TI 2001-073/4
Tinbergen Institute Discussion Paper

How to overcome the Jeffreys-Lindleys Paradox for invariant Bayesian Inference in Regression Models

Frank Kleibergen

Name and address of the coordinator of the TI Econometrics programme is:

Prof.dr. S.J. Koopman
Econometrics and Operations Research
Faculty of Economics and Business Administration
Vrije Universiteit Amsterdam
De Boelelaan 1105
1081 HV Amsterdam

s.j.koopman@econ.vu.nl

Department of Quantitative Economics, Faculty of Economics and Econometrics, University of Amsterdam, and Tinbergen Institute

How to overcome the Jeffreys-Lindleys Paradox for invariant Bayesian Inference in Regression Models

Frank Kleibergen[*,†]

August 8, 2001

Abstract

We obtain invariant expressions for prior probabilities and priors on the parameters of nested regression models that are induced by a prior on the parameters of an encompassing linear regression model. The invariance is with respect to specifications that satisfy a necessary set of assumptions. Invariant expressions for posterior probabilities and posteriors are induced in an identical way by the respective posterior. These posterior probabilities imply a posterior odds ratio that is robust to the Jeffreys-Lindleys paradox. This results because the prior odds ratio obtained from the induced prior probabilities corrects the Bayes factor for the plausibility of the competing models reflected in the prior. We illustrate the analysis, where we focus on the construction of specifications that satisfy the set of assumptions, with examples of linear restrictions, *i.e.* a linear regression model, and non-linear restrictions, *i.e.* a cointegration and ARMA(1,1) model, on the parameters of an encompassing linear regression model.

1 Introduction

In Bayesian model comparison, that uses Bayes factors and prior and posterior odds ratios, prior probabilities for competing models are assigned independently from the prior densities on the parameters of these models. When one of these models encompasses the others, the prior (density) of its parameters has a specific value at the locations of the parameters that correspond with the competing nested models. Hence, it can occur that this prior has a low value at the location of a competing nested model while that model has a prior probability equal to the prior probability of the encompassing model. These kind of instances lead to the Jeffreys-Lindleys paradox, see *e.g.* Lindley (1957), Bernardo and Smith (1994), O'Hagan (1994) and Poirier (1995).

The prior probabilities and densities are specified independently because the prior (density) on the parameters of an encompassing model does, because of the Borel-Kolmogorov paradox, see *e.g.* Kolmogorov (1950), Billingsley (1986), Wolpert (1995) and Drèze and Richard (1983), not imply unambiguous probabilities for the lower dimensional sub-sets of its parameter space that constitute the nested models. By restricting ourselves to regression models, we are, however, able to construct a necessary set of assumptions that allow us to define the (Hausdorff)

[*]Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Email: `kleiberg@fee.uva.nl`

[†]I thank participants of the conference on "Recent advances in Bayesian Econometrics" in June 2001 at GREQAM Marseille and especially Jacek Osiewalski for helpful comments and suggestions.

integral of the prior of the parameters of an encompassing linear regression (ELR) model over lower dimensional sub-sets of its parameters space that constitute nested regression models. The (Hausdorff) integral is invariant with respect to specifications of the nested regression models that satisfy the assumptions. When we make an additional assumption about the completeness of the set of regression models of interest, the (Hausdorff) integral implies prior probabilities for nested regression models that are induced by the prior on the parameters of the ELR model. Alongside this result, it also implies the priors on the parameters of the nested regression models that are induced by the prior on the parameters of the ELR model. Both the prior and prior probability are invariant with respect to specifications that satisfy the assumptions.

In an identical manner as for the priors, we also obtain posteriors of the parameters of nested regression models that are induced by the posterior of the parameters of the ELR model. In line with the Bayesian paradigm, identical expressions of these posteriors result when we update the prior with the likelihood. The same reasoning applies to the posterior probabilities that are induced by the (Hausdorff) integral of the posterior of the parameters of the ELR model over the lower dimensional sub-sets that constitute the nested regression models. These posterior probabilities lead to a posterior odds ratio that equals its specification of prior odds ratio times Bayes factor. This shows that the prior odds ratio, which is induced by the prior of the parameters of the ELR model, corrects the Bayes factor for the plausibility of the competing models reflected in the prior. The posterior odds ratio that results is therefore robust against the Jeffreys-Lindleys paradox because prior probabilities and densities are specified in line with one another.

The paper is organized as follows. In the second section, we make the assumption that allows us to obtain the invariant expression of the Hausdorff integral over lower dimensional sub-sets. Because we do not restrict ourselves to linear lower dimensional sub-sets, we use Hausdorff integration instead of Lebesque integration. We therefore define the Hausdorff measure of lower dimensional sets and the Hausdorff integral. In the third section, we make an assumption about the completeness of the analyzed set of regression models. Alongside with the assumption that concerns the invariance of the Hausdorff integral, this assumption implies the prior probability for each nested regression model that is induced by the prior on the parameters of the ELR model. The assumptions also imply invariant priors on the parameters of the nested regression models. We discuss why they avoid the Borel-Kolmogorov paradox. Section 4 shows that the analysis continues to the case of the posterior and posterior probabilities. In section 5, we discuss the Jeffreys-Lindleys paradox and show that the use of prior probabilities that are induced by the prior on the parameters of the ELR model leads to a posterior odds ratio that is robust to the Jeffreys-Lindleys paradox. Section 6 extends the analysis to regression models that are conditional on nuisance parameters. Section 7 discusses examples where we focus on the construction of specifications that satisfy the assumption that concerns the invariance of the Hausdorff integral. Because these specifications do not automatically result from the nested regression model of interest, they can be difficult to construct especially in case of non-linear restrictions. We illustrate this for some models that result from linear restrictions, *i.e.* linear regression models, and non-linear restrictions, the cointegration and autoregressive moving average model, on the parameters of an ELR model. Finally, the eight section concludes.

We use the following notation throughout the paper: $\text{vec}(A)$ stands for the column vectorization of the $T \times k$ matrix A such that $\text{vec}(A) = (a_1' \dots a_k')'$ when $A = (a_1 \dots a_k)$, $M_A = I_T - A(A'A)^{-1}A'$, with I_T the $T \times T$ dimensional identity matrix; $J(a,(b,c))$ is the jacobian of the transformation from a to (b,c) and $|_a$ stands for evaluated in a.

2 Hausdorff integrals for lower dimensional sub-sets

We consider the linear regression model,

$$\text{G} : y = X\beta + \varepsilon, \tag{1}$$

with y a $T \times 1$ vector of dependent variables, X a $T \times k$ matrix that contains the independent explanatory variables, β a $k \times 1$ vector of parameters and ε a $T \times 1$ vector of disturbances. The support of β is the $\mathbb{R}^k$. For expository purposes, we assume that $\varepsilon \sim N(0, I_T)$ but another distributional assumption can be made as well. We specify a prior on β in model G, $p_{\text{G}}(\beta)$, that is continuous and continuous differentiable. This prior $p_{\text{G}}(\beta)$ induces prior probabilities for convex k-dimensional sets $S \subset \mathbb{R}^k$,

$$\Pr_{\text{G}}[S] = \int_S p_{\text{G}}(\beta) d\beta, \tag{2}$$

where $d\beta$ is shorthand notation for $L_k(d\beta)$ because (2) is a Lebesque integral.

We use the prior density p_{G} and prior probability $\Pr_{\text{G}}$ in model G to induce prior probabilities and densities for the parameters of nested regression models

$$\text{G}_i : y = X f_i(\varphi_i) + \varepsilon, \qquad i = 1, \ldots, n, \tag{3}$$

with $\varphi_i \in \Theta_{\text{G}_i}$, Θ_{G_i} is an open convex set in the $\mathbb{R}^{m_i}$ and f_i is a k-dimensional continuous differentiable vector function of $\varphi_i : m_i \times 1$, $m_i \leq k$. The nested regression models (3) are represented by lower-dimensional sub-sets in the parameter space of β, the $\mathbb{R}^k$,

$$S_{\text{G}_i} = \{\varphi_i \in \Theta_{\text{G}_i} \subset \mathbb{R}^{m_i} | \beta = f_i(\varphi_i)\}, \qquad i = 1, \ldots, n. \tag{4}$$

The sub-sets S_{G_i} (4) are m_i-dimensional manifolds in the $\mathbb{R}^k$.

The prior probability and density of the regression models (3) have to be invariant with respect to the specification of β and f_i. In order to achieve this, we make an assumption with respect to the specification of β and f_i.

Assumption 1: *The $k \times 1$ dimensional vector β is an invertible function of the $m_i \times 1$ dimensional vector φ_i and the $(k - m_i) \times 1$ dimensional vector λ_i :*

$$\beta = f_i(\varphi_i) + g_i(\varphi_i, \lambda_i), \tag{5}$$

where $g_i(\varphi_i, \lambda_i)$ is a continuous differentiable $k \times 1$ vector function of (φ_i, λ_i) which is such that:

a. $g_i(\varphi_i, \lambda_i) = 0 \Leftrightarrow \lambda_i = 0.$

b. *The set of values of φ_i that lead to a unique value of $f_i(\varphi_i)$ is identical to the set of values of φ_i that lead to a unique value of $f_i(\varphi_i) + g_i(\varphi_i, \lambda_i)$, and the latter set does not depend on λ_i, such that $g_i(\varphi_i, \lambda_i)$ is a strictly monotonic function of λ_i for all values of φ_i.*[1]

[1] We note that this condition refers to the functional relationship $f_i(\varphi_i) + g_i(\varphi_i, \lambda_i)$. The spaces where φ_i, λ_i result from are therefore considered unrestricted, $\varphi_i \in \mathbb{R}^{m_i}$, $\lambda_i \in \mathbb{R}^{k-m_i}$, such that Θ_{H_i} is not involved and, for example, the set of values of φ_i that do not imply a unique value for $f_i(\varphi_i)$ can even be outside Θ_{H_i}.

c. $\left(\frac{\partial g_i(\varphi_i,\lambda_i)}{\partial \lambda_i'}|_{\lambda_i=0}\right)'\left(\frac{\partial g_i(\varphi_i,\lambda_i)}{\partial \lambda_i'}|_{\lambda_i=0}\right) \equiv A_i$ *for all values of* φ_i*, with* A_i *a fixed positive definite symmetric* $(k-m_i)\times(k-m_i)$ *matrix that does not depend on* φ_i*.*

To construct the prior probability of G_i induced by $p_{\mathrm{G}}(\beta)$, we evaluate the integral of $p_{\mathrm{G}}(\beta)$ over S_{G_i}. Because S_{G_i} does not have to be a linear set, Lebesque integration can not be used to evaluate this integral. We therefore use Hausdorff integration, see *e.g.* Billingsley (1986). Like Lebesque integrals, Hausdorff integrals result after we have defined the Hausdorff measure of sub-sets of S_{G_i}. We construct the Hausdorff measure by using a limit sequence of sets that converge monotonically to the sub-sets within S_{G_i}. The Hausdorff measure then results as the limit of the ratio of the measure of a k-dimensional set and the measure of the $(k-m_i)$-dimensional set that is eventually restricted. Assumptions 1a-b are necessary for the construction of the monotonically converging limit sequence while assumption 1c implies that the limit of the measure of the restricted set does not depend on φ_i. This shows the necessity of all three assumptions to obtain the invariant Hausdorff measure. The definition of the Hausdorff measure also shows how Hausdorff integrals of non-negative functions are constructed, see *e.g.* Billingsley (1986).

Definition 1: *When* $m_i < k$ *and assumption 1 holds, the Hausdorff-measure of* $W_{\mathrm{G}_i} \subset S_{\mathrm{G}_i}$*,*

$$W_{\mathrm{G}_i} = \{\varphi_i \in \Omega_{\mathrm{G}_i} \subset \Theta_{\mathrm{G}_i} | \beta = f_i(\varphi_i)\}, \qquad i = 1,\ldots,n, \tag{6}$$

with Ω_{G_i} *a convex open* m_i*-dimensional sub-set of* Θ_{G_i}*, reads*

$$H_{m_i}(W_{\mathrm{G}_i}) = \lim_{\rho\to 0}\left[\frac{L_k(W_{\mathrm{G}_i}(\rho))}{H_{k-m_i}(g(\varphi_i, B_{k-m_i}(0,\rho)))}\right] \tag{7}$$

where $L_k(W_{\mathrm{G}_i}(\rho))$ *is the Lebesque measure of the* k*-dimensional set* $W_{\mathrm{G}_i}(\rho)$*,*

$$W_{\mathrm{G}_i}(\rho) = \left\{\varphi_i \in \Omega_{\mathrm{G}_i},\ \lambda_i \in B_{k-m_i}(0,\rho) \subset \mathbb{R}^{k-m_i} | \beta = f_i(\varphi_i) + g_i(\varphi_i,\lambda_i)\right\}, \tag{8}$$

$H_{k-m_i}(g(\varphi_i, B_{k-m_i}(0,\rho)))$ *is the Hausdorff measure of the* $(k-m_i)$*-dimensional set* $g(\varphi_i, B_{k-m_i}(0,\rho))$ *and* $B_{k-m_i}(0,\rho)$ *is a* $(k-m_i)$*-dimensional sphere centered at zero with radius* ρ*.*

Definition 2: *When* $m_i < k$ *and assumption 1 holds, the Hausdorff integral of a non-negative function* $q(\beta)$ *over the* m_i*-dimensional set* W_{G_i} *reads*

$$\begin{aligned}\int_{W_{\mathrm{G}_i}} q(\beta) H_{m_i}(d\beta) &= \lim_{\rho\to 0}\left[\frac{\int_{W_{\mathrm{G}_i}(\rho)} q(\beta) L_k(d\beta)}{H_{k-m_i}(g(\varphi_i, B_{k-m_i}(0,\rho)))}\right] \\ &= \lim_{\rho\to 0}\left[\frac{\int_{W_{\mathrm{G}_i}(\rho)} q(\beta) d\beta}{H_{k-m_i}(g(\varphi_i, B_{k-m_i}(0,\rho)))}\right].\end{aligned} \tag{9}$$

Theorem 1 *When* $m_i < k$ *and assumption 1 holds, the Hausdorff measure* $H_{m_i}(W_{G_i})$ *(7) is equal to*

$$H_{m_i}(W_{G_i}) = \int_{\Omega_{G_i}} \left|\left(\frac{\partial f_i}{\partial \varphi_i'}\right)' M_{\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0}\right)} \left(\frac{\partial f_i}{\partial \varphi_i'}\right)\right|^{\frac{1}{2}} d\varphi_i, \tag{10}$$

and is invariant with respect to the specification of β *and* (φ_i, λ_i) *that satisfy assumption 1.*

Proof. see the appendix. ■

Theorem 2 *When $m_i < k$ and assumption 1 holds, the Hausdorff integral of the non-negative function $q(\beta)$ over W_{G_i} (9) is equal to*

$$\int_{W_{G_i}} q(\beta)H_{m_i}(d\beta) = \frac{1}{|A_i|^{\frac{1}{2}}} \int_{\Omega_{G_i}} q(\beta(\varphi_i, \lambda_i)|_{\lambda_i=0}) \left|J(\beta, (\varphi_i, \lambda_i))|_{\lambda_i=0}\right| d\varphi_i. \tag{11}$$

The Hausdorff integral (11) is invariant with respect to the specification of β and (φ_i, λ_i) that satisfy assumption 1.

Proof. see the appendix. ■

When $m_i = k$, the Hausdorff measure and integral are identical to the Lebesque measure and integral, see Billingsley (1986). In the next sections, we use the Hausdorff integrals to construct prior and posterior probabilities and densities.

3 Prior density and Prior probability

We construct the prior probability of G_i, $i = 1, \ldots, n$, that is induced by $p_{\mathrm{G}}(\beta)$. In order to obtain these probabilies we make an assumption about the completeness of the set of models G_i, $i = 1, \ldots, n$.

Assumption 2: *The true model is an element of $\{G_i,\ i = 1, \ldots, n\}$ such that the joint prior probability of the regression models G_i, $i = 1, \ldots, n$, is equal to one.*

Assumption 2 shows that we consider the models G_i, $i = 1, \ldots, N$, as separate events, unless they result from functions $f_i(\varphi_i)$ that are invertible transformations of one another, even when one of them equals G and encompasses all the other models. Hence, all sets S_{G_i} constitute a discrete separate event, the model G_i, although they are lower dimensional sets in the $\mathbb{R}^k$. The probabilities for these events then result from the Hausdorff integral over S_{G_i} with respect to the prior $p_{\mathrm{G}}(\beta)$ after an appropriate normalization for the completeness of the set of models G_i, $i = 1, \ldots, n$. The Hausdorff integral results from theorem 2.

Theorem 3 *When assumptions 1 and 2 hold, the invariant prior probability for model G_i, $i = 1, \ldots, n$, that is induced by $p_G(\beta)$ reads*

$$Pr_G\left[G_i\right] = \frac{Q_{G_i}}{Q} \qquad\qquad i = 1, \ldots, n, \tag{12}$$

with

$$Q_{G_i} = \int_{S_{G_i}} p_G(\beta)H_{m_i}(d\beta), \tag{13}$$

and

$$Q = \sum_{j=1}^{w} \int_{\cup_{i=1}^{n_j} S_{i_j}} p_G(\beta)H_{m_j}(d\beta), \tag{14}$$

with w the number of sets S_{G_i} that have a different function f_i, $w \leq n$, n_j is the number of sets that have the identical function f_j (or an invertible transformation thereof), m_j is the dimension of S_{G_j} and S_{i_j}, $i_j = 1, \ldots, n_j$ are the sets with the same function f_j involved.

Proof. follows directly from theorem 2. The specification of Q ensures the completeness that results from assumption 2. ■

When $m_i = k$, Hausdorff integrals are identical to Lebesque integrals and

$$Q_{\mathrm{G}_i} = \int_{S_{\mathrm{G}_i}} p_{\mathrm{G}}(\beta)d\beta \qquad m_i = k. \tag{15}$$

When $m_i < k$, we obtain from theorem 2 that

$$\begin{aligned} Q_{\mathrm{G}_i} &= \frac{\left[\frac{\partial \Pr_{\mathrm{G}}\left[\beta(\{\Theta_{\mathrm{G}_i},(-\infty,\lambda_i)\})\right]}{\partial\lambda_i}|_{\lambda_i=0}\right]}{\left|\frac{\partial\beta(\varphi_i,\lambda_i)}{\partial\lambda_i}|_{\lambda_i=0}\right|} && m_i < k \\ &= \frac{p_{\mathrm{G}}(\lambda_i)|_{\lambda_i=0}}{|A_i|^{\frac{1}{2}}}\left[\int_{\Theta_{\mathrm{G}_i}} p_{\mathrm{G}}(\varphi_i|\lambda_i)|_{\lambda_i=0}d\varphi_i\right] && m_i < k, \end{aligned} \tag{16}$$

where we haved used that

$$\begin{aligned} p_{\mathrm{G}}(\varphi_i,\lambda_i) &= p_{\mathrm{G}}(\beta(\varphi_i,\lambda_i))|J(\beta,(\varphi_i,\lambda_i))| \\ &= p_{\mathrm{G}}(\varphi_i|\lambda_i)p_{\mathrm{G}}(\lambda_i). \end{aligned} \tag{17}$$

The resulting specification of Q then corresponds with

$$Q = \sum_{j=1}^{w-1} \frac{p_{\mathrm{G}}(\lambda_j)|_{\lambda_j=0}}{|A_j|^{\frac{1}{2}}}\left[\int_{\cup_{i_j=1}^{n_j}\Theta_{\mathrm{G}_{i_j}}} p_{\mathrm{G}}(\varphi_j|\lambda_j)|_{\lambda_j=0}d\varphi_j\right] + \int_{\cup_{i=1}^{n_k}S_{\mathrm{G}_{i_w}}} p_{\mathrm{G}}(\beta)d\beta,$$

where n_k are the number of sets of dimension k. Because of theorem 2, the prior probability (12) is invariant with respect to the specification of β, (φ_i, λ_i) that satisfy assumption 1.

For expository purposes, we consider an example with $n = 2$. We use theorem 3 to obtain the prior probabilities of a nested (non-linear) regression model,

$$\mathrm{G}_1 : y = Xf_1(\varphi_1) + \varepsilon, \tag{18}$$

with $\varphi_1 \in \Theta_{\mathrm{G}_1} \subset \mathbb{R}^{m_1}$, such that

$$S_{\mathrm{G}_1} = \{\varphi_1 \in \Theta_{\mathrm{G}_1} \subset \mathbb{R}^{m_1} | \beta = f_1(\varphi_1)\}, \tag{19}$$

with $m_1 < k$ and $f_1(\varphi_1)$ continuous and continuous differentiable, and an encompassing linear regression model,

$$\mathrm{G}_2 : y = X\beta + \varepsilon, \tag{20}$$

with $\beta \in \mathbb{R}^k$ such that $S_{\mathrm{G}_2} = \left\{\beta \in \mathbb{R}^k\right\}$. The vital element of the applicability of theorem 3 is the existence of a function $g_1(\varphi_1, \lambda_1)$ which is such that β, (φ_1, λ_1) satisfy the conditions from assumption 1. Depending on the regression model of interest, the function $g_1(\varphi_1, \lambda_1)$ can be difficult to construct and we therefore in section 7 give some examples of its specification for some commonly used regression models. Alongside assumption 1, we also make assumption 2. Because $\int_{S_{\mathrm{G}_2}} p_{\mathrm{G}}(\beta)d\beta = 1$, we obtain the probabilities induced by $p_{\mathrm{G}}(\beta)$ for S_{G_1} and S_{G_2} from theorem 3,

$$\Pr_{\mathrm{G}}\left[S_{\mathrm{G}_1}\right] = \frac{Q_{\mathrm{G}_1}}{1+Q_{\mathrm{G}_1}}, \qquad \Pr_{\mathrm{G}}\left[S_{\mathrm{G}_2}\right] = 1 - \Pr_{\mathrm{G}}\left[S_{\mathrm{G}_1}\right], \tag{21}$$

with

$$Q_{\mathrm{G}_1} = \frac{p_{\mathrm{G}}(\lambda_1)|_{\lambda_1=0}}{|A_1|^{\frac{1}{2}}} \left[\int_{\Theta_{\mathrm{G}_1}} p_{\mathrm{G}}(\varphi_1|\lambda_1)|_{\lambda_1=0} d\varphi_1 \right]. \tag{22}$$

These prior probabilities imply the prior odds ratio (PROR):

$$\begin{aligned} \mathrm{PROR}\left[\mathrm{G}_1, \mathrm{G}_2\right] &= \frac{\mathrm{Pr}_{\mathrm{G}}\left[\mathrm{G}_1\right]}{\mathrm{Pr}_{\mathrm{G}}\left[\mathrm{G}_2\right]} \\ &= Q_{\mathrm{G}_1}. \end{aligned} \tag{23}$$

The prior probability from theorem 3 also implies a prior density of φ_i on Θ_{G_i}.

Theorem 4 *When assumption 1 holds, the prior probabilities (12) induce the prior densities*

$$\begin{aligned} p_{G_i}(\varphi_i) &= \lim_{\rho \to 0} \frac{Pr_G\left[G_i(\varphi_i, \rho)\right]}{L_{m_i}\left[B_{m_i}(\varphi_i, \rho)\right]} \qquad i = 1, \ldots, n, \\ &= \frac{p_G(\varphi_i|\lambda_i)|_{\lambda_i=0}}{\int_{\Theta_{G_i}} p_G(u|\lambda_i)|_{\lambda_i=0} du}, \end{aligned} \tag{24}$$

on Θ_{G_i}, where $B_{m_i}(\varphi_i, \rho)$ is a m_i-dimensional sphere with radius ρ centered at $\varphi_i \in \Theta_{G_i}$ and $G_i(\psi_i, \rho)$ is the model G_i where φ_i results from $B_{m_i}(\psi_i, \rho)$. The prior density (24) is invariant with respect to the specification of β, (φ_i, λ_i) that satisfy the conditions from assumption 1.

Proof. see the appendix. ■

For the example that we considered previously, theorem 4 implies the prior on φ_1 in G_1 :

$$p_{\mathrm{G}_1}(\varphi_1) = \frac{p_{\mathrm{G}}(\varphi_1|\lambda_1)|_{\lambda_1=0}}{\int_{\Theta_{\mathrm{G}_1}} p_{\mathrm{G}}(u|\lambda_1)|_{\lambda_1=0} du}, \tag{25}$$

and on β in G_2 :

$$p_{\mathrm{G}_2}(\beta) = p_{\mathrm{G}}(\beta). \tag{26}$$

Theorem 4 states that these priors are invariant with respect to the specification of (φ_1, λ_1) and β that satisfy assumption 1.

Assumption 1 ensures that λ_i represents the difference between G and G_i for all values of φ_i. This allows us to use λ_i to project β onto $f_i(\varphi_i)$. We can therefore also interpret conditioning on $\lambda_i = 0$ to result from integrating over λ_i with respect to the projection function $pj_i : \mathbb{R}^k \to S_{\mathrm{G}_i}$, $pj_i(\beta) = pj_i(f_i(\varphi_i) + g_i(\varphi_i, \lambda_i)) = f_i(\varphi_i)$, see McCulloch and Rossi (1992).

Borel-Kolmogorov Paradox

The Borel-Kolmogorov paradox, see *e.g.* Kolmogorov (1950), Billingsley (1986), Wolpert (1995) and Drèze and Richard (1983), implies that the probability of a lower dimensional set is not unambiguously defined. Definitions 1-2 state the Hausdorff measure and integral for lower dimensional sets. They lead to probabilities for these sets that are invariant with respect to their specification when this specification accords with assumption 1. Hence, assumption 1 gives a manner of unambiguously specifying probabilities on lower dimensional sets and therefore avoids the Borel-Kolmogorov paradox. There are two reasons why assumption 1 avoids the Borel-Kolmogorov paradox. First, assumptions 1a-b imply a specification in which one parameter, λ_i, completely and only reflects the imposed restriction. For all values of β, conditioning on a zero value of λ_i leads to the desired restriction, $\beta = f_i(\varphi_i)$. Assumptions

1a-b therefore identify the restriction and avoid the issue of non-conglomerability that is one element of the Borel-Kolmogorov paradox, see *e.g.* De Finetti (1972). Second, assumption 1c implies that the Hausdorff measure of the restricted space does not depend on φ_i. This allows us to separate the measures of the restricted and unrestricted space in a joint set of (φ_i, λ_i).

Assumption 1 excludes the traditional example of the Borel-Kolmogorov paradox that the restrictions $\lambda_i = 0$ and $\frac{\lambda_i}{\varphi_{i,1}} = 0$ lead to different densities and probabilities, see *e.g.* Wolpert (1995) and Drèze and Richard (1983). This traditional example is concerned with non-conglomerability and shows that the restriction can be represented in a non-denumerable infinite number of ways, see *e.g.* De Finetti (1972). Assumptions 1a-b impose a structure on the representation of the restriction, $g_i(\varphi_i, \lambda_i)$, such that it represents the restriction for all values of φ_i. This property does, for example, not hold for the specification $\frac{\lambda_i}{\varphi_{i,1}} = 0$. That specification is not uniquely determined when $\varphi_{i,1} = 0$. It does therefore not represent the restriction $\lambda_i = 0$ when $\varphi_{i,1} = 0$ and explains why assumption 1b is necessary. The structure imposed by assumptions 1a-b on $g_i(\varphi_i, \lambda_i)$ overcomes the issue of non-conglomerability and reduces the number of ways how the restriction can be represented to a denumberable infinite amount. Assumption 1c adds further structure and is concerned with the measure of the restricted space. It ensures that this measure does not depend on φ_i and does therefore also exclude the specification $\frac{\lambda_i}{\varphi_{i,1}} = 0$.

Theorems 3-4 show how we conduct Bayesian inference in regression models that are non-linear in the parameters in a manner that corresponds with Bayesian inference in linear regression models. The latter analysis is well-developed and theorems 3-4 show how we extend this analysis to regression models that are non-linear in the parameters. For example, sufficient statistics exist for the parameter β in G and we know how the prior influences the posterior. By specifying the prior on φ_i in G_i according to theorem 4, this property also holds for the prior and posterior of φ_i in G_i. We discuss this property for the posterior in the next section.

4 Posterior density and Posterior probability

The posterior in model G (1) is obtained by updating the prior with the likelihood:

$$p_{\mathrm{G}}(\beta|D) = \frac{p_{\mathrm{G}}(\beta)\mathcal{L}(D|\beta)}{\int_{\mathbb{R}^k} p_{\mathrm{G}}(u)\mathcal{L}(D|u)du}, \tag{27}$$

where $\mathcal{L}(D|\beta)$ is the likelihood function, which in our case of normal disturbances corresponds with

$$\mathcal{L}(D|\beta) = (2\pi)^{-\frac{1}{2}T} \exp\left[-\tfrac{1}{2}\left(y - X\beta\right)'\left(y - X\beta\right)\right], \tag{28}$$

but any other likelihood that is a continuous and continuous differentiable function of β can be used as well. Because the posterior (27) is a proper density function, and therefore non-negative, we can, analogous to theorem 3, construct posterior probabilities by usage of theorem 2.

Theorem 5 *When assumptions 1 and 2 hold, the invariant posterior probability for model* G_i, $i = 1, \dots, n$, *that is induced by* $p_G(\beta|D)$ *(27) reads*

$$Pr_G\left[G_i|D\right] = \frac{Q_{G_i|D}}{Q_D} \qquad i = 1, \dots, n, \tag{29}$$

with

$$Q_{G_i|D} = \int_{S_{G_i}} p_G(\beta|D)H_{m_i}(d\beta), \tag{30}$$

and

$$Q_D = \sum_{j=1}^{w} \int_{\bigcup_{i=1}^{n_j} S_{i_j}} p_G(\beta|D)H_{m_j}(d\beta). \tag{31}$$

Proof. results directly from the proofs of theorem 2. ■

When $m_i = k$, the Hausdorff integral is identical to the Lebesque integral and

$$Q_{\mathrm{G}_i|D} = \int_{S_{\mathrm{G}_i}} p_{\mathrm{G}}(\beta|D)d\beta \qquad m_i = k. \tag{32}$$

When $m_i < k$, we use theorem 2 to obtain that

$$Q_{\mathrm{G}_i|D} = \frac{p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0}}{|A_i|^{\frac{1}{2}}}\left[\int_{\Theta_{\mathrm{G}_i}} p_{\mathrm{G}}(\varphi_i|\lambda_i, D)|_{\lambda_i=0}d\varphi_i\right] \qquad m_i < k, \tag{33}$$

where

$$\begin{aligned} p_{\mathrm{G}}(\varphi_i, \lambda_i|D) &= p_{\mathrm{G}}(\beta(\varphi_i, \lambda_i)|D)|J(\beta, (\varphi_i, \lambda_i))| \\ &= p_{\mathrm{G}}(\varphi_i|\lambda_i, D)p_{\mathrm{G}}(\lambda_i|D). \end{aligned} \tag{34}$$

The accompanying specification of Q corresponds with

$$Q = \sum_{j=1}^{w-1} \frac{p_{\mathrm{G}}(\lambda_j|D)|_{\lambda_j=0}\left[\int_{\bigcup_{i_j=1}^{n_j}\Theta_{\mathrm{G}_{i_j}}} p_{\mathrm{G}}(\varphi_j|\lambda_j,D)|_{\lambda_j=0}d\varphi_j\right]}{|A_j|^{\frac{1}{2}}} + \int_{\bigcup_{i=1}^{n_k} S_{\mathrm{G}_{i_w}}} p_{\mathrm{G}}(\beta|D)d\beta. \tag{35}$$

We refer to theorem 3 for further clarification of the different symbols. Theorem 2 shows that the posterior probabilities are invariant to the specification of β, (φ_i, λ_i) that satisfy assumption 1.

Similar to theorem 4 also the posterior probabilities (29) imply a posterior density for φ_i on Θ_{G_i}.

Theorem 6 *When assumption 1 holds, the posterior probabilities (29) induce the posterior densities*

$$p_{G_i}(\varphi_i|D) = \frac{p_G(\varphi_i|\lambda_i,D)|_{\lambda_i=0}}{\int_{\Theta_{G_i}} p_G(u|\lambda_i,D)|_{\lambda_i=0}du} \qquad i = 1, \ldots, n, \tag{36}$$

on Θ_{G_i}, and these posterior densities are invariant with respect to the specification of β, (φ_i, λ_i) that satisfy the conditions from assumption 1.

Proof. results directly from the proof of theorem 4. ■

Naturally, the posterior densities (36) also result when we update the prior $p_{\mathrm{G}_i}(\varphi_i|D)$ with the likelihood:

$$p_{\mathrm{G}_i}(\varphi_i|D) = \frac{p_{\mathrm{G}_i}(\varphi_i)\mathcal{L}(D|\beta)|_{\beta=f_i(\varphi_i)}}{\int_{\Theta_{\mathrm{G}_i}} p_{\mathrm{G}_i}(\psi_i)\mathcal{L}(D|u)|_{u=f_i(\psi_i)}d\psi_i} \qquad i = 1, \ldots, n. \tag{37}$$

Similarly, the posterior probabilities (29) result from the equality between the posterior odds ratio (POR) and the prior odds ratio (PROR) times the Bayes factor (BF):

$$\mathrm{POR}(\mathrm{G}_i, \mathrm{G}_j) = \mathrm{PROR}(\mathrm{G}_i, \mathrm{G}_j) \times \mathrm{BF}(\mathrm{G}_i, \mathrm{G}_j) \tag{38}$$

where

$$\text{POR}(\text{G}_i,\text{G}_j) = \tfrac{\Pr_{\text{G}}[\text{G}_i|D]}{\Pr_{\text{G}}[\text{G}_j|D]}, \qquad \text{PROR}(\text{G}_i,\text{G}_j) = \tfrac{\Pr_{\text{G}}[\text{G}_i]}{\Pr_{\text{G}}[\text{G}_j]}, \qquad \text{BF}(\text{G}_i,\text{G}_j) = \tfrac{p_{\text{G}_i}(D)}{p_{\text{G}_j}(D)}, \tag{39}$$

and $p_{\text{G}_i}(D)$ is the marginal data density,

$$\begin{aligned} p_{\text{G}_i}(D) &= \textstyle\int_{\Theta_{\text{G}_i}} p_{\text{G}_i}(\varphi_i)\mathcal{L}(D|\beta)|_{\beta=f_i(\varphi_i)}d\varphi_i \\ &= c_\beta \times \tfrac{p_{\text{G}}(\lambda_i|D)|_{\lambda_i=0}}{p_{\text{G}}(\lambda_i)|_{\lambda_i=0}} \times \tfrac{\int_{\Theta_{\text{G}_i}} p_{\text{G}}(\varphi_i|\lambda_i,D)|_{\lambda_i=0}d\varphi_i}{\int_{\Theta_{\text{G}_i}} p_{\text{G}}(\varphi_i|\lambda_i)|_{\lambda_i=0}d\varphi_i}, \end{aligned} \tag{40}$$

with $c_\beta = \int_{\mathbb{R}^k} p_{\text{G}}(\beta)\mathcal{L}(D|\beta)d\beta$, for a proof of (40) we refer to the appendix, see also Verdinelli and Wasserman (1995).

The specification of the prior $p_{\text{G}_i}(\varphi_i)$ satisfies the conditions for the Bayes factor to equal the Savage-Dickey density ratio, see *e.g.* Dickey (1971) and Verdinelli and Wasserman (1995). The Bayes factor is therefore equal to the ratio of the posterior heights divided by the prior heights:

$$\text{BF}(\text{G}_i,\text{G}_j) = \frac{\left[\tfrac{p_{\text{G}}(\lambda_i|D)|_{\lambda_i=0}}{p_{\text{G}}(\lambda_i)|_{\lambda_i=0}}\right]\left[\tfrac{\int_{\Theta_{\text{G}_i}} p_{\text{G}}(\varphi_i|\lambda_i,D)|_{\lambda_i=0}d\varphi_i}{\int_{\Theta_{\text{G}_i}} p_{\text{G}}(\varphi_i|\lambda_i)|_{\lambda_i=0}d\varphi_i}\right]}{\left[\tfrac{p_{\text{G}}(\lambda_j|D)|_{\lambda_j=0}}{p_{\text{G}}(\lambda_j)|_{\lambda_j=0}}\right]\left[\tfrac{\int_{\Theta_{\text{G}_j}} p_{\text{G}}(\varphi_j|\lambda_j,D)|_{\lambda_j=0}d\varphi_j}{\int_{\Theta_{\text{G}_j}} p_{\text{G}}(\varphi_j|\lambda_j)|_{\lambda_j=0}d\varphi_j}\right]}. \tag{41}$$

Substituting this expression for the Bayes factor in (38) results in the posterior odds ratio that accords with the one that results directly from the posterior probabilities (29), *i.e.*

$$\text{POR}(\text{G}_i,\text{G}_j) = \frac{Q_{\text{G}_i|D}}{Q_{\text{G}_j|D}}. \tag{42}$$

For our example with $n = 2$, the Bayes factor for comparing G_1 with G_2 becomes

$$\text{BF}(\text{G}_1,\text{G}_2) = \left[\tfrac{p_{\text{G}}(\lambda_1|D)|_{\lambda_1=0}}{p_{\text{G}}(\lambda_1)|_{\lambda_1=0}}\right]\left[\tfrac{\int_{\Theta_{\text{G}_1}} p_{\text{G}}(\varphi_1|\lambda_1,D)|_{\lambda_1=0}d\varphi_1}{\int_{\Theta_{\text{G}_1}} p_{\text{G}}(\varphi_1|\lambda_1)|_{\lambda_1=0}d\varphi_1}\right] \tag{43}$$

and the posterior odds ratio for comparing G_1 and G_2 reads

$$\begin{aligned} \text{POR}(\text{G}_1,\text{G}_2) &= Q_{\text{G}_1|D} \\ &= \tfrac{p_{\text{G}}(\lambda_1|D)|_{\lambda_1=0}}{|A_1|^{\frac{1}{2}}}\left[\textstyle\int_{\Theta_{\text{G}_1}} p_{\text{G}}(\varphi_1|\lambda_1,D)|_{\lambda_1=0}d\varphi_1\right]. \end{aligned} \tag{44}$$

The first part in the Bayes factor (43) is the Savage-Dickey density ratio, see Dickey (1971) and Verdinelli and Wasserman (1995). The second part arises because the integral of the conditional densities $p_{\text{G}}(\varphi_1|\lambda_1,D)|_{\lambda_1=0}$ and $p_{\text{G}}(\varphi_1|\lambda_1)|_{\lambda_1=0}$ over Θ_{G_1} does not have to be equal to one. When $\Theta_{\text{G}_1} = \mathbb{R}^{m_1}$, the integrals of both conditional densities are equal to one and the Bayes factor simplifies to the Savage-Dickey density ratio.

5 Jeffreys-Lindleys Paradox

We consider the Bayes factor in our example of comparing G_1 and G_2 (43). Furthermore, we use $\Theta_{\text{G}_1} = \mathbb{R}^{m_1}$ such that the Bayes factor is equal to the Savage-Dickey density ratio,

$$\text{BF}(\text{G}_1,\text{G}_2) = \tfrac{p_{\text{G}}(\lambda_1|D)|_{\lambda_1=0}}{p_{\text{G}}(\lambda_1)|_{\lambda_1=0}}. \tag{45}$$

The Bayes factor (45) favors G_1 above G_2 when $\lambda_1 = 0$ is more likely in the posterior than in the prior. Hence, it shows whether the information in the data is more favorable for G_1 compared to G_2 relative to the prior. When the likelihood dominates the prior in the posterior, for example, because of a large number of observations, the Bayes factor remains sensitive to the specification of the prior while the posterior is hardly sensitive to the prior anymore. This is known as the Jeffreys-Lindleys paradox, see *e.g.* Lindley (1957), Bernardo and Smith (1994), O'Hagan (1994) and Poirier (1995). It indicates that when the value of the prior on λ_1 in $\lambda_1 = 0$, $p_{\mathrm{G}}(\lambda_1)|_{\lambda_1=0}$, decreases that the Bayes factor increases, especially when the value of the prior becomes so small that it hardly affects the posterior anymore. We can achieve this by decreasing the prior precision on λ_1 (or β) or increasing its prior variance.

The Bayes factor is popular for model comparison because it is equal to the posterior odds ratio when the prior odds ratio is equal to one, which implies equal prior probabilities for G_1 and G_2 of one-half, see *e.g.* Kass and Raftery (1995). This specification of the posterior odds ratio is then also affected by the Jeffreys-Lindleys paradox. The sensitivity to the Jeffreys-Lindleys paradox results because a prior odds ratio equal to one does not correct the Bayes factor for the plausibility of G_1 compared to G_2 reflected in the prior. When we instead use prior probabilities for the prior odds ratio that correct the Bayes factor for the plausibility of G_1 compared to G_2 reflected in the prior, which are the prior probabilities induced by $p_{\mathrm{G}}(\beta)$ stated in theorem 3, we obtain the posterior odds ratio (44),

$$\mathrm{POR}(\mathrm{G}_1, \mathrm{G}_2) = \frac{p_{\mathrm{G}}(\lambda_1|D)|_{\lambda_1=0}}{|A_1|^{\frac{1}{2}}}. \tag{46}$$

The posterior odds ratio (46) shows the *a posteriori* support for G_1 compared to G_2 when we use the prior $p_{\mathrm{G}}(\beta)$. The posterior odds ratio (46) is not affected by Jeffreys-Lindleys paradox. This results because a decrease of the prior precision (or an increase of the prior variance) on λ_1 does not directly influence the posterior odds ratio (46). Hence, this posterior odds ratio can also be used in case of an improper prior because only the posterior $p_{\mathrm{G}}(\varphi_1, \lambda_1|D)$, or put differently the posterior $p_{\mathrm{G}}(\beta|D)$, needs to be proper. The posterior odds ratio (46) is related to the posterior information criterium of Phillips and Ploberger, see *e.g.* Phillips and Ploberger (1994,1996) and Phillips (1996).

6 Nuisance Parameters

For expository purposes, sofar, we only discussed regression models that contain no nuisance parameters. When model G (1) is a linear regression model conditional on a realization of a $l \times 1$ vector of nuisance parameters η, we specify it as

$$\mathrm{G} : P_y(\eta)y = P_X(\eta)X\beta + \varepsilon, \tag{47}$$

and model G_i as

$$\mathrm{G}_i : P_y(\eta)y = P_X(\eta)Xf_i(\varphi_i) + \varepsilon, \qquad i = 1, \ldots, n, \tag{48}$$

where the $T \times T$ matrices $P_y(\eta)$ and $P_X(\eta)$ are observable given a realization of the nuisance parameter vector η. The matrices $P_y(\eta)$ and $P_X(\eta)$ scale out the nuisance parameters such that the disturbances $\varepsilon : T \times 1$ have a pre-defined distribution that does not depend on nuisance parameters. We specify a joint prior on (β, η),

$$p_{\mathrm{G}}(\beta, \eta) = p_{\mathrm{G}}(\beta|\eta)p_{\mathrm{G}}(\eta). \tag{49}$$

When assumption 1 is satisfied, where we note that A_i should be fixed and therefore independent of η, theorems 1 and 2 hold and the Hausdorff integrals of the conditional prior and posterior of β given η, $p_{\mathrm{G}}(\beta|\eta)$ and $p_{\mathrm{G}}(\beta|\eta, D)$, over S_{G_i} are invariant with respect to specifications that satisfy assumption 1. This implies that theorems 3-6 apply such that the conditional prior and posterior of β given η imply invariant probabilities and densities.

Theorem 7 *When assumptions 1 and 2 hold and model G (47) is a linear regression model given a realization of the nuisance parameter vector* η*, the expressions of the prior and posterior probabilities in theorem 3 and 5 induced by* $p_G(\beta, \eta)$ *and* $p_G(\beta, \eta|D)$ *remain unaltered when we replace* Q_{G_i} *and* $Q_{G_i|D}$ *by*

$$\begin{aligned} Q_{G_i} = & \int_{\Theta_\eta}\left[\int_{\Theta_{G_i}} p_G(\beta|\eta)H_{m_i}(d\beta)\right]p_G(\eta)d\eta, & i = 1, \ldots, n, \\ Q_{G_i|D} = & \int_{\Theta_\eta}\left[\int_{\Theta_{G_i}} p_G(\beta|\eta, D)H_{m_i}(d\beta)\right]p_G(\eta|D)d\eta, & i = 1, \ldots, n, \end{aligned} \tag{50}$$

where Θ_η *is the parameter region of* η*. Similarly, the joint prior and posterior densities of* (φ_i, η) *defined on* $\Theta_{G_i} \times \Theta_\eta$ *that result from theorems 4 and 5 read*

$$\begin{aligned} p_{G_i}(\varphi_i, \eta) = & \frac{p_G(\varphi_i, \lambda_i|\eta)|_{\lambda_i=0}p_G(\eta)}{\int_{\Theta_\eta}\left[\int_{\Theta_{G_i}} p_G(\varphi_i, \lambda_i|\eta)|_{\lambda_i=0}d\varphi_i\right]p_G(\eta)d\eta}, & i = 1, \ldots, n, \\ p_{G_i}(\varphi_i, \eta|D) = & \frac{p_G(\varphi_i, \lambda_i|\eta, D)|_{\lambda_i=0}p_G(\eta|D)}{\int_{\Theta_\eta}\left[\int_{\Theta_{G_i}} p_G(\varphi_i, \lambda_i|\eta, D)|_{\lambda_i=0}d\varphi_i\right]p_G(\eta|D)d\eta}, & i = 1, \ldots, n. \end{aligned} \tag{51}$$

The probabilities that result from (50) and the densities (51) are invariant with respect to specifications that satisfy assumption 1.

Proof. results directly from theorem 2. ■

Theorem 7 shows that we can extend the invariant probabilities and densities to restrictions on linear regression models that condition on nuisance parameters. These restrictions should, however, be such that they do not involve the nuisance parameters. This explains why we refer to these parameters as nuisance parameters. Assumption 1 should also not involve the nuisance parameters in any of its elements.

For many regression models a function $g_i(\varphi_i, \lambda_i)$ can be constructed such that the conditions for theorem 7 are satisfied. Amongst these models are not only linear regression models but also models that are non-linear in the parameters, like, for example, cointegration, instrumental variables and auto-regressive moving-average (ARMA) models. Hence, for all these models prior/posterior probabilities and densities result through theorem 7 from a prior specified on the parameters of an encompassing linear regression model. In the next section, we briefly discuss a few examples of these models and focus on the specification of $g_i(\varphi_i, \lambda_i)$.

The prior/posterior probabilities and densities stated in theorem 7 are invariant with respect to transformations that satisfy the conditions from assumption 1. They are not invariant to transformations that involve the nuisance parameter η. Invariance to these kind of transformations can be achieved by an appropriate specification of the prior $p_{\mathrm{G}}(\beta, \eta)$ (49).

7 Examples

We discuss a few examples of regression models that result from a restriction on the parameters of an encompassing linear regression model. The first example concerns linear restrictions that lead to a nested linear regression model. The second and third example are concerned with non-linear restrictions that lead to a cointegration model and an autoregressive moving average model of order (1,1) (ARMA(1,1)). The vital element of the applicability of theorems 1-7 is the construction of a function $g_i(\varphi_i, \lambda_i)$ that makes assumption 1 hold. This function is not given by the specification of the regression model of interest and can be difficult to obtain. We therefore focus on the construction of $g_i(\varphi_i, \lambda_i)$. Once $g_i(\varphi_i, \lambda_i)$ is obtained, theorems 1-7 can be used. For regression models that are non-linear in the parameters, these theorems then imply priors that differ from the priors that are traditionally used for the parameters in these models. Hence, these traditional Bayesian analyzes are not in line with the Bayesian analysis of a linear regression model. Besides the cointegration and ARMA model, other models which have this property are, for example, instrumental variables regression models, see Kleibergen and Zivot (1998), and simultaneous equation models, see Kleibergen (1997) and Kleibergen and van Dijk (1998).

7.1 Linear regression model

Our first example considers linear restrictions on the parameters of a linear regression model, see also Tiao, Tan and Chang (1977),

$$\mathrm{G}: y = (X\ Z)\beta + u, \tag{52}$$

where $y: T \times 1$, $X: T \times m$, $Z: T \times (k-m)$, $m < k$, $\beta: k \times 1$, $\beta \in \mathbb{R}^k$ and $u \sim N(0, \sigma^2 I_T)$. Our linear regression model of interest G_1,

$$\mathrm{G}_1: y = X\varphi + u, \tag{53}$$

where $\varphi: m \times 1$, $\varphi \in \mathbb{R}^m$, is nested in the encompassing model G (52). We therefore specify S_{G_1} as

$$S_{\mathrm{G}_1} = \left\{ \varphi \in \mathbb{R}^m | \beta = \begin{pmatrix} \varphi \\ 0 \end{pmatrix} \right\}. \tag{54}$$

The model with which we compare G_1, G_2, is identical to G, such that S_{G_2} reads

$$S_{\mathrm{G}_2} = \left\{ \beta \in \mathbb{R}^k \right\}. \tag{55}$$

Because $\sigma^2 \in \mathbb{R}^+$ is a nuisance parameter, we respecify G_1 and G_2 towards the notation used in theorem 7,

$$\begin{array}{ll} \mathrm{G}_1: & P(\sigma)y = P(\sigma)X\varphi + \varepsilon, \\ \mathrm{G}_2: & P(\sigma)y = P(\sigma)(X\ Z)\beta + \varepsilon, \end{array} \tag{56}$$

where $P(\sigma) = \sigma^{-1} I_T$ and $\varepsilon \sim N(0, I_T)$.

A specification of $g_i(\varphi_i, \lambda_i)$ that makes assumption 1 hold is in this case straightforward to construct

$$\beta = \begin{pmatrix} I_m \\ 0 \end{pmatrix} \varphi + \begin{pmatrix} 0 \\ I_{k-m} \end{pmatrix} \lambda, \tag{57}$$

with $\lambda: (k-m) \times 1$ and $g(\varphi, \lambda) = (0\ I_{k-m})'\lambda$. Specification (57) satisfies assumption 1 since it is an invertible relationship and (a.) $\beta = (I_m\ 0)'\varphi \Leftrightarrow \lambda = 0$, (b.) all values of φ lead to a unique value of β both when $\lambda = 0$ and when $\lambda \neq 0$, (c.) $\frac{\partial g(\varphi, \lambda)}{\partial \lambda'} = (0\ I_{k-m})'$ and does not depend on φ. A prior specified on (β, σ^2) in model G (52) therefore implies invariant prior/posterior probabilities for G_1 and G_2 and densities for φ when we apply theorem 7.

7.2 Cointegration model

The second example that we consider is a cointegration model. Cointegration implies a non-linear restriction on the parameters of a linear regression model. The restriction that cointegration implies is that the long run multiplier of a vector autoregressive model has a reduced rank value, see *e.g.* Engle and Granger (1987) and Johansen (1991). For a vector autoregressive model of order 1, cointegration with r cointegrating vectors implies therefore that we can specify it as

$$\mathrm{G}_r : \Delta y_t = \alpha_r' \beta_r' y_{t-1} + u_t, \qquad r = 1, \ldots, k-1, \tag{58}$$

where y_t, $u_t : k \times 1$; $\Delta y_t = y_t - y_{t-1}$, α_r', $\beta_r : k \times r$, $\beta_r = (I_r \; -\beta_{2,r}')'$, $\beta_{2,r} : (k-r) \times r$ and u_t, $t = 1, \ldots, T$, are independently and identically normal distributed with mean zero and covariance matrix Ω. We reflect the cointegration models G_r (58) in matrix notation

$$\mathrm{G}_r : Y = X \beta_r \alpha_r + U, \qquad r = 1, \ldots, k-1, \tag{59}$$

where $Y = (\Delta y_1 \ldots \Delta y_T)'$, $X = (y_0 \ldots y_{T-1})'$, $U = (u_1 \ldots u_T)'$. The cointegration models G_r (59) are nested in the multi-variate linear regression model

$$\mathrm{G} : Y = X\Pi + U, \tag{60}$$

where $\Pi : k \times k$. We specify the cointegration models G_r (59) and the encompassing linear regression model G (60) in line with theorem 7 as

$$\begin{array}{rll} \mathrm{G}_r : & P(\Omega)\mathrm{vec}(Y) = P(\Omega)(I_k \otimes X)\mathrm{vec}(\beta_r \alpha_r) + \mathrm{vec}(\varepsilon), & r = 1, \ldots, k-1, \\ \mathrm{G} : & P(\Omega)\mathrm{vec}(Y) = P(\Omega)(I_k \otimes X)\mathrm{vec}(\Pi) + \mathrm{vec}(\varepsilon), & \end{array} \tag{61}$$

where $P(\Omega) = (\Omega^{-\frac{1}{2}} \otimes I_T)$, $\varepsilon = U\Omega^{-\frac{1}{2}}$, $\mathrm{vec}(\varepsilon) \sim N(0, I_{kT})$. Equation (61) shows that G_r, $r = 1, \ldots, k-1$, is represented using the lower dimensional sets

$$S_{\mathrm{G}_r} = \left\{ \alpha_r \in \mathbb{R}^{k,r}, \; \beta_{2,r} \in \mathbb{R}^{(k-r),r} | \Pi = \begin{pmatrix} I_r \\ -\beta_{2,r} \end{pmatrix} \alpha_r \right\}, \qquad r = 1, \ldots, k-1. \tag{62}$$

The unrestricted full rank model G_k is identical to G (60) such that S_{G_k} reads

$$S_{\mathrm{G}_k} = \left\{ \Pi \in \mathbb{R}^{k,k} \right\}. \tag{63}$$

Because cointegration imposes a non-linear restriction on the parameters of a linear regression model, the specification of a function $g_i(\varphi_i, \lambda_i)$ that makes assumption 1 hold is rather difficult to obtain. In Kleibergen and Paap (2000) a specification of Π that, results from a singular value decomposition and, makes assumption 1 hold is given:

$$\begin{array}{rll} \Pi = & \beta_r \alpha_r + \beta_{r,\perp} \lambda_r \alpha_{r,\perp}, & r = 1, \ldots, k-1, \\ & \Leftrightarrow & \\ \mathrm{vec}(\Pi) = & \mathrm{vec}(\beta_r \alpha_r) + (\alpha_{r,\perp}' \otimes \beta_{r,\perp})\mathrm{vec}(\lambda_r), & r = 1, \ldots, k-1, \end{array} \tag{64}$$

where $\lambda_r : (k-r) \times (k-r)$; $\beta_{r,\perp}$, $\alpha_{r,\perp}' : k \times (k-r)$ and $\beta_{r,\perp}' \beta_r \equiv 0$, $\beta_{r,\perp}' \beta_{r,\perp} \equiv I_{k-r}$, $\alpha_r \alpha_{r,\perp}' \equiv 0$, $\alpha_{r,\perp} \alpha_{r,\perp}' \equiv I_{k-r}$, such that

$$g_r(\varphi_r, \lambda_r) = (\alpha_{r,\perp}' \otimes \beta_{r,\perp})\mathrm{vec}(\lambda_r), \qquad r = 1, \ldots, k-1, \tag{65}$$

with $\varphi_r = (\alpha_r, \beta_{2,r})$. The specification of Π (64) satisfies assumption 1 because, there is an invertible relationship between Π and $(\alpha_r, \beta_{2,r}, \lambda_r)$, see Kleibergen and Paap (2000), (a.) $\Pi = \beta_r \alpha_r \Leftrightarrow \lambda_r = 0$, (b.) $(\alpha_r, \beta_{2,r})$ imply a unique value of $\beta_r \alpha_r$ when α_r has full rank and identically $(\alpha_r, \beta_{2,r}, \lambda_r)$ imply a unique value of $\beta_r \alpha_r + \beta_{r,\perp} \lambda_r \alpha_{r,\perp}$ when α_r has full rank, (c.) $\frac{\partial g_r(\varphi_r, \lambda_r)}{\partial \mathrm{vec}(\lambda_r)'} = (\alpha'_{r,\perp} \otimes \beta_{r,\perp})$ such that $\left(\frac{\partial g_r(\varphi_r, \lambda_r)}{\partial \mathrm{vec}(\lambda_r)'}\right)' \left(\frac{\partial g_r(\varphi_r, \lambda_r)}{\partial \mathrm{vec}(\lambda_r)'}\right) = I_{(k-r)^2}$ and does not depend on φ_r. Hence, all conditions of assumption 1 are satisfied. Theorem 7 therefore applies and a prior specified on (Π, Ω) in G implies a prior probability for G_r, $r = 1, \ldots, k$, and a prior for $(\alpha_r, \beta_{2,r}, \Omega)$ in G_r that are invariant with respect to the specification of Π and $(\alpha_r, \beta_{2,r}, \lambda_r)$ that satisfy assumption 1. For more details about the resulting Bayesian analysis of the cointegration model, we refer to Kleibergen and Paap (2000).

7.3 ARMA(1,1)

As another example of a non-linear restriction on the parameters of a linear regression model, we consider the autoregressive moving average (ARMA) model of order (1,1), see *e.g.* Box, Jenkins and Reinsel (1994),

$$\mathrm{G}_1 : y_t = \rho y_{t-1} - \alpha u_{t-1} + u_t, \qquad t = 1, \ldots, T, \tag{66}$$

where the disturbances u_t are independently and identically distributed, $u_t \sim N(0, \sigma^2)$. When we recurrently substitute u_{t-1} in (66), we obtain

$$\mathrm{G}_1 : y_t = (\rho - \alpha) \sum_{j=1}^{T} \alpha^{j-1} y_{t-j} + u_t, \qquad t = 1, \ldots, T. \tag{67}$$

We specify (67) as a regression model that is non-linear in the parameters,

$$\mathrm{G}_1 : y = X f(\alpha, \rho) + u, \tag{68}$$

where $y = (y_1 \ldots y_T)'$, $X = (x_1 \ldots x_T)'$, $x_i = (y_{i-1} \ldots y_0 \; 0 \ldots 0 \;)' : T \times 1$, $i = 1, \ldots, T$; $u = (u_1 \ldots u_T)'$, and

$$f(\alpha, \rho) = (\rho - \alpha) \begin{pmatrix} 1 \\ \alpha \\ \vdots \\ \alpha^{T-1} \end{pmatrix} : T \times 1. \tag{69}$$

Model G_1 (68) is nested in the linear regression model

$$\mathrm{G} : y = X\beta + u, \tag{70}$$

with $\beta : T \times 1$. We specify both G_1 (68) and G (70) in the notation of theorem 7,

$$\begin{array}{rl} \mathrm{G}_1 : & P(\sigma) y = P(\sigma) X f(\alpha, \rho) + \varepsilon, \\ \mathrm{G} : & P(\sigma) y = P(\sigma) X \beta + \varepsilon, \end{array} \tag{71}$$

with $P(\sigma) = \sigma^{-1} I_T$ and $\varepsilon \sim N(0, I_T)$.

The ARMA(1,1) model imposes a non-linear restriction on the parameters of (70), $\beta = f(\alpha, \rho)$. This implies that it is not straightforward to obtain a specification of $g_i(\varphi_i, \lambda_i)$ that makes assumption 1 hold. A (unrestricted) specification of β that gives such a function $g_i(\varphi_i, \lambda_i)$ is

$$\beta = (\rho - \alpha)\begin{pmatrix} 1 \\ \alpha \\ \vdots \\ \alpha^{T-1} \end{pmatrix} + \begin{pmatrix} 0 \\ I_{T-2} \end{pmatrix}\lambda, \tag{72}$$

with $\lambda : (T-2) \times 1$ and $g(\varphi, \lambda) = (0 \ I_{T-2})'\lambda$ with $\varphi = (\alpha, \rho)$. Equation (72) satisfies the conditions from assumption 1 since: β has an invertible relationship with (α, ρ, λ), (a.) $\beta = f(\alpha, \rho) \Leftrightarrow \lambda = 0$, (b.) (ρ, α) implies a unique value of $f(\alpha, \rho)$ when $\rho - \alpha \neq 0$, identically (ρ, α, λ) implies a unique value of $f(\alpha, \rho) + (0 \ I_{T-2})'\lambda$ when $\rho - \alpha \neq 0$, (c.) $\frac{\partial g(\varphi,\lambda)}{\partial \lambda'} = (0 \ I_{T-2})'$ and independent of (α, ρ). Theorem 7 therefore applies and a prior that is specified on (β, σ^2) in G induces a prior probability for G_1 and a prior on (α, ρ, σ^2) that are invariant with respect to the specification of β and (α, ρ, λ) that satisfy assumption 1. For more details on the resulting Bayesian analysis of the ARMA(1,1) model, we refer to Kleibergen and Hoek (1999).

8 Conclusions

The paper obtains expressions for prior/posterior probabilities and densities of the parameters of nested regression models that are induced by the prior/posterior on the parameters of an encompassing linear regression model. The resulting probabilities and densities are invariant with respect to specifications that satisfy a necessary set of assumptions. Hence, by specifying a prior and a likelihood for the parameters of an encompassing linear regression model, we obtain a complete Bayesian analysis, that includes both prior/posterior probabilities and densities, for all of its nested regression models that allow for a specification that satisfies the set of assumptions. The resulting Bayesian analyzes of these nested regression models are in line with one another.

The Bayes factor in the resulting analysis corresponds with the Savage-Dickey density ratio and equals the ratio of the posterior and prior height in the hypothesized parameter point. When we multiply the Bayes factor with the prior odds ratio, we obtain the posterior odds ratio. Because both the prior probability and density result from the same prior on the parameters of the encompassing linear regression model, the prior odds ratio corrects the Bayes factor, for the plausibility of the competing models reflected in the prior, in the expression of the posterior odds ratio. The posterior odds ratio is therefore robust to the Jeffreys-Lindleys paradox.

Applications of the above results are especially important for regression models that result from non-linear restrictions on the parameters of encompassing linear regression models. In these models, the resulting analysis leads to priors and posteriors that are different from the ones that are used traditionally. The traditional Bayesian analysis leads to anomalies in these models, like, for example, in simultaneous equation, see Kleibergen (1997) and Kleibergen and van Dijk (1998), and cointegration models, see Kleibergen and van Dijk (1994). When we deduce the priors and posteriors of the parameters in these models from priors and posteriors on the parameters of encompassing linear regression models, these anomalies disappear, see *e.g.* Kleibergen (1997), Kleibergen and van Dijk (1998), Kleibergen and van Paap (1998). This further illustrates the importance of the analysis.

Appendix

Proof of Theorem 1.

Before we obtain the specification of the Hausdorff measure, we note the structure that assumption 1 imposes on the jacobian of the transformation from β to (φ_i, λ_i) :

$$J(\beta, (\varphi_i, \lambda_i)) = \left(\begin{array}{cc} \frac{\partial f_i}{\partial \varphi_i'} + \frac{\partial g_i}{\partial \varphi_i'} & \frac{\partial g_i}{\partial \lambda_i'} \end{array} \right).$$

Because $g_i(\varphi_i, \lambda_i)$ is a strictly monotonic function of λ_i and $g_i(\varphi_i, \lambda_i) = 0 \Leftrightarrow \lambda_i = 0$, $\frac{\partial g_i}{\partial \varphi_i'}|_{\lambda_i=0} = 0$. Hence, the jacobian in $\lambda_i = 0$ reads

$$J(\beta, (\varphi_i, \lambda_i))|_{\lambda_i=0} = \left(\begin{array}{cc} \frac{\partial f_i}{\partial \varphi_i'} & \frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \end{array} \right),$$

and

$$\begin{aligned} |J(\beta, (\varphi_i, \lambda_i))|_{\lambda_i=0}| = & \left| \left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right)' \left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right) \right|^{\frac{1}{2}} \left| \left(\frac{\partial f_i}{\partial \varphi_i'} \right)' M_{\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right)} \left(\frac{\partial f_i}{\partial \varphi_i'} \right) \right|^{\frac{1}{2}} \\ = & \left| \left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right)' M_{\left(\frac{\partial f_i}{\partial \varphi_i'} \right)} \left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right) \right|^{\frac{1}{2}} \left| \left(\frac{\partial f_i}{\partial \varphi_i'} \right)' \left(\frac{\partial f_i}{\partial \varphi_i'} \right) \right|^{\frac{1}{2}}, \end{aligned}$$

where $\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right)' \left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right) = A_i$.

The Hausdorff measure $\mathrm{H}_{m_i}(W_{\mathrm{G}_i})$ is obtained by considering that $g_i(\varphi_i, \lambda_i)$ is a strictly monotonic function of λ_i. We therefore consider a sequence of sets centered at $\lambda_i = 0$,

$$W_{\mathrm{G}_i}(\rho) = \left\{ \varphi_i \in \Omega_{\mathrm{G}_i} \subset \mathbb{R}^{m_i},\ \lambda_i \in B_{k-m_i}(0, \rho) \subset \mathbb{R}^{k-m_i} | \beta = f(\varphi_i) + g_i(\varphi_i, \lambda_i) \right\},$$

where $B_{k-m_i}(0, \rho)$ is a $(k - m_i)$-dimensional sphere with radius ρ centered at 0. We use a limiting sequence of $W_{\mathrm{G}_i}(\rho)$ that is obtained by letting ρ converge to zero,

$$\lim_{\rho \to 0} W_{\mathrm{G}_i}(\rho) = W_{\mathrm{G}_i}.$$

This results because $g_i(\varphi_i, \lambda_i)$ is a strictly monotonic function of λ.

Because $\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right)' \left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right) = A_i$ and $\frac{\partial g_i}{\partial \varphi_i'}|_{\lambda_i=0} = 0$, for values of ρ close to zero the Lebesque measure of $W_{\mathrm{G}_i}(\rho)$, $L_k(W_{\mathrm{G}_i}(\rho))$, can be specified as:

$$\begin{aligned} L_k(W_{\mathrm{G}_i}(\rho)) = & \int_{\Omega_{\mathrm{G}\,i}} \int_{B_{\mathrm{G}\,i}(0,\rho)} |J(\beta, (\varphi_i, \lambda_i))| \, d\lambda_i d\varphi_i \\ \approx & \int_{\Omega_{\mathrm{G}\,i}} \left| \left(\frac{\partial f_i}{\partial \varphi_i'} \right)' M_{\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right)} \left(\frac{\partial f_i}{\partial \varphi_i'} \right) \right|^{\frac{1}{2}} \int_{B_{k-m_i}(0,\rho)} |A_i|^{\frac{1}{2}} \, d\lambda_i d\varphi_i \\ \approx & \left\{ \int_{\Omega_{\mathrm{G}\,i}} \left| \left(\frac{\partial f_i}{\partial \varphi_i'} \right)' M_{\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right)} \left(\frac{\partial f_i}{\partial \varphi_i'} \right) \right|^{\frac{1}{2}} d\varphi_i \right\} \left\{ \int_{B_{k-m_i}(0,\rho)} |A_i|^{\frac{1}{2}} \, d\lambda_i \right\} \\ \approx & \left\{ \int_{\Omega_{\mathrm{G}\,i}} \left| \left(\frac{\partial f_i}{\partial \varphi_i'} \right)' M_{\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0} \right)} \left(\frac{\partial f_i}{\partial \varphi_i'} \right) \right|^{\frac{1}{2}} d\varphi_i \right\} |A_i|^{\frac{1}{2}} \left\{ \int_{B_{k-m_i}(0,\rho)} d\lambda_i \right\}. \end{aligned}$$

To construct the Hausdorff measure, we divide $L_k(W_{\mathrm{G}_i}(\rho))$ by the measure of the set of which we construct the limiting sequence $B_{\mathrm{G}_i}(0,\rho)$ transformed using the strictly monotonic function of λ_i, $g_i(\varphi_i,\lambda_i)$. In $\lambda_i=0$, the derivative $\frac{\partial g_i(\varphi_i,\lambda_i)}{\partial \varphi_i'}|_{\lambda_i=0}$ is equal to zero. This implies that, for the construction of the measure of the limiting sequence, $g_i(\varphi_i,\lambda_i)$ can be considered as a function of λ_i only. For values of ρ close to zero, the measure $H_{k-m_i}(g(\varphi_i,B_i(0,\rho)))$ does therefore not depend on φ_i and can be specified as, see Billingsley (1986):

$$\begin{aligned}
H_{k-m_i}(g(\varphi_i,B_i(0,\rho))) &= \int_{B_{k-m_i}(0,\rho)} \left|\frac{\partial g_i}{\partial \lambda_i}\right| H_{k-m_i}(d\lambda_i) \\
&= \int_{B_{k-m_i}(0,\rho)} \left|\left(\frac{\partial g_i}{\partial \lambda_i}\right)'\left(\frac{\partial g_i}{\partial \lambda_i}\right)\right|^{\frac{1}{2}} d\lambda_i \\
&\approx \int_{B_{k-m_i}(0,\rho)} \left|\left(\frac{\partial g_i}{\partial \lambda_i}|_{\lambda_i=0}\right)'\left(\frac{\partial g_i}{\partial \lambda_i}|_{\lambda_i=0}\right)\right|^{\frac{1}{2}} d\lambda_i \\
&\approx \int_{B_{k-m_i}(0,\rho)} |A_i|^{\frac{1}{2}}\, d\lambda_i \\
&\approx |A_i|^{\frac{1}{2}} \int_{B_{k-m_i}(0,\rho)} d\lambda_i.
\end{aligned}$$

The specification of the Hausdorff measure then becomes:

$$\begin{aligned}
H_{m_i}(W_{\mathrm{G}_i}) &= \lim_{\rho\to 0}\left[\frac{L_k(W_{\mathrm{G}_i}(\rho))}{H_{k-m_i}(g(\varphi_i,B_i(0,\rho)))}\right] \\
&= \lim_{\rho\to 0}\left[\frac{\left\{\int_{\Omega_{\mathrm{G}_i}}\left|\left(\frac{\partial f_i}{\partial \varphi_i'}\right)' M_{\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0}\right)}\left(\frac{\partial f_i}{\partial \varphi_i'}\right)\right|^{\frac{1}{2}} d\varphi_i\right\}|A_i|^{\frac{1}{2}}\left\{\int_{B_{k-m_i}(0,\rho)} d\lambda_i\right\}}{|A_i|^{\frac{1}{2}}\left\{\int_{B_{k-m_i}(0,\rho)} d\lambda_i\right\}}\right] \\
&= \int_{\Omega_{\mathrm{G}_i}}\left|\left(\frac{\partial f_i}{\partial \varphi_i'}\right)' M_{\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0}\right)}\left(\frac{\partial f_i}{\partial \varphi_i'}\right)\right|^{\frac{1}{2}} d\varphi_i.
\end{aligned}$$

To show the invariance of the Hausdorff measure, we consider an invertible function $h:\mathbb{R}^k\to\mathbb{R}^k$, $\mu=h(\beta)$. Because of assumption 1, we can specify β as

$$\beta = f_i(\varphi_i) + g_i(\varphi_i,\lambda_i)$$

and μ can therefore be specified as

$$\mu = l_i(\psi_i) + r_i(\psi_i,\theta_i),$$

with $l_i(\psi_i)=h(f_i(\varphi_i))$ and $r(\psi_i,\theta_i)=h(f_i(\varphi_i)+g_i(\varphi_i,\lambda_i))-h(f_i(\varphi_i))$. Because of assumption 1b, that $g_i(\varphi_i,\lambda_i)$ is a strictly monotonic function of λ_i, h has to be strict monotonic. This implies that $\left(\frac{\partial h}{\partial \beta'}\right)'\left(\frac{\partial h}{\partial \beta'}\right)$ is a positive definite symmetric matrix for all values of β and that θ_i is an invertible function of λ_i only. Because of assumption 1c, $\frac{\partial r_i}{\partial \theta_i'}=\frac{\partial h}{\partial \beta'}\frac{\partial g_i}{\partial \lambda_i'}\frac{\partial \lambda_i}{\partial \theta_i'}$ should be such that

$$\begin{aligned}
\left(\frac{\partial r_i}{\partial \theta_i'}|_{\theta_i=0}\right)'\left(\frac{\partial r_i}{\partial \theta_i'}|_{\theta_i=0}\right) &= B_i \Leftrightarrow \\
\left(\frac{\partial h}{\partial \beta'}\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0}\right)\left(\frac{\partial \lambda_i}{\partial \theta_i'}|_{\theta_i=0}\right)\right)'\left(\frac{\partial h}{\partial \beta'}\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0}\right)\left(\frac{\partial \lambda_i}{\partial \theta_i'}|_{\theta_i=0}\right)\right) &= B_i \Leftrightarrow \\
\left(\frac{\partial \lambda_i}{\partial \theta_i'}|_{\theta_i=0}\right)'\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0}\right)'\left(\frac{\partial h}{\partial \beta'}\right)'\left(\frac{\partial h}{\partial \beta'}\right)\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0}\right)\left(\frac{\partial \lambda_i}{\partial \theta_i'}|_{\theta_i=0}\right) &= B_i,
\end{aligned}$$

with B_i independent of ψ_i. Since θ_i is an invertible function of λ_i only and $\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0}\right)'\left(\frac{\partial g_i}{\partial \lambda_i'}|_{\lambda_i=0}\right) = A_i$, with A_i independent of φ_i, $\left(\frac{\partial h}{\partial \beta'}\right)'\left(\frac{\partial h}{\partial \beta'}\right)$ should be equal to some fixed positive definite symmetric matrix that is independent of β. Unlike g_i, h is an invertible function such that the only specification of h that satisfies all conditions is an invertible linear function. Hence every specification $\mu = l_i(\psi_i) + r_i(\psi_i, \theta_i)$ that satisfies assumption 1 is such that (1.) μ is an invertible linear function of β and (2.) θ_i is an invertible function of λ_i only and ψ_i is an invertible function of φ_i only. It is straightforward to show that these transformations lead to an identical Hausdorff measure.

Proof of Theorem 2.

For values of ρ close to zero, $\int_{W_{G_i}(\rho)} q(\beta)d\beta$ can be specified as

$$\begin{aligned}\int_{W_{G_i}(\rho)} q(\beta)d\beta = &\int_{\Omega_{G_i}}\int_{B_{k-m_i}(0,\rho)} q(\beta(\varphi_i,\lambda_i))\left|J(\beta,(\varphi_i,\lambda_i))\right| d\lambda_i d\varphi_i \\ \approx &\int_{\Omega_{G_i}}\int_{B_{k-m_i}(0,\rho)} q(\beta(\varphi_i,\lambda_i)|_{\lambda_i=0})\left|J(\beta,(\varphi_i,\lambda_i))|_{\lambda_i=0}\right| d\lambda_i d\varphi_i \\ \approx &\left\{\int_{\Omega_{G_i}} q(\beta(\varphi_i,\lambda_i)|_{\lambda_i=0})\left|J(\beta,(\varphi_i,\lambda_i))|_{\lambda_i=0}\right| d\varphi_i\right\}\left\{\int_{B_{k-m_i}(0,\rho)} d\lambda_i\right\}.\end{aligned}$$

To obtain the Haussdorf integral, we divide $\int_{W_{G_i}(\rho)} q(\beta)d\beta$ by $H_{k-m_i}(g(\varphi_i, B_{k-m_i}(0,\rho)))$ that we constructed previously,

$$\begin{aligned}\int_{W_{G_i}} q(\beta)H_{m_i}(d\beta) = &\lim_{\rho\to 0}\left[\frac{\int_{W_{G_i}(\rho)} q(\beta)d\beta}{H_{k-m_i}(g(\varphi_i,B_{k-m_i}(0,\rho)))}\right] \\ = &\lim_{\rho\to 0}\left[\frac{\left\{\int_{\Theta_{G_i}} q(\beta(\varphi_i,\lambda_i)|_{\lambda_i=0})\left|J(\beta,(\varphi_i,\lambda_i))|_{\lambda_i=0}\right|d\varphi_i\right\}\left\{\int_{B_{k-m_i}(0,\rho)} d\lambda_i\right\}}{|A_i|^{\frac{1}{2}}\left\{\int_{B_{k-m_i}(0,\rho)} d\lambda_i\right\}}\right] \\ = &\frac{1}{|A_i|^{\frac{1}{2}}}\left\{\int_{\Theta_{G_i}} q(\beta(\varphi_i,\lambda_i)|_{\lambda_i=0})\left|J(\beta,(\varphi_i,\lambda_i))|_{\lambda_i=0}\right| d\varphi_i\right\}.\end{aligned}$$

The proof of the invariance of the Hausdorff integral to specifications of β, (φ_i, λ_i) that satisfy assumption 1 is analogous to the proof for theorem 1.

Proof of Theorem 4.

Equation (24) gives the definition of a density function. The invariance of it follows from the proof of theorem 1. We have shown in this proof that when

$$\beta = f_i(\varphi_i) + g_i(\varphi_i, \lambda_i)$$

and

$$\mu = l_i(\psi_i) + r_i(\psi_i, \theta_i),$$

are two specifications that satisfy assumption 1 that ψ_i is an invertible function of φ_i only and θ_i is an invertible function of λ_i only. Hence, we can independently transform φ_i to ψ_i and λ_i to θ_i. This does not affect the specification of the prior from theorem 4.

Proof of equation (40)

$$\begin{aligned}
p_{\mathrm{G}_i}(D) = & \int_{\Theta_{\mathrm{G}_i}} p_{\mathrm{G}_i}(\varphi_i)\mathcal{L}(D|\beta)|_{\beta=f_i(\varphi_i)} d\varphi_i \\
= & \; p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0} \left\{ \int_{\Theta_{\mathrm{G}_i}} \left[\frac{p_{\mathrm{G}_i}(\varphi_i)\mathcal{L}(D|\beta)|_{\beta=f_i(\varphi_i)}}{p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0}} \right] d\varphi_i \right\} \\
= & \; p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0} \left\{ \int_{\Theta_{\mathrm{G}_i}} \left[\frac{p_{\mathrm{G}_i}(\varphi_i)\mathcal{L}(D|\beta)|_{\beta=f_i(\varphi_i)} p_{\mathrm{G}}(\varphi_i|\lambda_i,D)|_{\lambda_i=0}}{p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0} p_{\mathrm{G}}(\varphi_i|\lambda_i,D)|_{\lambda_i=0}} \right] d\varphi_i \right\} \\
= & \; p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0} \left\{ \int_{\Theta_{\mathrm{G}_i}} \left[\frac{p_{\mathrm{G}_i}(\varphi_i)\mathcal{L}(D|\beta)|_{\beta=f_i(\varphi_i)} p_{\mathrm{G}}(\varphi_i|\lambda_i,D)|_{\lambda_i=0}}{p_{\mathrm{G}}(\varphi_i,\lambda_i|D)|_{\lambda_i=0}} \right] d\varphi_i \right\} \\
= & \; p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0} \left\{ \int_{\Theta_{\mathrm{G}_i}} \left[\frac{p_{\mathrm{G}_i}(\varphi_i)\mathcal{L}(D|\beta)|_{\beta=f_i(\varphi_i)} p_{\mathrm{G}}(\varphi_i|\lambda_i,D)|_{\lambda_i=0}}{\frac{p_{\mathrm{G}}(\varphi_i,\lambda_i)|_{\lambda_i=0}\mathcal{L}(D|\beta)|_{\beta=f_i(\varphi_i)})}{c_\beta}} \right] d\varphi_i \right\} \\
= & \; c_\beta \times p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0} \left\{ \int_{\Theta_{\mathrm{G}_i}} \left[\frac{p_{\mathrm{G}_i}(\varphi_i) p_{\mathrm{G}}(\varphi_i|\lambda_i,D)|_{\lambda_i=0}}{p_{\mathrm{G}}(\varphi_i,\lambda_i)|_{\lambda_i=0}} \right] d\varphi_i \right\} \\
= & \; c_\beta \times \frac{p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0}}{\int_{\Theta_{\mathrm{G}_i}} p_{\mathrm{G}}(\varphi_i|\lambda_i)|_{\lambda_i=0} d\varphi_i} \left\{ \int_{\Theta_{\mathrm{G}_i}} \left[\frac{p_{\mathrm{G}}(\varphi_i|\lambda_i)|_{\lambda_i=0} p_{\mathrm{G}}(\varphi_i|\lambda_i,D)|_{\lambda_i=0}}{p_{\mathrm{G}}(\lambda_i)|_{\lambda_i=0} p_{\mathrm{G}}(\varphi_i|\lambda_i)|_{\lambda_i=0}} \right] d\varphi_i \right\} \\
= & \; c_\beta \times \frac{p_{\mathrm{G}}(\lambda_i|D)|_{\lambda_i=0}}{p_{\mathrm{G}}(\lambda_i)|_{\lambda_i=0}} \times \frac{\int_{\Theta_{\mathrm{G}_i}} p_{\mathrm{G}}(\varphi_i|\lambda_i,D)|_{\lambda_i=0} d\varphi_i}{\int_{\Theta_{\mathrm{G}_i}} p_{\mathrm{G}}(\varphi_i|\lambda_i)|_{\lambda_i=0} d\varphi_i}
\end{aligned}$$

where

$$\begin{aligned}
c_\beta = & \int_{\mathbb{R}^k} p_{\mathrm{G}}(\beta)\mathcal{L}(D|\beta) d\beta \\
= & \int_{\mathbb{R}^{m_i}} \int_{\mathbb{R}^{k-m_i}} p_{\mathrm{G}}(\varphi_i,\lambda_i)\mathcal{L}(D|\beta(\lambda_i,\varphi_i)) d\lambda_i d\varphi_i, \\
p_{\mathrm{G}_i}(\varphi_i) = & \frac{p_{\mathrm{G}}(\varphi_i|\lambda_i)|_{\lambda_i=0}}{\int_{\Theta_{\mathrm{G}_i}} p_{\mathrm{G}}(u|\lambda_i)|_{\lambda_i=0} du}.
\end{aligned}$$

References

[1] Bernardo, J.M. and A.F.M. Smith. *Bayesian Theory.* Wiley, New York, 1994.

[2] Billingsley, P. *Probability and Measure.* Wiley (New York), (1986).

[3] Box G.E.P., G.M. Jenkins and G.C. Reinsel. *Time Series Analysis: Forecasting and Control.* Prentice Hall, 1994.

[4] De Finetti, B. *Probability, Induction and Statistics.* Wiley (New York), 1972.

[5] Dickey, J.M. The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. *The Annals of Mathematical Statistics*, **42**:204–223, 1971.

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

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

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

[9] Kass, R.E. and A.E. Raftery. Bayes Factors. *Journal of the American Statistical Association*, **90**:773–795, 1995.

[10] Kleibergen, F. Bayesian Simultaneous Equations Analysis using Equality Restricted Random Variables. In *1997 Proceedings of the Section on Bayesian Statistical Science*, pages 141–146. American Statistical Association, 1997.

[11] Kleibergen, F. and H. Hoek. Bayesian Analysis of ARMA models. Tinbergen Institute Discussion Paper TI 2000-027/4, 1999.

[12] Kleibergen, F. and R. Paap. Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration. *Journal of Econometrics*, 1998. Forthcoming, Econometric Institute Report 9821/A.

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

[14] Kleibergen, F. and H.K. van Dijk. Bayesian Simultaneous Equation Analysis using Reduced Rank Structures. *Econometric Theory*, **14**:701–743, 1998.

[15] Kleibergen F. and E. Zivot. Bayesian and Classical Approaches to Instrumental Variable Regression. *Journal of Econometrics*, 1998. Forthcoming, Econometric Institute Report 9835/A.

[16] Kolmogorov, A.N. *Foundations of the Theory of Probability.* Chelsea (New York), (1950).

[17] Lindley, D.V. A Statistical Paradox. *Biometrika*, **44**:187–192, (1957).

[18] McCulloch, R.E., and P.E. Rossi. Bayes factors for nonlinear hypotheses and likelihood distributions. *Biometrika*, **79**:663–676, 1992.

[19] O'Hagan, A. *Bayesian Theory*, volume **2B** of *Kendall's Advanced Theory of Statistics*. Arnold, 1994.

[20] Phillips, P.C.B. Econometric Model Determination. *Econometrica*, **64**:763–812, (1996).

[21] Phillips, P.C.B. and W. Ploberger. Posterior Odds Testing for a Unit Root with Data-based Model Selection. *Econometric Theory*, **10**:774–808, (1994).

[22] Phillips, P.C.B. and W. Ploberger. An Asymptotic Theory of Bayesian Inference for Time Series. *Econometrica*, **64**:381–412, (1996).

[23] Poirier, D.J. *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press, (Cambridge, MA), (1995).

[24] Tiao, G.C., W.-Y. Tan and Y.-C. Chang. Some Aspects of Bivariate Regression Subject to Linear Constraints. *Journal of Econometrics*, **5**:13–35, 1977.

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

[26] Wolpert, R.J. Comment on: Inference for a Deterministic Population Dynamics Model for Bowhead Whales, by, A.E. Raftery, G.H. Rivens and J.E. Zeh. *Journal of the American Statistical Association*, **90**:426–427, (1995).