_x y)\,/(T-1)$. Hence the following holds:

$$p\,(\beta, \pi \mid y, x, z) \propto \left(v^{'} v\right)^{-\frac{T}{2}} \left((T-1)\, s_{\hat{\beta}}^2\right)^{-\frac{T}{2}} \left(1 + \frac{\left(\beta - \hat{\beta}\right)^{'} \left(\beta - \hat{\beta}\right)}{(T-1)\, s_{\hat{\beta}}^2 /\,(x^{'} M_v x)}\right)^{-\frac{T}{2}}. \tag{A.25}$$

From (A.25), conditional posterior of β after integrating out Ω is:

$$p\,(\beta \mid \pi, y, x, z) \sim t(\hat{\beta}, s_{\hat{\beta}}^2/(x^{'} M_v x), T-1). \tag{A.26}$$

.

From (A.25) and (A.26), the marginal density for π is:

$$p(\pi \mid y, x, z) = \int p(\beta, \pi \mid y, z)\, d\beta \propto (x' M_v x)^{-\frac{1}{2}} \left(v' v\right)^{-\frac{T}{2}} \left((T-1)\, s_{\hat{\beta}}^2\right)^{-\frac{T-1}{2}}. \quad \text{(A.27)}$$

We simplify the first and the last terms in equation (A.27) using the properties of projection:

$$x' M_v x = \frac{v' M_x v x' x}{v' v} \propto \frac{v' M_x v}{v' v} \quad \text{(A.28)}$$

The last term in (A.27) can be written as:

$$(T-1)\, s_{\hat{\beta}}^2 \propto (M_x y)' \, M_{M_v x} (M_x y) = (M_x v)' \, M_{M_y x} (M_x v) \frac{y' M_x y}{v' M_x v} \propto v' M_{(y\ x)} v \, (v' M_x v)^{-1} \quad \text{(A.29)}$$

Hence (A.27) becomes:

$$p(\pi \mid y, x, z) \propto \left(v' v\right)^{-T/2} \left((v' M_x v) / v' v\right)^{-1/2} \left(v' M_{(y\ x)} v \, (v' M_x v)^{-1}\right)^{\frac{T-1}{2}}, \quad \text{(A.30)}$$

where the following holds for the projections on the right-hand side of (A.30):

$$M_x v = M_x (x - z\pi) = M_x z\pi \quad \text{(A.31)}$$

$$M_{(y\ x)} v = M_{(y\ x)} (x - z\pi) = M_{(y\ x)} z\pi. \quad \text{(A.32)}$$

Inserting (A.31) and (A.32) in (A.30):

$$p(\pi \mid y, x, z) = (v' v)^{-\frac{T-1}{2}} (\pi' z' M_x z \pi)^{-\frac{1}{2}}, \left(\frac{\pi' z' M_x z \pi}{\pi' z' M_{(y\ x)} z \pi}\right)^{\frac{T-1}{2}} \quad \text{(A.33)}$$

is a t-density form times a rational function, see Kleibergen and Van Dijk (1994, 1998) for the derivation of this posterior density. It is an improper density for an exactly identified model ($k = 1$); and a proper density for an overidentified model ($k > 1$). It is not trivial to simulate from this density.

Possibly improper posterior densities and the effect of the number of instruments For the structural form representation of the model in (A.1) and (A.2), we employ a simple change of variables to show that the joint posterior of β, π, Ω has a ridge under the flat prior. Consider the transformation of variables in the parameter space leading to the OSF representation in (A.3) and (A.4): $\eta = \omega_{12}/\omega_{22}$, $\sigma_{11} = \omega_{11} - \omega_{12}^2/\omega_{22}$, $\sigma_{22} = \omega_{22}$, $\beta = \beta$, $\pi = \pi$. Note that, by definition, $(\eta, \beta, \pi) \in \mathbb{R}^{k+2}$, $(\sigma_{11}, \sigma_{22}) \in \mathbb{R}_+^2$. The determinant of the Jacobian of this transformation is $|J| = 1/\sigma_{22}$.

Define the diagonal matrix $\Sigma = \left(\begin{smallmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{smallmatrix}\right)$ such that $\Omega = \left(\begin{smallmatrix} 1 & \eta \\ 0 & 1 \end{smallmatrix}\right) \Sigma \left(\begin{smallmatrix} 1 & 0 \\ \eta & 1 \end{smallmatrix}\right)$, and let $\phi(.; \mu, V)$ denote the (multivariate) normal density with mean μ and covariance V. The likelihood for observation i is:

$$p(y_i, x_i \mid \beta, \pi, \sigma_{11}, \eta, \sigma_{22}, z_i) = \phi\left(\begin{pmatrix} 1 & -\eta \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y_i - x_i\beta \\ x_i - z_i\pi \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right).$$

Using the independence assumption and the properties of the diagonal matrix Σ, the likelihood for all data points is:

$$p(y, x \mid \beta, \pi, \sigma_{11}, \eta, \sigma_{22}, z) = \phi\left(y - x\beta - x\eta + z\pi\eta; 0, \sigma_{11}I\right)\phi\left(x - z\pi; 0, \sigma_{22}I\right). \tag{A.34}$$

Consider a (general) flat prior: $|\Omega|^{-h/2}$. In the transformed parameter space we have

$$p(\beta, \pi, \eta, \sigma_{11}, \sigma_{22}) \propto |\Sigma|^{-h/2}/\sigma_{22}, \tag{A.35}$$

where we use the Jacobian of the transformation and the following equality: $|\Sigma| = \sigma_{11}\sigma_{22} = |\Omega|$.

Combining the prior in (A.35) and the likelihood in (A.34), the posterior density is:

$$\begin{aligned} p(\beta, \pi, \sigma_{11}, \eta, \sigma_{22} \mid y, x, z) \propto & |\Sigma|^{-(T+h)/2}/\sigma_{22}\,\phi\left(y - x\beta - x\eta + z\pi\eta; 0, \sigma_{11}I\right) \\ & \times \phi\left(x - z\pi; 0, \sigma_{22}I\right). \end{aligned} \tag{A.36}$$

We next show that the joint posterior in (A.36) has a ridge in the parameter subspace $\pi = 0$, and $\beta + \eta = C$ for a constant $C \in \mathbb{R}$:

$$\begin{aligned} p(\beta, \pi, \sigma_{11}, \eta, \sigma_{22} \mid y, x, z, \pi = 0, \beta + \eta = C) = & |\Sigma|^{-(T+h)/2}/\sigma_{22}\phi\left(y - xC; 0, \sigma_{11}\right) \\ & \times \phi\left(x; 0, \sigma_{22}\right), \end{aligned} \tag{A.37}$$

where, given $(\sigma_{11}, \sigma_{22})$, the right-hand-side is a non-zero constant for (infinitely many) points $(\beta, \eta, \pi) \in \mathbb{R}^3$ on satisfying $\beta + \eta = C, \pi = 0$. That is, in the 3-dimensional space of the OSF parameters (β, η, π) this corresponds to the straight line with $\{\beta + \eta = C, \pi = 0\}$; in the 5-dimensional space of $(\beta, \eta, \pi, \sigma_{11}, \sigma_{22})$ this obviously amounts to a 3-dimensional subspace of infinite volume. For the original parameter space of the structural form, the subspace of this ridge corresponds to a more complex, non-trivially curved subspace. The effect of this ridge on the properness of the joint posterior depends on the number of instruments.

In line with the ridge of the posterior density under flat priors, the joint density of β, π, Ω is possibly improper. We focus on the properness (i.e., the integrability) of the marginal posterior of π for the case of a single instrument and for $k > 1$ instruments. Under flat priors, the marginal posterior of π is:

$$p(\pi \mid y, x, z) = (v'v)^{-\frac{T-1}{2}}\left(\frac{\pi' z' M_x z\pi}{\pi' z' M_{(y\ x)} z\pi}\right)^{\frac{T-1}{2}}(\pi' z' M_x z\pi)^{-1/2}, \tag{A.38}$$

where the first term on the right-hand-side of (A.38) is a (k-dimensional) t-density form with $T-1-k$ degrees of freedom, the second factor is a rational function and also the last term is a (simple) rational function. For the first factor we only assume that our number of observations is not extremely small (e.g. $T > k+1$ suffices, which is obviously typically satisfied). The second factor in (A.38) is a Rayleigh quotient, bounded from above and below. Further, the first factor in (A.38) is bounded from above and below by definition. It can be shown that for $\pi \to -\infty$ or $\pi \to +\infty$, the first factor in (A.38) goes to 0 quickly. Hence the density is rather well-behaved in these regions. The main problem of integrability in (A.38) is the behavior of the last factor in (A.38) around the space $\pi = 0_k$ where 0_k is the $k \times 1$ vector of zeros. For the IV model, this space corresponds to irrelevant instruments, where the endogenous variable cannot be explained by any of the instruments.

We have

$$\int_{\pi \in A} (\pi' z' M_x z \pi)^{-1/2} d\pi = |z' M_x z|^{-1/2} \int_{\{\pi^* | \pi^{*\prime} \pi^* \leq 1\}} (\pi^{*\prime} \pi^*)^{-1/2} d\pi^*, \tag{A.39}$$

where A is a certain subspace of $\mathbb{R}^k$ of which 0 is an interior point and where $\pi^* \equiv (z' M_x z)^{1/2} \pi$. For $k = 1$ the non-integrability of the posterior is clear, as the integral on the right hand side of (A.39) amounts to

$$\int_{-1}^{1} \frac{1}{|\pi^*|} \, d\pi^* = \infty. \tag{A.40}$$

For $k = 2$ we have

$$\int_{\{\pi^* | \pi^{*\prime} \pi^* \leq 1\}} (\pi^{*\prime} \pi^*)^{-1/2} d\pi^* = \pi + \int_{1}^{\infty} \pi \, \frac{1}{f^2} \, df = \pi + \left[-\pi \frac{1}{f} \right]_{1}^{\infty} = \pi + (0 - (-\pi)) = 2\pi, \tag{A.41}$$

with $\pi = 3.14159\ldots$ (as opposed to the $k \times 1$ vector of coefficients at the instruments throughout the text) on the right hand side. Here we used that the volume under the graph of $f = (\pi^{*\prime} \pi^*)^{-1/2}$ at $\{\pi^* | \, \pi^{*\prime} \pi^* \leq 1\}$ can be computed by integrating the surfaces $\pi \frac{1}{f^2}$ of circles with radius $\frac{1}{f}$ for $1 \leq f < \infty$ and the surfaces π of circles with radius 1 for $0 \leq f < 1$. For $k \geq 3$ a similar derivation involving an integral over k-dimensional balls yields a different *finite* value.

Hence the existence of the marginal posterior of π for $k \geq 2$, so that the joint posterior of β, π, Ω depends on the number of instruments.

We finally focus on the properness of the marginal posterior of β for the case of a single instrument and for $k > 1$ instruments. Under flat priors, the marginal posterior of β is:

$$p\left(\beta \mid x, y, z\right) \propto \left(\frac{u' u}{u' M_z u} \right)^{-\frac{T-1}{2}} \left(u' M_z u\right)^{-\frac{k-1+1}{2}}, \tag{A.42}$$

where the first factor in (A.42) is smaller than 1, since $u'u > u' M_z u$. The second factor is a t-density form with $k - 1$ degrees of freedom. For $k = 1$ the tails are too fat to have a finite integral, so that the kernel in (A.42) does not correspond to a finite density. For $k \geq 2$ the tails correspond to a proper t-density, so that (A.42) defines a proper density.