Bayesian Approach and Identification

Kociecki, Andrzej

National Bank of Poland

25 April 2013

BAYESIAN APPROACH AND IDENTIFICATION

Andrzej Kocięcki
National Bank of Poland
e–mail: andrzej.kociecki@nbp.pl

First version: April 2013
This version: April 2013

Abstract: The paper aims at systematic placement of identification concept within Bayesian approach. Pointing to some deficiencies of the standard Bayesian language to describe identification problem we propose several useful characterizations that seem to be intuitively sound and attractive given their potential applications. We offer comprehensive interpretations for them. Moreover we introduce the concepts of uniform, marginal and faithful identification. We argue that all these concepts may have practical significance. Our theoretical development is illustrated with a number of simple examples and one real application i.e. Structural VAR model.

I. INTRODUCTION

Studying classical papers from early sixties (by Drèze, Rothenberg, Zellner and others) one has unambiguous impression that attractiveness of Bayesian inference in econometrics was linked to the possibility of imposing inexact (approximate) restrictions in the Simultaneous Equations Models (SEM) instead of exact ones. Hence one may say that identification problem was a fuel for the first prototypes of Bayesian engines in econometrics. Presently there is some ambiguity concerning the legitimacy of approximate restrictions introduced within Bayesian framework. Although Drèze (1972,1974,1976), Zellner (1971), Rothenberg (1973) seem to stand by this opinion, Kadane (1974) disagrees. However it appears that most Bayesians reached a consensus that identification is a property of the structural model likelihood function and probabilistic restrictions (i.e. non–degenerate prior) can not

be a substitute for exact restrictions that are needed to identify the sampling model[1]. Although the same position is upheld in our paper we propose (hopefully) an interesting view on this subject.

We ask the same research questions as the pioneers of Bayesian econometrics did. Since most restrictions employed in economic modeling should be treated not as genuine statements but only as useful approximations, can we use in some way the prior which expresses our uncertainty concerning "proper" restrictions in order to "objectivize" or "robustify" our inference? Do we really believe that exclusion restrictions used in many econometric models are literally true or they rather may be interpreted that impact of some variable is negligible? Isn't it rational to treat approximate knowledge as explicitly approximate? Is there any difference between identification of the sampling and Bayesian model? More fundamentally: Can Bayesian contribute to successful dealing with non–identified sampling model?

Answers to these questions may interest not only the people (like the author) who have wrestled with them for many years but also practitioners given the recent resurgence of Bayesian analysis in the field of quantitative macroeconomics, where identification issue occupies central position, see e.g. Canova and Sala (2009).

II. A Critical Report on Bayesian Identification Struggles

As claimed in the introduction, the problem of identification was intimately connected with Simultaneous Equations Models (SEM). It should not be surprising that early Bayesian notions of identification in the econometric literature were fitted into SEM. The seminal results are due to J. Drèze's unpublished (1962) work, which reappear in Drèze (1972,1974,1976). Fortunately the intrinsic properties of SEM turned out to be characteristic of many other models in a variety of fields of science. Hence this framework quickly became a cornerstone for discussion of Bayesian identification in every model within econometric literature (see e.g. Hsiao (1983), Poirier (1998)) and statistics literature (see e.g. Dawid (1979), Gelfand and Sahu (1999), Gustafson (2005)). In a nutshell these results are as follows. Assume that the parameter space may be decomposed as variation free i.e. $\Theta = \Theta_1 \times \Theta_2$, in such a way that the likelihood depends only on $\theta_2 \in \Theta_2$ i.e. $p(y \mid \theta_1, \theta_2) = p(y \mid \theta_2)$ for every $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$, where y denotes the data. If this is the case we will say that the

[1] However the recent aricle by Kraay (2012) is a clear sign that the intuition of Drèze is still alive and deeply rooted in minds of Bayesian users.

likelihood is θ_2 −oriented. Assume we have a joint prior $p(\theta_1,\theta_2)$. Then the simple consequence is that the conditional prior of θ_1 given θ_2 is not updated by the data i.e. $p(\theta_1 \mid \theta_2, y) = p(\theta_1 \mid \theta_2)$. However one caveat is that unless $p(\theta_1,\theta_2) = p(\theta_1)p(\theta_2)$, we generally have $p(\theta_1 \mid y) \neq p(\theta_1)$. Moreover the assumption that parameter space is variation free is crucial. Indeed, even if $p(\theta_1,\theta_2) = p(\theta_1)p(\theta_2)$ but the parameter space is not variation free we can have $p(\theta_1 \mid y) \neq p(\theta_1)$. For future reference when $p(\theta_1 \mid y) \neq p(\theta_1)$ we will say that there is the Bayesian learning.

Unfortunately the above reasoning hinges on 1) existence and 2) finding a variation free decomposition $\Theta = \Theta_1 \times \Theta_2$ such that the likelihood is θ_2 −oriented. In SEM the matter was simple: Θ_1 was the space of nonsingular matrices and Θ_2 the space of reduced form parameters. Usually the basic model parameterization does not allow to decompose the parameter space $\Theta = \Theta_1 \times \Theta_2$ so as the likelihood is θ_2 −oriented. Some efforts might be needed to find such a decomposition. The more fundamental question is under what conditions a model possesses the reparameterization which results in θ_2 −oriented likelihood. To realize that it may not be always attainable note that $\Theta = \Theta_1 \times \Theta_2$ means that for all $\theta \in \Theta$ there is a bijective decomposition $\theta \leftrightarrow (\theta_1,\theta_2)$ for some $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$. What matters is the universal quantifier "for all". There is no complete theory with this respect. Some partial (but quite general) results are given in Oulhaj and Mouchart (2003) and implicitly in Kocięcki (2011). As a matter of fact it is not unusual in practice that the most natural or sensible parameterization of a model precludes existence of the variation free decomposition $\Theta = \Theta_1 \times \Theta_2$. See Koop et al. (2011) for important examples from DSGE literature, and Gustafson (2005,2009) and references therein for other ones from many branches of applied statistics. You should be also aware that even if the variation free decomposition exists the parameter of interest may be a function of the parameters comprising both Θ_1 and Θ_2, see Poirier (1998) p. 488–489 for similar caution.

Among other treatments of the Bayesian identification we should primarily mention that of Kadane (1974). However his main result establishes the fact that if the likelihood is θ_2 −oriented the whole Bayesian analysis may be confined to the space Θ_2, which is obvious form the above informal considerations. His other concept i.e. informativeness of experiment, which essentially compares the behavior of the prior to the posterior are condemned to failure for the reasons outlined above and in the sequel of our paper. The same is true for similar concepts used by Leamer (1978) pp. 191–192, i.e. identification in probability, personal and public informativeness. In

fact the relationship between prior and the posterior for given data captures no information about identification of a model. You may be surprised but this relationship is data–driven. That is if you are unlucky and you get not the "right" sample then the posterior may be equal to the prior even if there is no identification problem. Conversely, the prior may be quite distinct from the prior even if the model is not identified. To realize this we suggest to consult highly instructive, but unfortunately little known, article by Drèze and Mouchart (1990). See also Sims (2007). This paragraph may be summarized by saying that Bayesian learning and identification are different concepts with the elusive or even haphazard mutual relationship. Attempts such as Müller (2012) aiming at capturing Bayesian learning, although very welcome, do not bring us closer to understanding the Bayesian identification. In our opinion measuring the Bayesian learning is meaningful and free of some pathologies only if the sampling model is identified. Then the Bayesian learning may capture "informativeness" of the particular sample. This opinion is by no means shared among all people within economic community, see e.g. extensive discussion with references in Koop et al. (2011).

Florens et al (1990) abstract treatment concentrates on the existence θ_2-oriented likelihood, where θ_2 is minimal sufficient in the sense of Barankin (1960), and its consequences for the identification.

The concept of observational equivalence of the marginal data densities discussed in Zellner (1971) pp. 254–256, although pedagogically useful, does not introduce any positive element into Bayesian conceptualization of identification problem. The conception proper is also based on interplay between prior and posterior distributions.

Some people (see e.g. Morales (1971), p. 20, Rothenberg (1973) p. 158) tended to think that there is a link between concentration of the posterior and identification of a model. Moreover for Drèze (1972) "*classical identification theory is really concerned with local uniqueness of posterior modes*". Though it is known that these views are in general false this will be more evident after studying the present paper[2]. Indeed even uniqueness of the maximum likelihood estimator (MLE) computed with very large number of observations does not imply that the model is identified for all

[2] However it is true that both Morales and Rothenberg were not fully satisfied with this conceptualization of the Bayesian identification. For example Rothenberg (1973), p. 159, concludes: "*we shall use the word "estimable" rather than "identifiable" when referring to a posterior density concentrated around its mean and leave unanswered the question of an appropriate Bayesian definition of identification*".

parameter points, but only that it is identified at the MLE. However keeping historical perspective this suggests that the concept of Bayesian identification was not altogether clear even after fundamental results of Drèze (1962) were widely known among Bayesians (including Drèze himself!). We note in passing that the reminiscences of Morales' and Rothenberg's crude view of the Bayesian identification reappears in Canova and Sala (2009). Similarly for Ríos–Rull et al. (2012) the curvature of the likelihood has always something to do with identification (see Koop et al. (2011) for some discussion on this point with further relevant references). We are very skeptical about the sense of such a practice. In this respect see Maddala (1976) for devastating counterexample.

Lastly there is one important drawback of the approach to Bayesian identification based on the variation free decomposition such that the likelihood is θ_2 −oriented (to be called just the decomposition), which seems to be overlooked in the econometric literature. If the decomposition is available many people think that in such a framework θ_1 captures all the non–identifiable aspects of a model. This is in general not true. In fact the whole reasoning based on the decomposition is silent about identification of the parameter θ_2. Identification of the latter is completely other issue. If the likelihood does not depend on θ_1 (given θ_2) then every sensible identification criterion (with or without Bayesian flavor) must suggest that θ_1 is not identifiable. But θ_2 may be non–identified by standard non–Bayesian arguments. Hence there is an implicit additional demand for Bayesian verdict concerning the identifiability of θ_2. This suggests that the purpose of the whole Bayesian approach based on the decomposition is somewhat pointless. Hence contrary to the common view the decomposition itself is not a very useful point of departure to discuss identification in the Bayesian framework.

Leaving the above safe framework (i.e. $\Theta = \Theta_1 \times \Theta_2$ and the likelihood is θ_2 −oriented) it is hard to find Bayesian words to discuss the identification problem. We think it is useful to have adequate words to do that. To this end we shall introduce mostly informal vocabulary to make the paper more readable.

III. Identification of the Sampling Model

Let $\mathcal{Y}$ denote the sample space, which is a set of all $y \in \mathcal{Y}$ attainable by at least one structure within a model. Each structure is indexed by the parameter $\theta \in \Theta$. For simplicity we assume that the data are continuous with the corresponding

probability density function (pdf) $p(y \mid \theta)$ with respect to some measure on $\mathcal{Y}$. For future reference $p(y \mid \theta)$ will be called interchangeably the data sampling density or the likelihood. A (parametric) structural model is a set $\mathcal{M}^S = \{p(y \mid \theta) \mid \theta \in \Theta, y \in \mathcal{Y}\}$. Since $\mathcal{M}^S$ consists of the data sampling density only, we shall call it the sampling model.

We use the standard definition of identification of the sampling model in terms of pdf's, see e.g. Haavelmo (1944), Rothenberg (1971). A sampling model is globally identified at $\theta \in \Theta$ if and only if (ifif) $\forall \overline{\theta} \in \Theta$ $[(\, p(y \mid \theta) = p(y \mid \overline{\theta})$ for all $y \in \mathcal{Y}\,)$ $\Rightarrow \theta = \overline{\theta}\,]$. In algebraic terms we can rewrite it: A sampling model is globally identified at $\theta \in \Theta$ ifif $\bigcap_{y \in \mathcal{Y}} \{\overline{\theta} \in \Theta \mid p(y \mid \theta) = p(y \mid \overline{\theta})\} = \{\theta\}$. We can state one more equivalent definition of the global identification

Lemma 1: *A sampling model is globally identified at* $\theta \in \Theta$ *ifif* $\forall \overline{\theta} \in \Theta$, $\exists y \in \mathcal{Y}$ $[\, p(y \mid \theta) = p(y \mid \overline{\theta}) \Rightarrow \theta = \overline{\theta}\,]$.

Proof: This follows by the rule of shifting and altering a quantifier, see e.g. Hamilton (1978) p. 84.

For economical reasons we shall drop "global" and when we refer to identification it should be understood that we mean the <u>global</u> identification.

A simple example will be useful to validate lemma 1 and algebraic definition of identification. Let $y \in \mathbb{R}$ have a univariate normal distribution with mean μ and variance 1 i.e. $p(y \mid \mu) \equiv N(\mu,1)$ (hence $\theta = \mu$). Then $p(y \mid \mu) = p(y \mid \overline{\mu})$ implies $\overline{\mu} = \mu$ or $\overline{\mu} = 2y - \mu$. Putting $y = \mu$ we get: $p(y \mid \mu) = p(y \mid \overline{\mu})$ implies $\overline{\mu} = \mu$. Hence for all $\overline{\mu}$ there exists y such that $p(y \mid \mu) = p(y \mid \overline{\mu}) \Rightarrow \overline{\mu} = \mu$. On the other hand our algebraic identification condition works well too since $\bigcap_{y \in \mathcal{Y}} \{\overline{\mu} \in \Theta \mid p(y \mid \mu) = p(y \mid \overline{\mu})\} = \{\mu\} \cap (\bigcap_{y \in \mathcal{Y} \setminus \{\mu\}} \{\mu, 2y - \mu\}) = \{\mu\}$, as required.

Finally let us define the equivalence class of the sampling model as

$$C_\theta^S = \bigcap_{y \in \mathcal{Y}} \{\overline{\theta} \in \Theta \mid p(y \mid \theta) = p(y \mid \overline{\theta})\} \tag{1}$$

IV. Identification of the Bayesian Model

In this section we pose and answer the following question: How can we interpret and rewrite the standard identification condition for the sampling model using Bayesian paradigm? To do that we have to take into account two obvious

things. In Bayesian approach we augment the data sampling density with the prior. For simplicity we assume that the prior is absolutely continuous with respect to Lebesgue measure and $p(\theta)$ is a prior pdf for parameters. Hence the joint density of the data and parameters is defined as $p(y,\theta) = p(y \mid \theta)p(\theta)$. Further, there is an implicit additional ingredient in the Bayesian approach, namely: the support of the prior. Let us denote it as $\Theta_{prior} = \{\theta \in \Theta \mid p(\theta) > 0\}$. This leads to definition of the Bayesian model i.e. $\mathcal{M}^B = \{p(y,\theta) \mid \theta \in \Theta_{prior}, y \in \mathcal{Y}\}$.

Since Bayesian analysis is conditioned on the data it is natural to look for identification condition that involves the likelihood (i.e. a sampling data density for the given data). To this end we have

Lemma 2: *If* $\exists y \in \mathcal{Y}$, $\forall \overline{\theta} \in \Theta$ $[\, p(y \mid \theta) = p(y \mid \overline{\theta}) \Rightarrow \overline{\theta} = \theta \,]$ *then the sampling model is globally identified at* $\theta \in \Theta$.

Proof: This follows by interchanging universal and existential quantifier in lemma 1.

Since the sufficient condition from lemma 2 involves the likelihood one may be tempted to think that it may constitute a good basis for Bayesian identification condition but it is not so. This sufficient condition requires that there exists at least one sample such that standard identification condition in terms of the likelihood holds. Finding such a sample may be difficult in practice (i.e. it may entail serious theoretical obstacles). Hence to securely define Bayesian notion of identification we should stick to the standard definition. Following identification definition for the sampling model as close as possible we define the equivalence class of the Bayesian model i.e. C^B_θ, as follows

$$
\begin{aligned}
C^B_\theta &= \bigcap\nolimits_{y\in\mathcal{Y}} \{\overline{\theta} \in \Theta \mid p(y \mid \theta) = p(y \mid \overline{\theta}) \text{ and } \theta, \overline{\theta} \in \Theta_{prior}\} \\
&= \bigcap\nolimits_{y\in\mathcal{Y}} \{\overline{\theta} \in \Theta \mid \tfrac{p(y,\theta)}{p(\theta)} = \tfrac{p(y,\overline{\theta})}{p(\overline{\theta})} \text{ and } \theta, \overline{\theta} \in \Theta_{prior}\} \\
&= \bigcap\nolimits_{y\in\mathcal{Y}} \{\overline{\theta} \in \Theta \mid \tfrac{p(y,\theta)}{p(y)p(\theta)} = \tfrac{p(y,\overline{\theta})}{p(y)p(\overline{\theta})} \text{ and } \theta, \overline{\theta} \in \Theta_{prior}\} \qquad (2) \\
&= \bigcap\nolimits_{y\in\mathcal{Y}} \{\overline{\theta} \in \Theta \mid \tfrac{p(\theta|y)}{p(\theta)} = \tfrac{p(\overline{\theta}|y)}{p(\overline{\theta})} \text{ and } \theta, \overline{\theta} \in \Theta_{prior}\} \\
&= (\bigcap\nolimits_{y\in\mathcal{Y}} \{\overline{\theta} \in \Theta \mid \tfrac{p(\theta|y)}{p(\theta)} = \tfrac{p(\overline{\theta}|y)}{p(\overline{\theta})}\}) \cap \Theta_{prior} \equiv (\bigcap\nolimits_{y\in\mathcal{Y}} \{\overline{\theta} \in \Theta \mid p(y \mid \theta) = p(y \mid \overline{\theta})\}) \cap \Theta_{prior}
\end{aligned}
$$

where $p(y) = \int_{\Theta_{prior}} p(y \mid \theta) p(\theta)\, d\theta$. Note that in the above $p(\square \mid y)$ means evaluation of the posterior at given $\square$ and y (so $p(\square \mid y)$ is a point–valued)[3]. Analogously, $p(\square)$ is the prior evaluated at given $\square$. It is useful to rewrite C_θ^B in classical terms noting that $\Theta_{prior} \subseteq \Theta$

Definition 1A: *The Bayesian model is (globally) identified at* $\theta \in \Theta_{prior}$ *ifif* $\forall \overline{\theta} \in \Theta_{prior}$ $[(\frac{p(\theta|y)}{p(\theta)} = \frac{p(\overline{\theta}|y)}{p(\overline{\theta})}$ *for all* $y \in \mathcal{Y}) \Rightarrow \theta = \overline{\theta}\,]$.

Alternatively, using arguments from the proof of lemma 1 one may state

Definition 1B: *The Bayesian model is (globally) identified at* $\theta \in \Theta_{prior}$ *ifif* $\forall \overline{\theta} \in \Theta_{prior}$, $\exists y \in \mathcal{Y}$ $[\frac{p(\theta|y)}{p(\theta)} = \frac{p(\overline{\theta}|y)}{p(\overline{\theta})} \Rightarrow \theta = \overline{\theta}\,]$.

From (2) we have the basic relationship

$$C_\theta^B = C_\theta^S \cap \Theta_{prior} \tag{3}$$

We immediately realize that if the underlying sampling model is not identified, the prior defined on the whole support Θ (in the sense $\Theta_{prior} = \Theta$) can not "identify" the Bayesian model because then we get $C_\theta^B = C_\theta^S$. This happens irrespective of whether the prior is highly concentrated on the subset of Θ on which the sampling model would be identified as the following example explains

Example 1 (Artificial but commonly stated to explain the identification problem)[4]:

$$y_t = \beta_1 + \beta_2 + \varepsilon_t \tag{4}$$

where y_t is a one–dimensional endogenous variable and $\varepsilon_t : (1 \times 1) \sim$ *i.i.d.* $N(0, \sigma^2)$. Let $\theta = (\beta_1, \beta_2, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+ = \Theta$, then $C_\theta^S = \{\beta_1 + g, \beta_2 - g, \sigma^2 \mid g \in \mathbb{R}\}$. Evidently the sampling model (4) is not identified (since $C_\theta^S \neq \{\beta_1, \beta_2, \sigma^2\}$). One possible way to identify (4) is to assume $\beta_2 = 0$. Then $\beta_2 - g = 0 - g = 0 \Rightarrow g = 0$, so as $C^S_{\beta_1, 0, \sigma^2} = \{\beta_1, 0, \sigma^2\}$. Can we replace exact restriction $\beta_2 = 0$ with its probabilistic

[3] We follow De Finetti (1974), p. 18, and $\square$ denotes the "placename" (i.e. "*something which awaits filling in*").

[4] In fact this example is not so far from reality. Similar form of non–identification appears in the following model (see e.g. Prakasa Rao (1992), p. 159). Suppose X_1 and X_2 are independently distributed with the exponential density $p(x) = \lambda_i \exp\{-\lambda_i x\}$ (for $i = 1,2$ and $x > 0$). Then $Y = \max\{X_1, X_2\}$ has density $p(y) = \lambda \exp\{-\lambda y\}$, where $\lambda = \lambda_1 + \lambda_2$. Clearly, $\lambda_1 + g$ and $\lambda_2 - g$ ($g \in \mathbb{R}$) result in the same distribution.

counterpart i.e. the prior $p(\beta_2)$ highly concentrated around $\beta_2 = 0$, with the hope that the Bayesian model would be then identified? Certainly no. Suppose $\Theta_{prior} = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+ = \Theta$ with the marginal prior $p(\beta_2)$ peaked around $\beta_2 = 0$. Then no matter how tightly $p(\beta_2)$ is concentrated around $\beta_2 = 0$, from (3) we get $C_\theta^B = C_\theta^S \cap \Theta_{prior} = C_\theta^S = \{\beta_1 + g, \beta_2 - g, \sigma^2 \mid g \in \mathbb{R}\}$. Hence the Bayesian model is not identified. What if we drastically shorten the support of the marginal prior $p(\beta_2)$? Assume that this support is $(-\frac{1}{1000}, \frac{1}{1000})$. This implies $\beta_2 - g \in (-\frac{1}{1000}, \frac{1}{1000})$, hence $C_\theta^B = \{\beta_1 + g, \beta_2 - g, \sigma^2 \mid \beta_2 - g \in (-\frac{1}{1000}, \frac{1}{1000})\}$. Let us check the identification of the Bayesian model at $(\beta_1, 0, \sigma^2)$: $C^B_{\beta_1,0,\sigma^2} = \{\beta_1 + g, 0 - g, \sigma^2 \mid 0 - g \in (-\frac{1}{1000}, \frac{1}{1000})\} =$ $= \{\beta_1 + g, g, \sigma^2 \mid g \in (-\frac{1}{1000}, \frac{1}{1000})\} \neq \{\beta_1, 0, \sigma^2\}$. Hence the Bayesian model is still non–identified. To see the problem from other perspective assume $\sigma^2 = 1$ and $\Theta_{prior} = \Theta$. Then each equivalence class $C^S_{\beta_1,\beta_2}$ is a downward sloping 45° straight line through (β_1, β_2) in $\mathbb{R}^2$. You can introduce the prior for all (β_1, β_2) in $\mathbb{R}^2$, but the data can only tell you on which line (β_1, β_2) lies. Everything else is a product of your imagination which is out of control by the design of the model. Thus part of your prior is intrinsically personal that can not be "objectivized" by the data. For example using the prior $p(\beta_1, \beta_2) = p(\beta_1)p(\beta_2)$, where $p(\beta_1) \coloneqq N(0, \tau^2)$ and $p(\beta_2) \coloneqq N(0, \omega^2)$, the Bayesian learning takes place so as $p(\beta_1) \neq p(\beta_1 \mid y)$ and $p(\beta_2) \neq p(\beta_2 \mid y)$ (see e.g. Poirier (1998)), but this is because both β_1 and β_2 are functions of the identified "parameter" $\psi = \beta_1 + \beta_2$ and the non–identified one i.e. $\beta_1 = \psi - \beta_2$, $\beta_2 = \psi - \beta_1$. Evidently the value of measuring Bayesian learning in such a case is limited since e.g. transforming $p(\beta_1)$ into $p(\beta_1 \mid y)$ encompasses both the impact of something that can be subject to scientific cognition i.e. ψ, and something metaphysical i.e. β_2. The lesson is that commonly shared intuition that "soft" probabilistic restrictions in the Bayesian model may be a substitute for the exact restrictions in the sampling model is plainly false.

On the other hand if the sampling model is identified at some $\theta \in \Theta$ then, provided that $\theta \in \Theta_{prior}$, the Bayesian model will be also identified at θ by (2):

Lemma 3: *Let $\theta \in \Theta_{prior}$. Then identification of the sampling model at θ implies identification of the Bayesian model at θ.*

The converse of lemma 3 does not necessarily hold. In example 1 even radical restriction of the prior support (but not degenerated one) would not make the Bayesian model identified. However when $\Theta_{prior} \neq \Theta$ than in some circumstances the Bayesian model may be identified even if the sampling model is not identified. The simplest example is when $\Theta = \mathbb{R}$, $C_\theta^S = \{-\theta, \theta\}$ with $\theta > 0$ (i.e. non–identification of the sampling model) and $\Theta_{prior} = (0, \infty)$ which results in $C_\theta^B = C_\theta^S \cap (0, \infty) = \{\theta\}$ (i.e. identification of the Bayesian model). Clearly what makes this simple example work is the support of the prior and not the shape of the prior. If $\Theta_{prior} = \Theta$ then identification of the Bayesian model implies identification of the sampling model.

Although (2) seems trivial, to the best of our knowledge, it did not appear in the literature. In fact (2) addresses successfully several cumbersome issues in Bayesian identification about which there arose much misunderstanding. To prepare the ground note that since $\Theta_{prior} \subseteq \Theta$ we can write $C_\theta^B = \bigcap_{y \in \mathcal{Y}} \{\overline{\theta} \in \Theta_{prior} \mid \frac{p(\theta|y)}{p(\theta)} = \frac{p(\overline{\theta}|y)}{p(\overline{\theta})}\}$.

First consider the case when the likelihood does not depend on θ. Then $p(\theta \mid y) = p(\theta)$ and we have $C_\theta^B = \bigcap_{y \in \mathcal{Y}} \{\overline{\theta} \in \Theta_{prior} \mid 1 = 1\} = \Theta_{prior}$. Hence the equivalence class for each $\theta \in \Theta_{prior}$ is the whole support of the prior and we get the most severe form of non–identification. This happens irrespective of whether we use flat or highly peaked prior $p(\theta)$. This is consistent with intuitive minimal requirements concerning "good" Bayesian definition of identification.

Second consider the flat prior in the sense $p(\theta) = p(\overline{\theta})$ for all $\theta, \overline{\theta} \in \Theta_{prior}$ (of course we assume that $p(y) = \int_{\Theta_{prior}} p(y \mid \theta)\, d\theta < \infty$ for all $y \in \mathcal{Y}$). In such a case $C_\theta^B = \bigcap_{y \in \mathcal{Y}} \{\overline{\theta} \in \Theta_{prior} \mid p(\theta \mid y) = p(\overline{\theta} \mid y)\}$. There are two points worth emphasizing. First of all it is not the degree of the posterior concentration that matters but whether and how posterior shape and/or location changes when we get different samples. To illustrate it consider the univariate case e.g. $\Theta = \mathbb{R}$. For global identification at θ it is necessary and sufficient that for one sample $y_1 \in \mathcal{Y}$ we have $p(\theta \mid y_1) = p(\overline{\theta} \mid y_1)$ at some single $\overline{\theta} \neq \theta$ (this will be the case e.g. when the posterior pdf is bell–shaped) and for other sample $y_2 \in \mathcal{Y}$ we have $p(\theta \mid y_2) = p(\overline{\overline{\theta}} \mid y_2)$ where $\overline{\theta} \neq \overline{\overline{\theta}}$. Then $C_\theta^B = \{\theta\}$ i.e. identification of the Bayesian model at θ. Now consider the general parameter space Θ. If there is some $y \in \mathcal{Y}$ such that the posterior $p(\theta \mid y)$ (under the flat prior) will be uniquely maximized at given $\theta = \overline{\theta}$ then the Bayesian model will be identified at $\overline{\theta}$. It means that uniqueness of the mode in the

posterior under the flat prior is sufficient for identification of the Bayesian model at the mode. In general, adopting any prior the analogous result seems to be hard to establish. That is whether uniqueness of the mode in the posterior under any prior is sufficient for identification of the Bayesian model at the mode is an open question (at least for the author). All we can say is

Lemma 4: *Define* $\theta^* = \arg\max_{\theta\in\Theta_{prior}} \frac{p(\theta|y)}{p(\theta)}$. *Uniqueness of* θ^* *implies that the Bayesian model is identified at* θ^*.

Proof: Let θ^* be obtained with given data $y^* \in \mathcal{Y}$. By definition the Bayesian model is identified at θ^* iff $C^B_{\theta^*} = \bigcap_{y\in\mathcal{Y}} \{\overline{\theta} \in \Theta_{prior} \mid \frac{p(\theta^*|y)}{p(\theta^*)} = \frac{p(\overline{\theta}|y)}{p(\overline{\theta})}\} = \{\theta^*\}$. But $C^B_{\theta^*} = \{\overline{\theta} \in \Theta_{prior} \mid \frac{p(\theta^*|y^*)}{p(\theta^*)} = \frac{p(\overline{\theta}|y^*)}{p(\overline{\theta})}\} \cap \left\{\bigcap_{y\in\mathcal{Y}\setminus\{y^*\}} \{\overline{\theta} \in \Theta_{prior} \mid \frac{p(\theta^*|y)}{p(\theta^*)} = \frac{p(\overline{\theta}|y)}{p(\overline{\theta})}\}\right\}$. If θ^* is unique then $\{\overline{\theta} \in \Theta_{prior} \mid \frac{p(\theta^*|y^*)}{p(\theta^*)} = \frac{p(\overline{\theta}|y^*)}{p(\overline{\theta})}\} = \{\theta^*\}$ hence $C^B_{\theta^*} = \{\theta^*\}$ (i.e. the Bayesian model is identified at θ^*).

Now consider the general case (any prior $p(\theta)$). In a common opinion when the prior is not much updated so as the posterior $p(\theta \mid y)$ looks like the prior $p(\theta)$ this points to some identification problems[5]. According to our definition of the Bayesian model identification the similarity (or non–similarity) $p(\theta \mid y)$ with $p(\theta)$ conveys no message. The important thing is whether the alternative (hypothetical) samples would revise the prior or not. That is we may have $\frac{p(\theta|y_1)}{p(\theta)} = 1$ for given $y_1 \in \mathcal{Y}$ and all $\theta \in \Theta$ but attain identification at some point $\overline{\theta} \in \Theta$ because alternative samples would change the shape and/or the location of the posterior so as $C^B_{\overline{\theta}} = \{\overline{\theta}\}$. This is a formal justification for insights in Drèze and Mouchart (1990).

Although the next lemma evidently resembles theorem 4 in Rothenberg (1971) its new feature is that it explicitly takes into account the fact that we apply Bayesian inference. It says that whatever prior you choose, the moments of the sampling density are globally identified (trivial as it may sound it was not formally demonstrated).

Lemma 5: *Let* $f : \mathcal{Y} \to \mathbb{R}$ *be any function of the data. Then* $\forall\theta,\overline{\theta} \in \Theta_{prior}$ $[(\frac{p(\theta|y)}{p(\theta)} = \frac{p(\overline{\theta}|y)}{p(\overline{\theta})}$ *for all* $y \in \mathcal{Y}) \Rightarrow E_\theta(f(y)) = E_{\overline{\theta}}(f(y))]$.

[5] This is in general false for well known reasons, see section II. However the discussion to follow is relevant even if we take these facts into account.

Proof: $\frac{p(\theta|y)}{p(\theta)} = \frac{p(\overline{\theta}|y)}{p(\overline{\theta})}$ for all $y \in \mathcal{Y}$ implies $\int_{\mathcal{Y}} f(y) \frac{p(\theta|y)}{p(\theta)} p(y)dy = \int_{\mathcal{Y}} f(y) \frac{p(\overline{\theta}|y)}{p(\overline{\theta})} p(y)dy$ $\Leftrightarrow$ $\int_{\mathcal{Y}} f(y)p(y \mid \theta)dy = \int_{\mathcal{Y}} f(y)p(y \mid \overline{\theta})dy$.

V. Further Considerations of the Bayesian Identification

Using proposition 1 in Kocięcki (2011) it can be shown that

$$C_{\theta}^{B} = \bigcap\nolimits_{y \in \mathcal{Y}} \{\overline{\theta} \in \Theta_{prior} \mid \tfrac{p(\theta|y)}{p(\theta)} = \tfrac{p(\overline{\theta}|y)}{p(\overline{\theta})}\} = \bigcap\nolimits_{y \in \mathcal{Y}} \{\overline{\theta} \in \Theta_{prior} \mid f(\tfrac{p(\theta|y)}{p(\theta)}) = f(\tfrac{p(\overline{\theta}|y)}{p(\overline{\theta})})\} \quad (5)$$

where f is any bijective function. In particular putting $f \equiv \ln$ we get

$$\begin{aligned} C_{\theta}^{B} &= \bigcap\nolimits_{y \in \mathcal{Y}} \{\overline{\theta} \in \Theta_{prior} \mid \ln(\tfrac{p(\theta|y)}{p(\theta)}) = \ln(\tfrac{p(\overline{\theta}|y)}{p(\overline{\theta})})\} \\ &= \{\overline{\theta} \in \Theta_{prior} \mid \ln(\tfrac{p(\theta|y)}{p(\theta)}) = \ln(\tfrac{p(\overline{\theta}|y)}{p(\overline{\theta})}) \text{ for all } y \in \mathcal{Y}\} \\ &\subseteq \{\overline{\theta} \in \Theta_{prior} \mid \int_{\mathcal{Y}} \ln(\tfrac{p(\theta|y)}{p(\theta)}) p(y)\, dy = \int_{\mathcal{Y}} \ln(\tfrac{p(\overline{\theta}|y)}{p(\overline{\theta})}) p(y)\, dy\} \end{aligned} \quad (6)$$

where $p(y) = \int_{\Theta_{prior}} p(y \mid \theta)p(\theta)\, d\theta$. On the other hand we also have:

$$\begin{aligned} C_{\theta}^{B} &= \bigcap\nolimits_{y \in \mathcal{Y}} \{\overline{\theta} \in \Theta_{prior} \mid \ln p(y \mid \theta) = \ln p(y \mid \overline{\theta})\} \\ &\subseteq \{\overline{\theta} \in \Theta_{prior} \mid \int_{\mathcal{Y}} \ln(p(y \mid \theta)) p(y \mid \theta)\, dy = \int_{\mathcal{Y}} \ln(p(y \mid \overline{\theta})) p(y \mid \theta)\, dy\} = \\ &= \{\overline{\theta} \in \Theta_{prior} \mid \int_{\mathcal{Y}} \ln(\tfrac{p(y|\theta)}{p(y|\overline{\theta})}) p(y \mid \theta)\, dy = 0\} \end{aligned} \quad (7)$$

Hence (6) and (7) allow for the following statements:

1) A Bayesian model is <u>not</u> identified at θ ifif $\exists \overline{\theta} \in \Theta_{prior}$ $[\theta \neq \overline{\theta}$ and $\ln(\frac{p(\theta|y)}{p(\theta)}) = \ln(\frac{p(\overline{\theta}|y)}{p(\overline{\theta})})$ for all $y \in \mathcal{Y}]$.

2) If $\forall \overline{\theta} \in \Theta_{prior}$ $[\int_{\mathcal{Y}} \ln(\frac{p(\theta|y)}{p(\theta)}) p(y)\, dy = \int_{\mathcal{Y}} \ln(\frac{p(\overline{\theta}|y)}{p(\overline{\theta})}) p(y)\, dy \Rightarrow \theta = \overline{\theta}\,]$ then the Bayesian model is identified at θ .

3) If $\forall \overline{\theta} \in \Theta_{prior}$ $[\int_{\mathcal{Y}} \ln(\frac{p(\theta|y)}{p(\overline{\theta}|y)}) p(y \mid \theta)\, dy = \ln(\frac{p(\theta)}{p(\overline{\theta})}) \Rightarrow \theta = \overline{\theta}\,]$ then the Bayesian model is identified at θ .

4) If $\forall\overline{\theta} \in \Theta_{prior}$ $[\int_y \ln(\frac{p(y|\theta)}{p(y|\overline{\theta})})p(y \mid \theta)\,dy = 0 \Rightarrow \theta = \overline{\theta}\,]$ then the Bayesian model is identified at θ.

We shall interpret these statements. To this end we use concepts developed by Good (1960,1966) and from standard information theory.

Good (1960), introducing axioms based on K. Popper's desiderata, proved that the only possible notion of "explanatory power" of y with respect to θ is a monotonic function of $\frac{p(\theta|y)}{p(\theta)}$ (where both θ and y refer to single values). The latter was also called by Good the amount of information concerning θ provided by y. Indeed, Good (1966) argued i.e. proved introducing some axioms, that the amount of information must be a monotonic function of $\frac{p(\theta|y)}{p(\theta)}$. Thus according to Good, $\ln(\frac{p(\theta|y)}{p(\theta)})$ is valid both for measuring "explanatory power" and amount of information. In light of these facts condition 1) is interpreted: the Bayesian model is <u>not</u> identified at θ ifif whatever sample y you may get there exists at least one $\overline{\theta} \neq \theta$ such that this sample will have the same explanatory power with respect to $\overline{\theta}$ and θ. Hence in all circumstances the data can not tell us whether θ or $\overline{\theta}$ is better explained by the observables. On the other hand, we may also say that amount of information for θ and $\overline{\theta}$ provided by (any) y is the same. Hence the information from the data provide no evidence in order to differentiate between two hypotheses θ and $\overline{\theta}$. Of course $\ln(\frac{p(\theta|y)}{p(\theta)})$ averaged with respect to $p(y)$, may be called 1) the prior amount of information concerning θ provided by potential sample from the given model or 2) potential explanatory power of the data with respect to θ.

To get further insight into the condition 2) we make use of some concepts from information theory. In particular note that $\ln(\frac{p(\theta|y)}{p(\theta)})$ for the given y is a gain in information about θ provided by y. Putting it other way, since $\ln(\frac{p(\theta|y)}{p(\theta)}) = -\ln p(\theta) - (-\ln p(\theta \mid y))$, this measures a reduction in uncertainty about θ that results from learning about the data y. Moreover since $\ln(\frac{p(\theta|y)}{p(\theta)}) = \ln(\frac{p(y|\theta)}{p(y)})$ the latter quantifies the predictability of y given that we know θ or simply usefulness of knowing θ for predicting y. Given this, condition 2) may be easily reinterpreted (which is omitted).

Condition 4) is just corollary in Bowden (1973) where detailed discussion of its relationship with local and global identification may be found. Note however that sufficient condition 4) relates to the Bayesian model (not the sampling model as in Bowden (1973)).

VI. Uniform and Marginal Identification

In this section we define a concept of the uniform identification in the sampling model. Further, exploiting the merits of the Bayesian approach we study (with the help of many examples) the marginal identifiability of the Bayesian model. In particular we demonstrate usefulness of the Bayesian approach in exploring the identification problem. We begin with the motivating considerations on the basis of

Example 1 (*cont.*): Evidently $\psi = \beta_1 + \beta_2$ is (globally) identified. Moreover imposing one restriction $\beta_1 = \beta_1^*$ or $\beta_2 = \beta_2^*$ we can uniquely retrieve β_2 or β_1, respectively. The problem arises when both β_1 and β_2 have economic interpretation. Though excluding restrictions are used frequently they usually may be treated not as genuine statements but only as useful approximations. For example we may put $\beta_2 = 0$ as before, but what this truly expresses is that β_2 is almost negligible. In such a case we may put, say any $\beta_2 \in (-\frac{1}{1000}, \frac{1}{1000})$, without disturbing the economic content of a model. So far it sounds like the classic motivation for imposing probabilistic restrictions instead of exact ones i.e. to introduce a prior distribution and apply Bayesian inference. As we explained this strategy falls short. The slight (but important) difference is that fixing $\beta_2 = k$, where $k \in (-\frac{1}{1000}, \frac{1}{1000})$, is not the same as adopting the prior distribution for β_2 with the support on $(-\frac{1}{1000}, \frac{1}{1000})$ and running the usual Bayesian procedure. In the latter case the identification of β_1 is lost, whereas in the former case β_1 is identified both in the sampling and the Bayesian model. Indeed in such a case $C^S_{\beta_1,k,\sigma^2} = \{\beta_1 + g, k - g, \sigma^2 \mid k - g = k\} = \{\beta_1, k, \sigma^2\}$ and by lemma 3 this equals to $C^B_{\beta_1,k,\sigma^2}$. Note that $C^S_{\beta_1,k,\sigma^2} = \{\beta_1, k, \sigma^2\}$ for all $k \in \mathbb{R}$ i.e. uniformly.

The above example leads to the formulation of the following concept. Assume that the parameter space may be decomposed $\Theta = \Theta_1 \times \Theta_2$, where Θ_2 comprise parameters that are subject to restrictions (in order to identify the sampling model), but the likelihood need not be $\theta_2 -$oriented or $\theta_1 -$oriented. For the ease of exposition we assume that we just fix elements of θ_2 at some value k i.e. $\theta_2 = k \in \Theta_2$.

Definition 2: *A sampling model is uniformly identified at* $\theta_1 \in \Theta_1$ *ifif* $\forall k \in \Theta_2$, $\forall \overline{\theta}_1 \in \Theta_1$ $[(\, p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)$ *for all* $y \in \mathcal{Y}\,) \Rightarrow \theta_1 = \overline{\theta}_1\,]$.

The uniform identifiability is the strongest possible condition for identification of θ_1 when we restrict the remaining parameters θ_2 i.e. θ_1 will be identified no matter what $k \in \Theta_2$ we choose. Needless to say, it is the most preferable situation. However the uniform identifiability may not be present. In general we can think of the cases when 1) identifiability holds for almost all $k \in \Theta_2$ and 2) identifiability holds for one particular $k \in \Theta_2$. Of course the case 1) is much more common but we may also face the case 2) as the following example of great practical importance illustrates

Example 2 (*identifiability holds for one particular* $k \in \Theta_2$):

We consider the bivariate Structural VAR (SVAR) model

$$A_0 y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t; \qquad t = 1,\ldots,T.$$

where $y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} \in \mathbb{R}^{2\times 1}$, $A_0 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{R}^{2\times 2}$ nonsingular, $A_i : (2\times 2)$, $c \in \mathbb{R}^{2\times 1}$ and $\varepsilon_t \mid y_{t-1},\ldots \sim N(0_{2\times 1}, \mathrm{I}_2)$. As is well known, the model as it stands is not identified since $C^S_{A_0,c,A_1,\ldots,A_p} = \{gA_0, gc, gA_1,\ldots,gA_p \mid g \in O_2\} \neq \{A_0, c, A_1,\ldots,A_p\}$, where O_2 denotes the space of (2×2) orthogonal matrices. Equally known is that putting one off–diagonal element in A_0 to zero, say $a_{12} = 0$, and imposing normalization $a_{11} > 0$, $a_{22} > 0$ will identify SVAR model. The restriction $a_{12} = 0$ means that $y_{1,t}$ does not respond instantaneously to changes in $y_{2,t}$. This may not be literally true. Instead $a_{12} = 0$ would be rather a synonym for "instantaneous response of $y_{1,t}$ to changes in $y_{2,t}$ is probably negligible". In such a case we may be also interested in examining robustness of some other conclusions based on SVAR (i.e. Impulse Response Functions) to the restriction $a_{12} = 0$. Specifically we may e.g. impose restriction $a_{12} = k$, where $k \in (-1,1)$ or even $k \in \mathbb{R}$, and trace its effects on many inferential objects from SVAR (i.e. variance decompositions, Impulse Response Functions, etc.). Actually exactly such a methodology was applied to testing long–run money–output neutrality in the bivariate SVAR by King and Watson (1997). The problem with this methodology is that when we replace excluding restriction $a_{12} = 0$ with $a_{12} = k \neq 0$ the SVAR is no longer identified under the normalization $a_{11} > 0$, $a_{22} > 0$. To show this note that in the case the restrictions are confined to A_0 only we get $C^S_{A_0,c,A_1,\ldots,A_p} \equiv C^S_{A_0} = \{gA_0 \mid g \in O_2\}$. Let $g = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \in O_2$. Then the SVAR will be identified at arbitrary A_0 ifif gA_0 implies $g = \mathrm{I}_2$. Since the only restriction is $a_{12} = k \neq 0$ this is equivalent to $g_{11}k + g_{12}a_{22} = k$ implies $g_{11} = 1$ and $g_{12} = 0$. However the equation has two pairs of solutions: ($g_{11} = 1$ and $g_{12} = 0$) or ($g_{11} = \frac{k^2 - a_{22}^2}{k^2 + a_{22}^2}$

and $g_{12} = \frac{2ka_{22}}{k^2+a_{22}^2}$). Let $g_1^{\perp} : (1 \times 2)$ be an orthonormal complement of $g_1 = (\frac{k^2-a_{22}^2}{k^2+a_{22}^2}, \frac{2ka_{22}}{k^2+a_{22}^2})$ i.e. $g_1^{\perp} g_1' = 0$ and $g_1^{\perp} g_1^{\perp\prime} = 1$. Define $g^* = \begin{bmatrix} g_1 \\ g_1^{\perp} \end{bmatrix}$. Then provided that $a_{12} = k \neq 0$, $C^S_{A_0,c,A_1,...,A_p}$ will not be a singleton but comprise two elements: $A_0, c, A_1, ..., A_p$ and $g^*A_0, g^*c, g^*A_1, ..., g^*A_p$. Of course when $a_{12} = k = 0$ then $g_1 = (\frac{k^2-a_{22}^2}{k^2+a_{22}^2}, \frac{2ka_{22}}{k^2+a_{22}^2}) = (-1, 0)$ and the assumed normalization $a_{11} > 0$, $a_{22} > 0$ will exclude this solution. A "normalization" in the case $a_{12} = k \neq 0$ would be much more tricky. Needless to say the nature of the problem will reappear also in $n-$dimensional SVAR (for $n > 2$).

Although we can list many examples when identifiability holds for almost all $k \in \Theta_2$ (think of coefficients entering model in a multiplicative way like in our forthcoming example 4) we shall mention the one from Koop et al. (2011) (which is based on Galí at al (2005))

Example 3 (*identifiability holds for almost all* $k \in \Theta_2$):

Consider a Hybrid New Keynesian Phillips Curve (HNKPC):

$\pi_t = \beta_b \pi_{t-1} + \beta_f E_{t-1}\pi_{t+1} + \gamma x_t + \varepsilon_t$; with $x_t = \rho x_{t-1} + \upsilon_t$

where π_t denotes inflation, x_t the output gap, $\varepsilon_t \sim i.i.d.\ N(0, \sigma_\varepsilon^2)$, $\upsilon_t \sim i.i.d.\ N(0, \sigma_\upsilon^2)$ (ε_t and υ_t uncorrelated) and β_b, β_f, γ are "intermediate" parameters (i.e. which are functions of deep parameters from DSGE model, see Galí at al (2005)). For simplicity we consider only the identification of the "intermediate" parameters (though what is really interesting is the identifiability of the deep parameters). Under certain conditions (see Koop at al. (2011)) the unique solution of HNKPC is

$\pi_t = \alpha_1 \pi_{t-1} + \alpha_2 x_{t-1} + u_t$ together with $x_t = \rho x_{t-1} + \upsilon_t$

where: $\alpha_1 = \frac{1}{2\beta_f}(1 - \sqrt{1 - 4\beta_f\beta_b})$, $\alpha_2 = \gamma\rho\,/(1 - \beta_f(\alpha_1 + \rho))$ and $u_t = (\varepsilon_t + \gamma\upsilon_t)/(1 - \alpha_1\beta_f)$. Let us denote $\sigma_u^2 = \mathrm{var}(u_t)$. The question is: Having $\alpha_1, \alpha_2, \rho, \sigma_u^2, \sigma_\upsilon^2$ can we uniquely retrieve $\beta_b, \beta_f, \gamma, \rho, \sigma_\varepsilon^2, \sigma_\upsilon^2$? A simple counting exercise suggests we can not. To identify the model first assume we impose the restriction $\beta_b = k \neq 0$. Noting that ρ and σ_υ^2 can be unambiguously taken from $x_t = \rho x_{t-1} + \upsilon_t$ (i.e. ρ and σ_υ^2 are identified) and assuming $\alpha_1 = \frac{1}{2\beta_f}(1 - \sqrt{1 - 4\beta_f k})$ has a unique solution for β_f (possibly in the restricted support for β_f), we can uniquely get all the remaining $\beta_f, \gamma, \sigma_\varepsilon^2$. Hence we identify the model. Now suppose $\beta_b = 0$. Then $\alpha_1 = 0$ which implies $\alpha_2 = \gamma\rho\,/(1 - \beta_f\rho)$ and $u_t = \varepsilon_t + \gamma\upsilon_t$. Since ρ and σ_υ^2 are free of identification problem, ultimately we have two equations $\alpha_2 = \gamma\rho\,/(1 - \beta_f\rho)$,

$\sigma_u^2 = \sigma_\varepsilon^2 + \gamma^2\sigma_v^2$ in three unknowns $\gamma, \beta_f, \sigma_\varepsilon^2$. As a result the model becomes non–identified. Similar analysis may be conducted by restricting γ. In the latter case if $\gamma = k \neq 0$ then the model is identified whereas setting $\gamma = 0$ destroys the identification of the remaining parameters.

Evidently examples 2 and 3 suggest that there is a room for the concepts weaker than the uniform identifiability. Can the Bayesian contribute to designing those weaker notions? We think so. Natural basis for this is the following reasoning. In practice we do not consider all restrictions $k \in \Theta_2$ as equally reasonable. It amounts to attaching some weights to possible restrictions. Staying within Bayesian framework suppose that we have a marginal (proper) prior measure π on Θ_2. Since we have a prior we must define the relevant identification concept in terms of the Bayesian model. How can we incorporate our subjective beliefs about the reasonable restrictions for θ_2? As we showed we can not replace the exact restriction $\theta_2 = k$ with the prior $p(\theta_2)$ (possibly highly concentrated around $\theta_2 = k$) and conduct the Bayesian analysis. In such a case neither θ_1 nor θ_2 will be identified in the Bayesian model. We must look for the other possibility. To this end let us define the integrated likelihood as $p(y \mid \theta_1) = \int_{\Theta_2} p(y \mid \theta_1, \theta_2)\pi(d\theta_2)$. The idea is that integrating θ_2 out from the likelihood, the ultimate inference based on an "integrated likelihood" model $p(y \mid \theta_1)$ will take into account our subjective uncertainty about the reasonable restrictions for θ_2.

Definition 3: *Let* π *be the marginal* (*proper*) *prior measure on* Θ_2. *Define* $p(y \mid \theta_1) = \int_{\Theta_2} p(y \mid \theta_1, \theta_2)\pi(d\theta_2)$. *The Bayesian model is marginally identified at* $\theta_1 \in \Theta_1$ *ifif* $\forall \overline{\theta}_1 \in \Theta_1$ $[(p(y \mid \theta_1) = p(y \mid \overline{\theta}_1)$ *for all* $y \in \mathcal{Y}) \Rightarrow \theta_1 = \overline{\theta}_1]$.

Example 1 (*cont.*): Suppose the parameter β_2 has economic meaning. Instead of putting dogmatic restriction $\beta_2 = 0$, let us assume $\pi(\beta_2) := N(0, \delta^2)$, where δ^2 is fixed (β_2 plays the role of θ_2 from definition 3). Having one observation $y \in \mathbb{R}$ we get $p(y \mid \beta_1, \sigma^2) = \int p(y \mid \beta_1, \beta_2, \sigma^2)\pi(d\beta_2) := N(\beta_1, \sigma^2 + \delta^2)$. As δ^2 tends to 0, the modified sampling model $p(y \mid \beta_1, \sigma^2)$ tends to the original sampling model with the exact restriction $\beta_2 = 0$. Since δ^2 is fixed (i.e. the prior hyperparameter), it is easy to observe that $p(y \mid \beta_1, \sigma^2) = p(y \mid \overline{\beta}_1, \overline{\sigma}^2)$ for all $y \in \mathbb{R}$ implies $(\beta_1, \sigma^2) = (\overline{\beta}_1, \overline{\sigma}^2)$. Hence the modified model is marginally identified. In this example uncertainty concerning

the restriction manifests itself in an intuitive way i.e. increasing the variance of the disturbance from σ^2 in the original sampling model to $\sigma^2 + \delta^2$.

It arises the natural question. What is the relationship between marginal identification of the Bayesian model and the standard identification of the sampling model $p(y \mid \theta_1, \theta_2 = k)$? For reliability of the concept of marginal identification we must require that if the sampling model is identified with the help of restriction $\theta_2 = k$ then the same must be true for this model but marginalized with respect to θ_2 using the prior concentrated around $\theta_2 = k$. To achieve this goal we employ the following reasoning. Being consistent with intuitive desideratum of early Bayesian econometricians we shall assume that if the implicit variance components in π (i.e. a prior measure on Θ_2) approach zero then the Bayesian "marginal likelihood" model becomes the sampling model with exact restrictions (which is identified). To this end let us formulate the internal coherence condition (ICC):

Assumption (ICC): If $\pi_n(\theta_2) \Rightarrow \delta(\theta_2 = k)$ then $\int_{\Theta_2} F(y \mid \theta_1, \theta_2)\pi_n(d\theta_2) \Rightarrow F(y \mid \theta_1, \theta_2 = k)$

where "$\Rightarrow$" denotes weak convergence of distribution functions, $\delta(\theta_2 = k)$ is a (degenerated) measure having a single saltus at $\theta_2 = k$ and $F(y \mid \theta_1, \theta_2)$ is the distribution function corresponding to $p(y \mid \theta_1, \theta_2)$. We note that ICC is nothing more than the mathematical statement of the opinion expressed by Fisher (1966), p. 184. Sufficient condition for ICC is that for all $\theta_1 \in \Theta_1$ and all $y \in \mathcal{Y}$; $F(y \mid \theta_1, \theta_2)$ must be continuous at $\theta_1, \theta_2 = k$. Needless to say the ICC will be met in standard models.

One caveat is that continuity and identifiability are not the same

Example 4: Let a single observation be $y \in \mathbb{R}$ such that $y \sim N(\alpha\beta, \sigma^2)$. Clearly if we restrict $\beta = k \neq 0$, the sampling model is identified at any α, σ^2. Suppose we take the prior $\pi(\beta) \coloneqq N(0,1)$. Now it may be shown that $p(y \mid \alpha, \sigma^2) = \int_{\mathbb{R}} p(y \mid \alpha, \sigma^2, \beta)\pi(d\beta) \equiv N(0, \sigma^2 + \alpha^2)$. Evidently $p(y \mid \alpha, \sigma^2)$ can not identify α, σ^2. Hence the identification is lost in the integrated likelihood.

It should not be surprising that the origin of the possible complications is the fact that we center the prior on the parameter value at which a sampling model is non–identified. For we have

Example 4 (cont.): Instead of $\pi(\beta) \coloneqq N(0,1)$ let us assume that $\pi(\beta) \coloneqq N(k,\delta^2)$ with $k \neq 0$. Hence we center the prior on value at which there is no identification problem in the sampling model. In such a case $p(y \mid \alpha,\sigma^2) = \int_{\mathbb{R}} p(y \mid \alpha,\sigma^2,\beta)\pi(d\beta)$ $\equiv N(\alpha k, \sigma^2 + \alpha^2\delta^2)$. This time everything goes right. As δ^2 tends to 0, $\pi(\beta)$ weakly converges to $\delta(\theta_2 = k)$ and the integrated likelihood $p(y \mid \alpha,\sigma^2)$ converges to $N(\alpha k,\sigma^2)$, which is the sampling model identified by imposing restriction $\beta = k \neq 0$. Hence the ICC is met. Importantly since k,δ^2 are fixed (i.e. hyperparameters of the prior), the Bayesian model is marginally identified i.e. $p(y \mid \alpha,\sigma^2) = p(y \mid \overline{\alpha},\overline{\sigma}^2)$ for all $y \in \mathbb{R}$ implies $(\alpha,\sigma^2) = (\overline{\alpha},\overline{\sigma}^2)$. Note that prior uncertainty about the "proper" value for the restriction $\beta = k$ nicely mixes with the sampling model by enlarging sampling variability of the latter from σ^2 to $\sigma^2 + \alpha^2\delta^2$.

In fact in example 4 it is sufficient to assume $\pi(\beta) \coloneqq N(k,1)$, with $k \neq 0$, to avoid problems exemplified with adoption of $\pi(\beta) \coloneqq N(0,1)$. In particular the expected value should be a point where the identification holds. The following powerful example strengthens this argument since it shows that even if a sampling model is identified for almost all parameter values we should be careful not to center the prior on the parameter value at which a sampling model is non–identified.

Example 5: Let us generalize example 4 and assume $y \in \mathbb{R}^2$ and $y \sim N(\mu,\Sigma)$, where $\mu = (\beta,\alpha\beta)'$. Clearly if $\beta \neq 0$ the sampling model is identified (without any restrictions). Let us choose the prior centered on $\beta = 0$, the parameter point at which the model ceased to be identified. Assume e.g. $\pi(\beta) \coloneqq N(0,1)$. If we marginalize the likelihood with respect to β using the prior measure $\pi(\beta) \coloneqq N(0,1)$ we get $p(y \mid \alpha,\Sigma) = \int_{\mathbb{R}} p(y \mid \alpha,\Sigma,\beta)\pi(d\beta) \equiv N(0,\Sigma + [1\,\alpha]'[1\,\alpha])$. Hence the identification of α,Σ is lost in the integrated likelihood. This example is instructive since it shows that even if there is no identification problem in the sampling model it may show up in the integrated likelihood. The lesson is that we can not judge the identification problem on the basis of the marginal posterior behavior. Particularly when we do not pay sufficient attention to careful specification of the prior. For example, suppose the parameter space is 3–dimensional. Then the visible ridges or flat regions in the 2–dimensional marginal posterior need not point to identification problems. Needless to say if we employ the measure $\pi(\beta) \coloneqq N(k,\delta^2)$ (or even $N(k,1)$)

with $k \neq 0$, then the above problems disappear since (using $\pi(\beta) \coloneqq N(k, \delta^2)$) we obtain $p(y \mid \alpha, \Sigma) \equiv N(k[1\,\alpha]', \Sigma + \delta^2[1\,\alpha]'[1\,\alpha])$.

VII. Faithful Identification

In this section we study one possible extension to the basic identification condition from the Bayesian point of view. As we showed, uniform identifiability is not a universal property of many economic models. In particular identifiability almost everywhere seems to be characteristic of many economic models. In what follows we define a concept of faithful identification of the Bayesian model and uncover its relationship with the marginal identification introduced in the previous section.

Definition 4: *Let π be the marginal (proper) prior measure on Θ_2. The Bayesian model is faithfully identified at $\theta_1 \in \Theta_1$ ifif $\forall \overline{\theta}_1 \in \Theta_1$ [($p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)$ for all $y \in \mathcal{Y}$) a.e. (π) implies $\theta_1 = \overline{\theta}_1$].*

Perhaps the main motivation for definition 4 relies on the following reasoning. According to definition 4, a Bayesian model is <u>not</u> faithfully identified at $\theta_1 \in \Theta_1$ ifif there exists at least one $\overline{\theta}_1 \neq \theta_1$ such that $p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)$ for all possible $y \in \mathcal{Y}$ and almost all (π) k's. Hence whatever sample you may have and for almost all restrictions $k \in \Theta_2$ there will be unsolvable ambiguity concerning whether $(\overline{\theta}_1, k)$ or (θ_1, k) could generate the sample. On the other hand the relevance of faithful identification question may be appreciated if we think of DSGE modeling when some parameters are fixed by calibration and the remaining ones are estimated. Then we can ask if other calibration would preserve the identification of the model. If a model is faithfully identified then almost all calibrations are acceptable in terms of its impact on identification of estimated parameters.

To better understand the faithful identifiability let us collect its equivalent definitions

Lemma 6: *The following expressions are equivalent*:

a) ($p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)$ for all $y \in \mathcal{Y}$) a.e. (π) implies $\theta_1 = \overline{\theta}_1$

b) $\pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k) \text{ for all } y \in \mathcal{Y}\}) = 1 \Rightarrow \theta_1 = \overline{\theta}_1$

c) ($p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)$ a.e. (π)) for all $y \in \mathcal{Y}$ implies $\theta_1 = \overline{\theta}_1$

d) $\pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)\}) = 1$ for all $y \in \mathcal{Y} \Rightarrow \theta_1 = \overline{\theta}_1$

e) $\pi(\{k \in \Theta_2 \mid \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)\}) = 0 \Rightarrow \theta_1 = \overline{\theta}_1$

f) $\theta_1 \neq \overline{\theta}_1 \Rightarrow \pi(\{k \in \Theta_2 \mid \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)\}) > 0$

g) $\theta_1 \neq \overline{\theta}_1 \Rightarrow \pi(\{k \in \Theta_2 \mid \forall y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)\}) < 1$

Proof:

b) This is just equivalent mathematical form of a)

c) We have

$1 = \pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k) \text{ for all } y \in \mathcal{Y}\}) =$

$= \pi(\bigcap_{y \in \mathcal{Y}} \{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)\})$

$\leq \pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)\})$

thus $\pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)\}) = 1$ for arbitrary $y \in \mathcal{Y}$, hence for every $y \in \mathcal{Y}$. But $\pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)\}) = 1$ for all $y \in \mathcal{Y}$ is equivalent to $(p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)$ a.e. $(\pi))$ for all $y \in \mathcal{Y}$.

d) This is just equivalent mathematical form of c)

e) $\pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k) \text{ for all } y \in \mathcal{Y}\}) = 1 \Leftrightarrow$

$0 = 1 - \pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k) \text{ for all } y \in \mathcal{Y}\}) =$

$= \pi(\{k \in \Theta_2 \mid \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)\})$

f) Take the contrapositive of e)

g) $\pi(\{k \in \Theta_2 \mid \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)\}) > 0 \Leftrightarrow$

$1 > 1 - \pi(\{k \in \Theta_2 \mid \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)\}) =$

$= \pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k) \text{ for all } y \in \mathcal{Y}\})$

Definitions based on f) and g) are particularly useful for interpretation. Standard identification says that the sampling model is identified at $\theta_1 \in \Theta_1$ (given the restriction $\theta_2 = k$) iff $\forall \overline{\theta}_1 \in \Theta_1$ $[\theta_1 \neq \overline{\theta}_1 \Rightarrow \exists y \in \mathcal{Y}, \; p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)]$. Exploiting the fact that we apply the Bayesian inference we can attach the weighting function expressing our belief concerning the most probable restrictions for θ_2 (i.e. the marginal prior measure π). Then we can ask if the standard identification holds for non–negligible set (i.e. of strictly positive π measure) in Θ_2. If the answer is in the affirmative then the Bayesian model is faithfully identified. On the other hand standard identification condition states that the sampling model is <u>not</u> identified at $\theta_1 \in \Theta_1$ (given the restriction $\theta_2 = k$) iff $\exists \overline{\theta}_1 \in \Theta_1$ $[\theta_1 \neq \overline{\theta}_1$ and $p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)$ for all $y \in \mathcal{Y}]$. But (on the basis of g)) the model is <u>not</u> faithfully identified at $\theta_1 \in \Theta_1$ iff $\exists \overline{\theta}_1 \in \Theta_1$ $[\theta_1 \neq \overline{\theta}_1$ and $\pi(\{k \in \Theta_2 \mid p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k) \text{ for all } y \in \mathcal{Y}\}) = 1]$. The intuition of the

latter in the context of standard non–identification condition (given the restriction $\theta_2 = k$) is self–explaining.

The next proposition gives the sufficient condition for faithful identifiability and is useful since it shows that all we have to check is the usual identifiability in terms of the integrated likelihood.

Proposition 1: *If the Bayesian model is marginally identified at* θ_1 *then it is faithfully identified at* θ_1.

Proof: Using formulation from lemma 6 c)

$\{\overline{\theta}_1 \in \Theta_1 \mid (p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k) \text{ a.e. } (\pi)) \text{ for all } y \in \mathcal{Y}\}$

$\subseteq \{\overline{\theta}_1 \in \Theta_1 \mid \int_{\Theta_2} p(y \mid \theta_1, \theta_2)\pi(d\theta_2) = \int_{\Theta_2} p(y \mid \overline{\theta}_1, \theta_2)\pi(d\theta_2) \text{ for all } y \in \mathcal{Y}\}$

$= \{\overline{\theta}_1 \in \Theta_1 \mid p(y \mid \theta_1) = p(y \mid \overline{\theta}_1) \text{ for all } y \in \mathcal{Y}\} = \{\theta_1\}$

where the last equality sign is by the hypothesis of proposition.

The converse of proposition 1 need not hold for we have the following

Example 4 (cont.): Recall that $y \in \mathbb{R}$ and $y \sim N(\alpha\beta, \sigma^2)$. Suppose we take the prior $\pi(\beta) \coloneqq N(0,1)$. One way to demonstrate that the Bayesian model is faithfully identified at (α, σ^2) is to use definition from lemma 6 g). Let $(\alpha, \sigma^2) \neq (\overline{\alpha}, \overline{\sigma}^2)$ and note that

$\pi(\{k \in \mathbb{R} \mid \forall y \in \mathcal{Y}, p(y \mid \alpha, \sigma^2, \beta = k) = p(y \mid \overline{\alpha}, \overline{\sigma}^2, \beta = k)\})$

$\leq \pi(\{k \in \mathbb{R} \mid \int_{\mathcal{Y}} yp(y \mid \alpha, \sigma^2, \beta = k)dy = \int_{\mathcal{Y}} yp(y \mid \overline{\alpha}, \overline{\sigma}^2, \beta = k)dy\})$

$= \pi(\{k \in \mathbb{R} \mid \alpha k = \overline{\alpha} k\})$

We have: $(\alpha, \sigma^2) \neq (\overline{\alpha}, \overline{\sigma}^2)$ implies $\pi(\{k \in \mathbb{R} \mid \alpha k = \overline{\alpha} k\}) = 0$ implies

$\pi(\{k \in \mathbb{R} \mid \forall y \in \mathcal{Y}, p(y \mid \alpha, \sigma^2, \beta = k) = p(y \mid \overline{\alpha}, \overline{\sigma}^2, \beta = k)\}) = 0 < 1$

i.e. faithful identifiability. On the other hand it may be shown that $p(y \mid \alpha, \sigma^2) = \int_{\mathbb{R}} p(y \mid \alpha, \sigma^2, \beta)\pi(d\beta) \equiv N(0, \sigma^2 + \alpha^2)$. Evidently the integrated likelihood $p(y \mid \alpha, \sigma^2)$ can not identify α, σ^2.

Proposition 2: *If the sampling model is uniformly identified at* θ_1 *then the Bayesian model is faithfully identified at* θ_1.

Proof:

$\forall k \in \Theta_2 \ [\theta_1 \neq \overline{\theta}_1 \Rightarrow \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)]$

ifif $\theta_1 \neq \overline{\theta}_1 \Rightarrow \forall k \in \Theta_2 \ [\exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)]$

implies $\theta_1 \neq \overline{\theta}_1 \Rightarrow \pi(\{k \in \Theta_2 \mid \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)\}) = 1$
implies $\theta_1 \neq \overline{\theta}_1 \Rightarrow \pi(\{k \in \Theta_2 \mid \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)\}) > 0$
it follows that
$\forall \overline{\theta}_1 \in \Theta_1, \forall k \in \Theta_2 \ [\theta_1 \neq \overline{\theta}_1 \Rightarrow \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)]$ implies
$\forall \overline{\theta}_1 \in \Theta_1 \ [\theta_1 \neq \overline{\theta}_1 \Rightarrow \pi(\{k \in \Theta_2 \mid \exists y \in \mathcal{Y}, p(y \mid \theta_1, \theta_2 = k) \neq p(y \mid \overline{\theta}_1, \theta_2 = k)\}) > 0]$
By lemma 6 the latter is equivalent to
$\forall \overline{\theta}_1 \in \Theta_1 \ [(p(y \mid \theta_1, \theta_2 = k) = p(y \mid \overline{\theta}_1, \theta_2 = k)$ for all $y \in \mathcal{Y})$ a.e. (π) implies $\theta_1 = \overline{\theta}_1]$
which proves the proposition.

Thus either uniform or marginal identification implies faithful identification. Unfortunately we could not establish any relationship between uniform and marginal identification.

VIII. Application: SVAR and Marginal Identification

In this section we provide one possible application of the marginal identification in the context of the Structural VAR (SVAR) model. We consider the following SVAR

$$A_0 y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + c + \varepsilon_t; \qquad t = 1, \ldots, T. \tag{8}$$

where $y_t \in \mathbb{R}^{m \times 1}$, $A_0 : (m \times m)$ is nonsingular, $A_i : (m \times m)$ for $i = 1, \ldots, p$, $c \in \mathbb{R}^{m \times 1}$ is a vector of constants and $\varepsilon_t \mid y_{t-1}, \ldots \sim N(0, \mathrm{I}_m)$.

Let us think about validity of the assumption $\mathrm{cov}(\varepsilon_t) = \mathrm{I}_m$. The ready defense is that it follows from the fact that SVAR methodology interprets the disturbances as being structural (with economic content). Indeed the nature of structural shocks in SVAR is well captured by Sims and Zha (2006) who express their attitude to the disturbances covariance matrix as follows: "*We assume that all structural disturbances are mutually uncorrelated. While this goes against the traditional simultaneous equations specification, it seems natural and avoids some conceptual conundrums … Our view is that a good multiple–equation model should not leave unexplained relations among variables in the error terms … we believe that what requires explicit discussion and economic interpretation is the presence of correlations among structural disturbances, not its absence*". More elaborating discussion on this point appears in Leeper et al. (1996). It should be mentioned that the above

interpretation is in striking contrast to the classical interpretation of disturbances in SEM. Hence SVAR is not just the SEM with identity covariance matrix.

However calling disturbances "structural" is fully justified if and only if those disturbances are exactly the shocks that appear in economic model (e.g. DSGE). This will not be in general true. There are (at least) two sources of complications 1) Ad–hoc (as measured by modern standards imposed by the DSGE methodology) identifying restrictions in SVAR and 2) non–existence of the SVAR representation and the problem of lag truncation. We argue that interpretation of structural shocks in SVAR is unclear and blurred.

As for 1) one should be aware that economic models rarely induce any "zeros" for the parameters of SVAR models[6]. But excluding identifying restrictions, mostly confined to A_0, are the most popular ones in SVAR modeling. One may say that theoretical models of the economy do not provide enough excluding restrictions in order to identify the underlying structural shocks within SVAR methodology. Since the "true" contemporaneous relations contain more variables that are allowed to be estimated within SVAR model (due to identification problems), researchers are forced to exclude more variables than in fact comprise these relations. Consequently, the omission of some variables from contemporaneous relations makes the bias problem for those that are included. This happens because the included and inappropriately excluded variables are often correlated. Hence omission of relevant variables in the contemporaneous relations may be a source of residuals serial correlation – see also Liu (1960). Moreover Cooley and LeRoy (1985) argued that since (just) identifying schemes are untestable, one can not be sure whether structural shocks identified (orthogonalized) by one particular identifying scheme are really exogenous (primitive) shocks. If they are not, the derived structural shocks are in fact a combination of the real exogenous shocks. For example, what we broadly identify as private sector shock may be in fact a combination of the taste and the technology shock. On this point see also Cooley and Dwyer (1998).

Concerning 2), as is known, the log–linear approximations to virtually all DSGE models may be put in the vector ARMA framework. Since VAR econometrics implicitly assume that such a VARMA model has invertible MA component, the question whether VAR modeling properly identifies the structural shocks arises quite

[6] Lucas and Stokey (1987) "*... with specific parameterization of preferences the theory would place many restrictions on the behavior of endogenous variables. But these predictions do not take the form of locating blocks of zeros in a VAR description of these variables*".

naturally. Hansen and Sargent (1991) looking at some specific economic models noted that the condition for invertibility of MA component "*fails to be met*" for a class of models that "*is not thin in any natural sense*". If this is the case they showed that VAR analysis works very poorly e.g. leads to completely distorted impulse responses which are very different from theoretical (economic model's) impulse responses. See also Fernández–Villaverde et al. (2005). On the other hand, under certain conditions log–linear approximations to some general class of DSGE models may be cast in the infinite order SVAR model, see e.g. Christiano et al. (2006), Fernández–Villaverde et al. (2005). Note however that the matrix of contemporaneous relations i.e. A_0, will be then, in general, a very nonlinear function of deep economic parameters that usually do not induce any excluding restrictions on this matrix, as motivated in point 1). Anyway, even if SVAR(∞) representation is valid we must truncate the lags to make inference feasible. This will necessarily introduce a portion of cross–correlation into "structural" disturbances in your finite order SVAR. More complete discussion relevant to 2) is available in Ravenna (2007).

The above critique of the assumption $\mathrm{cov}(\varepsilon_t) = \mathrm{I}_m$ logically hinges on the methodological stance that for the time–series model to be valid it must be consistent with theoretical model. Although we can not philosophically object to the reverse reasoning, we think that even proponents of the latter view do not treat the assumption $\mathrm{cov}(\varepsilon_t) = \mathrm{I}_m$ as literally true but only as a useful approximation. In particular see Leeper and Zha (2002) for specific calculations of correlations between structural shocks that suggest $\mathrm{cov}(\varepsilon_t) \approx \mathrm{I}_m$ (at best). In what follows we apply the marginal identification concept in the context of the restriction $\mathrm{cov}(\varepsilon_t) = \mathrm{I}_m$.

To this end we must introduce additional notation. Let $\underset{m\times(mp+1)}{B} = [A_1\ A_2\ \dots\ A_p\ c]$, $\underset{m\times T}{y} = [y_1\ y_2\ \dots\ y_T]$ and

$$\underset{(mp+1)\times T}{X'} = \begin{bmatrix} y_0 & y_1 & \cdots & y_{T-1} \\ y_{-1} & y_0 & \cdots & y_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{-p+1} & y_{-p+2} & \cdots & y_{T-p} \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

The original SVAR imposes $\mathrm{cov}(\varepsilon_t) = \mathrm{I}_m$. Suppose for a moment that $\mathrm{cov}(\varepsilon_t) = \Omega > 0$ [7]. Then the data sampling density would read

$$p(y \mid A_0, B, \Omega, X) = (2\pi)^{-\frac{1}{2}mT} \left|\det(A_0)\right|^T |\Omega|^{-\frac{1}{2}T} \mathrm{etr}\{-\tfrac{1}{2}\Omega^{-1}(A_0 y - BX')(A_0 y - BX')'\} \qquad (9)$$

[7] For any square symmetric matrix X, $X > 0$ signifies that X is positive definite.

where $\text{etr}\{\cdot\} \equiv e^{tr\{\cdot\}}$ and $tr\{\cdot\}$ is the matrix trace operator. Of course putting $\text{cov}(\varepsilon_t) = \text{I}_m$ we arrive at the likelihood corresponding to (8)

$$p(y \mid A_0, B, \Omega = \text{I}_m, X) = (2\pi)^{-\frac{1}{2}mT} \left|\det(A_0)\right|^T \text{etr}\{-\tfrac{1}{2}(A_0 y - BX')(A_0 y - BX')'\} \tag{10}$$

Anyway being consistent with our earlier notation let $\theta_1 = (A_0, B)$ and $\theta_2 = \Omega$. Our goal is to introduce the marginal prior measure for $\theta_2 = \Omega$ and use the latter to integrate Ω out from the likelihood (9). For the ease of exposition it is useful to assume inverted Wishart distribution for Ω. In particular using the notation from Zellner (1971), p. 395, we take $\pi(\theta_2) \equiv \pi(\Omega) = IW((\nu - m - 1)\cdot \text{I}_m, \nu, m)$. The parameters of $\pi(\Omega)$ were chosen so as $E(\Omega) = \text{I}_m$ (provided that $\nu > m + 1$). Then it is easy to show

$$\begin{aligned} &p(y \mid A_0, B, X) = \int_{\Omega>0} p(y \mid A_0, B, \Omega, X)\pi(d\Omega) = \\ &= \pi^{-\frac{1}{2}mT} \prod_{i=1}^{m} \Gamma(\tfrac{T+\nu+1-i}{2})[\Gamma(\tfrac{\nu+1-i}{2})]^{-1} \cdot \mid \Sigma \mid^{\frac{1}{2}T} \ (\nu - m - 1)^{-\frac{1}{2}mT} \cdot \\ &\qquad \cdot \mid \text{I}_T + (y - \Pi X')'(\tfrac{1}{\nu-m-1}\Sigma)(y - \Pi X') \mid^{-\frac{1}{2}(T+\nu)} \end{aligned} \tag{11}$$

where $\Sigma = A_0' A_0$ and $\Pi = A_0^{-1} B$. Hence $p(y \mid A_0, B, X)$ is the pdf of the matricvariate Student distribution i.e. $p(y \mid A_0, B, X) = f_{Mt}^{m \times T}(y \mid \Pi X', \text{I}_T, \frac{1}{\nu-m-1}\Sigma, T + \nu - m)$ (notation taken from Drèze and Richard (1983)). Now since as $\nu \to \infty$ the prior $\pi(\Omega)$ becomes degenerated at $\Omega = \text{I}_m$ we will check if the ICC holds. To this end note that

$$\begin{aligned} &\lim_{\nu\to\infty} \mid \text{I}_T + \tfrac{1}{\nu-m-1}(y - \Pi X')'\Sigma(y - \Pi X') \mid^{-\frac{1}{2}(T+\nu)} = \text{etr}\{-\tfrac{1}{2}(y - \Pi X')'\Sigma(y - \Pi X')\} = \\ &= \text{etr}\{-\tfrac{1}{2}(A_0 y - BX')(A_0 y - BX')'\} \end{aligned} \tag{12}$$

$$\lim_{\nu\to\infty} (\nu - m - 1)^{-\frac{1}{2}mT} \prod_{i=1}^{m} \Gamma(\tfrac{T+\nu+1-i}{2})[\Gamma(\tfrac{\nu+1-i}{2})]^{-1} = 2^{-\frac{1}{2}mT} \tag{13}$$

Hence as $\nu \to \infty$, $p(y \mid A_0, B, X)$ converges to $p(y \mid A_0, B, \Omega = \text{I}_m, X)$ i.e. the ICC holds. Moreover since we center the prior $\pi(\Omega)$ on I_m (i.e. a point at which the usual identification theory for the model (10) works) we should not expect any problems (as indicated in example 4). Indeed we can easily demonstrate

Proposition 3: *Define* $\Sigma = A_0' A_0$, $\Pi = A_0^{-1} B$, $\overline{\Sigma} = \overline{A}_0' \overline{A}_0$, $\overline{\Pi} = \overline{A}_0^{-1} \overline{B}$. *Provided that* X' *has full row rank we obtain*:

$f_{Mt}^{m\times T}(y \mid \Pi X', \mathrm{I}_T, \frac{1}{\nu-m-1}\Sigma, T+\nu-m) = f_{Mt}^{m\times T}(y \mid \overline{\Pi} X', \mathrm{I}_T, \frac{1}{\nu-m-1}\overline{\Sigma}, T+\nu-m)$ *for all* $y \in \mathbb{R}^{m\times T}$ *implies* $(\overline{A}_0 = HA_0$ *and* $\overline{B} = HB)$, *where* $H : (m \times m)$ *is any orthogonal matrix i.e.* $HH' = H'H = \mathrm{I}_m$.

Proof: $f_{Mt}^{m\times T}(y \mid \Pi X', \mathrm{I}_T, \frac{1}{\nu-m-1}\Sigma, T+\nu-m) = f_{Mt}^{m\times T}(y \mid \overline{\Pi} X', \mathrm{I}_T, \frac{1}{\nu-m-1}\overline{\Sigma}, T+\nu-m)$ for all $y \in \mathbb{R}^{m\times T}$ implies that expected values of these densities are the same hence: $E(y \mid A_0, B, X) = A_0^{-1}BX' = \overline{A}_0^{-1}\overline{B}X' = E(y \mid \overline{A}_0, \overline{B}, X)$. Provided that X' is full row rank it possesses the right inverse hence $A_0^{-1}BX' = \overline{A}_0^{-1}\overline{B}X'$ implies $A_0^{-1}B = \overline{A}_0^{-1}\overline{B}$. Moreover the hypothesis of the proposition implies that the covariance matrices of the two matricvariate Student pdf's are also the same i.e. $\mathrm{cov}(vec(y) \mid A_0, B, X) = \mathrm{I}_T \otimes \Sigma^{-1} = \mathrm{I}_T \otimes \overline{\Sigma}^{-1} = \mathrm{cov}(vec(y) \mid \overline{A}_0, \overline{B}, X) \Leftrightarrow \Sigma^{-1} = \overline{\Sigma}^{-1}$ hence $A_0^{-1}A_0'^{-1} = \overline{A}_0^{-1}\overline{A}_0'^{-1}$. By Vinograd's theorem $A_0^{-1}A_0'^{-1} = \overline{A}_0^{-1}\overline{A}_0'^{-1}$ ifif $\overline{A}_0^{-1} = A_0^{-1}H$ ifif $\overline{A}_0 = H'A_0$ for some orthogonal $H : (m \times m)$ matrix. Inserting $\overline{A}_0 = H'A_0$ into $A_0^{-1}B = \overline{A}_0^{-1}\overline{B}$ we get $A_0^{-1}B = A_0^{-1}H\overline{B}$ ifif $B = H\overline{B}$ ifif $\overline{B} = H'B$.

Proposition 3 states that equivalence class of the sampling model $p(y \mid A_0, B, \Omega = \mathrm{I}_m, X)$ is the same as the equivalence class of the marginal (Bayesian) model $p(y \mid A_0, B, X) = \int_{\Omega>0} p(y \mid A_0, B, \Omega, X)\pi(d\Omega)$. Hence all results on identification given in Rubio–Ramírez et al (2010) in the context of $p(y \mid A_0, B, \Omega = \mathrm{I}_m, X)$ are also valid for the marginal model.

As a final remark we note that Bayesian inference on marginal model may be easily conducted. In particular using a joint flat prior for A_0, B we can decompose the posterior as

$$p(A_0, B \mid X, y) = p(B \mid A_0, X, y)p(A_0 \mid X, y) \tag{14}$$

in which:

$$p(B \mid A_0, X, y) = f_{Mt}^{m\times(mp+1)}(B \mid A_0 yX(X'X)^{-1}, (X'X)^{-1}, ((\nu-m-1)\cdot \mathrm{I}_m + A_0 QA_0')^{-1}, T+\nu-m)$$

$$p(A_0 \mid X, y) \propto \mid A_0 Q^* A_0' \mid^{\frac{1}{2}T} \mid \mathrm{I}_m + A_0 Q^* A_0' \mid^{-\frac{1}{2}(T+\nu-mp-1)}$$

where $Q = y[\mathrm{I}_T - X(X'X)^{-1}X']y'$ and $Q^* = \frac{1}{\nu-m-1}Q$.

Since the factorization (14) was derived under a flat prior, the question about existence of the posterior is highly relevant. The following lemma will be instrumental in answering this question

Lemma 7:

$$\int_{\mathbb{R}^{m\times m}} |\, A_0Q^*A_0' \,|^{\frac{1}{2}T} |\, \mathrm{I}_m + A_0Q^*A_0' \,|^{-\frac{1}{2}(T+\nu-mp-1)}\, (dA_0) =$$

$$= |\, Q^* \,|^{-\frac{1}{2}m}\ \pi^{\frac{1}{2}m^2}[\Gamma_m(\tfrac{m}{2})]^{-1}\Gamma_m(\tfrac{T+m}{2})\Gamma_m(\tfrac{\nu-mp-m-1}{2})[\Gamma_m(\tfrac{T+\nu-mp-1}{2})]^{-1}$$

where $\Gamma_m(\cdot)$ *is the multivariate gamma function defined as* $\Gamma_m(a) = \pi^{\frac{1}{4}m(m-1)}\prod_{i=1}^{m}\Gamma(a-\frac{i-1}{2})$.

Proof: If $rank(y' \vdots X) = m + mp + 1$ (i.e. $(y' \vdots X)$ is of full column rank) then Q^* is positive definite and we may decompose $Q^* = RR'$, where R is lower triangular. Changing variables as $X = A_0R$ with the Jacobian $J(A_0 \to X) = |\, R \,|^{-m}$ we get

$$\int_{\mathbb{R}^{m\times m}} |\, A_0Q^*A_0' \,|^{\frac{1}{2}T} |\, \mathrm{I}_m + A_0Q^*A_0' \,|^{-\frac{1}{2}(T+\nu-mp-1)}\, (dA_0)$$

$$= |\, R \,|^{-m} \int_{\mathbb{R}^{m\times m}} |\, XX' \,|^{\frac{1}{2}T} |\, \mathrm{I}_m + XX' \,|^{-\frac{1}{2}(T+\nu-mp-1)}\, (dX) =$$

$$= |\, R \,|^{-m}\ \pi^{\frac{1}{2}m^2}[\Gamma_m(\tfrac{m}{2})]^{-1}\int_{W>0} |\, W \,|^{\frac{1}{2}(T-1)} |\, \mathrm{I}_m + W \,|^{-\frac{1}{2}(T+\nu-mp-1)}\, (dW)$$

where the last equality follows by Hsu's lemma (see e.g. theorem 1.4.10. in Gupta and Nagar (2000)). But the integrand is the kernel of matricvariate beta type II pdf (see e.g. Gupta and Nagar (2000) p. 166). Hence the integrating constant of the integral (with respect to W) is readily available. Multiplying all the constant terms and noting $|\, R \,|^{-m} = |\, Q^* \,|^{-\frac{1}{2}m}$ we arrive at

$$\int_{\mathbb{R}^{m\times m}} |\, A_0Q^*A_0' \,|^{\frac{1}{2}T} |\, \mathrm{I}_m + A_0Q^*A_0' \,|^{-\frac{1}{2}(T+\nu-mp-1)}\, (dA_0)$$

$$= |\, Q^* \,|^{-\frac{1}{2}m}\ \pi^{\frac{1}{2}m^2}[\Gamma_m(\tfrac{m}{2})]^{-1}\Gamma_m(\tfrac{T+m}{2})\Gamma_m(\tfrac{\nu-mp-m-1}{2})[\Gamma_m(\tfrac{T+\nu-mp-1}{2})]^{-1}.$$

Clearly for the existence of $p(A_0 \mid X, y)$ we need all arguments of multivariate gamma functions to be greater than $\frac{m-1}{2}$ (i.e. the integrating constant derived in lemma 7 is well defined). In particular $\frac{\nu-mp-m-1}{2} > \frac{m-1}{2}$ is the most stringent which amounts to $\nu > mp + 2m$. Hence the latter is the condition for existence of the joint posterior under the joint flat prior for A_0, B [8] in the marginal model.

[8] More precisely the condition $\nu > mp + 2m$ is necessary and sufficient for existence of the posterior of a sampling model without any identifying restrictions i.e. posterior with the support $(A_0, B) \in \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times(mp+1)}$. Note however that with any other support being a subset $\mathbb{R}^{m\times m} \times \mathbb{R}^{m\times(mp+1)}$ (induced by identifying restrictions) the condition $\nu > mp + 2m$ is still sufficient for existence of the posterior defined on this subset.

Of course to use in practice the Bayesian inference based on the "integrated likelihood" model we should provide identifying restrictions. This will require a design of the efficient sampling methods to draw from $p(A_0 \mid X, y)$, which is however beyond the scope of the present paper.

IX. Concluding Remarks

We provided (hopefully) a fresh clarifying view on identification problem seen from Bayesian perspective. Although most results concerning the core of Bayesian identification presented in the paper seem to be trivial their interpretations suggest that they constitute a proper basis for discussing the identification problem. One major message is that relationship between the prior and the posterior for the given data is silent about identification of Bayesian or sampling model. What is important is how this relationship behaves when alternative samples are available. Moreover the hope that inexact (probabilistic) restrictions laid out within Bayesian approach could replace exact restrictions should be abandoned. Probabilistic restriction can not identify either sampling or Bayesian model.

We discussed three concepts: uniform, marginal and faithful identification. We think that the concept of marginal identification is the most important. First, it implies faithful identification. Second, it constitutes a Bayesian contribution to the general concept of identification. Something which does not have any counterpart in non–Bayesian framework. However in practice the use of marginal identification may be a challenge. In particular integrating out "uncertain" restrictions from the model may not be analytically attainable. That is it may entail numerical integration. For example we can not analytically integrate out parameters comprising contemporaneous relations in SVAR (though this seems to be the most natural application of marginal identification in the context of SVAR). However this may not be considered as a serious obstacle in a world where Monte Carlo methods became the synonym of applied Bayesian analysis. As a potential interesting application of the marginal identification we should mention the DSGE modeling. In particular if one is uncertain about sensible calibration of a group of deep parameters we can just integrate them out from a model. Then such a prepared "integrated likelihood" model will be subject to the usual Bayesian inference.

REFERENCES:

Barankin, E.W. (1960), "Sufficient Parameters: Solution of the Minimal Dimensionality Problem", *Annals of the Institute of Statistical Mathematics*, 12, pp. 91–118.

Bowden, R. (1973), "The Theory of Parametric Identification", *Econometrica*, 41, pp. 1069–1074.

Canova, F. and L. Sala (2009), "Back to Square One: Identification Issues in DSGE Models", *Journal of Monetary Economics*, 56, pp. 431–449.

Christiano, L.J., M. Eichenbaum and R. Vigfusson (2006), "Assessing Structural VARs", in. D. Acemoglu, K. Rogoff and M. Woodford, eds., *NBER Macroeconomics Annual 2006*, *vol. 21*, MIT Press.

Cooley, T.F. and M. Dwyer (1998), "Business Cycle Analysis Without Much Theory: A Look at Structural VARs", *Journal of Econometrics*, 83, pp. 57–88.

Cooley, T.J. and S.F. LeRoy (1985), "Atheoretical Macroeconomics: A Critique", *Journal of Monetary Economics*, 16, pp. 283–308.

Dawid, A. P. (1979), "Conditional Independence in Statistical Theory" (with discussion), *Journal of the Royal Statistical Society*, *Ser. B*, 41, pp. 1–31.

De Finetti, B. (1974), *Theory of Probability*, *vol. I*, John Wiley & Sons, Chichester.

Drèze, J.H. (1962), "The Bayesian Approach to Simultaneous Equations Estimation", *ONR Research Memorandum 67*, *The Technological Institute*, *Northwestern University*.

Drèze, J.H. (1972), "Econometrics and Decision Theory", *Econometrica*, 40, pp. 1–17.

Drèze, J.H. (1974), "Bayesian Theory of Identification in Simultaneous Equations Models", in: S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics*, North–Holland Pub. Co., Amsterdam.

Drèze, J.H. (1976), "Bayesian Limited Information Analysis of the Simultaneous Equations Model", *Econometrica*, 44, pp. 1045–1075.

Drèze, J.H. and M. Mouchart (1990), "Tales of Testing Bayesians", in: R.A.L. Carter, J. Dutta and A. Ullah, eds., *Contributions to Econometric Theory and Application*, Springer–Verlag, New York.

Drèze, J.H. and J–F. Richard (1983), "Bayesian Analysis of Simultaneous Equation Systems", in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, vol. I*, North–Holland Pub. Co., Amsterdam.

Fernández–Villaverde, J., J.F. Rubio–Ramírez and T.J. Sargent (2005), "A,B,C's (and D)'s for Understanding VARS", *New York University*, *Working Paper*.

Fisher, F.M. (1966), *The Identification Problem in Econometrics*, McGraw–Hill, New York.

Florens, J–P., M. Mouchart and J–M. Rolin (1990), *Elements of Bayesian Statistics*, Marcel Dekker, Inc., New York.

Galí, J., M. Gertler and J.D. López–Salido (2005), "Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve", *Journal of Monetary Economics*, 52, pp. 1107–1118.

Gelfand, A.E. and S.K. Sahu (1999), "Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models", *Journal of the American Statistical Association*, 94, pp. 247–253.

Good, I.J. (1960), "Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments", *Journal of the Royal Statistical Society*, *series B*, 22, pp. 319–331.

Good, I.J. (1966), "A Derivation of the Probabilistic Explication of Information", *Journal of the Royal Statistical Society*, *series B*, 28, pp. 578–581.

Gupta, A.K. and D.K. Nagar (2000), *Matrix Variate Distributions*, Chapman & Hall/CRC, Boca Raton.

Gustafson, P. (2005), "On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables", *Statistical Science*, 20, pp. 111–129.

Gustafson, P. (2009), "What Are the Limits of Posterior Distributions Arising from Nonidentied Models, and Why Should We Care?", *Journal of the American Statistical Association*, 104, pp. 1682–1695.

Haavelmo, T. (1944), "The Probability Approach in Econometrics", *Econometrica*, 12, pp. 1–115.

Hamilton, A.G. (1978), *Logic for Mathematicians*, Cambridge University Press, Cambridge.

Hansen, L.P. and T.J. Sargent (1991), "Two Difficulties in Interpreting Vector Autoregressions", in: L.P. Hansen and T.J. Sargent, *Rational Expectations Econometrics*, Westview Press.

Hsiao, C. (1983) "Identification", in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, vol. I*, North–Holland Pub. Co., Amsterdam.

Kadane, J.B. (1974), "The Role of Identification in Bayesian Theory", in: S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics*, North–Holland Pub. Co., Amsterdam.

King, R.G. and M.W. Watson (1997), "Testing Long–Run Neutrality", *Federal Reserve Bank of Richmond Economic Quarterly*, *vol. 83/3*.

Kocięcki, A. (2011), "Algebraic Theory of Identification in Parametric Models", *National Bank of Poland*, *Working Paper no. 88*.

Koop, G., H.M. Pesaran and R.P. Smith (2011), "On Identification of Bayesian DSGE Models", *Journal of Business and Economic Statistics*, forthcoming.

Kraay, A. (2012), "Instrumental Variables Regressions with Uncertain Exclusion Restrictions: A Bayesian Approach", *Journal of Applied Econometrics*, 27, pp. 108–128.

Leamer, E.E. (1978), *Specification Searches*: *Ad Hoc Inference with Nonexperimental Data*, John Wiley & Sons, New York.

Leeper, E.M., C.A. Sims and T. Zha (1996), "What Does Monetary Policy Do?" with discussion, *Brookings Papers on Economic Activity*, 2, pp. 1–78.

Leeper, E.M. and T. Zha (2002), "Empirical Analysis of Policy Interventions", *NBER working paper no. 9063*.

Liu, T–C. (1960), "Underidentification, Structural Estimation and Forecasting", *Econometrica*, 28, pp. 855–865.

Lucas, R.E. and N.L. Stokey (1987), "Money and Interest in a Cash–in–Advance Economy", *Econometrica*, 55, pp. 491–513.

Maddala, G.S. (1976), "Weak Priors and Sharp Posteriors in Simultaneous Equation Models", *Econometrica*, 44, pp. 345–251.

Morales, J–A. (1971), *Bayesian Full Information Structural Analysis*, Springer–Verlag, Berlin.

Müller, U.K. (2012), "Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models", *Journal of Monetary Economics*, 59, pp. 581–597.

Oulhaj A. and M. Mouchart (2003), "Partial Sufficiency with Connection to the Identification Problem", *Metron*, 61, pp. 267–283.

Poirier, D.J. (1998), "Revising Beliefs in Nonidentified Models", *Econometric Theory*, 14, pp. 483–509.

Prakasa Rao, B.L.S. (1992), *Identifiability in Stochastic Models*: *Characterization of Probability Distributions*, Academic Press, Inc.

Ravenna, F. (2007), "Vector Autoregressions and Reduced Form Representations of DSGE Models", *Journal of Monetary Economics*, 54, pp. 2048–2064.

Ríos–Rull, J–V., F. Schorfheide, C. Fuentes–Albero, M. Kryshko and R. Santaeulàlia–Llopis (2012), "Methods versus Substance: Measuring the Effects of Technology Shocks", *Journal of Monetary Economics*, 59, pp. 826–846.

Rothenberg, T.J. (1971), "Identification in Parametric Models", *Econometrica*, 39, pp. 577–591.

Rothenberg, T.J. (1973), *Efficient Estimation with A Priori Information*, Yale University Press, New Heaven and London.

Rubio–Ramírez, J.F, D.F. Waggoner and T. Zha (2010), "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference", *The Review of Economic Studies*, 77, pp. 665–696.

Sims, C.A. (2007), "Thinking about Instrumental Variables", *manuscript*, available at http://sims.princeton.edu/yftp/IV/IV.pdf.

Sims, C.A. and T. Zha (2006), "Does Monetary Policy Generate Recessions?", *Macroeconomic Dynamics*, 10, pp. 231–272.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, New York.