Discussion of Prior Choice and DSGE Model Comparisons

by Del Negro and Schorfheide

Simon Potter

New York Fed

August 2006

The views expressed are not necessarily those of FRBNY or FRS.

Overview of Paper

- Priors matter for Bayesian analysis of DSGE models
- Standard approaches to choosing priors in DSGE models as more frictions are introduced is dangerous
- Some of the dangers can be avoided by using dummy observation priors

Assessment

- Paper addresses a crucial issue in Bayesian analysis (of DSGE models): Construction of Prior
- Makes major contribution to "dummy observation" approach of construction of prior
- Focus comments on the effects of prior choice/construction on marginal likelihood calculations in simple example AR(1) from paper

Simplified Version

Consider two models:

- Model 0 with parameter vector ψ, prior distribution $p_0(\psi)$ and likelihood $\ell(Y|\psi)$. For example, the neoclassical model without frictions
- Model 1 with parameter vector (ψ, θ),prior distribution $p_1(\psi, \theta)$ and likelihood $\ell(Y|\psi, \theta)$. For example, neoclassical model with adjustment costs, mark up shocks etc. captured by θ
- Assume the prior distribution for Model 0 is given by other sample information
- Prior distribution for Model 1 will depend on subjective views of investigator and "more objective prior information"

The data Y allows updating on prior weights and parameters of each model.
The marginal likelihood is defined by

$$m(Y) = \int \ell(Y|\chi)p(\chi)d\chi,$$

$$\frac{m_0(Y)}{m_1(Y)} = \frac{\int \ell(Y|\psi)p_0(\psi)d\psi}{\int\int \ell(Y|\psi,\theta)p_1(\psi,\theta)d\psi d\theta}$$

This can be adjusted by prior model weights to form posterior model weights to construction predictions that average out over model uncertainty.

Priors on parameters and models might vary in model averaging exercise depending on the nature of the prediction: pure vs. policy projection

Nested Models

Assume (with great loss of generality) that model 0 is nested within other model:

$$\ell(Y|\psi) = \ell(Y|\psi, \theta^*) = \ell(Y|\psi, \phi^*)$$

for $\theta = \theta^*$

Bayes factor comparisons with nested models simplify to Savage Dickey Density Ratio

Loss of generality because in practice models nested at boundary of parameter space and simplifications of Savage Dickey Density ratio might not be valid

Basic Marginal Likelihood Identity

Basic marginal likelihood identity (Chib or rearranging Bayes' Rule):

$$m(Y) = \frac{\ell(Y|\chi)p(\chi)}{p(\chi|Y)} \quad \text{for all } \chi$$

Then:

$$\frac{m_0(Y)}{m_1(Y)} p_0(\psi|Y) = \frac{\ell(Y|\psi)p_0(\psi)}{\ell(Y|\psi,\theta)p_1(\psi,\theta)} p_1(\psi,\theta|Y)$$

Using nested property we have

$$\begin{aligned} \frac{m_0(Y)}{m_1(Y)} p_0(\psi|Y) &= \frac{p_0(\psi)}{p_1(\psi,\theta^*)} p_1(\psi,\theta^*|Y) \\ \frac{m_0(Y)}{m_1(Y)} \frac{p_1(\psi|\theta^*)}{p_0(\psi)} p_0(\psi|Y) &= \frac{p_1(\theta^*|Y)}{p_1(\theta^*)}. \end{aligned}$$

Savage Dickey Density Ratio

Define the correction factor:

$$c_1 = \int \frac{p_1(\psi|\theta^*)}{p_0(\psi)} p_0(\psi|Y) d\psi,$$

then

$$\frac{m_0(Y)}{m_1(Y)} = \frac{p_1(\theta^*|Y)}{p_1(\theta^*)c_1}.$$

If $p_1(\psi|\theta^*) = p_0(\psi)$ then $c_1 = 1$ and

$$\frac{m_0(Y)}{m_1(Y)} = \frac{p_1(\theta^*|Y)}{p_1(\theta^*)} \qquad \text{Savage Dickey Density Ratio}$$

Denominator Very Similar to Frequentist Test Statistics

Bartlett "Paradox"

If $p_1(\theta)$ is improper, the odds in favor of model 0 are infinite. Simple example:

$$y_t = \mu + \varepsilon_t,$$

where *a priori* $\mu \sim N(0, 1/\tau^2)$ then Bayes Factor for model centered at zero vs centered at μ is given by

$$\frac{\sqrt{\tau^2 + T}}{\tau} \exp\left[-\frac{T^2\overline{y}^2}{2(\tau^2 + T)}\right].$$

Generalized Savage Dickey Density Ratio

$$\frac{m_0(Y)}{m_1(Y)} = \frac{p_1(\theta^*|Y)}{p_1(\theta^*)c_1}.$$

While $p_1(\theta^*|Y)$ depends on the choice for $p_1(\psi)$ and $p_1(\theta|\psi)$, the data as viewed through the likelihood for model 1 will dominate as the sample size grows

For example, Prescott would choose a $p_1(\theta)$ with most mass around zero. In contrast if $p_1(\psi|\theta^*) \neq p_0(\psi)$ then the Bayes Factor will be effected by prior choices over ψ as viewed through the likelihood of model 0 in the value of c_1

Prescott would choose independence but would not assume equality of two priors on ψ

Simple Example: Mean of AR(1)

Consider following example from paper

$$y_t = \psi + \theta y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim N(0, 1).$$

At $\theta = \theta^* = 0$ models are nested.
Interested in assessing whether the data has dependence

If we assume that ψ and θ are a priori independent then $p_1(\psi|\theta^*) = p_1(\psi)$

If we further assume that $p_1(\psi) = p_0(\psi)$ then sample information on the location μ of Y rather than persistence of the observed sample can dominate the Bayes Factor

Message

Use of same priors across different models can have unintended effects on Marginal Likelihoods\Bayes Factors

Thus priors should be chosen so that prior predictive distributions of location and scale are close

Message on Autocorrelation Function

Most analysis of DSGE models focuses on linear approximation with Gaussian errors

That is, only relevant information in data for Bayesian is location, scale and autocorrelation function

It seems sensible to standardize prior predictive distributions for location and scale of real variables, but not for autocorrelation function in such contexts

Prior Construction to Avoid Problem

In standard example from paper involving change of variable argument assume common belief across models about location,

$$\frac{\psi}{1-\theta} = \mu \sim N(\underline{\mu}, \underline{\lambda}^2),$$

and $\theta \sim U(-1, 1)$. Thus

$$\psi|\theta \sim N(\underline{\mu}(1-\theta), \underline{\lambda}^2(1-\theta)^2),$$

Now ψ and θ are *a priori* DEPENDENT and

$$p_1(\psi|\theta) \neq p_0(\psi) \text{ if } \theta \neq \theta^*,$$

BUT with equality at θ^*.

Call this prior: $p_1^*(\psi, \theta)$

Practical Prior Construction (FM)

Practical (i.e., computationally feasible) approach in paper is to match prior mean view on location for a specific value of θ

Contribution of paper is to given a practical method of doing this for complex nonlinear DSGE models

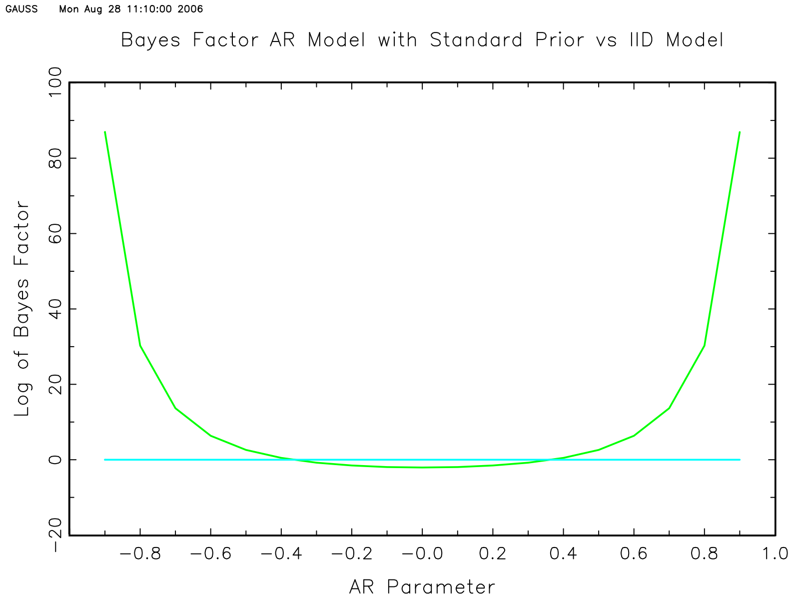
In simple example under model 1 we have $\psi \sim N(\underline{\mu}(1-\underline{\theta}), \underline{\lambda}^2(1-\underline{\theta})^2)$ where again the marginal $p_1(\theta)$ is not affected. Call this prior: $p_1^{**}(\psi, \theta)$

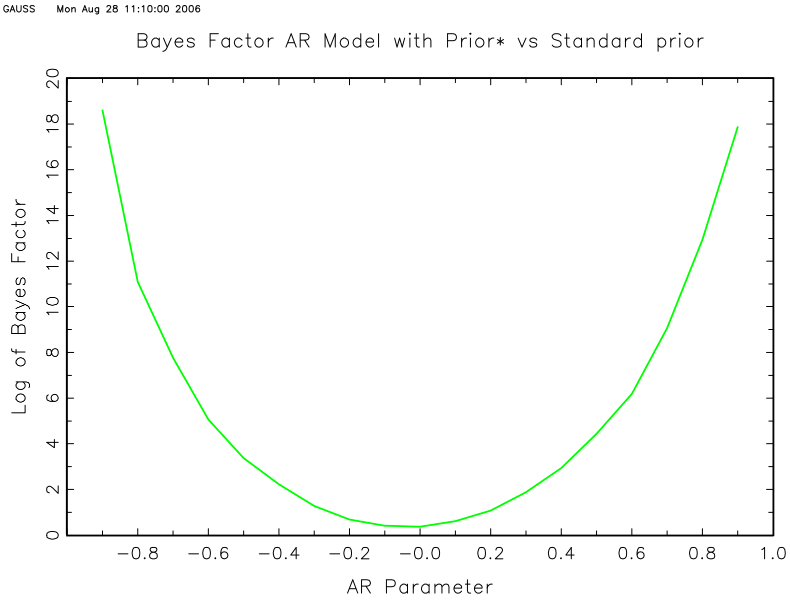
In the more general case of paper the marginal $p_1(\theta)$ will change. Some Bayesians would find this controversial

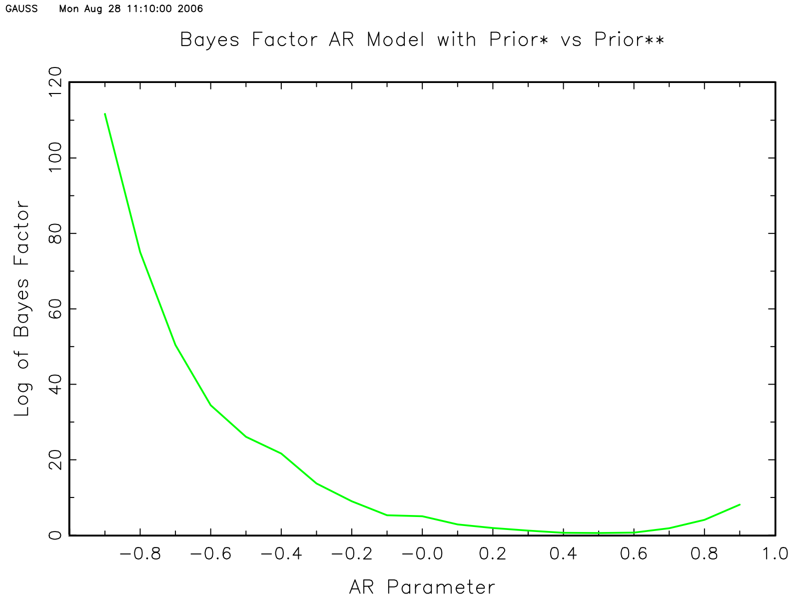
Experiment

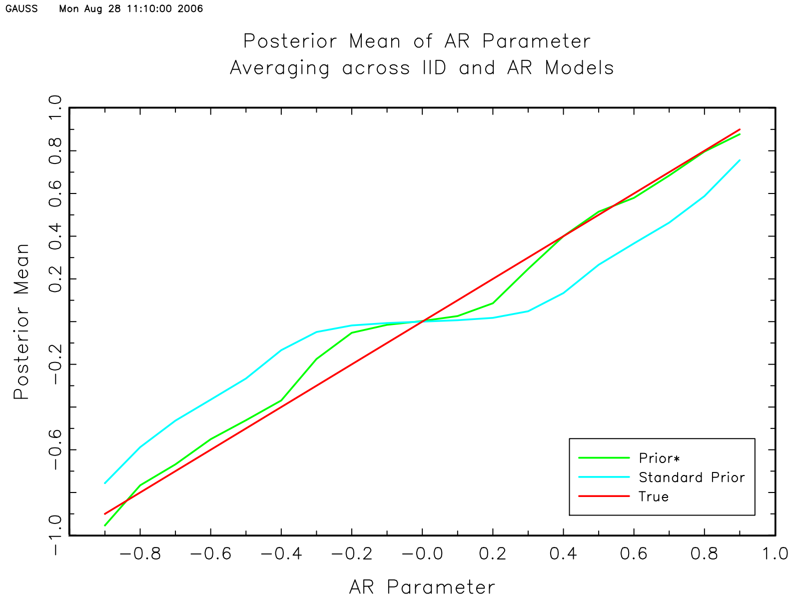
- $\underline{\mu} = 3$, $\underline{\lambda} = 0.4$, $T = 50$, $\underline{\theta} = 0.4$

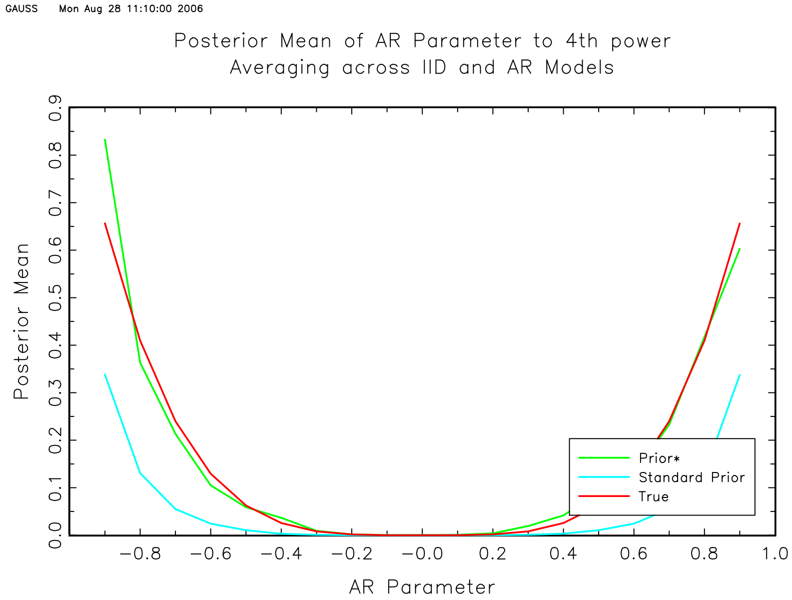
- Vary θ from -0.9 to $+0.9$
- Use population moments for sufficient statistics in updating prior to posterior
- Use posterior under standard prior for AR model to average importance weights of other priors

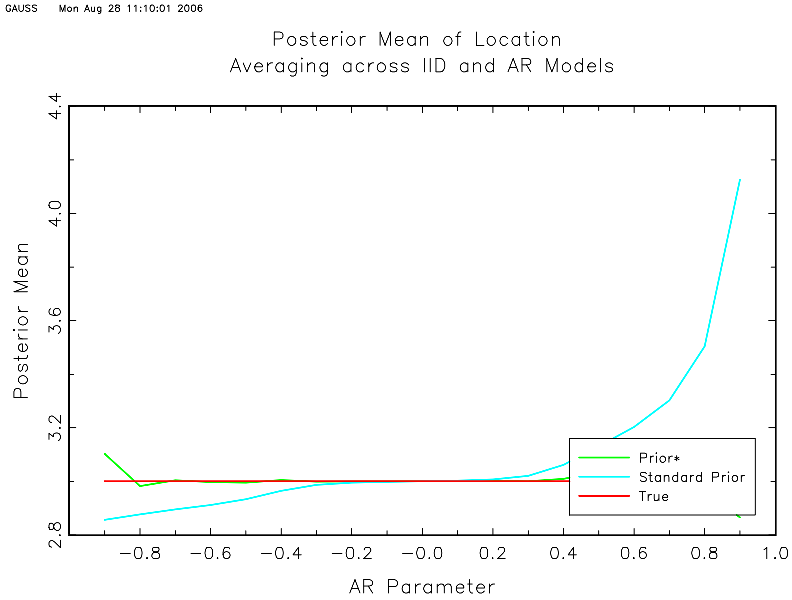

GAUSS Mon Aug 28 11:10:00 2006
Bayes Factor AR Model with Standard Prior vs IID Model
Log of Bayes Factor
−20
0
20
40
60
80
100
−0.8
−0.6
−0.4
−0.2
−0.0
0.2
0.4
0.6
0.8
1.0
AR Parameter


GAUSS Mon Aug 28 11:10:00 2006
Bayes Factor AR Model with Prior* vs Standard prior
Log of Bayes Factor
0
2
4
6
8
10
12
14
16
18
20
−0.8
−0.6
−0.4
−0.2
−0.0
0.2
0.4
0.6
0.8
1.0
AR Parameter


GAUSS Mon Aug 28 11:10:00 2006
Bayes Factor AR Model with Prior* vs Prior**
Log of Bayes Factor
0
20
40
60
80
100
120
−0.8
−0.6
−0.4
−0.2
−0.0
0.2
0.4
0.6
0.8
1.0
AR Parameter


GAUSS Mon Aug 28 11:10:00 2006
Posterior Mean of AR Parameter
Averaging across IID and AR Models
Posterior Mean
AR Parameter
Prior*
Standard Prior
True


GAUSS Mon Aug 28 11:10:00 2006
Posterior Mean of AR Parameter to 4th power
Averaging across IID and AR Models
Posterior Mean
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
−0.8
−0.6
−0.4
−0.2
−0.0
0.2
0.4
0.6
0.8
1.0
AR Parameter
Prior*
Standard Prior
True


GAUSS Mon Aug 28 11:10:01 2006
Posterior Mean of Location
Averaging across IID and AR Models
Posterior Mean
AR Parameter
4.4
4.0
3.6
3.2
2.8
−0.8
−0.6
−0.4
−0.2
−0.0
0.2
0.4
0.6
0.8
1.0
Prior*
Standard Prior
True

Conclusions

- Priors matter for Bayesian analysis of DSGE models
- Much more care needs to be taken with constructing priors
- Frank and Marco have introduced a method that allows more care based on a new approach to dummy observation priors