Bayesian Analysis (2017) **12**, Number 3, pp. 879–917

Bayesian Analysis of Boundary and Near-Boundary Evidence in Econometric Models with Reduced Rank*

Nalan Baştürk[†], Lennart Hoogerheide[‡], and Herman K. van Dijk[§¶]

"Why econometrics should always and everywhere be Bayesian"
— C. Sims (2007)

Abstract. Weak empirical evidence near and at the boundary of the parameter region is a predominant feature in econometric models. Examples are macroeconometric models with weak information on the number of stable relations, microeconometric models measuring connectivity between variables with weak instruments, financial econometric models like the random walk with weak evidence on the efficient market hypothesis and factor models for investment policies with weak information on the number of unobserved factors. A Bayesian analysis is presented of the common issue in these models, which refers to the topic of a reduced rank. Reduced rank is a boundary issue and its effect on the shape of the posteriors of the equation system parameters with a reduced rank is explored systematically. These shapes refer to ridges due to weak identification, fat tails and multimodality. Discussing several alternative routes to construct regularization priors, we show that flat posterior surfaces are integrable even though the marginal posterior tends to infinity if the parameters tend to the values corresponding to local non-identification. We introduce a lasso type shrinkage prior combined with orthogonal normalization which restricts the range of the parameters in a plausible way. This can be combined with other shrinkage, smoothness and data based priors using training samples or dummy observations. Using such classes of priors, it is shown how conditional probabilities of evidence near and at the boundary can be evaluated effectively. These results allow for Bayesian inference using mixtures of posteriors under the boundary state and the near-boundary state. The approach is applied to the estimation of education-income effect in all states of the US economy. The empirical results indicate that there exist substantial differences of this effect between almost all states. This may affect important national and state-wise policies on required length of education. The use of the proposed approach may, in general, lead to more accurate forecasting and decision analysis in other problems in economics, finance and marketing.

*This paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. The authors are indebted to Lukasz Gatarek and Richard Kleijn for expert research assistance. We also thank the editor, two anonymous referees and Mattias Villani for their comments on an earlier version of this paper.

[†]Maastricht University, n.basturk@maastrichtuniversity.nl
[‡]VU University Amsterdam, l.f.hoogerheide@vu.nl
[§]Econometric Institute, Erasmus University Rotterdam and Norges Bank, hkvandijk@ese.eur.nl
[¶]Corresponding author: hkvandijk@ese.eur.nl.

 DOI: 10.1214/17-BA1061

1 Introduction

Inference near and at the boundary of the parameter space of a probability model is occurring frequently in the field of econometrics. We list three economic and financial topics where (near-)boundary evidence became empirically relevant in the second halve of the twentieth century and it led to important econometric research. In micro-econometrics the estimation of the effect of length of education on earned income encountered the (near-)boundary of weak or no endogeneity and/or weak or no identification. In macro-econometrics investigating which and how many stable relations exist between macroeconomic time series has been extensively explored in order to estimate forecast and policy uncertainty. Here moving to the boundary refers to going from near-nonstationarity to unit roots. In financial econometrics efficient data reduction using large cross sectional data on stocks was investigated using a certain number of unobserved factors which affect, for instance, equity momentum strategies. Weak information on the number of factors is a near-boundary issue. To motivate our analysis, we provide in Section 2 several illustrative examples also for more general model structures.

The literature dealing with these issues is substantial and an extensive overview is outside the scope of this paper. In the frequentist econometric literature the focus has been largely on testing whether one's view is at the boundary and on assessing what is the sensitivity of the test when one is near the boundary. We restrict ourselves to listing three classic tests: the Anderson–Rubin test for (over-)identification which is regularly used in the literature on the education-income analysis (Anderson and Rubin, 1950); the Johansen test used for determining the number of stable relations in macro-economic time series (Johansen, 1991); and the Anderson–Rubin test for determining the number of factors (Anderson and Rubin, 1956)

The major message of the present paper is that many modeling, forecasting and policy problems in non-experimental empirical econometrics are not about asymptotically valid parameter estimation and testing near or at a boundary. Given several different sources of information on features of economic processes, the relevant issue is to use this information and average over the available evidence on the different states of the economy, near and at the boundary, where the evidence on these states is measured using posterior probability weights. The Bayesian approach is eminently suitable for this. We take the viewpoint that the scientific evidence should be reported in such a way that the information specified in the likelihood dominates with respect to other sources of information, see Baştürk et al. (2014a) for a historical background. Thus our approach to specifying prior information is one where relatively weak information is used compared to that of the likelihood.

In order to back-up the general message, this paper makes four points. The first is to show that there exists a common structure in the three issues mentioned and that the effect of the boundary issue on the shape of the posterior densities of the model parameters can be studied within the context of a standard *reduced rank* regression model under different restrictions on the parametric structure and alternative choices of weak priors. It is well-known that the shape of the likelihood, and therefore the shape of the posterior with a flat prior, in the standard multivariate regression model is bell-shaped or elliptical. As a consequence, credibility regions of parameters can be simply determined using second order moments. However, the posterior density of the matrix of equation system parameters in a reduced rank model is non-elliptical. We provide in Section 2 several motivating examples. This nonstandard shape refers to several typical features. We focus on two features that have an effect on the existence of posterior moments: a ridge or, more generally, flat parts in the surface and heavy tails. A ridge refers to weak or non-identification of parameters and it makes a marginal posterior density unbounded, while very heavy tails make the use of first and higher order moments unsuitable for all inference. We will show in Section 3 that the posterior in a standard or *workhorse* reduced rank model, which in our case is a cointegration model, is locally integrable even in the case of a flat prior with flat parts in the posterior surface and the tails are heavy but also integrable. Therefore, the search for plausible restrictions on the parameter space has become an important topic of research. Apart from this research line, we also show that using triangular restrictions on the parameters modify the workhorse model into an instrumental variable regression model and that a normal prior on some equation parameters together with a diagonal covariance matrix on the disturbances modify the workhorse model into a static factor model. We will show that these typical restrictions help in making a posterior with a flat prior more regular with existence of first and higher order moments.

We note that, given the structure of our three types of reduced rank models, multimodality and skewness (of multiple parameters) are more computational problems about numerical evaluation of the posterior but not about the existence. More complex mixture models may give existence problems due to weak empirical identification of a component of the mixture, see for instance Frühwirth-Schnatter (2006).

A second purpose of the paper is to discuss alternative ways that appeared in the literature of specifying *prior regularization* information. This is helpful for determining model weights. One way is to use a more technical econometric approach. That is, construct priors that are based on information or reference theory concepts connected to the identification issue. However, we shall argue that these priors are in many cases not sufficient for making posteriors proper. We add in Section 4 a new result on the existence of the posterior distribution of model parameters with a reduced rank where the regularizing prior information is based on weak and plausible restrictions on the *range* of the parameters of interest. We introduce a *lasso type shrinkage prior combined with orthogonal normalization*. We also, briefly, explore several other routes that deal with regularizing prior information. The focus is then more on prior information that makes *economic models behave more reasonably*, see Sims (2008). That is, one may be more interested in regular behavior of a nonlinear function of the equation system parameters like the impulse response function of a model after a shock. Here the implications of prior information for posterior and predictive analysis are important. Other examples are the effect of prior information on multipliers of an econometric model, which is prior-predictive analysis and such an effect on posterior estimates of stability of a model, which refers to posterior-predictive analysis.

A third purpose of this paper is to show how the evaluation of conditional probabilities on the evidence of different states of an econometric model can be made operational when the prior information is weak. That is, although the issue of weak identification is not an impediment for obtaining a proper probability, weak prior information and a nearly flat posterior do play a major role in the evaluation of posterior and predictive probabilities of evidence near and at a boundary of non-identification and irrelevant instruments. Given the bounded regions of integration, the Bartlett/Jeffreys/Lindley paradox, see Jeffreys (1939), Lindley (1957) and Bartlett (1957) does not show up as a mathematical statistical result, but it appears as a serious practical problem for model evaluation when prior probabilities are assumed over regions where there is weak or no data information. Here the use of a training sample and weak economic information is recommended. Second, a sensitivity analysis is recommended in order to obtain more robustness in the results. We explore several routes that are described in Section 5. Once a model weight is obtained, Bayesian inference can proceed with model averaging in order to estimate mixtures of models suitable for forecasting and policy analysis.

As a final contribution, in Section 6 we explore the regional differences between all states of the US with respect to the effect of length of education on earned income using an instrumental variables model and a mixture of endogenous and weakly exogenous states of the model. We obtain strong empirical evidence that the financial income returns of education vary substantially between almost all states in the USA. This may affect important state and national policies on the requires length of education.

We emphasize that there is much more done on the topic of model averaging in Bayesian econometrics, a recent example in the field of macroeconomics is given in Strachan and Van Dijk (2013). We refer to the Handbook of Bayesian Econometrics, Geweke et al. (2011), and to the Supplementary Material (Baştürk et al., 2017) in the Online Appendix for more examples in the fields of economics, finance and marketing. In Section 7 several perspectives for further research are presented.

Remark 1. Given the length of this paper which is due to a combination of its survey character as well as presentation of new results, the material is divided into a main text and Supplementary Material which is in the Online Appendix.

Remark 2. The development of efficient computational procedures using simulation-based methods has been essential and an active area of research in Bayesian econometrics but it is a topic beyond the scope of this paper. For a historical analysis of the development of this topic since the early nineteen-seventies we refer to Baştürk et al. (2014a). Modern hardware and software including parallel computation allow detailed analysis of many of the issues listed in this paper.

Remark 3. Bayesian inference of mixture processes is extensively studied in the statistical literature, see *e.g.* Frühwirth-Schnatter (2006) and Mengersen et al. (2011). In this paper we focus on the issues that refer to the evidence near and at the boundary of *econometric* models and how to average over these states.

2 Motivating examples

In this section we provide several motivating examples of the boundary and near-boundary issues and the irregular likelihoods resulting in these examples. One econometric model, the cointegration model, serves as workhorse model for reduced rank analysis in this paper. Two other models, the instrumental variable and the factor model, are

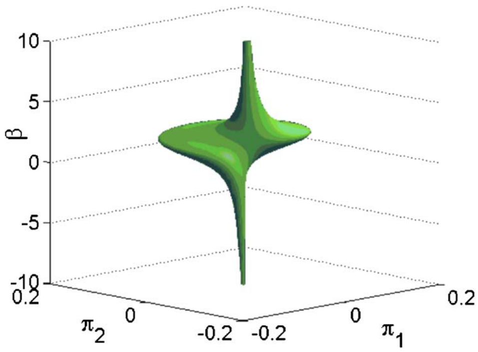


Figure 1: 95% HPD credible set for π_1, π_2, β for simulated data from the IV model.

special cases of the workhorse model. We illustrate the boundary and near-boundary issue for the cointegration and instrumental variable models using simulated and real data. In addition to these motivating examples, we provide three other empirical applications where the boundary issue is evident in the Supplementary Material.

Posteriors of an instrumental variables (IV) model The restricted reduced form of an IV model for data y_i with one explanatory variable x_i and two instruments (z_{1i}, z_{2i}) can be written as follows:

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} = \begin{pmatrix} \beta \\ 1 \end{pmatrix} \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} z_{1i} \\ z_{2i} \end{pmatrix} + \begin{pmatrix} u_i \\ v_i \end{pmatrix}, \tag{1}$$

where β, π_1 and π_2 are scalar model parameters, and disturbances $(u_i \quad v_i)'$ have e.g. an iid normal distribution. This restricted reduced formulation of the model clearly shows the reduced rank structure within this class of models.

Under flat priors, the posterior distribution of the model parameters for the above IV model has a ridge at the region implying 'a move from weak to irrelevant instruments', where $\pi_1 = \pi_2 = 0$. We illustrate this issue in Figure 1. More details are given in the Supplementary material, in Hoogerheide et al. (2007) and Zellner et al. (2014).

Posteriors of a cointegration model The second model we consider is a cointegration model, specifically a Vector Error Correction Model (VECM), with data $y_{1,t}, y_{2,t}$:

$$\begin{pmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 & -\beta \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \tag{2}$$

where $(\alpha_1, \alpha_2, \beta)$ are the model parameters, the disturbances $(\varepsilon_{1,t}\, \varepsilon_{2,t})'$ have iid normal distributions. Similar to the earlier IV model formulation, the reduced rank issue is evident in the matrix multiplication on the right hand side of this model.

The boundary issue for the posterior distributions for the cointegration model under diffuse priors is illustrated in Figure 2. In this case, the 'boundary' corresponds to the

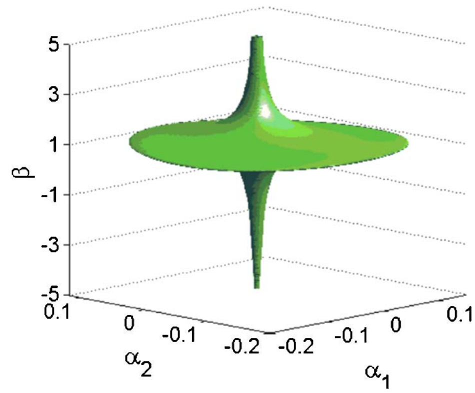


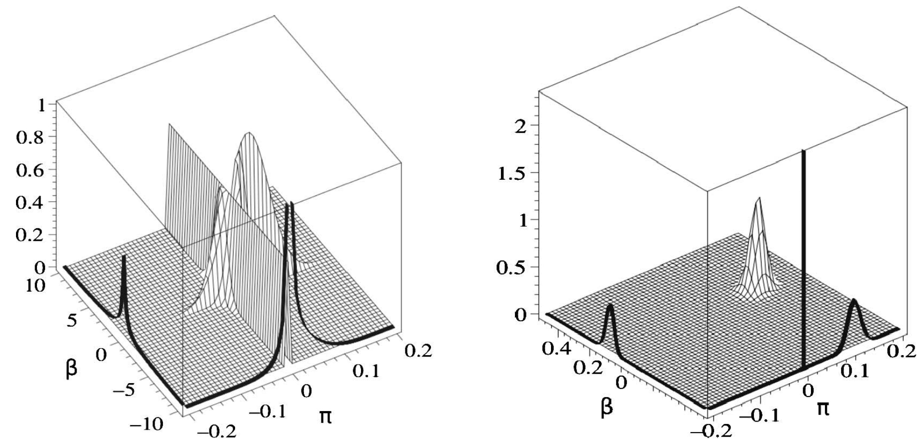
Figure 2: 95% HPD credible set for $\alpha_1, \alpha_2, \beta$ for simulated data from the VECM.

case where there is no dynamic adjustment in the model towards an equilibrium, i.e. $\alpha_1 = \alpha_2 = 0$.

In the Supplementary material the set-up of the experiments for Figures 1 and 2 is given.

Education-income analysis using the IV model As a first empirical motivating example, we present the posterior density of the parameters of an instrumental variables model for education and income data from individuals living in the US, which are analyzed in Angrist and Krueger (1991) and Hoogerheide and Van Dijk (2008) among others. The fundamental issue is that years of education in these data are instrumented with a dummy variable for individuals born in quarters 2–4 of a year. Quarter of birth had an effect on the years of compulsory schooling, due to the compulsory schooling laws. These data represent a typical 'weak instrument' case since the explanatory power of quarter of birth on education is expected to be present only for individuals whose years of education were affected by the compulsory schooling requirement. We refer to the Supplementary Material in Appendix A.1 for an introduction and more explanations of the instrumental variable model.

Figure 3 illustrates the boundary issue which refers to local non-identification of the posteriors under flat priors for the income-education data of the state of New York and the whole US. The two figures of the joint posterior kernels in the model with the effect of education on income (β) and the effect of quarter of birth differences on education (Π) show a substantial 'ridge' in the posterior. For New York data, this ridge is visible at $\Pi = 0$, which dominates the marginal posterior of Π. On the other hand, for the US data, the shapes are nearly elliptical, which reflects that in this case the quarter-of-birth instrument is less weak. The peak around the posterior mode is high compared with the ridge around $\pi = 0$, so that the latter is not visible in the joint posterior density kernel (even though the marginal posterior of π tends to ∞ for $\pi \to 0$). We will show in Section 3 and the Supplementary material A.3.2 that the ridge is integrable but the bimodality is a serious issue for simple inference using only a second moment to



data of New York (29015 observations) data of USA (329509 observations)

Figure 3: Posterior density kernels for simple instrumental variables models for the effects of education on income (β) using the difference in mean education between men born in quarters 2–4 and quarter 1 (π). The model is applied to Angrist and Krueger (1991) data on income and education.

measure estimation uncertainty. We refer here also to the Supplementary Material for more empirical examples.

We end this section by summarizing the issue: our motivation for more methodological analysis is that non-elliptical shapes appear in much of the non-experimental empirical econometric analysis. Possible causes of typical shapes need to be studied.

As an important note we emphasize that is it not easy and probably not a good strategy to perform a conjugate analysis when the likelihood is not regular. Since conjugacy would involve some prior irregularity in this context.

3 Basic model structures, nonstandard likelihood shapes and posterior existence

3.1 Common structure of three reduced rank regression models and summary of posterior existence results

In this section we start to investigate the effect of a reduced rank on the likelihood shape and existence of a posterior within the context of a cointegration model. This model serves as our workhorse model since it can be interpreted as a multivariate regression model where the matrix of equation parameters has reduced rank, see the middle of Figure 4. Using an improper flat prior and linear normalization, it is clear from the cointegrated equation system that a value of $\alpha = 0$ results in a ridge in the parameter space. We will show that this feature leads to an unbounded marginal posterior that is

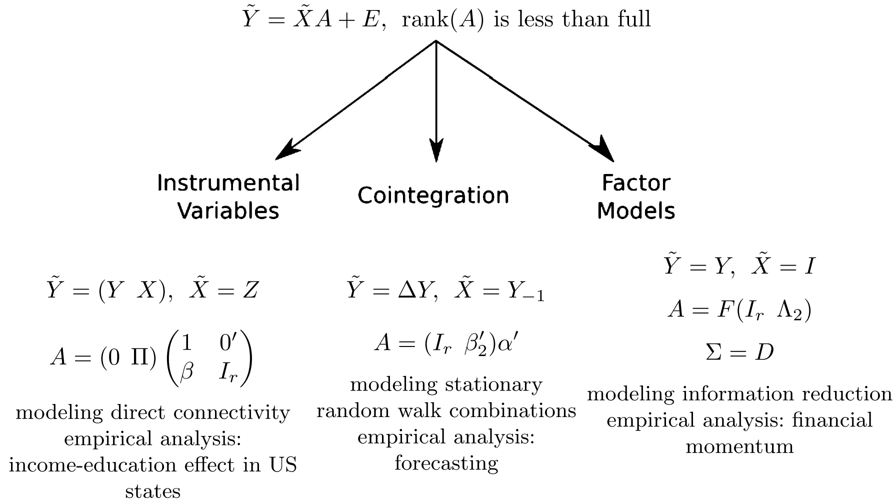


Figure 4: Common structure of three reduced rank econometric models: General structure of reduced rank regression models with linear normalization/identification.

however integrable on a finite region around $\alpha = 0$. We further show that the posterior of α has heavy tails but the density is proper. We note that all conditional distributions are proper with first and higher order moments. We emphasize that the posterior of this cointegration model has the same features as the posterior of a full system Simultaneous Equations Model, an Error in Variables model, and a Static Factor model with no prior information on the factors.

We investigate in the Supplementary Material A.3.3 the effect of imposing a lower triangular structure on the equation system parameters. It is interesting to observe that we can then move from the workhorse model to the so-called Instrumental Variable (IV) regression model, see the left side of Figure 4. Given this triangular structure, we show that the posterior with a flat prior, which leads to a ridge in the posterior surface when the matrix $\Pi = 0$, is a proper density for the case of enough instrumental variables. A large number of instruments makes the tail behavior of the posterior more regular with existence of first and higher order moments. Thus an improper prior yields in this situation a much more regular posterior. The case of many instruments and that of weak endogeneity versus strong endogeneity together with weak and strong identification are all analyzed. We note that there exists an analogy with a triangular cointegraton system, see Martin and Martin (2000).

Thirdly, we explore, also in the Supplementary Material A.3.4, the case where the covariance matrix of the disturbances is diagonal together with the assumption of a standard normal prior on the matrix β. Now, we can move from the workhorse model to a static factor model, see the model on the right of Figure 4. Here the matrix of the unobserved factors F plays the same role as the matrix β in the cointegration model. Similarly the matrix Λ in the factor model has the same role as the matrix α

in the cointegration model. When one adds the normal assumption and the one of a diagonal covariance of the disturbances then the posterior with a flat prior is proper. We emphasize that the effect of a diagonal covariance matrix within an IV model yields well behaved student t posterior densities.[1]

There exist several lines of criticisms on our use of flat priors and linear normalization. It is well-known that the posterior results using a linear normalization may, in an empirical analysis, be sensitive for the ordering of the variables. In the case of IV this ordering is natural since one is mainly interested in the effect that a possibly endogenous explanatory variable may have on the left hand side endogenous variable (years of education on earned income). But in cointegration and factor models one is often symmetric between variables or factors. Then orthogonal or orthonormal normalization is interesting to explore. We investigate that in Section 4. Second, a uniform prior on parameters is not invariant to a transformation. It is very important that one specifies the prior information on the parameter that reflects the issue of interest. We will also explore this issue more in Section 4 and in the Supplementary material.

3.2 Likelihood shape and existence of posterior in a workhorse reduced rank model: the case of cointegration

A cointegration model constitutes a general class of a reduced rank regression model. Special cases with different restrictions on the parametric structure are covered in the Supplementary Material for the instrumental variable regression model and the static factor model.

3.3 Posterior of a standard cointegration model under linear normalization and a diffuse prior

A Vector AutoRegressive (VAR) model of lag order 1 is usually specified as

$$y_t = \Phi y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \sim NID(0, \Sigma), \text{ for } t = 1, \ldots, T, \tag{3}$$

where y_t is $k \times 1$ dimensional vector of observations on economic variables (in deviation from their mean) at time t; Φ is a $k \times k$ matrix of parameters belonging to the observations on the lagged endogenous variables; the disturbances ε_t for $t = 1, \ldots, T$ have independent Gaussian distributions with Σ a positive definite symmetric (PDS) parameter matrix. Observations on y_0 are given as initial values. A basic paper on this VAR model is Sims (1980). For a general introduction to the class of models we refer also to Johansen (1995).

The VAR model (3) can be cast into the Vector Error Correction Model (VECM) as follows:

$$\Delta y_t = \Pi' y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \sim NID(0, \Sigma), \text{ for } t = 1, \ldots, T, \tag{4}$$

[1] We note that due to the similarity of three model structures, one can prove the equivalence of the Anderson–Rubin test for overidentification and the Johansen test for cointegration. For details, see Hoogerheide and Van Dijk (2001).

where $\Pi' = \Phi - I_k$. In matrix notation, this error correction model can be specified as:

$$\Delta Y = Y_{-1}\Pi + E, \tag{5}$$

where ΔY is a $T \times k$ matrix of observations Δy_1 to Δy_T in its rows and similarly, Y_{-1} is a $T \times k$ matrix of observations containing y_0 to y_{T-1} in its rows. The $T \times k$ random matrix E has a matrix-variate distribution, $E \sim MN\left(0, I_T, \Sigma\right)$.

Stationarity of the process corresponds to Π having full rank. Then all series converge to a finite long run mean and have a bounded variance in the long run. When Π has rank 0, a k-dimensional random walk occurs. The long run mean is equal to the next period mean and long run variance tends to infinity. The more interesting case is where the process $\{y_t\}$ has a so-called *cointegrating* rank r, that is, when Π has rank $r < k$. In this case one has r cointegrating or otherwise stated r stable relations between k economic variables and the matrix Π can be specified as the product of two $k \times r$ matrices α and β with full column rank and $\Pi = \beta\alpha'$.

The resulting model is called a *cointegrating* VECM, which in matrix notation takes the following form:

$$\Delta Y = Y_{-1}\beta\alpha' + E. \tag{6}$$

The number of parameters in α and β together may be larger than the number of free parameters in Π under a rank restriction. For the case of k variables and $r \leq k$ cointegrating relations, it holds for any $(r \times r)$ non-singular matrix R that:

$$\Pi = \beta\alpha' = (\beta R)(\alpha R^{-1})',$$

with $\text{rank}(\beta) = \text{rank}(\beta R)$ and $\text{rank}(\alpha) = \text{rank}(\alpha R^{-1})$. That is, the parameters β and α are non-identified. A straightforward way of identifying the parameters is by using a linear normalization on β as restriction:

$$\beta = \begin{pmatrix} I_r \\ \beta_2 \end{pmatrix}, \tag{7}$$

where β_2 is a $(k - r) \times r$ matrix, see Kleibergen and Van Dijk (1994); Kleibergen and Paap (2002) among others. We will consider as an alternative in Section 4.2 the case of orthogonal normalization.

Consider a diffuse class of priors defined on the space of (α, β_2) and on the space of positive definite matrices Σ given as $p(\alpha, \beta_2, \Sigma) \propto |\Sigma|^{-h/2}, h > 1$. We make use of the prior value $h = k + 1$, which gives an equivalence between the marginal posterior of (α, β_2) and their, so-called, concentrated likelihood function. We discuss the effect of a more general choice of h later.

The posterior density (apart from the integrating constant) under the normalization is obtained by multiplying the likelihood and the diffuse prior which yields:

$$p\left(\alpha, \beta_2, \Sigma \mid Y\right) \propto |\Sigma|^{-(T+k+1)/2} \exp\left[-\frac{1}{2}\text{tr}\left\{\Sigma^{-1}\left(\Delta Y - Y_{-1}\beta\alpha'\right)'\left(\Delta Y - Y_{-1}\beta\alpha'\right)\right\}\right]. \tag{8}$$

We note that for notational convenience, we make use of only the symbol Y to denote the data $(\Delta Y, Y_{-1})$.

In the previous section it is shown empirically that the shape of such a posterior (more precisely the marginal one after integrating out Σ) is such that there exists a ridge in the surface when $\alpha = 0$. We will show analytically that this feature leads to an unbounded marginal posterior that is however integrable and, further, that the tails are heavy but the posterior remains integrable. It is noteworthy that all conditionals are proper density function with first and higher order moments.

Marginal and conditional posterior densities We consider marginal and conditional posterior density functions of the parameters under a diffuse prior and discuss existence conditions for the posterior distributions and their first and higher order moments. A summary of the derivations and results is presented in Figure 5. For details on the derivation we refer to the online Appendix A.3.2. We note that our results are quite general and several are, to best of our knowledge, novel.

Marginal densities of α and β_2 after integrating out Σ Application of the inverse-Wishart integration step yields the joint posterior distribution of (α, β_2) with density:

$$p\left(\alpha, \beta_2 \mid Y\right) \propto \left|\left(\Delta Y - Y_{-1}\beta\alpha'\right)'\left(\Delta Y - Y_{-1}\beta\alpha'\right)\right|^{-T/2}. \tag{9}$$

Exact expressions of the conditional densities which are of the *matrix-t class* are presented in Appendix A.3.2.

Marginal posterior of β_2 and existence of moments From (9), using a matrix-t density step on α and applying a matrix decomposition and properties of the projection matrix, as presented in Appendix A.3.1 and A.3.2, one can obtain the following result:

Proposition. *Given the standard form of a cointegration model under linear normalization and using a diffuse class of priors, the marginal posterior of the cointegration parameters β_2 is proportional to a matrix-t density times a polynomial in β_2. This density is proper, independent of the cointegrating rank r, but no first or higher order moments exist.*

It is noteworthy that this result is also independent of the difference $k - r$. We come back in the case of the IV model, presented in the Online Appendix. This result extends the analysis and results of Kleibergen and Van Dijk (1994). We further note that the choice of the prior parameter h does not play a role in the existence condition for the distribution function.

Marginal posterior of α and existence of moments It is shown in Appendix A.3.1 and A.3.2 that using a matrix-t density step on β_2 and applying a matrix decomposition and properties of the projection matrix presented in that appendix, one can obtain the marginal posterior density of α.

Proposition. *Given the standard form of a cointegration model under linear normalization and using a diffuse class of priors, the marginal posterior density of the adjustment parameters α is a rational function in α and this density is not proportional to a known form of densities.*

Model and posterior

$$\Delta Y = Y_{-1}\beta\alpha' + E, \;\; E \sim N(0, \Sigma \otimes I_T)$$

Identification restriction is linear normalization on β

$$\Delta Y = Y_{-1}\begin{pmatrix} I_r \\ \beta_2 \end{pmatrix}\alpha' + E, \; \beta_2 \text{ is } (k-r)\times r, \; \alpha \text{ is } k \times r$$

posterior has ridge at $\alpha = 0$*, but joint density is proper*

Conditional posteriors

$p(\alpha, \beta_2, \Sigma \mid Y)$, data $= \{\Delta Y, Y_{-1}\}$ *is summarized as* Y

↓	↓	↓
complete sum of squares in α	complete sum of squares in β_2	use Inverse-Wishart dist.
↓	↓	↓
$p(\alpha \mid \beta_2, \Sigma, Y) \propto$ matrix Normal density	$p(\beta_2 \mid \alpha, \Sigma, Y) \propto$ matrix Normal density	$p(\Sigma \mid \beta_2, \alpha, Y) \propto$ inverse-Wishart density
↓	↓	↓

Conditional moments of $p(\beta_2 \mid \alpha, \Sigma, data)$*,* $p(\alpha \mid \beta_2, \Sigma, Y)$ *and* $p(\Sigma \mid \beta_2, \alpha, Y)$ *exist for all values of the conditioning parameters in their domains and for all finite* k *and* r*.*

Marginal posteriors of α **and** β_2

$p(\alpha, \beta_2, \Sigma \mid Y)$

↓

Inverse-Wishart step on Σ

↓

$$p(\alpha, \beta_2 \mid Y) \propto \left|(\Delta Y - Y_{-1}\beta\alpha')'(\Delta Y - Y_{-1}\beta\alpha')\right|^{-T/2}$$

↓	↓
complete sum of squares on α	complete sum of squares on β_2 in three steps
↓	↓
$p(\alpha \mid \beta_2, Y) \propto$ a matrix t density *conditional moment exist for all values of* β_2 *in its domain*	$p(\beta_2 \mid \alpha, Y) \propto$ a matrix t density *conditional moments exist for all values of* α *in its domain*
↓	↓
matrix t-density step on α	matrix t-density step on β_2
↓	↓
use matrix decomposition and properties of the projection matrix:	use matrix decomposition and properties of the projection matrix:
↓	↓
$p(\beta_2 \mid Y)$ *is proportional to a matrix* t *density times a polynomial in* β_2*. It is a proper density independent of the cointegrating rank r, but no first or higher order moments exist.*	$p(\alpha \mid Y)$ *is a rational polynomial function in* α *and not a member of a known class of densities. It is integrable despite having an asymptote at* $\alpha = 0$*. The tails are heavy but integrable.*

Figure 5: Derivation Scheme for Posterior Densities of a Cointegration model with k variables and $r < k$ cointegrating relations under a diffuse prior.

Existence of the marginal posterior of $\alpha|Y$ It is shown in Appendix A.3.2 that a sufficient condition for the existence of the posterior of α at $\alpha = 0_{(k\times r)}$ is:

$$\int \left|\alpha' D^{-1}\alpha\right|^{-(k-r)/2} d\alpha < \infty, \tag{10}$$

where D is a matrix which only depends on data.

We next analyze two shape features: the asymptote in the interior when $\alpha = 0_{(k\times r)}$ and the tail behavior when α tends to infinity. We show that the determinant in (10) is integrable around $\alpha = 0$ despite the asymptote at $\alpha = 0_{(k\times r)}$ and we show that the tails are heavy but integrable.

2-dimensional vector case $r = 1, k = 2$ For simplicity, consider the integral on a ball A_k with radius R for the special case, $k = 2$, $r = 1$ where for ease of exposition we assume that the data matrices have been scaled and rotated such that $Y_{-1}'Y_{-1} = I_k$:

$$\int_{A_k} |\alpha'\alpha|^{-(k-r)/2} d\alpha. = \iint_{\alpha_1^2+\alpha_2^2\leq R^2} (\alpha_1^2 + \alpha_2^2)^{-1/2} d\alpha_1 d\alpha_2. \tag{11}$$

We perform a polar coordinate transformation of α_1, α_2 to show that the above integral is finite but depends on the value of R. Consider the change of variables:

$$\begin{aligned} \alpha_1 &= \lambda\cos\theta,\ \ \alpha_2 = \lambda\sin\theta \\ \lambda^2 &= \alpha_1^2 + \alpha_2^2,\ \ \theta = \tan^{-1}(\alpha_2/\alpha_1), \end{aligned}$$

where $\theta \in (0, 2\pi]$, $\lambda > 0$ and the determinant of the Jacobian for this change of variables is

$$|J| = \begin{vmatrix} \cos\theta & -\lambda\sin\theta \\ \sin\theta & \lambda\cos\theta \end{vmatrix} = \lambda(\cos^2\theta + \sin^2\theta) = \lambda. \tag{12}$$

With the change of variables, the integral in (11) becomes:

$$\int_{\theta=0}^{2\pi}\int_{\lambda=0}^{R} (\lambda^2)^{-1/2}\lambda d\lambda d\theta = \int_{\theta=0}^{2\pi}\int_{\lambda=0}^{R} 1 d\lambda d\theta = 2\pi R. \tag{13}$$

The integral corresponds to the volume under the graph of $f(\alpha) = (\alpha'\alpha)^{-1/2}$. The volume over the region $\{\alpha|\alpha'\alpha \leq 1\}$ can be computed by integrating the surfaces of circles with radius $f(\alpha)$ for $1 \leq f(\alpha) < \infty$ and the surfaces α of circles with radius 1 for $0 \leq f(\alpha) < 1$. Figure 6 illustrates this: for each function value $f(\alpha) = (\alpha'\alpha)^{-1/2}$ with $f(\alpha)$ as the horizontal 'slice' through the graph is a circle with radius $1/f(\alpha)$. For any finite R the integral is bounded from which we conclude that the asymptote poses no problems. A proof that the asymptote poses no problem for the general vector and the matrix case is presented in the online Appendix A.3.2.

If however R tends to ∞ the integral in (13) also goes to ∞ at a rate R, so that the *sufficient* condition is not satisfied then. However, the tails are integrable and the marginal posterior of α is proper. The easiest way to see this is as follows. We show in

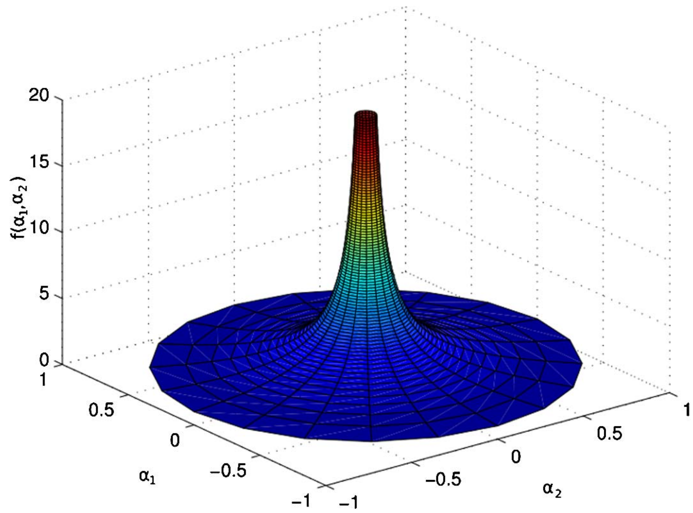


Figure 6: $f(\alpha) = (\alpha'\alpha)^{-1/2}$ for $\alpha'\alpha \leq 1$, where $\alpha = (\alpha_1, \alpha_2)'$.

Appendix 3.2 that the marginal posterior of β_2 is proper but it has no first or higher order moments, see equation (A.66). Further, the conditional posterior of α given β_2 is proper for each value of β_2, see (A.39) and (A.58). Therefore, the joint posterior of (α, β_2) is proper. We could simulate α from its (marginal) posterior by simulating β_2 from its marginal posterior and simulating α given the draw of β_2. We emphasize that the line of reasoning to show that the tails are integrable is a general one. That is, it holds for the bivariate case, the general vector case and the matrix case.

All this leads to the following proposition:

Proposition. *Given the standard form of a cointegration model under linear normalization and using a diffuse class of priors, the marginal posterior density of α, given in Appendix A.3.2, equation (A.72), is integrable despite the fact that it has an asymptote at $\alpha = 0$. The tails are heavy but integrable, so that the marginal posterior density of α is proper.*

This result also holds for the Simultaneous Equations Model when there exist only a few restrictions on the structure, the Errors-in-Variable model and the Static Factor Model with no information on the factors.

General conclusion of Section 3 In this section we have shown that, using a flat prior, Bayesian analysis of a general reduced rank model yields non-elliptical shapes of posteriors that can be classified as: flatness and unboundedness due to weak or non-identification and weak or irrelevant instruments. We further showed that unbounded posteriors are locally integrable under weak conditions and posterior tails are heavy but integrable. These results are to the best of our knowledge new. We will show in the Supplementary Material that by making use of extra restrictions such as a lower triangular matrix of β one can obtain proper posteriors with more desired properties (existence of higher order posterior moments). This is shown in the Supplementary Material for the Instrumental Variable Model. Alternatively, one may use a weakly informative prior such as a normal prior $N(0, cI)$ with c a large constant on α which makes the tails of the posterior of α more regular. This can be seen in the class of factor models, see for instance Geweke (1996).

We note that, given the structure of our three types of models, multi-modality and skewness (of multiple parameters) are more a computational problem about numerical evaluation. More complex mixture models may give an existence problem due to weak empirical identification of a component of the mixture but this is a topic beyond the scope of this paper. In the next section we investigate how regularization priors deal with the two issues of flat regions (unbounded marginals) and heavy tails.

4 Regularization priors

Since the early nineteen-seventies there has been a strong tradition in Bayesian econometrics of studying the shape and integrability of posteriors of parameters of models with a reduced rank under flat priors. The first class of models studied was the Simultaneous Equations Model (SEM) where the issue of endogeneity of explanatory variables was analyzed. One of the early important papers is Drèze (1976) where a posterior density is presented of the parameters of a single SEM equation, marginalized with respect to all parameters in the remaining part of the SEM where no restrictions were imposed. For a detailed explanation of the shape of the likelihood of the full model and of one single equation we refer to Bauwens and Van Dijk (1990). Next, the so-called Incomplete Simultaneous Equations (INSEM) model, see Zellner et al. (1988), was studied from a Bayesian point of view. This model was shown to be a triangular SEM model and to be identical to an IV model. Bauwens and Van Dijk (1990) present a derivation of the marginal posterior of the single equation parameters but do not discuss in detail under what conditions this is a proper density.

In the present section we present a set of priors that are potentially suitable for making posterior densities proper. First, in Section 4.1 we follow an econometric methodological or statistical approach to specifying weak prior information that is intended to make an unbounded posterior more regular by using the information matrix and an other reference approach. In Section 4.2 we present a new result on a lasso type shrinkage prior combined with orthogonal normalization that serves this purpose well. Furthermore, in Section 4.3 we specify prior information that is meant to make economic models behave 'reasonably'. A motivation for the latter property was given by Sims (2008) for the case of macroeconomic models. This can be applied more generally to all economic models.

A final point of this section is that in order to obtain robust results for posterior and predictive analysis with weak prior information, it is recommended to use a sequence of priors with increasing amounts of information starting from very weak prior information. Therefore the contents of this section are organized with listing regularization priors in increasing amount of information.

4.1 Information matrix, subspace and reference priors

Information Matrix and Embedding priors An alternative to using a flat prior on the parameters of a cointegration model (as workhorse model for a reduced rank) is provided by the Information Matrix prior, also known as Jeffreys prior. It is proportional to the square root of the determinant of the information matrix and it can be specified as:

$$p(\Sigma) \propto |\Sigma|^{-(k+1)/2} \tag{14}$$

$$\begin{aligned}
p(\alpha, \beta_2|\Sigma) \propto & |\mathcal{I}(\alpha, \beta_2|\Sigma)|^{\frac{1}{2}} \\
= & \left| \left(\frac{\partial \operatorname{vec}(\Pi)}{\partial(\operatorname{vec}(\alpha)' \ \operatorname{vec}(\beta_2)')} \right)' \mathcal{I}(\Pi|\Sigma) \left(\frac{\partial \operatorname{vec}(\Pi)}{\partial(\operatorname{vec}(\alpha)' \ \operatorname{vec}(\beta_2)')} \right) \right|^{\frac{1}{2}} \\
= & \left| \left(I_n \otimes \beta \quad \alpha' \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \right)' (\Sigma^{-1} \otimes Y_{-1}' Y_{-1}) \right. \\
& \left. \times \left(I_n \otimes \beta \quad \alpha' \otimes \begin{pmatrix} 0 \\ -I_{n-r} \end{pmatrix} \right) \right|^{\frac{1}{2}} \\
\propto & |\beta' Y_{-1} Y_{-1} \beta|^{\frac{1}{2}(k-r)} |\alpha \Sigma^{-1} \alpha'|^{\frac{1}{2}(k-r)} |\Sigma|^{-\frac{1}{2}(k+1)},
\end{aligned} \tag{15}$$

where k is the dimensionality. For a derivation and more details on Jeffreys prior see, Kleibergen and Van Dijk (1994), Uhlig (1994), Kleibergen and Van Dijk (1998), Martin and Martin (2000) and Martin (2001). Both $\mathcal{I}(\alpha, \beta_2|\Sigma)$ and $\mathcal{I}(\Pi|\Sigma)$ denote the conditional information matrices. The distinctive feature of this prior is its ability to annihilate probability mass at points where the identification problem occurs. This result also holds for the instrumental variable model, see the example in Figure 3 in Section 2. To visualize the effects of applying the Information Matrix prior to the likelihood of the cointegration model we present the shape of this prior and the shape of credible sets and the posterior distribution in Figure 7. In the Figures of the prior and posterior density of (α_1, α_2) the activity of Information matrix prior is evident around point $(0, 0)$.

It is clear from the equations and from the figure that Jeffreys prior relates to strength of information on β (long term equilibrium) and α (speed of adjustment). This prior gives no weight to the state where the model is not identified (where the likelihood exhibits a ridge) and it gives more weight to values of the parameters α and β when the likelihood also has some weight. More formally, the Information Matrix or Jeffreys prior is a polynomial in these parameters and the prior density kernel tends to infinity when the parameters tend to infinity. Therefore this class of priors is not suitable as regularization prior in the general case of a reduced rank model where the problem is with the tail behavior. However, this class of priors can be used for the case of the Instrumental Variable regression model where the tail behavior of the likelihood is very regular for a large number of instrumental variables, see the analysis in the Online Appendix A.2.3.

We emphasize that there exists an equivalence between the Jeffreys prior and the prior that stems from the **embedding approach**, see, for instance, Kleibergen and Zivot (2003). In the embedding approach one specifies a flat prior on the unrestricted reduced form and makes use of a transformation of random variables to the parameter

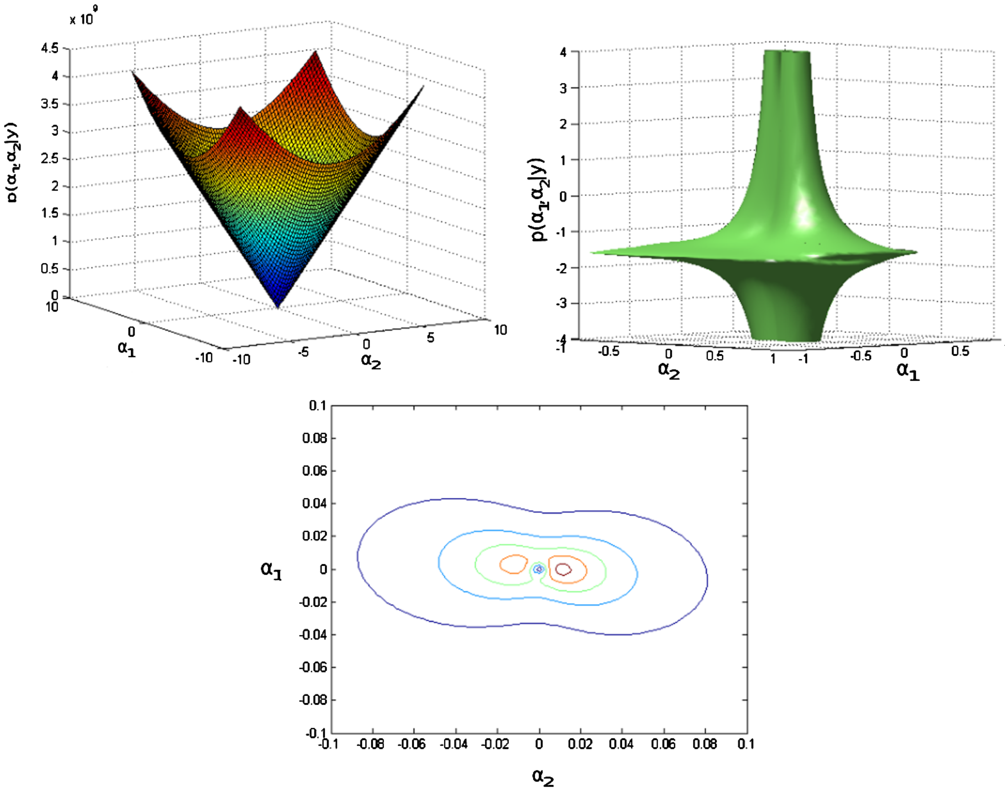


Figure 7: Shape of the Information Matrix or Jeffreys prior, credible sets and posterior distributions under this prior. Data generated from one unit root cointegration model (eigenvalues $\lambda = (0.6074, 1.0)$) with $\alpha = (0.5, -0.0561)'$, $\beta = (-0.6640, 1.0799)'$; $\Pi_1 = \Pi + I = (0.6680,\ 0.5399;\ 0.0373,\ 0.9394)$.

of the structural form. This approach has been used to specify priors for a simultaneous equations model and a co-integration model, see Kleibergen and Van Dijk (1998) and Kleibergen and Paap (2002). For the embedding approach the same conclusion holds as for Jeffreys prior approach. We present an empirical analysis in the Supplementary Material, Appendix A.1.3. Another interesting analysis is presented for this IV model comparing Bayes and GMM by Sims (2007). We refer to that paper for details.

Subspace/Reference based priors Villani (1998), see also Villani (2000), proposed a prior on the subspace spanned by the columns of the matrix with reduced rank using the concept of a Grassmann manifold. This prior was then transformed to a prior on the parameters α and β in the linear normalization case, treated in Section 3, in order to perform Gibbs sampling. Villani (2005) continued this line of work, now labeled as a reference approach but still based on the subspace approach. It gave proper posteriors that are invariant to the ordering of variables.

Strachan and Inder (2004) and Strachan and Van Dijk (2004) applied the subspace approach to the case of orthonormal normalization. This led to a prior of the parameters β defined on a bounded region. These authors developed a sampling algorithm that allowed to sample from the orthonormal normalization. We refer to the survey by Koop et al. (2006) for a more detailed analysis of the subspace/reference approach.

Conclusion Although the technical approaches listed so far are elegant and 'repair' some or all anomalies of the likelihood function of a reduced rank regression model, we take a different direction in the present paper. The reason being that we intend to work with several states of the econometric model, that is, near the boundary of a reduced rank as well as at the boundary. We want to specify a convenient class of priors that yield proper posteriors which can be used to effectively evaluate posterior and/or predictive probabilities at and near a boundary. Further, we discuss priors that explore implications for posterior and predictive probabilities that may be used for prediction and decision analysis, that is, prior- and posterior-predictive and -decision analysis.

4.2 Orthogonal normalization and lasso type shrinkage prior

Given a diffuse prior and under linear normalization we have shown that the marginal posteriors of the parameters of interest of a workhorse reduced rank regression model are not regular in the sense that they do not belong to a known class of densities like the matrix-t densities. We took the cointegration model as an example. We note that in the case of such a model, when the parameter matrix has everywhere full rank the posterior is regular. That occurs when the data in the cointegration model are all stationary. Also in the case when the rank is zero, that is, when all data series are random walks one encounters regular posteriors. We now explore an approach where weak regularizing prior information is introduced that makes use of restrictions, in particular, plausible restrictions on the range of the parameters. For expository purposes we continue with the cointegration model but emphasize that our results hold also for the instrumental variable and factor model with sometimes slight modifications.

Identification and orthogonal normalization In general an $n \times k$ matrix of rank r has $(n+k)r - r^2$ free elements, that is $(n-r)(k-r)$ restrictions. In our case, the $k \times k$ matrix Π has rank r and therefore it has $2kr - r^2$ independent free elements and $(k-r)^2$ restrictions. The matrices α and β in the parametrization $\Pi = \beta\alpha'$ with $\text{rank}(\Pi) = r$ together have $2kr$ elements, which are r^2 too many to identify α and β. The normalization $\beta_1 = I_r$ that we used in the previous sections exactly accounts for the additional r^2 required restrictions. The parametrization $\Pi = \beta\alpha'$ can be linked to the singular value decomposition $\Pi = USV'$, where the rectangular $k \times r$ matrix U is an element of the Stiefel manifold $U'U = I_r$ and the square $r \times r$ matrix V is an element of the manifold of orthogonal matrices $V'V = I_r$. S is a diagonal $r \times r$ matrix with positive diagonal entries equal to the singular values of Π. We denote the vector of these diagonal elements as $\lambda = (\lambda_1, \ldots, \lambda_r)'$. Note that the manifolds on which U and V are defined have finite volume. The manifold on which λ is defined is not bounded and we shall come back to that later.

E.g. Kleibergen and Van Dijk (1998) and Kleibergen and Paap (2002) explicitly link their parametrization to the singular value decomposition and they combine it with the linear restriction $\beta_1 = I_r$. This linear normalization subsequently implies a mapping from these manifolds to Cartesian coordinates in Euclidean space, that is $\alpha \in \mathbb{R}^{k\times r}$ and $\beta_2 \in \mathbb{R}^{(k-r)\times r}$. This mapping thus transforms from manifolds with finite volume (except λ) to unbounded spaces.

Another common normalization of β used in the literature is $\beta'\beta = I_r$. A major motivation for the choice of this orthogonal normalization of the matrix β is that in this case no preferred ordering of the variables is imposed and the region of integration for β is bounded. In the case of a VAR these may be reasonable assumptions in several situations, in particular, when one considers a set of similar price indices or quantity series.

We emphasize that this normalization alone is not sufficient to identify both α and β. This normalization imposes only $r(r+1)/2$ unique restrictions, because of the symmetry of $\beta'\beta$, so an additional $r(r-1)/2$ restrictions are required. One could impose these on β but this should be done with caution in order to avoid the issue of imposing too much structure through the combination of ordering, restricting and assigning a flat prior. For a more information on normalization and identification, we refer to Hamilton et al. (2007).

Lasso type shrinkage prior under orthogonal normalization We propose an approach that more directly uses the structure of the singular value decomposition and also makes use of the concept of lasso type shrinkage priors, see Tibshirani (1996).

As specified above, the singular value decomposition is not uniquely defined. Any simultaneous permutation of the columns of U, S and V also constitutes a singular value decomposition. A common way to avoid this ambiguity is by ordering the singular values that occur on the diagonal of S as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$. We shall use this ordering. Ordering the singular values is also more straightforward than devising an ordering of the columns of U and V directly (or the columns of α and β for that matter).

Because of this ordering each element λ_{i+1} for $i = 1, \dots, r-1$ is bounded by λ_i. Only λ_1 remains unbounded towards $+\infty$. Integrability is thus determined by the behaviour of λ_1.

Having fixed the ordering of the singular values the uniqueness of the singular value decomposition when all λ_i's are different is up to simultaneous sign changes of corresponding columns of U and V which could be mitigated for instance by imposing a positive sign for the first non-zero entry in each column of U. Finally, if a singular value occurs more than once, then the columns of U and V corresponding to these singular values are not uniquely defined. Any other orthonormal basis that spans the same space will also do. Although in this particular case the transformation between the matrix Π and its singular value decomposition (U, S, V) is still not invertible everywhere, this is however an event with zero measure and we observe that the Jacobian of this transformation equals 0 whenever a repeated singular value occurs because then the factor $\lambda_i^2 - \lambda_j^2$ will be 0 for some $i < j$.

We analyse the specification in which we combine $\beta'\beta = I_r$ with $\alpha'\alpha = I_r$ in the parametrization $\Pi = \beta\Lambda\alpha'$ with Λ diagonal. This corresponds directly to the singular value decomposition $\Pi = USV'$ with $\beta = U$, $\alpha = V$ and $\Lambda = S = \text{diag}(\lambda)$. The restriction $\alpha'\alpha = I_r$ imposes $r(r+1)/2$ restrictions which amount to r restrictions more than required, but λ subsequently provides these extra r degrees of freedom.

Λ and α in this parametrization combine into α in the usual parametrization $\Pi = \beta\alpha'$ as in the previous bullet.

The advantage of this specification is that now both α and β have finite support. If the issue of non-integrability arises it will be in the parameter λ, and if so it is also clear they will also have to be repaired in λ.

Regarding the econometric interpretation of the parametrization $\Pi = \beta\Lambda\alpha'$ we may think of $\beta' y_t$ as the deviation from the r cointegrating relations $\beta' y_t = 0$ between the k variables y_t, which is similar to the role of β in the more usual parametrization $\Pi = \beta\alpha'$. The interpretation of λ is that of the rate of adjustment of the system towards each of the r cointegrating relations. α in the parametrization $\Pi = \beta\Lambda\alpha'$ describes the contribution of each of the k variables y_t to the adjustment towards each of these r cointegrating relations. This has advantages over the more usual parametrization $\Pi = \beta\alpha'$ in which the speed of adjustment towards the cointegrating relations is amalgamated with the distribution of these adjustments over the variables into one single parameter matrix (also denoted α).

Each data vector y_t defines a vector in k-dimensional space. The geometric interpretation is that β defines r directions in the space of the data. Λ scales in these directions and α rotates the result to a r dimensional subspace of the data.

To distinguish the parameter matrix α in $\Pi = \beta\Lambda\alpha'$ from the parameter matrix α in the usual parametrization we shall denote the latter by α^* such that $\Pi = \beta\alpha^{*\prime}$ in the remainder of this section. In order to translate results on α and $\Pi = \beta\Lambda\alpha'$ back and forth to α^* and $\Pi = \beta\alpha^{*\prime}$ we now briefly describe how they are related. Both parametrizations are linked by the relation $\alpha^* = \alpha\Lambda$. This can be seen when we combine $\beta'\beta = I_r$ with $\alpha^{*\prime}\alpha^* = S$ in the parametrization $\Pi = \beta\alpha^{*\prime}$ where S is a $r \times r$ diagonal matrix with λ_i, $i = 1, \ldots, r$, as diagonal elements. The relation with the singular value decomposition $\Pi = USV'$ is $\beta = U$, $\alpha^* = VS = \alpha\Lambda$. This also gives exactly the number of required restrictions: all off-diagonal elements of $\alpha^{*\prime}\alpha^*$ are constrained to 0 and because of the symmetry of $\alpha^{*\prime}\alpha^*$ each off-diagonal element occurs twice which results in $r(r-1)/2$ unique restrictions. In terms of the columns α_i^* of α^*: $\alpha_i^{*\prime}\alpha_i^* = \lambda_i^2$ for $i = 1, \ldots, r$ and $\alpha_i^{*\prime}\alpha_j^* = 0$ for $i \neq j$.

Prior choice and existence of posterior moments In Appendix A.3.2 we present a derivation where given that diffuse priors are specified for α and β on their respective Stiefel manifolds and a usual diffuse prior on Σ one can derive proper posteriors and existence of first and higher order moments.

For convenience we present here the reasoning, which proceeds as follows. Using the parametrization $\Pi = \beta\Lambda\alpha'$ and the normalizing restrictions $\alpha'\alpha = I_r$, $\beta'\beta = I_r$

and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$ all parameters except λ_1 are defined on bounded sets (conditionally upon the (finite) value of λ_1). A natural choice for an uninformative prior is the uniform prior over these sets. Only λ_1 is defined on an infinite interval. A natural choice for λ_1 that is consistent with the uniform prior on the simplex for $\lambda_2, \ldots, \lambda_r|\lambda_1$ is the exponential distribution. Another way to look at this, is that although $\lambda \in [0, \infty)$ has infinite support, it can also be transformed to the unit interval on which a uniform prior can be specified. By doing so, all model parameters (except the covariance matrix Σ) are bounded to finite areas. Specifically, when either the transformation $\lambda^\flat = \exp(-\lambda) \in (0, 1]$ or $\lambda^\sharp = 1 - \exp(-\lambda) \in [0, 1)$ is used and a standard uniform density is specified on $\lambda^\flat$ or $\lambda^\sharp$ then λ also has a standard exponential distribution. Using a similar argument the rate parameter θ could be included by specifying a uniform prior on e.g. $\exp(-\theta\lambda)$. A note refers to the rate θ of the exponential distribution. By choosing θ to a value close to 0, the exponential distribution tends towards a flat distribution over the positive real numbers.

We can summarize the results from this section as follows.

Proposition. *Given the standard form of a cointegration model and using a lasso type shrinkage prior under orthogonal normalization on the parameters of the matrix with reduced rank, the marginal posteriors of these parameters are proper with finite first and higher order moments.*

We emphasize that the cointegration model serves as an example of a general reduced rank model but our result holds generally for this class of models. That is, one may also apply it to the instrumental variable model and the factor model when in these latter models one does not want to impose specific restrictions like triangularity and/or diagonality.

4.3 Short survey of other regularization priors

Inequality conditions where data and economic information matters As explained in the previous subsection area restrictions play a useful role in formulating prior information. Baumeister and Hamilton (2015) have carried this issue further. These authors explore the effect of sign restrictions, coming from broad economic considerations, on vector autoregressive models under different identification conditions. They also explore the effect of weak prior information on implied impulse response functions. Apart from restrictions based on economic relationships and characteristics, there exist data based inequality conditions that can also be relevant as prior information. A simple example of this is the restriction that autoregressive parameters in a dynamic model should not be taken outside the unit interval since explosive time series are highly unlikely for the long run because the occurrence of a regime change is then very likely. An analogous point can be made for values of the autoregressive parameters close to zero. From stylized facts of macroeconomic and financial time series it is well-known that the relevant range of the autoregressive parameters is a subinterval of the unit interval close to the unit root. For more details of the locally uniform prior where the data play a role, we refer to the next section and to Schotman and Van Dijk (1991b).

Dummy observations and training sample priors One popular way to make use of weak data-based prior information is to split the data into two parts: a training set and a 'hold-out' set of data. In the first part the weak prior is transformed to an informative posterior which serves as a prior for the second part of the data and this leads to model validation and forecasting. For an illustrative example we refer to the next section and for background to, e.g. Berger et al. (2004). Another approach is to construct a so-called imaginary sample by introducing a set of dummy observations. It yields a pragmatic class of priors, proposed by Sims (2004, 2005). This approach can be combined with a more informative prior approach, see below.

Dynamic patterns for parameters Given the dynamic nature of many models in economics, it is very natural to allow not only the variables but also the parameters of such models to move through time. However, simply adding a subscript of time to an equation system parameter yields an intractable likelihood since a T-dimensional integral is added to the Bayesian inferential problem. The well-known Normal or Kalman Filter imposes in such a case a structure on the dynamic parameters and it forms a pattern which yields a tractable likelihood and posterior. The literature on this topic is abundant and we refer only to two basic textbooks for more background: Pole et al. (1994) and Harvey (1990). A related approach is the Minnesota prior, see Doan et al. (1992), which may be characterized as a smoothness priors. This class of priors is meant to improve forecasting properties by making use of stylized facts of macroeconomic time series.

One may also explore the predictive implications of a prior. For instance, does a weak prior on the equation system parameters give plausible prior values of multipliers, impulse response function and/or periods of oscillations from an implied business cycle. For an early reference we refer to Van Dijk and Kloek (1980). The literature on this *prior-predictive approach* is substantial and a more detailed analysis is outside the scope of the present paper.

We also mention an approach where the priors are anchored to some long run plausible values. A basic approach was taken by Schotman and Van Dijk (1991a,b) for the unit root case. It was extended by Villani (2009) to refer to long run plausible values and recently again extended to be combined with a dummy variable prior by Giannone et al. (2015). A similar idea is to connect the prior to a plausible posterior-predictive analysis, see Gelman et al. (1996) and Baştürk et al. (2014b).

Economic structural information We end this brief survey by mentioning the approach to add economic structural information like so-called Dynamic Stochastic General Equilibrium (DSGE) priors due to Del Negro and Schorfheide (2004), while Strachan and Van Dijk (2013) combine economic information and technical econometric information.

We conclude that there are many useful approaches to explore the sensitivity of the posterior and predictive results with respect to a sequence of weak priors where the amount of prior information is gradually increasing. This will be illustrated in the next section. Finally, we note that the issue of sensitivity of weak priors and also prior choice is very much studied in the Bayesian literature, see for instance Tuyl et al. (2008).

5 Model probabilities under regularization priors and possibly irregular likelihoods

This section forms a bridge between the more theoretical analysis of the shape of posterior densities for the reduced rank regression model with possibly irregular likelihoods and the empirical analysis of a micro-econometric problem on the education-income effect where we make use of mixtures of models. An important concern is how to give probabilistic weights to evidence that is near and at the boundary of the parameter region of reduced rank models using Bayesian methods. We show that, although the issue of weak identification is not an impediment for showing posterior existence of distributions, very weak prior information does play a major role for the evaluation of posterior and predictive probabilities of evidence near and at a boundary of identification and relevance of instruments. We illustrate that the Bartlett/Jeffreys/Lindley paradox is not only a mathematical or statistical result but it shows up as a problem when flat prior density kernels are assumed over regions where there is little empirical evidence like near a boundary with weak instruments. This issue was pointed out by Hoogerheide and Van Dijk (2013). Here a training sample and weak economic information on area restrictions is recommended together with a sensitivity analysis in order to obtain more robustness in the results. We present two examples. One refers to a basic time series model where the likelihood is regular but the prior interval contains many irrelevant parameter values. In order to save space, the issue of model evaluation without and with regularization priors is discussed for this class of models in the Supplementary material in Online Appendix, Section A.4. The second example studies the effect that an irregular likelihood due to a lack of identification and the presence of weak instruments has on model probability evaluation within the context of an IV model. These results are reported in Section 5.1. Armed with these results, we continue in the next section with an empirical analysis using a mixture of models with mixing probabilities coming from evidence near and at a boundary.

5.1 IV model probabilities under alternative identification and endogeneity structures using training sample priors

In this subsection we apply the predictive likelihood approach, see Gelfand and Dey (1994) and Eklund and Karlsson (2007), to simulated data from the IV model. Our purpose is show that, although the posterior densities in an IV model with diffuse type priors and weak instruments/identification are very non-regular and require special simulation based procedures to evaluate their shape, it is relatively easy to evaluate posterior/predictive probabilities near and at the boundary using reasonable area restrictions and training sample priors. In the next section a mixture of posteriors under endogeneity and exogeneity for the IV model is estimated using US data.

In the present subsection we investigate the robustness of the results on estimating predictive probabilities for the case of no endogeneity for different levels of endogeneity, different levels of empirical identification and different lengths of training samples, where the total number of observations is 1000 for each simulated dataset. We will use the basic structural IV model from Section 3, see also Appendix A.3.3. For simplicity and for computational convenience we take the case of one endogenous variable and one instrument, where $\beta = 0$ and $\Sigma = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$ and the parameter ρ indicates the degree of endogeneity with $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$. We restrict the parameters to a plausible finite region.

Left panel of Table 1 For the cases of strong and medium instruments/identification and strong and medium level of endogeneity the posterior probability $Pr(\rho = 0 \mid \text{data})$ is correctly chosen as zero, given the 50% training sample. That is, let $y^\star$ be the training sample and $\tilde{y}$ be the validation sample, then $Pr(\rho = 0 \mid y^\star, \tilde{y})$ is much smaller than $Pr(\rho = 0 \mid y^\star)$, since the data $\tilde{y}$ contain much evidence about ρ being not equal to zero. For the bottom row, it holds that $\pi = 0$ implies that β, Σ, ρ are not identified. That is, the data contain no information on ρ and thus the posterior probability $Pr(\rho = 0 \mid \text{data})$ is equal to the prior probability $Pr(\rho = 0) = 50\%$.

For the right hand column one would expect that $Pr(\rho = 0 \mid \text{data}) = 1$. However, the situation is as follows. Given $y^\star$ and $\tilde{y}$, $Pr(\rho = 0 \mid y^\star)$ using 50% of data is already rather precisely located around $\rho = 0$, with a standard deviation only about $\sqrt{2}\times$ larger than for $Pr(\rho = 0 \mid y^\star, \tilde{y})$. This implies $Pr(\rho = 0 \mid y^\star, \tilde{y}) = \sqrt{2}Pr(\rho = 0 \mid y^\star)$ which leads to $Pr(\rho = 0 \mid \text{data}) = \sqrt{2}/(1 + \sqrt{2}) \approx 0.586$.[2]

Middle panel of Table 1 The results in the upper left corner are as expected: $\approx 0\%$. Similarly, the results in the bottom row are: $\approx 50\%$. The results in the right column follow from $Pr(\rho = 0 \mid \text{data}) \approx \sqrt{1/m}/(1 + \sqrt{1/m}) = \sqrt{1/10}/(1 + \sqrt{1/10}) \approx 0.760$.

Right panel of Table 1 Again the results in the upper left corner are as expected: $\approx$ 0%. Next, the advantage of very small training sample m is shown at the top of the right column: $Pr(\rho = 0 \mid \text{data})$ is close to 1, which is the true value given that $\rho = 0$. The disadvantage of a very small m is recognized as a case of Bartlett/Jeffreys/Lindley paradox. That is, the false null hypothesis $\rho = 0$ is wrongly favored in the bottom row and in the third column of results. The reason is that $Pr(\rho = 0 \mid y^\star)$ after only 5 of 1000 observations is still very diffuse. That is, more diffuse than $Pr(\rho = 0 \mid y^\star, \tilde{y})$ after 1000 observations.

The conclusions of Table 1 may be summarized as follows.

In the interior of the parameter region For the cases of strong and medium instruments and strong and medium level of endogeneity the posterior probability $Pr(\rho = 0 \mid \text{data})$ is correctly chosen as zero for several values of the length of the training sample.

At the boundaries of the parameter region For the bottom row, which refers to the case of no identification/irrelevant instruments, the estimated posterior probability $Pr(\rho = 0 \mid \text{data})$ is sensitive for the length of the training sample. A training sample of less than 10 percent should not be selected. For the right hand columns in all three panels, which refers to the case of no endogeneity, the estimated posterior probability

[2]Given a training sample fraction equal to m and given a normal distribution with mean $\rho = 0$, stdev = const / $\sqrt{\#data}$, one has $Pr(\rho = 0 \mid y^\star, \tilde{y}) = \sqrt{1/m} \times Pr(\rho = 0 \mid y^\star)$ and $Pr(\rho = 0 \mid \text{data}) = \sqrt{1/m}/(1 + \sqrt{1/m}) = 1/(1 + \sqrt{1/m})$.

Table 1: Simulation experiments for the IV model: $Pr(\rho = 0 \mid \text{data})$ for different levels of endogeneity (ρ), different instrument strength (π) and different prior data percentages (m). Standard deviations and numerical standard errors of $Pr(\rho = 0 \mid \text{data})$ based on 10 sets of simulated data are reported in parentheses and square brackets, respectively.

	percentage m of observations that are used as training sample											
	$m = 50\%$				$m = 10\%$				$m = 0.5\%$			
	level of endogeneity				level of endogeneity				level of endogeneity			
identification level / instrument strength	strong $\rho = 0.9$	medium $\rho = 0.5$	weak $\rho = 0.1$	no $\rho = 0$	strong $\rho = 0.9$	medium $\rho = 0.5$	weak $\rho = 0.1$	no $\rho = 0$	strong $\rho = 0.9$	medium $\rho = 0.5$	weak $\rho = 0.1$	no $\rho = 0$
	0.00	0.00	0.27	0.57	0.00	0.00	0.32	0.75	0.00	0.00	0.66	0.97
strong π =1	(0.00)	(0.00)	(0.18)	(0.04)	(0.00)	(0.00)	(0.26)	(0.03)	(0.00)	(0.00)	(0.31)	(0.00)
	[0.00]	[0.00]	[0.06]	[0.01]	[0.00]	[0.00]	[0.08]	[0.01]	[0.00]	[0.00]	[0.10]	[0.00]
	0.00	0.00	0.44	0.57	0.00	0.00	0.53	0.73	0.00	0.00	0.89	0.97
medium π =0.5	(0.00)	(0.00)	(0.14)	(0.04)	(0.00)	(0.00)	(0.23)	(0.04)	(0.00)	(0.00)	(0.09)	(0.01)
	[0.00]	[0.00]	[0.04]	[0.01]	[0.00]	[0.00]	[0.07]	[0.01]	[0.00]	[0.00]	[0.03]	[0.00]
	0.00	0.41	0.53	0.54	0.00	0.39	0.71	0.73	0.00	0.58	0.88	0.88
weak π =0.1	(0.00)	(0.18)	(0.05)	(0.04)	(0.00)	(0.20)	(0.16)	(0.11)	(0.00)	(0.27)	(0.05)	(0.05)
	[0.00]	[0.06]	[0.02]	[0.01]	[0.00]	[0.06]	[0.05]	[0.03]	[0.00]	[0.09]	[0.01]	[0.02]
	0.50	0.48	0.49	0.48	0.49	0.49	0.52	0.53	0.72	0.70	0.70	0.70
irrelevant π =0	(0.05)	(0.05)	(0.06)	(0.05)	(0.09)	(0.14)	(0.15)	(0.13)	(0.09)	(0.11)	(0.14)	(0.14)
	[0.02]	[0.02]	[0.02]	[0.02]	[0.03]	[0.04]	-0.14	[0.04]	[0.03]	[0.03]	[0.04]	[0.04]

$Pr(\rho = 0 \mid \text{data})$ is also very sensitive to the length of the training sample. A small training sample and a large validation sample are to be recommended in this case.

Near the boundaries of the parameter region This refers to the third column and third row in each of the three panels. Here there also exists a trade-off between the case of weak instruments/identification and the case of weak and no endogeneity. In the case of weak instruments/identification one would prefer a large training sample to get informative priors while for the case of no endogeneity a small training sample so that most of the data can be used for validation.

It is clear evidence from the results of the Table that the choice of a 'prior data' percentage m is important. The advantage of the predictive likelihood approach is that m is a scalar. This may be easier to choose than specifying an entire not 'too non-informative or not too informative' prior density. The problem of predictive likelihood remains: How to choose this scalar m? A practical sensitivity analysis is: simply show results for multiple values of m and find the interval of m values where results are 'similar'.

General conclusion of Section 5 The evaluation of predictive model probabilities under weak prior information and near a boundary of the parameter region gives correct results which are relatively robust under the condition of choosing the right training sample. A sensitivity analysis is recommended for the length of the training sample. In extreme cases very near and at the boundary with weak identification one should be very careful with strong conclusions. More informative priors are then to be recommended.

6 Bayesian mixtures to analyze the effect of length-of-education on earned income in US states

In this section, we present and apply a predictive likelihood approach for model comparison or model combination to the Angrist and Krueger (1991) data on income and education, which are also analyzed in Hoogerheide and Van Dijk (2008). Angrist and Krueger (1991) data consist of men born in the US during the periods 1920–1929, 1930–1939 and 1940–1949, where the data for the first group are collected in 1970, and the data for the last two groups are collected in 1980.[3] We use a subset of their data, consisting of men born during the period 1930–1939, including the data on weekly wages, number of completed years of education and instruments consisting of quarter of birth dummies. The data include 51 states and 329.509 observations.[4] The IV model applied to data from each state is[5]:

[3]For an introduction to a Bayesian analysis of an IV model using real and simulated data, we refer to the Supplementary Material in the Online Appendix A.1.

[4]The source of the data is the 1980 Census, 5 percent public sample, also available from `econ-www.mit.edu/faculty/angrist/data1/data/angkru1991`. We refer to the Online Appendix for a summary of these data.

[5]In order to keep the notation simple, we do not define an index for each state, but note that the described IV model is applied to each US state separately.

$$\tilde{y}_i = \alpha_1 + \tilde{x}_i\beta + \sum_{t=1}^{9} D_{t,i}\delta_t + \tilde{\epsilon}_i, \tag{16}$$

$$\tilde{x}_i = \alpha_2 + \sum_{q=2}^{4} D_{q,i}\Pi_q + \sum_{t=1}^{9} D_{t,i}\delta_t + \tilde{\nu}_i, \tag{17}$$

where $\tilde{y}_i$ and $\tilde{x}_i$ are the natural logarithm of the weekly wage and the number of completed years of education for the person i in 1979, respectively.

In (16) and (17), $D_{t,i}$ for $t = \{1, \ldots, 9\}$ are the dummy variables for year of birth which take the value 1 if individual i was born in year $1929 + t$, and 0 otherwise. $D_{q,i}$ for $q = \{2, 3, 4\}$ are the quarter of birth dummy variables which take the value 1 if individual i was born in quarter q, and 0 otherwise. α_1 and α_2 are constants, and $\tilde{\epsilon}_i$ and $\tilde{\nu}_i$ are disturbances assumed to be normally distributed, and independent across individuals.

The model in (16) and (17) is similar to the model of Hoogerheide and Van Dijk (2006). For simplicity, we do not consider interactions of year dummies and quarter of birth dummies as instruments. Furthermore, the model employed here does not include state dummies, as each state is analyzed separately. We simplify the IV model in (16) and (17) correcting for the constant terms and exogenous year of birth dummies. Using this simplification, the IV model becomes:

$$y_i = x_i\beta + \epsilon_i, \tag{18}$$
$$x_i = Z_i\Pi + \nu_i, \tag{19}$$

where y_i, x_i are the residuals from regressing the log weekly wage and years of education on a constant and year of birth dummies, respectively. Z_i is the 3×1 vector of instruments, obtained from regressing quarter of birth dummies on a constant and the year of birth dummies. ϵ_i and ν_i are the error terms that have a joint normal distribution, and are independent across individuals.

6.1 Bayesian model mixtures using predictive model probabilities

In order to calculate the predictive model probabilities, we define two models M_0 and M_1, where M_0 is a nested model compared to M_1. In the IV model example, M_1 corresponds to the IV model while the nested model M_0 corresponds to M_1 with a parameter restriction: $\rho = 0$. The posterior odds ratio K_{01} for comparing M_0 with model M_1 is the product of the Bayes factor and the prior odds ratio:

$$K_{01} = \frac{p\left(Y \mid M_0\right)}{p\left(Y \mid M_1\right)} \times \frac{p\left(M_0\right)}{p\left(M_1\right)}, \tag{20}$$

where Y is the observed data, and the prior model probabilities $(p(M_1), p(M_0)) \in (0,1) \times (0,1)$ and $p(M_1) + p(M_0) = 1$.

It is often difficult to compute K_{01} since the marginal likelihoods are given by the following integrals: $p(Y \mid M_0) = \int_{\theta_{-\rho}} \ell(\rho = 0, \theta_{-\rho}) p_0(\theta_{-\rho}) d(\theta_{-\rho})$ and $p(Y \mid M_1) = \int_{\theta_{-\rho},\rho} \ell(\rho, \theta_{-\rho}) p(\rho, \theta_{-\rho}) d(\rho) d(\theta_{-\rho})$, where $\theta_{-\rho}$ are the model parameters apart from ρ.

We therefore calculate model probabilities using the Savage-Dickey Density Ratio (SDDR). Dickey (1971) shows that 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. $p_1(\theta_{-\rho} \mid \rho = 0) = p_0(\theta_{-\rho})$[6]. In this case, (20) becomes:

$$K_{01} = \frac{p(\rho = 0 \mid Y, M_1)}{p(\rho = 0 \mid M_1)} \times \frac{p(M_0)}{p(M_1)}, \tag{21}$$

where $p(\rho \mid Y) = \int p(\rho, \theta_{-\rho} \mid Y) d\theta_{-\rho}$ and $p(\rho) = \int p(\rho, \theta_{-\rho}) d\theta_{-\rho}$[7]. We perform the model averaging scheme using the model probabilities in Section 5. Specifically, given the posterior odds ratio, it is possible to weight the evidence of alternative models using Bayesian Model Averaging (BMA). We consider the effect of model uncertainty on the estimation of the parameter β, as this parameter is the main focus in most cases. The information about β is summarized by the following posterior:

$$p(\beta \mid Y) \quad = \quad p(\beta \mid Y, M_0)\, p(M_0 \mid Y) + p(\beta \mid Y, M_1)\, p(M_1 \mid Y). \tag{22}$$

Furthermore, functions of parameters, i.e. $g(\beta)$ in the IV model are estimated by:

$$E[g(\beta \mid Y)] = E[g(\beta \mid Y, M_0)] p(M_0 \mid Y) + E[g(\beta \mid Y, M_1)] p(M_1 \mid Y). \tag{23}$$

Hence both models under consideration should be estimated, and the inference on parameters is simply the weighted average of the results in both models. The weights in averaging the results are the posterior model probabilities.

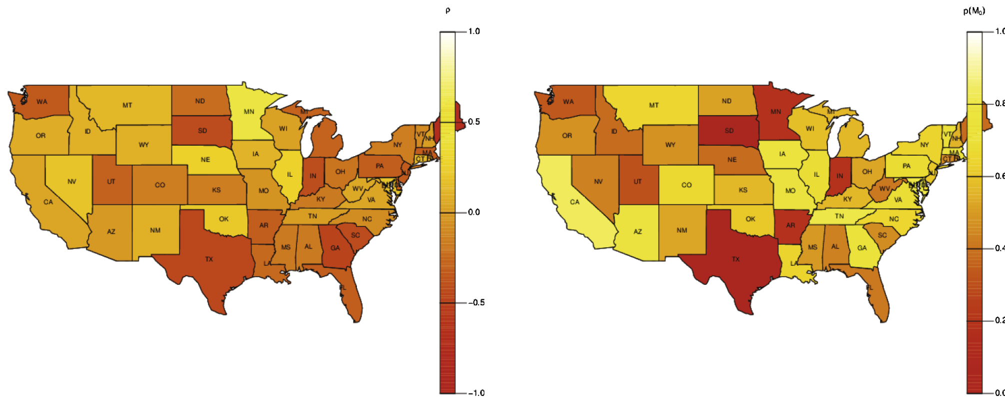
6.2 Empirical results

The degree of instrument strength (indicated by posterior densities of Π_2, Π_3 and Π_4) differs substantially across states, as reported in Hoogerheide and Van Dijk (2006). A second source of heterogeneity across states is the degree of endogeneity (indicated by posterior ρ). For some states, such as Maine, Minnesota and Texas, 95% intervals for posterior ρ densities do not include point 0, while for the rest of the states 95% posterior intervals of ρ include the value 0. Besides the finding of heterogeneity across states, we conclude that the use of instruments may not be necessary for most states. For further details of these estimation results we refer to the Online Appendix.

Posterior means for the degree of endogeneity of ρ in the IV model, and the predictive model probabilities for model M_0, corresponding to the model with $\rho = 0$ in which the instruments are not used for the estimation of β, are given in Figure 8. For the predictive model probabilities, the training sample consists of a randomly chosen subset of 10% of the observations, prior model probabilities in (21) are chosen to be equal. Furthermore,

[6]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.

[7]As a generalization, Verdinelli and Wasserman (1995) show that K_{01} is equal to the Savage-Dickey density ratio in (21) times a correction factor when the prior condition fails.



Posterior mean ρ for US states

Model probabilities for the model M_0 ($\rho = 0$) for US states

Figure 8: Degree of endogeneity in the US states and predictive model probabilities for model M_0. M_0 denotes the model with $\rho = 0$ in which no instruments are used for the estimation of β.

the effect of the training sample choice is partially eliminated by averaging predictive model probabilities from 20 different random training samples.[8]

Model probabilities are quite close to 0.5 and do not show a clear preference for either model, except for a few states such as Texas and Tennessee. For Texas, model probabilities indicate that the IV model is necessary. For Tennessee on the other hand, we find strong evidence against the need for the IV model. We conclude that choosing one of the alternative models according to these probabilities can be quite inaccurate, and employ model averaging to infer the state-specific effects of income on education.

We next present how *average* effects of education on income can be inferred using the model probabilities. The *average* estimated effects of education on income for the US states, i.e. the posterior distributions resulting from BMA, are summarized in Table 2. Model probabilities are achieved by using training sample with 10% of the observations, averaged over 20 repetitions. The main advantage of model averaging is the improved efficiency of the estimates. Standard deviations of posterior β draws are less than half of those achieved by the IV model only.

Regional patterns in income-education relationship – analysis of US divisions: We further analyze the income-education relationship in US divisions. We apply the IV model in (18) and (19) to 9 divisions for the Angrist and Krueger (1991) data according to the Census Bureau designated areas. The purpose of this analysis is to compare the results in terms of instrument strength with those of Hoogerheide and Van Dijk (2006),

[8] The results with 5% and 25% training sample sizes and a single training sample were similar, except for some states with very small number of observations, such as South Dakota.

who show that quarter of birth dummies are strong instruments mainly in southern states. Furthermore, we document the effect of averaging the data within divisions or regions.

Table 2 reports posterior results of the IV model for US divisions. Similar to the state-specific results, census regions show heterogeneity both in terms of instrument strength and the degree of endogeneity. Posterior results for education effects on income are quite different across divisions. Especially for the West North Central division, the posterior standard deviation is quite high, indicating the relatively weak instruments in this division. Figure 9 presents posterior mean ρ and model probabilities for M_0, the model with $\rho = 0$ in which no instruments are used for the estimation of β. The training sample consists of 25% of the observations[9]. Predictive model probability for the nested model without instruments is far from 0.5 only for two regions: East North Central Division, and West South Central Division. In East North Central division, the model without instruments is favored by model probabilities. Notice that the states within this region are quite heterogeneous in terms of predictive model probabilities reported in Figure 8. The IV model is clearly necessary only for two states in this region, namely Arkansas (AR) and Texas (TX). Hence 'average' income-education relationship within this region is determined mainly by these two states. This problem is also seen in East North Central division. According to posterior model probabilities in Figure 8, this region consists of states where an IV model is clearly preferred, such as Minnesota (MN) and South Dakota (SD), and also states where the IV model is not necessary, such as Iowa (IA). Hence the 'average' model probability for this region reported in Figure 9 is misleading.

For the US data on the income-education relationship, we conclude that there is substantial heterogeneity in the effect of the length of years of education on earned income. We document that differences between states are characterized by different instrument strengths, as reported by Hoogerheide and Van Dijk (2006). Our results also show that the degree of endogeneity is different across states and regions.

Using this data set we have shown different, and mostly weak power of quarter of birth in explaining education. This finding, in combination with the not so severe problem of endogeneity makes it hard to assess whether the IV model should be preferred over a simpler and more parsimonious linear regression model without instruments. Hence we conclude that averaging over these alternative models is a reasonable way to deal with model uncertainty.

General conclusion of Section 6 We have shown that the effect of length of education on earned income differs considerably among almost all US states. This may have important policy implication of determining the length of required schooling. This issue should be investigated in more detail.

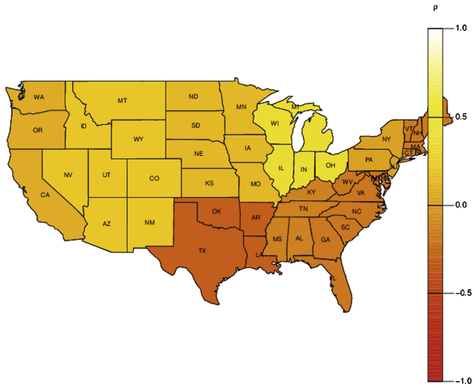
[9] We experimented with the model using smaller training samples, and the results are quite insensitive to the training sample size.

Average effects of education on income								
State	Mean	Std. Dev.	State	Mean	Std. Dev.	State	Mean	Std. Dev.
AL	0.11	0.03	LA	0.09	0.04	OH	0.08	0.04
AZ	0.11	0.03	ME	0.09	0.03	OK	0.04	0.03
AR	0.07	0.02	MD	0.05	0.03	OR	0.05	0.10
CA	0.05	0.01	MA	0.21	0.09	PA	0.11	0.03
CO	0.07	0.02	MI	0.08	0.03	RI	0.07	0.03
CT	0.06	0.04	MN	-0.06	0.08	SC	0.12	0.03
DE	0.02	0.05	MO	0.07	0.03	SD	0.16	0.07
DC	0.10	0.04	MS	0.09	0.04	TN	0.07	0.01
FL	0.13	0.05	MT	0.04	0.04	TX	0.16	0.06
GA	0.12	0.02	NC	0.08	0.02	UT	0.09	0.07
HI	0.08	0.04	NC	0.09	0.04	VT	0.06	0.03
ID	0.05	0.06	NE	0.03	0.09	VA	0.08	0.04
IL	0.05	0.08	NH	0.09	0.04	WA	0.10	0.09
IN	0.04	0.03	NJ	0.09	0.03	WV	0.06	0.03
IA	0.15	0.12	NM	0.05	0.05	WI	0.05	0.03
KS	0.08	0.03	NV	0.03	0.06	WY	0.04	0.06
KY	0.07	0.01	NY	0.08	0.03			

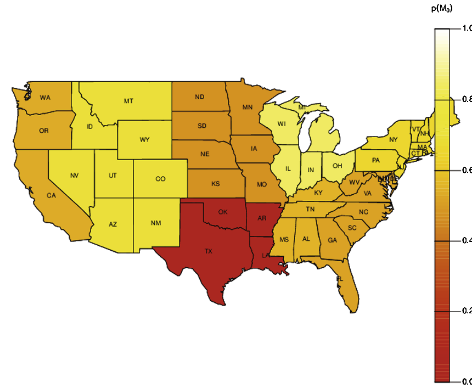
Parameter estimates					
	β	Π_2	Π_3	Π_4	ρ
New England Division	0.11	0.09	0.17	0.21	-0.16
	(0.05)	(0.04)	(0.04)	(0.04)	(0.23)
Middle Atlantic Division	0.07	0.07	0.03	0.03	0.03
	(0.07)	(0.02)	(0.02)	(0.03)	(0.31)
East North Central Division	-0.03	0.07	0.02	0.08	0.36
	(0.08)	(0.02)	(0.02)	(0.02)	(0.25)
West North Central Division	0.02	-0.06	0.01	0.02	0.15
	(0.13)	(0.04)	(0.04)	(0.04)	(0.40)
South Atlantic Division	0.11	-0.01	0.14	0.22	-0.18
	(0.03)	(0.03)	(0.03)	(0.03)	(0.16)
East South Central Division	0.09	0.03	0.27	0.41	-0.13
	(0.02)	(0.04)	(0.04)	(0.04)	(0.12)
West South Central Division	0.12	-0.04	0.20	0.30	-0.29
	(0.02)	(0.04)	(0.04)	(0.03)	(0.11)
Mountain Division	0.01	0.20	0.14	0.18	0.21
	(0.08)	(0.05)	(0.06)	(0.06)	(0.30)
Pacific Division	0.04	0.23	0.21	0.11	0.08
	(0.05)	(0.04)	(0.04)	(0.04)	(0.23)

Note: The top panel reports means and standard deviations of effect of education on income, resulting from BMA, for the US states. The bottom panel reports posterior means for 9 US divisions. All results are based on 30000 draws (3000 burn-in). Estimated posterior standard deviations are reported in parentheses.

Table 2: Income-education effects in US states.



Posterior mean ρ for US divisions



Model probabilities for the model M_0 ($\rho = 0$) for US divisions

Figure 9: Degree of endogeneity in the US divisions and predictive model probabilities for model M_0. M_0 denotes the model with $\rho = 0$ in which no instruments are used for the estimation of β.

7 Conclusions and perspectives

We have sketched in this paper an approach using Bayesian mixtures to average over those states of an econometric model which are known as near a boundary and at a boundary with the purpose to obtain more precise structural inference, accurate forecasting and effective policy analysis. In order to do this several results have been established. There exists a common structure in three well-known econometric models where the matrix of equation system parameters has reduced rank. The case of a reduced rank can be interpreted as a boundary in the parameter space. The econometric models are the cointegration, instrumental variables and factor model. Using a flat prior, the effect that the reduced rank has on the shape of the likelihood/posterior has been studied for a general workhorse model, that is equal to a cointegration model. Marginal posterior densities of equation parameters are of the student-t type times a polynomial or rational function. Their shapes may contain ridges due to weak identification, be bimodal and have very fat tails. These posteriors can get nicer properties (such as finite higher order moments) when extra restrictions are imposed like triangular ones for the instrumental variable model and a diagonal covariance matrix for the factor model. But their shapes may still be strongly non-elliptical.

In order to obtain meaningful posterior and predictive probabilities of states near and at the boundary of a reduced rank, weak regularization priors are discussed and compared. These are dealing with area restrictions, smoothness properties and training samples. As a novel class we introduce a lasso type shrinkage prior combined with orthogonal normalization which restricts the range of the parameters in a plausible way. A sensitivity analysis with respect to a sequence of weak priors is recommended.

The conditional posterior and predictive probabilities of different states of the econometric model near and at the boundary can then be used to estimate Bayesian mixture processes of several relevant economic and econometric issues.

We end this paper with listing some perspectives. The Bayesian approach to econometrics is now dominant in the field of macroeconomics. This is due to the pioneering work by Sims and his co-workers on Bayesian analysis of vector autoregressive models. The basic paper is Sims (1980) and an incomplete list of a few recent references are Sims and Zha (1998), Primiceri (2005) and Del Negro and Schorfheide (2011). An extension is to use more complex economic model structures like Dynamic Stochastic Equilibrium models, see Herbst and Schorfheide (2015). Complex cointegration models and inferential issues of these models have been studied extensively as well. Within this literature, we refer to Strachan (2003) for parameter instability, Jochmann et al. (2013) and Sugita et al. (2016) for regime-switching models, Jochmann and Koop (2015) for structural breaks, Koop et al. (2011) for time-varying parameter models and Chan et al. (2017) for cointegrating rank variations. Also in the fields of finance and marketing the Bayesian approach is becoming the dominant one. More details are presented in the Handbook of Bayesian Econometrics, Geweke et al. (2011).

We emphasize another perspective. There exists already much research to extend the analysis of this paper to models with time varying mixtures and to connections with expert systems and machine learning. See, among others, Frühwirth-Schnatter (2006), Chan et al. (2012), Billio et al. (2013), Casarin et al. (2015) and Baştürk et al. (2016).

All these extensions require a much more algorithmic approach to evaluating posterior probabilities of parameters and unknown unobserved states. Simulation based Bayesian Econometrics (SBBE) should be developed even more than already done so using modern software like parallel algorithms, filtering methods and modern hardware like clusters of machines of Graphical Processing Unit (GPU) processing. Developing operational methods useful for Bayesian empirical econometrics will lead to more insight in structural analysis, more accurate forecasting and more effective policy analysis with implied probabilistic components.

Supplementary Material

Supplementary Material for "Bayesian Analysis of Boundary and Near-Boundary Evidence in Econometric Models with Reduced Rank" (DOI: 10.1214/17-BA1061SUPP; .pdf).

References

Anderson, T. W. and Rubin, H. (1950). "The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations." *The Annals of Mathematical Statistics*, 570–582. MR0039188. 880

Anderson, T. W. and Rubin, H. (1956). "Statistical inference in factor analysis." In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, volume 5, 1. MR0084943. 880

Angrist, J. D. and Krueger, A. B. (1991). "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, 106(4): 979–1014. 884, 885, 904, 907

Bartlett, M. S. (1957). "A Comment on D. V. Lindley's Statistical Paradox." *Biometrika*, 44(3/4): 533–534. MR0086727. doi: http://dx.doi.org/10.1093/biomet/44.1-2.27. 882

Baştürk, N., Çakmaklı, C., Ceyhan, S., and Van Dijk, H. K. (2014a). "On the Rise of Bayesian Econometrics after Cowles Foundation Monographs 10, 14." *Œconomia*, 4: 381–447. 880, 882

Baştürk, N., Çakmaklı, C., Ceyhan, S. P., and Van Dijk, H. K. (2014b). "Posterior-Predictive Evidence on US Inflation using Extended Phillips Curve Models with non-filtered Data." *Journal of Applied Econometrics*, 29(7): 1164–1182. MR3293139. doi: http://dx.doi.org/10.1002/jae.2411. 900

Baştürk, N., Grassi, S., Hoogerheide, L. F., and Van Dijk, H. K. (2016). "Time-varying combinations of Bayesian dynamic models and equity momentum strategies." Technical Report 16-099/III, Tinbergen Institute. 911

Baştürk, N., Hoogerheide, L. F., and Van Dijk, H. K. (2017). "Supplementary Material for "Bayesian Analysis of Boundary and Near-Boundary Evidence in Econometric Models with Reduced Rank"." *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/17-BA1061SUPP. 882

Baumeister, C. and Hamilton, J. D. (2015). "Sign restrictions, structural vector autoregressions, and useful prior information." *Econometrica*, 83(5): 1963–1999. MR3414197. doi: http://dx.doi.org/10.3982/ECTA12356. 899

Bauwens, L. and Van Dijk, H. K. (1990). "Bayesian limited information analysis revisited." In Gabszewicz, J., Richard, J., and Wolsey, L. (eds.), *Economic Decision-Making: Games, Econometrics and Optimisation: Contributions in Honour of Jacques H. Drèze*, chapter 18, 385–424. Amsterdam: North-Holland. 893

Berger, J. O., Pericchi, L. R., et al. (2004). "Training samples in objective Bayesian model selection." *The Annals of Statistics*, 32(3): 841–869. MR2065191. doi: http://dx.doi.org/10.1214/009053604000000238. 900

Billio, M., Casarin, R., Ravazzolo, F., and Van Dijk, H. K. (2013). "Time-varying combinations of predictive densities using nonlinear filtering." *Journal of Econometrics*, 177(2): 213–232. MR3118557. doi: http://dx.doi.org/10.1016/j.jeconom.2013.04.009. 911

Casarin, R., Grassi, S., Ravazzolo, F., and Van Dijk, H. K. (2015). "Dynamic predictive density combinations for large data sets in economics and finance." Technical Report 2015-084/III, Tinbergen Institute. 911

Chan, J., Leon-Gonzalez, R., and Strachan, R. W. (2017). "Invariant inference and efficient computation in the static factor model." *Journal of the American Statistical Association*, (just-accepted). 911

Chan, J. C., Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2012). "Time varying dimension models." *Journal of Business & Economic Statistics*, 30(3): 358–367. MR2969219. doi: http://dx.doi.org/10.1080/07350015.2012.663258. 911

Del Negro, M. and Schorfheide, F. (2004). "Priors from general equilibrium models for VARs." *International Economic Review*, 45(2): 643–673. MR2053752. doi: http://dx.doi.org/10.1111/j.1468-2354.2004.00139.x. 900

Del Negro, M. and Schorfheide, F. (2011). "Bayesian macroeconometrics." In Geweke, J., Koop, G., and Van Dijk, H. K. (eds.), *The Oxford handbook of Bayesian econometrics*, 293–389. Oxford University Press: Oxford. MR1807781. doi: http://dx.doi.org/10.1016/S0304-4076(00)00046-4. 911

Dickey, J. (1971). "The weighted likelihood ratio, linear hypothesis on normal location parameters." *The Annals of Mathematical Statistics*, 42: 240–223. MR0309225. doi: http://dx.doi.org/10.1214/aoms/1177693507. 906

Doan, T., Litterman, R., and Sims, C. (1992). "Forecasting and conditional projections using realistic prior distributions." *Econometric Reviews*, 3: 1–100. 900

Drèze, J. H. (1976). "Bayesian limited information analysis of the simultaneous equations model." *Econometrica: Journal of the Econometric Society*, 1045–1075. MR0464532. doi: http://dx.doi.org/10.2307/1911544. 893

Eklund, J. and Karlsson, S. (2007). "Forecast combination and model averaging using predictive measures." *Econometric Reviews*, 26(2–4): 329–363. MR2364365. doi: http://dx.doi.org/10.1080/07474930701220550. 901

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media. MR2265601. 881, 882, 911

Gelfand, A. E. and Dey, D. K. (1994). "Bayesian model choice: asymptotics and exact calculations." *Journal of the Royal Statistical Society. Series B (Methodological)*, 501–514. MR1278223. 901

Gelman, A., Meng, X., and Stern, H. (1996). "Posterior predictive assessment of model fitness via realized discrepancies." *Statistica Sinica*, 733–760. MR1422404. 900

Geweke, J. (1996). "Bayesian reduced rank regression in econometrics." *Journal of Econometrics*, 75(1): 121–146. MR1414507. doi: http://dx.doi.org/10.1016/0304-4076(95)01773-9. 893

Geweke, J., Koop, G., and Van Dijk, H. K. (2011). *The Oxford handbook of Bayesian econometrics*. Oxford University Press. 882, 911

Giannone, D., Lenza, M., and Primiceri, G. (2015). "Prior selection for vector autoregressions." *The Review of Economics and Statistics*, 97(2): 436–451. 900

Hamilton, J., Waggoner, D., and Zha, T. (2007). "Normalization in econometrics." *Econometric Reviews*, 26(2–4): 221–252. MR2364362. doi: http://dx.doi.org/10.1080/07474930701220329. 897

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press. 900

Herbst, E. P. and Schorfheide, F. (2015). *Bayesian estimation of DSGE models*. Princeton University Press. MR3467463. doi: http://dx.doi.org/10.1515/9781400873739. 911

Hoogerheide, L. and Van Dijk, H. K. (2013). "Bridging two key issues in Bayesian inference: The relationship between the Lindley paradox and non-elliptical credible sets." In Singpurwalla, N., Dawid, P., and O'Hagan, A. (eds.), *Festschrift for Dennis Lindley's ninetienth birthday*, volume 2, 511–530. Blurb Publishers. 901

Hoogerheide, L. F., Kaashoek, J. F., and 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. MR2380763. doi: http://dx.doi.org/10.1016/j.jeconom.2006.06.009. 883

Hoogerheide, L. F. and Van Dijk, H. K. (2001). "Comparison of the Anderson–Rubin test for overidentification and the Johansen test for cointegration." Technical Report 2001-04, Econometric Institute Research Papers. 887

Hoogerheide, L. F. and Van Dijk, H. K. (2006). "A reconsideration of the Angrist-Krueger analysis on returns to education." *Econometric Institute Report*. 905, 906, 907, 908

Hoogerheide, L. F. and Van Dijk, H. K. (2008). "Possibly ill-behaved posteriors in econometric models." Technical Report 08-036/4, Tinbergen Institute. 884, 904

Jeffreys, H. (1939). *The theory of probability*. Oxford: The Clarendon Press, 1st edition. MR0000924. 882

Jochmann, M. and Koop, G. (2015). "Regime-switching cointegration." *Studies in Nonlinear Dynamics & Econometrics*, 19(1): 35–48. MR3315033. doi: http://dx.doi.org/10.1515/snde-2012-0064. 911

Jochmann, M., Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2013). "Stochastic search variable selection in vector error correction models with an application to a model of the UK macroeconomy." *Journal of Applied Econometrics*, 28(1): 62–81. MR3042827. doi: http://dx.doi.org/10.1002/jae.1238. 911

Johansen, S. (1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models." *Econometrica*, 59(6): 1551–1580. MR1135641. doi: http://dx.doi.org/10.2307/2938278. 880

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press on Demand. MR1487375. doi: http://dx.doi.org/10.1093/0198774508.001.0001. 887

Kleibergen, F. and Paap, R. (2002). "Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration." *Journal of Econometrics*, 111(2): 223–249. MR1943122. doi: http://dx.doi.org/10.1016/S0304-4076(02)00105-7. 888, 895, 897

Kleibergen, F. and Van Dijk, H. K. (1994). "On the shape of the likelihood/posterior in cointegration models." *Econometric Theory*, 10(3–4): 514–551. MR1309109. doi: http://dx.doi.org/10.1017/S0266466600008653. 888, 889, 894

Kleibergen, F. and Van Dijk, H. K. (1998). "Bayesian simultaneous equations analysis using reduced rank structures." *Econometric Theory*, 14(06): 701–743. MR1666708. doi: http://dx.doi.org/10.1017/S0266466698146017. 894, 895, 897

Kleibergen, F. and Zivot, E. (2003). "Bayesian and classical approaches to instrumental variable regression." *Journal of Econometrics*, 114(1): 29–72. MR1962372. doi: http://dx.doi.org/10.1016/S0304-4076(02)00219-1. 894

Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2011). "Bayesian inference in a time varying cointegration model." *Journal of Econometrics*, 165(2): 210–220. MR2846645. doi: http://dx.doi.org/10.1016/j.jeconom.2011.07.007. 911

Koop, G., Strachan, R., Van Dijk, H. K., and Villani, M. (2006). "Bayesian approaches to cointegration." In *The Palgrave Handbook of Theoretical Econometrics*, 871–898. Palgrave Macmillan. 896

Lindley, D. V. (1957). "A statistical paradox." *Biometrika*, 44(1/2): 187–192. MR0087273. doi: http://dx.doi.org/10.1093/biomet/44.1-2.179. 882

Martin, G. M. (2001). "Bayesian analysis of a fractional cointegration model." *Econometric Reviews*, 20(2): 217–234. MR1861180. doi: http://dx.doi.org/10.1081/ETC-100103824. 894

Martin, G. M. and Martin, V. L. (2000). "Bayesian inference in the triangular cointegration model using a Jeffreys prior." *Communications in Statistics – Theory and Methods*, 29(8): 1759–1785. MR1787917. doi: http://dx.doi.org/10.1080/03610920008832577. 886, 894

Mengersen, K., Robert, C., and Titterington, M. (2011). *Mixtures: estimation and applications*, volume 896. John Wiley & Sons. MR2867716. doi: http://dx.doi.org/10.1002/9781119995678. 882

Pole, A., West, M., and Harrison, J. (1994). *Applied Bayesian forecasting and time series analysis*. CRC press. 900

Primiceri, G. E. (2005). "Time varying structural vector autoregressions and monetary policy." *The Review of Economic Studies*, 72(3): 821–852. MR2148143. doi: http://dx.doi.org/10.1111/j.1467-937X.2005.00353.x. 911

Schotman, P. and Van Dijk, H. K. (1991a). "A Bayesian analysis of the unit root in real exchange rates." *Journal of Econometrics*, 49: 195–238. 900

Schotman, P. C. and Van Dijk, H. K. (1991b). "On Bayesian routes to unit roots." *Journal of Applied Econometrics*, 6(4): 387–401. 899, 900

Sims, C. (1980). "Macroeconomics and reality." *Econometrica*, 48(1): 1–48. 887, 911

Sims, C. A. (2004). "Econometrics for Policy Analysis: Progress and Regress."

De Economist, 152(2): 167–175. MR2261528. doi: http://dx.doi.org/10.1093/0199267693.001.0001. 900

Sims, C. A. (2005). "Dummy Observation Priors Revisited." Manuscript, Princeton University. http://sims.princeton.edu/yftp/DummyObs/DumObsPrior.pdf 900

Sims, C. A. (2007). "Thinking about instrumental variables." Manuscript, Princeton University. http://sims.princeton.edu/yftp/IV/IV.pdf 895

Sims, C. A. (2008). "Making macro models behave reasonably." Manuscript, Princeton University. http://sims.princeton.edu/yftp/ReasonableModels/MakingReasonable.pdf 881, 893

Sims, C. A. and Zha, T. (1998). "Bayesian methods for dynamic multivariate models." *International Economic Review*, 949–968. 911

Strachan, R. and Van Dijk, H. K. (2004). "Valuing structure, model uncertainty and model averaging in vector autoregressive processes." Technical Report EI 2004-23, Econometric Institute. 896

Strachan, R. and Van Dijk, H. K. (2013). "Evidence on features of a DSGE business cycle model from Bayesian model averaging." *International Economic Review*, 54(1): 385–402. MR3020351. doi: http://dx.doi.org/10.1111/j.1468-2354.2012.00737.x. 882, 900

Strachan, R. W. (2003). "Valid Bayesian estimation of the cointegrating error correction model." *Journal of Business & Economic Statistics*, 21(1): 185–195. MR1950386. doi: http://dx.doi.org/10.1198/073500102288618883. 911

Strachan, R. W. and Inder, B. (2004). "Bayesian analysis of the error correction model." *Journal of Econometrics*, 123(2): 307–325. MR2100416. doi: http://dx.doi.org/10.1016/j.jeconom.2003.12.004. 895

Sugita, K. et al. (2016). "Bayesian inference in Markov switching vector error correction model." *Economics Bulletin*, 36(3): 1534–1546. 911

Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288. MR1379242. 897

Tuyl, F., Gerlach, R., and Mengersen, K. (2008). "A comparison of Bayes–Laplace, Jeffreys, and other priors: the case of zero events." *The American Statistician*, 62(1): 40–44. MR2416895. doi: http://dx.doi.org/10.1198/000313008X267839. 900

Uhlig, H. (1994). "On Jeffreys prior when using the exact likelihood function." *Econometric Theory*, 633–644. MR1309114. doi: http://dx.doi.org/10.1017/S0266466600008707. 894

Van Dijk, H. K. and Kloek, T. (1980). "Further experience in Bayesian analysis using Monte Carlo integration." *Journal of Econometrics*, 14(3): 307–328. 900

Verdinelli, I. and Wasserman, L. (1995). "Computing Bayes factors using a generalization of the Savage-Dickey Density Ratio." *Journal of the American Statistical Association*, 90: 614–618. MR1340514. 906

Villani, M. (1998). "On the representation of ignorance regarding a matrix of reduced rank." Technical report, Stockholm University, Department of Statistics. 895

Villani, M. (2000). "Aspects of Bayesian cointegration." Ph.D. thesis, Stockholm University, Sweeden. 895

Villani, M. (2005). "Bayesian reference analysis of cointegration." *Econometric Theory*, 21(02): 326–357. MR2179541. doi: http://dx.doi.org/10.1017/S026646660505019X. 895

Villani, M. (2009). "Steady-state priors for vector autoregressions." *Journal of Applied Econometrics*, 24(4): 630–650. MR2675198. doi: http://dx.doi.org/10.1002/jae.1065. 900

Zellner, A., Ando, T., Baştürk, N., Hoogerheide, L., and Van Dijk, H. K. (2014). "Bayesian analysis of instrumental variable models: Acceptance-rejection within direct Monte Carlo." *Econometric Reviews*, 33: 3–35. MR3170839. doi: http://dx.doi.org/10.1080/07474938.2013.807094. 883

Zellner, A., Bauwens, L., and Van Dijk, H. K. (1988). "Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods." *Journal of Econometrics*, 38(1): 39–72. MR0946171. doi: http://dx.doi.org/10.1016/0304-4076(88)90026-7. 893