Journal of Econometrics 175 (2013) 132–141

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Methods for computing marginal data densities from the Gibbs output

Cristina Fuentes-Albero [a,*], Leonardo Melosi [b]

[a] *Department of Economics, 75 Hamilton Street, Rutgers University, New Brunswick, NJ 08901, United States*

[b] *Federal Reserve Bank of Chicago, 230 S LaSalle St., Chicago, IL 60604, United States*

ARTICLE INFO

Article history:
Received 4 August 2011
Received in revised form
2 January 2013
Accepted 8 March 2013
Available online 16 March 2013

JEL classification:
C11
C15
C16
C32

Keywords:
Marginal likelihood
Gibbs sampler
Time series econometrics
Bayesian econometrics
Reciprocal importance sampling

ABSTRACT

We introduce two estimators for estimating the Marginal Data Density (MDD) from the Gibbs output. Our methods are based on exploiting the analytical tractability condition, which requires that some parameter blocks can be analytically integrated out from the conditional posterior densities. This condition is satisfied by several widely used time series models. An empirical application to six-variate VAR models shows that the bias of a fully computational estimator is sufficiently large to distort the implied model rankings. One of the estimators is fast enough to make multiple computations of MDDs in densely parameterized models feasible.

1. Introduction

Modern macroeconometric methods are based on densely parameterized models such as vector autoregressive models (VARs) or dynamic factor models (DFMs). Densely parameterized models deliver a better in-sample fit. It is well-known, however, that such models can deliver erratic predictions and poor out-of-sample forecasts due to parameter uncertainty. To address this issue, Sims (1980) suggested to use priors to constrain parameter estimates by "shrinking" them toward a specific point in the parameter space. Provided that the direction of shrinkage is chosen accurately, it has been shown that densely parameterized models are extremely successful in forecasting. This explains the popularity of largely parameterized models in the literature (Stock and Watson, 2002; Forni et al., 2003; Koop and Potter, 2004; Korobilis, 2008; Banbura et al., 2010; Koop, 2011).

The direction of shrinkage is often determined by maximizing the marginal likelihood of the data (see Carriero et al., 2010; Giannone et al., 2012), also called marginal data density (MDD). The marginal data density is defined as the integral of the likelihood function with respect to the prior density of the parameters. In few cases, the MDD has an analytical representation. When an analytical solution for this density is not available, we need to rely on computational methods, such as Chib's method (Chib, 1995), Importance Sampling estimators (Hammersley and Handscomb, 1964; Kloek and van Dijk, 1978; Geweke, 1989), estimators based on the Reciprocal Importance Sampling principle (Gelfand and Dey, 1994), importance sampling based on mixture approximations (Frühwirth-Schnatter, 1995), the Bridge Sampling estimator (Meng and Wong, 1996), or the Warp Bridge Sampling estimator (Meng and Shilling, 2002). Since all these methods rely on computational methods to integrate the model parameters out of the posterior density, their accuracy deteriorates as the dimensionality of the parameter space grows large. Hence, there is a tension between the need for using broadly parameterized models for forecasting and the accuracy in estimating the MDD which influences the direction of shrinkage.

This paper aims at mitigating this tension by introducing two estimators (henceforth, Method 1 and Method 2) that exploit the information about models' analytical structure. While Method 1 can be considered as a refinement of the approach proposed by Chib (1995), Method 2 is based upon the Reciprocal Importance Sampling principle as in Gelfand and Dey (1994). Conversely to

* Corresponding author.
E-mail addresses: cfuentes@econ.rutgers.edu (C. Fuentes-Albero), lmelosi@frbchi.org (L. Melosi).

0304-4076/$ – see front matter
http://dx.doi.org/10.1016/j.jeconom.2013.03.002

fully computational methods, Method 1 and Method 2 rely on the analytical integration of some parameter blocks.[1]

The proposed estimators can be applied to econometric models satisfying two conditions. The first condition (henceforth, *sampling condition*) requires that the posterior density can be block-partitioned so as to be approximated via the Gibbs sampler. The second condition (henceforth, *analytical tractability condition*) states that there exists an integer $\tau \geq 2$ such that the conditional posterior $p(\theta_1, \ldots, \theta_\tau | \theta_{\tau+1}, \ldots, \theta_s, D, Y)$ can be analytically derived, where Y is the sample data, D is a set of unobservable model variables, and s is the total number of parameter blocks θ_i, $i \in \{1, \ldots, s\}$. These two conditions are met by a wide range of models, such as Vector AutoRegressive Models (VARs), just-identified Structural VAR models (SVARs), Reduced Rank Regression Models such as Vector Equilibrium Correction Models (VECMs), unrestricted Markov-Switching VAR models (MS VARs), Dynamic Factor Models (DFMs), Factor Augmented VAR models (FAVARs), and Time-Varying Parameter (TVP) VAR models.

By means of a Monte Carlo experiment, we show that exploiting the *analytical tractability condition* leads to sizable gains in accuracy and computational burden, which quickly grow with the dimensionality of the parameter space of the model. We consider VAR(p) models, in the form studied by Villani (2009) and Del Negro and Schorfheide (2010) (i.e., the so-called mean-adjusted VAR models), from one up to four lags, $p = 1, \ldots, 4$. We fit these four VAR models, under a single-unit-root prior (Sims and Zha, 1998), to data sets with increasing number of observable variables. It is compelling to focus on mean-adjusted VAR models because the true conditional predictive density[2] can be analytically derived in closed form. We can compare the performance of our estimators with their fully computational counterparts; that is to say the estimator proposed by Chib (1995) and that introduced by Gelfand and Dey (1994). Method 1 and Chib's method only differ in the computation of the conditional predictive density when applied to mean-adjusted VAR models. While Method 1 evaluates the exact analytical expression for the conditional predictive density, Chib's method approximates this density computationally via Monte Carlo integration. Therefore, we can quantify the accuracy gains associated with exploiting the *analytical tractability condition* by comparing the conditional predictive density estimated by Chib's method with its true value. This assessment would have not been possible, if we had based our Monte Carlo experiment on models that require data augmentation to approximate the posterior, such as DFMs, or on other estimators rather than Chib's method, such as the Bridge Sampling estimator.

The main findings of the experiment are: (i) the fully-computational estimators that neglect the *analytical tractability condition* lead to an estimation bias that severely distorts model rankings; (ii) our two methods deliver very similar results in terms of posterior model rankings, suggesting that their accuracy is of the same order of magnitude in the experiment; (iii) exploiting the analytical tractability condition prevents our estimators from being affected by the curse of dimensionality. Related to this last finding, we argue that Method 2 is suitable for performing model selection and model averaging across a large number of models, as it is the fastest.

The paper is organized as follows. Section 2 introduces the conditions that a model has to satisfy in order to apply our two estimators. In this section, we describe the two methods proposed in this paper for computing the MDD. Section 3 performs the Monte Carlo application. Section 4 concludes.

2. Methods for computing the marginal data density

The marginal data density (MDD), also known as the marginal likelihood of the data, is defined as the integral taken over the likelihood with respect to the prior distribution of the parameters. Let Θ be the parameter set of an econometric model and Y be the sample data. Then, the marginal data density is defined as

$$p(Y) = \int p(Y|\Theta)p(\Theta)d\Theta \qquad (1)$$

where $p(Y|\Theta)$ and $p(\Theta)$ denote the likelihood and the prior density, respectively.

In Section 2.1, we describe the two methods proposed in this paper in a canonical situation consisting of four vector blocks. In Section 2.2, we present the two estimators applied to the general case of s vector blocks. Finally, Section 2.3 deals with the scope of application of the proposed estimators.

2.1. Four vector blocks

Let us consider a model whose set of parameters and latent variables is denoted by $\Theta^D = \{D, \Theta\}$ where D stands for the latent variables and Θ for the parameters of the model, where $\Theta = \{\theta_1, \theta_2, \theta_3\}$. We denote the prior for model's parameters as $p(\Theta)$, which is assumed to have a known analytical representation. Furthermore, the likelihood function, $p(Y|\Theta)$, is assumed to be known in closed form or easy to evaluate. We focus on models satisfying the following two conditions:

(i) It is possible to draw from the conditional posterior distributions $p(\theta_1|\theta_2, \theta_3, D, Y)$, $p(\theta_2|\theta_1, \theta_3, D, Y)$, $p(\theta_3|\theta_1, \theta_2, D, Y)$, and from the posterior predictive density, $p(D|\theta_1, \theta_2, \theta_3, Y)$.

(ii) The conditional posterior distribution $p(\theta_1, \theta_2|\theta_3, D, Y)$ is analytically tractable.

Condition (i) implies that we can approximate the joint posterior $p(\Theta|Y)$ and the predictive density $p(D|Y)$ through the Gibbs sampler. We label this condition as the *sampling condition*. Condition (ii) is the *analytical tractability condition* and is most likely to be satisfied through a wise partitioning of the parameter space and the specification of a conjugate prior.

Method 1 is based on interpreting the MDD as the normalizing constant of the joint posterior distribution

$$p(Y) = \frac{p(Y|\Theta)\,p(\Theta)}{p(\theta_1|\theta_2, \theta_3, Y)\,p(\theta_2|\theta_3, Y)\,p(\theta_3|Y)} \qquad (2)$$

where the numerator is the product of the likelihood and the prior, with all integrating constants included, and the denominator is the posterior density of Θ. Denote the posterior mode as $\widetilde{\Theta} = [\widetilde{\theta}_1, \widetilde{\theta}_2, \widetilde{\theta}_3]$. Hereafter, let $p(\cdot)$ denote a density for which an analytical expression is available and $\widehat{p}(\cdot)$ denote a density that needs to be

[1] Fiorentini et al. (2011) use Kalman filtering and Gaussian quadrature to integrate scale parameters out of the likelihood function for dynamic mixture models.

[2] If one partitions the parameter space Θ into s vector blocks; that is $\Theta = \{\theta_1, \ldots, \theta_s\}$, the conditional predictive density $p(Y|\theta_{\tau+1}, \ldots, \theta_s)$ is defined as

$$p(Y|\theta_{\tau+1}, \ldots, \theta_s) \equiv \int p(Y|\theta_1, \ldots, \theta_s)\, p(\theta_1, \ldots, \theta_\tau|\theta_{\tau+1}, \ldots, \theta_s)\, d\theta_1 \ldots d\theta_\tau$$

where $p(Y|\theta_1, \ldots, \theta_s)$ is the likelihood function and $p(\theta_1, \ldots, \theta_\tau|\theta_{\tau+1}, \ldots, \theta_s)$ is the prior for the first τ parameter blocks (conditional on the remaining blocks). Note that the conditional predictive density is a component of the MDD, $p(Y)$, that can be expressed as follows:

$$p(Y) = \int p(Y|\theta_{\tau+1}, \ldots, \theta_s)\, p(\theta_{\tau+1}, \ldots, \theta_s)\, d\theta_{\tau+1}, \ldots, d\theta_s$$

where $p(\theta_{\tau+1}, \ldots, \theta_s)$ is the prior for the parameter blocks that cannot be analytically integrated out.