Levelling the Playing Field: Prior Choice and DSGE Model Comparisons

Marco Del Negro
Federal Reserve Bank of Atlanta

Frank Schorfheide*
University of Pennsylvania

This Draft: May 8, 2006

Abstract

What frictions are important in a DSGE model? In the Bayesian DSGE literature (e.g., Smets and Wouters 2003) this question is answered by computing the posterior odds of the model with and without the friction of interest. The prior distribution for the deep parameters plays a key role in these model comparisons. For some of the parameters, like the autocorrelations and the standard deviations of the structural shocks, the choice of the prior is not straightforward, and can make a difference for model comparison. We provide an approach for choosing the prior aimed at levelling the playing field for DSGE model comparisons.

JEL CLASSIFICATION: C32

KEY WORDS: Bayesian Analysis, DSGE Models,

---

*Correspondence: Marco Del Negro: Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St NE, Atlanta GA 30309-4470. Email: marcodelnegro@frbatlanta.org. Frank Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu.

The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

1 Introduction

2 A Simple Example

Consider the following two models, in which y_t is the observed endogenous variable and u_t is an unobserved shock process. In model $\mathcal{M}_1$, the u_t's are serially correlated:

$$\mathcal{M}_1 : \quad y_t = \frac{1}{\alpha}\mathbb{E}_t[y_{t+1}] + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t \sim iid(0, \sigma^2). \tag{1}$$

In model $\mathcal{M}_2$ the shocks are serially uncorrelated, but we introduce a backward-looking term ϕy_{t-1} on the right-hand-side as is often done in the New Keynesian Phillips Curve literature:

$$\mathcal{M}_2 : \quad y_t = \frac{1}{\alpha}\mathbb{E}_t[y_{t+1}] + \rho y_{t-1} + u_t, \quad u_t = \epsilon_t \sim iid(0, \sigma^2). \tag{2}$$

This example is taken from Lubik and Schorfheide (2006). Under $\mathcal{M}_1$ the equilibrium law of motion becomes

$$\mathcal{M}_1 : \quad y_t = \rho y_{t-1} + \frac{1}{1 - \rho/\alpha}\epsilon_t, \tag{3}$$

whereas under the 'backward looking' specification[1]

$$\mathcal{M}_2 : \quad y_t = \frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4\rho\alpha})y_{t-1} + \frac{2\alpha}{\alpha + \sqrt{\alpha^2 - 4\rho\alpha}}\epsilon_t. \tag{4}$$

Models $\mathcal{M}_1$ and $\mathcal{M}_2$ are observationally equivalent. The model with the 'backward looking' component is distinguishable from the purely 'forward looking' specification only under a strong *a priori* restriction on the exogenous component, namely $\rho = 0$. Although $\mathcal{M}_1$ and $\mathcal{M}_2$ will generate identical reduced form forecasts, the effect of changes in α on the law of motion of y_t is different in the two specifications.

Subsequently, we will illustrate the consequences of seemingly innocuous choices for prior distributions on posterior model odds. In the DSGE model literature, we can, broadly speaking, distinguish two types of parameters: 'deep' taste and technology parameters and 'auxiliary' parameters that determine the law of motion of the exogenous processes. Priors for the deep parameters are often chosen based on micro-econometric evidence, whereas the priors for the auxiliary parameters are either chosen arbitrarily or they are chosen based on some beliefs about the serial correlation and volatility of the endogenous variables. In our example we assume that α is a deep parameter whose prior has been specified based on micro-econometric evidence, whereas ρ and σ are auxiliary parameters. The baseline prior is denoted as Prior 1 and summarized in Table 1.

To start, we use the same prior distribution for models $\mathcal{M}_1$ and $\mathcal{M}_2$. We generate 200 draws from the prior predictive distribution of the sample autocorrelation and standard deviation of y_t. These draws are depicted in Panels (1,1) and (2,1) of Figure 1. Notice that according to model $\mathcal{M}_1$ the prior mean of the autocorrelation of y_t corresponds to the mean of ρ and is approximately 0.5. Under $\mathcal{M}_2$ the reduced-form autocorrelation coefficient is a nonlinear function of both ρ and α. It turns out that the prior mean of the predictive distribution of the autocorrelation is about 0.7. Hence, $\mathcal{M}_1$ and $\mathcal{M}_2$ have seemingly different implications for the observables.

We now generate a sample of $T = 80$ observations and compute the posterior for the two models under Prior 1. Draws from the posterior predictive distribution of the sample moments are plotted in Panels (1,2) and (2,2) of Figure 1. The intersection of the solid lines depict the actual sample moments. Given the fairly tight prior on α and ρ the estimated version of $\mathcal{M}_2$ still over-predicts the sample correlation of the data, whereas $\mathcal{M}_1$ captures it quite well. Log marginal data densities are reported in Table 2. The Bayes factor in favor of $\mathcal{M}_1$ is approximately 6.5. Whether this value provides a good summary of our post-data model uncertainty depends crucially on how confident we are about the specification of Prior 1. If the prior reflects our intrinsic uncertainty about the parameters then the Bayes factor is appropriate and we are ready to conclude that the 'backward-looking' specification is less desirable than the specification with serially correlated shocks. If, on the other hand, the prior for the auxiliary parameters was fairly arbitrary, then the Bayes factors might be regarded as misleading. After all, the two models are observationally equivalent. Therefore we might regard a Bayes factor of 1 a more reasonable result than a Bayes factor of 6.5.

We re-estimate model $\mathcal{M}_1$ under an alternative prior, denoted as Prior 2, that puts more weight on large values of ρ. The prior predictive distribution of the sample moments under this prior is depicted in Panel (3,1) of Figure 1. The draws are virtually indistinguishable from those obtained with model $\mathcal{M}_2$ and Prior 1. Indeed, under this modified prior distribution the Bayes factor of $\mathcal{M}_1$ versus $\mathcal{M}_2$ is essentially 1. Draws from the prior and posterior distribution of the $\mathcal{M}_1$ parameters under Priors 1 and 2 are depicted in Figures 2 and 3. Under Prior 1 the joint prior distribution of ρ and α is virtually indistinguishable from the posterior, whereas the distribution of σ is much more concentrated. Under Prior 2, the mean of ρ shifts from 0.73 (prior) to about 0.55 (posterior). The fairly tight Prior 2 prevents $\mathcal{M}_1$ from correctly capturing the fairly low autocorrelation in the data.

3 Adjusting Prior Distributions for Model Comparisons

Suppose we would like to compare structural models $\mathcal{M}_i$, $i = 1, \ldots, J$ with parameter vectors $\theta^{(i)}$. We split each parameter vector into two components: $\theta = [\theta_1', \theta_2]'$. Roughly speaking, θ_1 collects the "deep" parameters for which we can solicit prior distributions, say, based on micro-econometric evidence, and θ_2 is a sub-vector of auxiliary parameters for which we choose prior distributions such that the model implied autocovariances of the endogenous variables are "realistic" and comparable across models.

Our adjusted prior distributions are based on quasi-likelihood functions for the DSGE models. These quasi-likelihood functions are based on vector autoregressions (VAR) of the form

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma), \tag{5}$$

where y_t is an $n \times 1$ vector of observables. Let x_t be the $k \times 1$ vector $[1, y_{t-1}', \ldots, y_{t-p};]'$. We re-write the VAR in matrix notation as

$$Y = X\Phi + U. \tag{6}$$

Here Y is the $T \times n$ matrix with rows y_t', X is the $T \times k$ matrix with rows x_t', U is composed of u_t' and $\Phi = [\Phi_0, \Phi_1, \ldots, \Phi_p]'$. In general, we use Γ_{YY}, Γ_{YX} and Γ_{XX} to denote population autocovariances $\mathbb{E}[y_t y_t']$, $\mathbb{E}[y_t x_t']$, and $\mathbb{E}[x_t x_t']$, respectively. Sample autocovariances are signified by a hat, e.g., $\widehat{\Gamma}_{YY}$. If the population autovariances are calculated from a DSGE model conditional on a particular parameterization, we use the notation $\Gamma_{YY}^{(i)}(\theta^{(i)})$, $\Gamma_{YY}^{(i)}(\theta_1^{(i)}, \theta_2^{(i)})$ or, if no ambiguity arises, $\Gamma_{YY}^{(i)}$. We also define the VAR approximation of the DSGE model given by

$$\Phi_*^{(i)} = [\Gamma_{XX}^{(i)}]^{-1}\Gamma_{XY}^{(i)}, \quad \Sigma_*^{(i)} = \Gamma_{YY}^{(i)} - \Gamma_{YX}^{(i)}[\Gamma_{XX}^{(i)}]^{-1}\Gamma_{XY}^{(i)}. \tag{7}$$

Omitting the (i) superscripts, the adjusted prior distributions are of the form

$$\begin{aligned} & p(\theta|\Gamma_{YY}, \Gamma_{XY}, \Gamma_{XX}) \qquad (8)\\ \propto\; & \mathcal{L}(\theta|\Gamma_{YY}, \Gamma_{XY}, \Gamma_{XX}) \\ =\; & |\Sigma_*(\theta)|^{-(T^*+n+1)/2} \times \exp\left\{ -\frac{T^*}{2} tr\left[\Sigma_*(\theta)^{-1}(\Gamma_{YY} - 2\Phi_*(\theta)\Gamma_{XY} + \Phi_*'(\theta)\Gamma_{XX}\Phi_*(\theta) \right] \right\}, \end{aligned}$$

where the autocovariance matrices Γ_{YY}, Γ_{XY}, Γ_{XX} are either constructed from introspection, a pre-sample of actual observations, or an alternative candidate model. The prior (8) places low probability on values of θ for which the DSGE model implied autocovariances strongly differ from the Γ's. The larger T^*, the more concentrated the prior density. The approaches discussed subsequently mainly differ in their choice of Γ matrices. To simplify the notation, we abbreviate the conditioning "$\cdot|\Gamma_{YY}, \Gamma_{XY}, \Gamma_{XX}$" by "$\cdot|\Gamma$" and we use $\mathcal{L}(\theta_1, \theta_2|\Gamma)$ as shorthand for $\mathcal{L}([\theta_1', \theta_2']'|\Gamma)$.

3.1 Adjustments Based on Pre-samples

Throughout this subsection we are omitting the (i) superscript. We use a proper prior for the "deep" parameters θ_1 and employ the quasi-likelihood function to construct a prior for θ_2. The Γ matrices are sample autocovariances estimated from a pre-sample. Hence,

$$p(\theta_1, \theta_2) = c_0 \mathcal{L}(\theta_1, \theta_2|\hat{\Gamma})\pi(\theta_1)\pi(\theta_2). \tag{9}$$

Thus, starting from some initial distributions $\pi(\theta_1)$ and $\pi(\theta_2)$ the quasi-likelihood function is used to construct the prior for the actual estimation, $p(\theta_1, \theta_2)$. Here, $\pi(\theta_2)$ could be chosen to be diffuse. The constant c_0 ensures that the prior is properly normalized. Notice that the marginal prior of the "deep" parameters θ_1 is in general different from $\pi(\theta_1)$, that is,

$$p(\theta_1) = \int p(\theta_1, \theta_2) d\theta_2 = c_0 \left[\int \mathcal{L}(\theta_1, \theta_2|\hat{\Gamma}) p(\theta_2) d\theta_2 \right] \pi(\theta_1) \neq \pi(\theta_1). \tag{10}$$

In order to preserve $\pi(\theta_1)$ as the marginal prior distribution for the "deep" parameters we have to use the alternative prior

$$p_*(\theta_1, \theta_2) = c_1(\theta_1) \mathcal{L}(\theta_1, \theta_2|\hat{\Gamma})\pi(\theta_1)\pi(\theta_2), \tag{11}$$

where $c_1(\theta_1)$ is chosen such that

$$\frac{1}{c_1(\theta_1)} = \int \mathcal{L}(\theta_1, \theta_2|\hat{\Gamma})\pi(\theta_2) d\theta_2 \quad \text{for all} \quad \theta_1.$$

In practice prior $p_*(\theta_1, \theta_2)$ is more difficult to implement since it depends on the generally unknown normalization function $c_1(\theta_1)$. A more easily implementable prior is obtained by fixing θ_1 at its prior mean $\underline{\theta}_1$:

$$p_{**}(\theta_1, \theta_2) = c_1(\underline{\theta}_1) \mathcal{L}(\underline{\theta}_1, \theta_2|\hat{\Gamma})\pi(\theta_1)\pi(\theta_2). \tag{12}$$

3.2 Adjustments Based on a Benchmark DSGE

Alternatively, we can adjust the prior for a model $\mathcal{M}_2$ to make it comparable to a prior for benchmark model $\mathcal{M}_1$. The notation is slightly more delicate. We use $\Gamma^{(1)}(\theta^{(1)})$ to denote the implied autocovariances from model $\mathcal{M}_1$ conditional on the parameter $\theta^{(1)}$. The prior distribution for $\theta^{(1)}$ is $p(\theta^{(1)})$. Moreover, $\mathcal{L}(\theta|\Gamma)$ will denote the quasi-likelihood function associated with model $\mathcal{M}_2$ and $\pi(\theta_1)$, $\pi(\theta_2)$ some initial distribution for its parameters. For simplicity we will omit the (2)-superscript from both the likelihood function and the parameter vectors associated with $\mathcal{M}_2$. A natural choice for a prior distribution for the $\mathcal{M}_2$ parameters conditional on the $\mathcal{M}_1$ parameters would be

$$p_*(\theta_1, \theta_2|\theta^{(1)}) = c_1(\theta_1, \theta^{(1)})\mathcal{L}(\theta_1, \theta_2|\Gamma^{(1)}(\theta^{(1)}))\pi(\theta_1)\pi(\theta_2), \tag{13}$$

where

$$\frac{1}{c_1(\theta_1, \theta^{(1)})} = \int \mathcal{L}(\theta_1, \theta_2|\Gamma^{(1)}(\theta^{(1)}))\pi(\theta_2)d\theta_2 \quad \text{for all} \quad \theta_1,\ \theta^{(1)}.$$

As in the previous subsection, a practical difficulty is the calculation of the normalization function $c_1(\theta_1, \theta^{(1)})$ for every value of the conditioning parameters. A shortcut would be to condition on the prior means $\underline{\theta}^{(1)}$ and $\underline{\theta}_1$:

$$p_{**}(\theta_1, \theta_2) = c_1(\underline{\theta}_1, \underline{\theta}^{(1)})\mathcal{L}(\underline{\theta}_1, \theta_2|\Gamma^{(1)}(\underline{\theta}^{(1)}))\pi(\theta_1)\pi(\theta_2). \tag{14}$$

4 The Simple Example Revisited

We regard α as "deep" parameter θ_1 and interpret ρ and σ as auxiliary parameters θ_2. In this simple example, the quasi-likelihood function coincides with the actual likelihood functions for models $\mathcal{M}_1$ and $\mathcal{M}_2$. We will focus on the adjustments based on a pre-sample and assume that $\Gamma_{XX} = \Gamma_{YY} = 4$ and $\Gamma_{XY} = 0.7$. Draws from the adjusted prior distributions are plotted in Figure 4. While the priors for α (by construction) and σ are nearly identical across the two models, the priors for ρ are very different. However, the prior predictive distribution for the sample autocorrelation and standard deviation is now very similar under Prior 3 as can be seen from Panels (1,1) and (1 2) of Figure 5. Consequently the Bayes factor for the two models is very close to one, 1.6 to be exact. Hence, the model odds are not distorted due to a careless choice of prior distribution, as under Prior 1.

Table 1: EXAMPLE 1 – PRIOR DISTRIBUTIONS

Name	Domain		Prior 1		Prior 2	
		Density	Para (1)	Para (2)	Para (1)	Para (2)
α	$I\!R^+$	Gamma	2.00	0.10	2.00	0.10
ρ	$[0,1)$	Beta	0.50	0.05	0.73	0.10
σ	$I\!R^+$	InvGamma	1.00	4.00	1.00	4.00

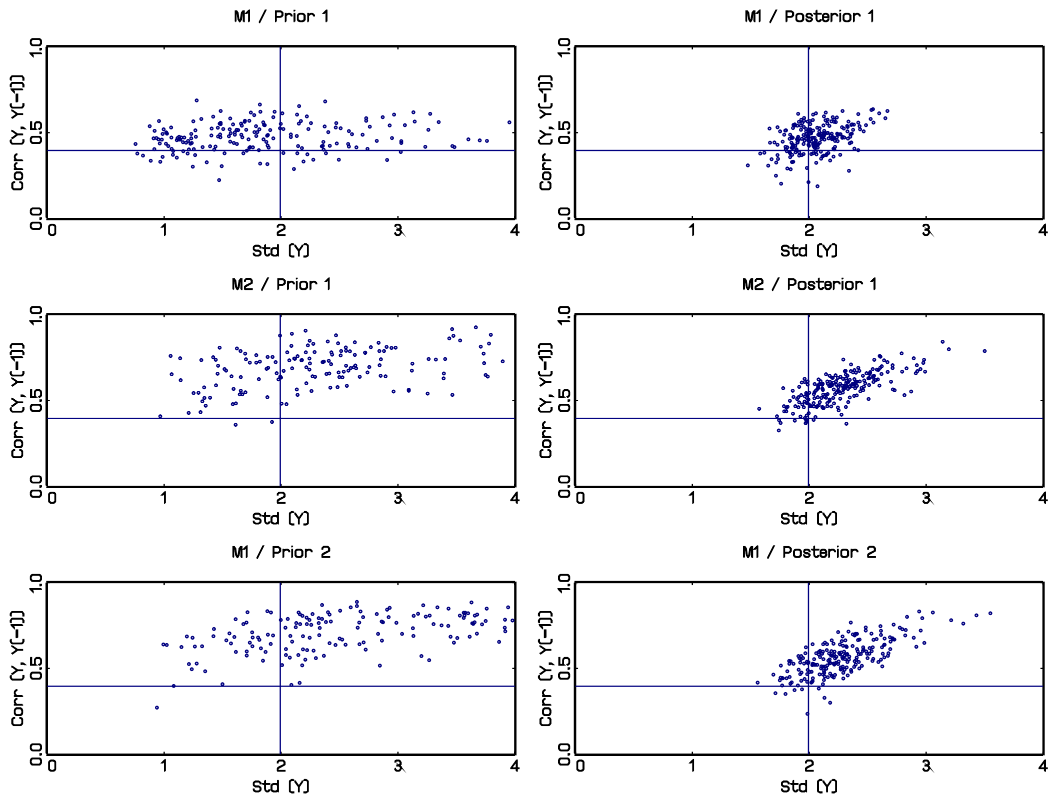
Notes: Para (1) and Para (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; s and ν for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu,s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region.

Table 2: EXAMPLE 1 – LOG MARGINAL DATA DENSITIES

Specification	$\ln p(Y)$
Model $\mathcal{M}_1$, Prior 1	-161.27
Model $\mathcal{M}_1$, Prior 2	-163.21
Model $\mathcal{M}_2$, Prior 1	-163.16

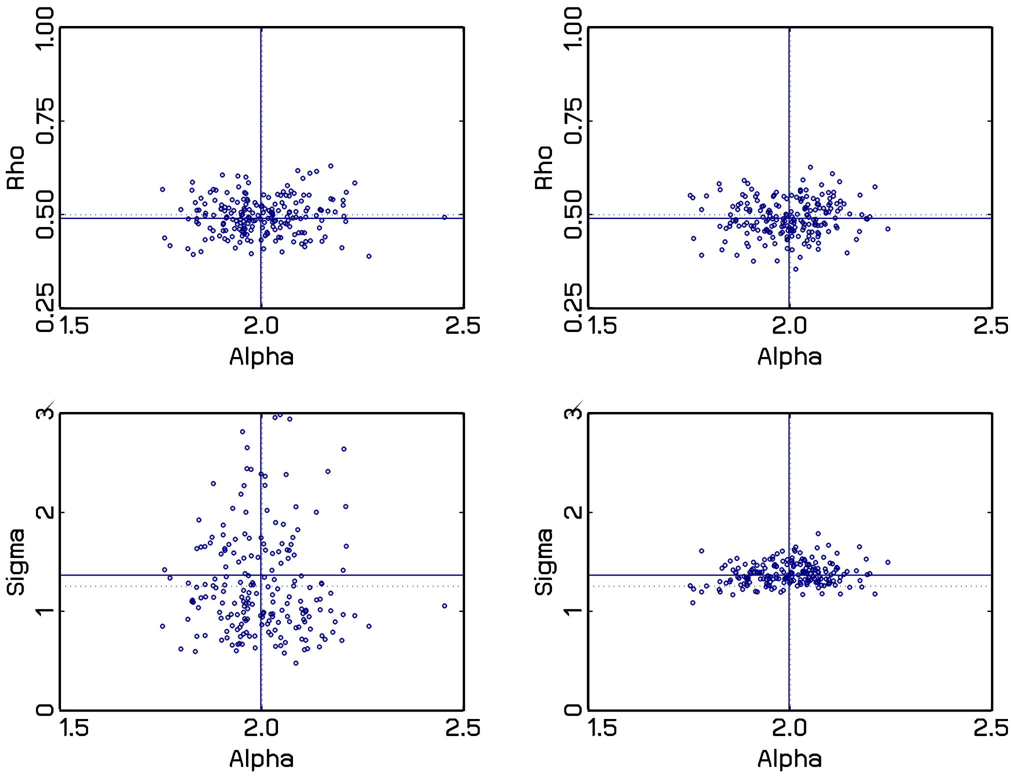
Notes: We truncate the prior distribution of α, ρ, σ at the boundary of the indeterminacy region. The marginal data densities have been adjusted accordingly.

Figure 1: Example 1 - Predictive Distributions of Sample Moments



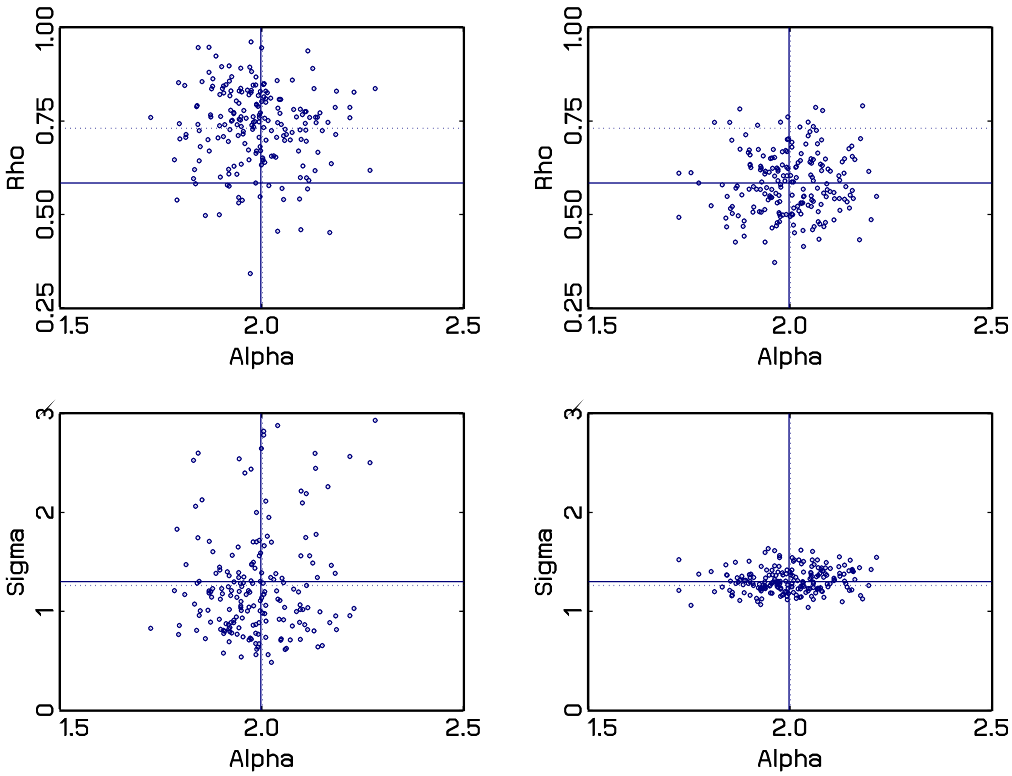
Notes: Each panel depicts 200 draws from predictive distribution for the sample autocorrelation and standard deviation. Intersection of solid lines signifies the actual sample moment.

Figure 2: Example 1 – Parameter Draws from Model $\mathcal{M}_1$, Prior 1



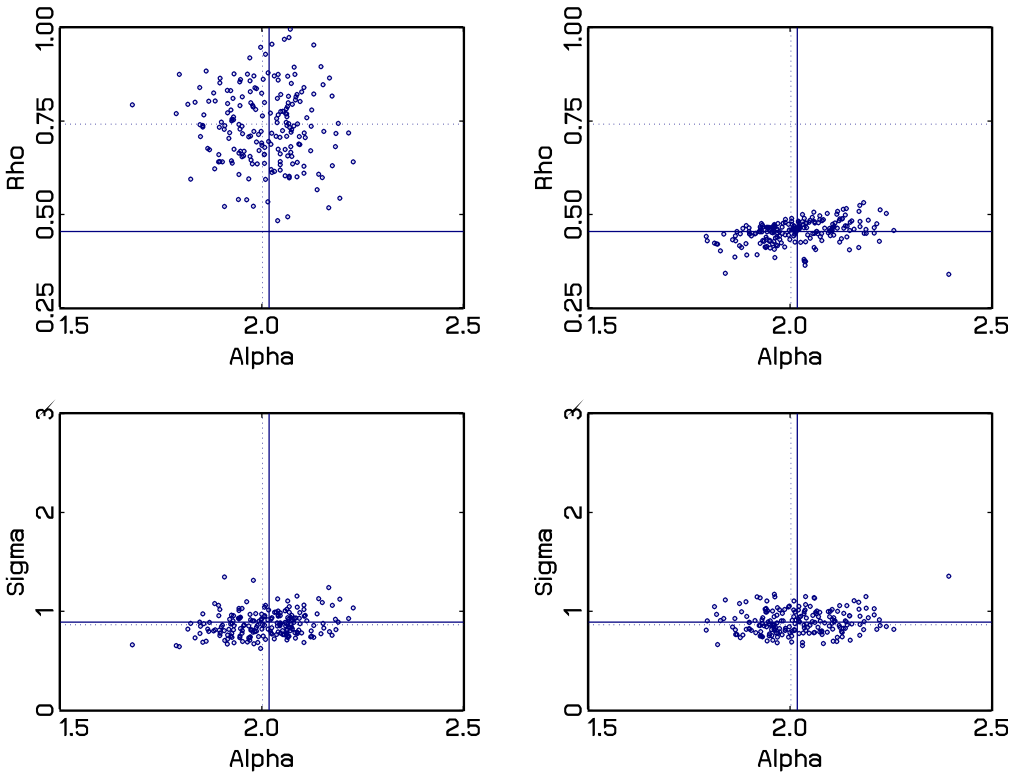
Notes: Left (right) column of panels depicts 200 draws from prior (posterior) distribution. Intersection of dotted (solid) lines indicates prior (posterior) means.

Figure 3: Example 1 – Parameter Draws from Model $\mathcal{M}_1$, Prior 2



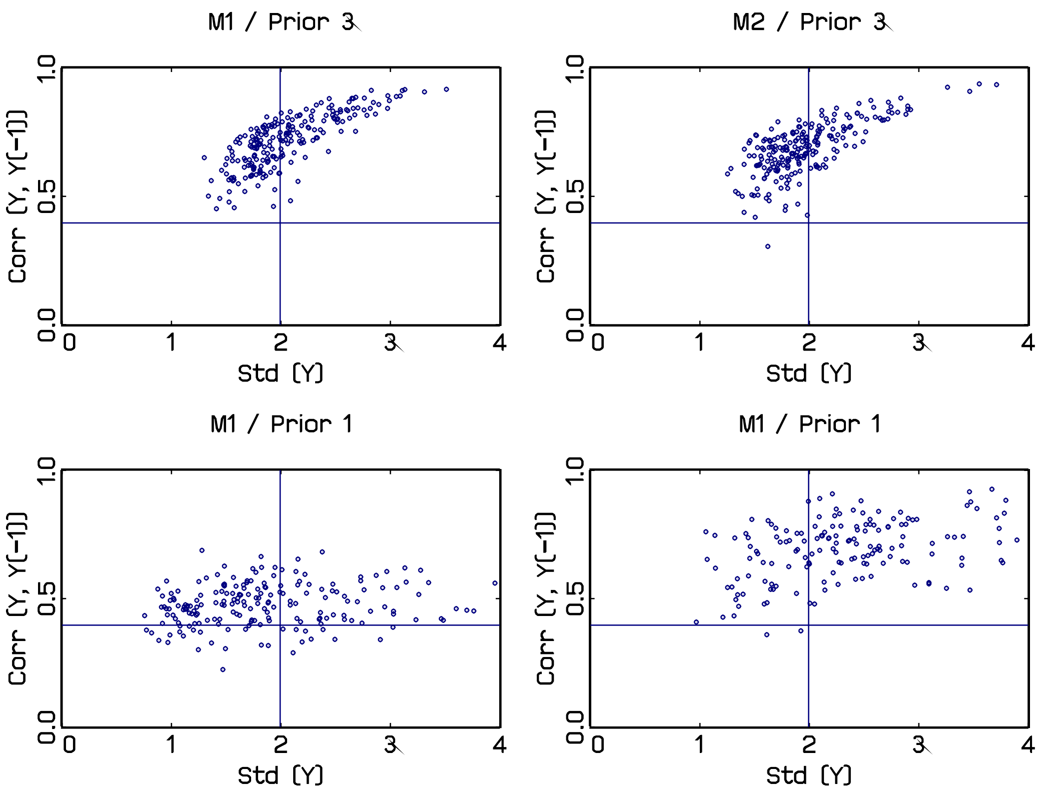
Notes: Left (right) column of panels depicts 200 draws from prior (posterior) distribution. Intersection of dotted (solid) lines indicates prior (posterior) means.

Figure 4: EXAMPLE 1 – PARAMETER DRAWS FROM MODEL $\mathcal{M}_1$, Prior 3



Notes: Left (right) column of panels depicts 200 draws from $\mathcal{M}_1$ ($\mathcal{M}_2$) prior distribution. Intersection of dotted (solid) lines indicates prior means.

Figure 5: Example 1 - Predictive Distributions of Sample Moments



Notes: Each panel depicts 200 draws from predictive distribution for the sample autocorrelation and standard deviation. Intersection of solid lines signifies the actual sample moment.