Journal of Monetary Economics 59 (2012) 581–597

Contents lists available at SciVerse ScienceDirect

Journal of Monetary Economics

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

Measuring prior sensitivity and prior informativeness in large Bayesian models

Ulrich K. Müller*

Princeton University, Department of Economics, Princeton, NJ 08544, United States

ARTICLE INFO

Article history:
Received 30 September 2011
Received in revised form
16 August 2012
Accepted 14 September 2012
Available online 25 September 2012

ABSTRACT

In large Bayesian models, such as modern DSGE models, it is difficult to assess how much the prior affects the results. This paper derives measures of prior sensitivity and prior informativeness that account for the high dimensional interaction between prior and likelihood information. The basis for both measures is the derivative matrix of the posterior mean with respect to the prior mean, which is easily obtained from Markov Chain Monte Carlo output. We illustrate the approach by examining posterior results in the small model of Lubik and Schorfheide (2004) and the large model of Smets and Wouters (2007).

1. Introduction

Especially in empirical macroeconomics, Bayesian inference has become a popular estimation method. For instance, the rapidly growing empirical literature of dynamic stochastic general equilibrium (DSGE) models is largely Bayesian (Smets and Wouters, 2003; Fernández-Villaverde and Rubio-Ramírez, 2007; Justiniano and Primiceri, 2008, etc.—see An and Schorfheide, 2007 for a survey), and also structural and reduced form time varying parameter models are often approached with Bayesian techniques (Kim and Nelson, 1999; Primiceri, 2005; Cogley and Sargent, 2005; Sims and Zha, 2006, among others). These models contain a moderate or large number of unknown parameters, requiring the specification of a corresponding prior. The empirical conclusions are then typically based on the center and spread of the resulting posterior.

At least to some extent, the results depend on the prior. This is, of course, not a problem as such—one key advantage of the Bayesian approach is that it allows the (coherent and optimal) incorporation of a priori information, which is useful and maybe even necessary in some large scale macroeconomic applications. It is nevertheless helpful for the interpretation of the results to try to disentangle the role of prior and likelihood information.

This task is substantially harder when there are many unknown parameters. While one might often have a reasonably good sense of what constitutes an informative marginal prior for an individual parameter, the combined effect of these (typically independent) marginal priors is more difficult to think about: The likelihood information about different parameters can be far from independent, so that marginal posterior distributions critically depend on the interaction of the likelihood with the whole prior. And with a high dimensional parameter space, it is simply not feasible to plot or otherwise describe in detail the shape of the likelihood, let alone to leave it to the reader to combine the likelihood with his or her prior beliefs.

* Tel.: +609 258 4026; fax +609 258 6419.
E-mail address: umueller@princeton.edu

0304-3932/$ - see front matter
http://dx.doi.org/10.1016/j.jmoneco.2012.09.003

Current standard practice is to provide two sets of numbers: (i) comparisons of marginal prior and posterior distributions; (ii) comparisons of posterior results over a small number of prior variations, such as an increase of the prior variance on all parameters. But these statistics are not necessarily very informative about the relative importance of the prior and likelihood. As a simple illustration, consider a model with a two-dimensional parameter $\theta=(\theta_1,\theta_2)'$, with θ_1 being the parameter of interest. Suppose the likelihood has the shape of a steep ridge along $\theta_1=\theta_2$. Thus, the data tells us that θ_1 is approximately equal to θ_2, but without further information, it does not pin down its value. Now a tight prior on the nuisance parameter θ_2 effectively selects a point on the ridge, and thus leads to a tight posterior for θ_1. Thus, the marginal posterior on θ_1 can be very different from the marginal prior, even though the data alone contains little information about θ_1. Furthermore, only particular variations of the prior on $(\theta_1,\theta_2)'$ will reveal the full extent of this effect, and one can construct similar higher-dimensional examples where a common increase in the prior variance only leads to a very moderate change in the marginal posterior distribution of the parameter of interest.

The goal of this paper is thus to develop additional, easily computed statistics that help to clarify the role of prior and likelihood information in Bayesian inference of large models. We ask two related questions. First, how sensitive are the posterior results to variations in the prior? Second, what is the relative importance of prior and likelihood information for individual parameters, that is how informative is the multivariate prior for individual parameters?

Both questions may be approached by analyzing how the posterior mean varies locally as a function of the prior mean. The idea is that the mean is a measure for the center of a distribution, so that the prior mean reflects the a priori information about predominant parameter values, and variations of the posterior mean are a key aspect of posterior sensitivity. What is more, if the likelihood is very peaked relative to the prior (so that the prior is not very informative compared to the data) then the posterior is dominated by the likelihood, and variations of prior means will have almost no impact on posterior means. In contrast, with an approximately flat likelihood (so the prior is relatively informative), the posterior is similar to the prior, and prior mean changes are pushed through one-for-one to the posterior mean. It thus makes sense to consider the *derivative* of the posterior mean with respect to the prior mean as a starting point for both questions.

To make this operational one must take a stand on how exactly the prior distribution changes along with its mean. The suggestion is to embed the baseline prior distribution in an exponential family. This choice has a certain theoretical appeal, as discussed below. But what is more, this embedding leads to a simple expression for the derivative matrix

$$J=\Sigma_\pi\Sigma_p^{-1} \tag{1}$$

where Σ_p and Σ_π are the prior and posterior covariance matrices of $\theta=(\theta_1,\ldots,\theta_k)'$, respectively, and the ith row of J contains the partial derivatives of the posterior mean of θ_i with respect to the prior mean of θ. It is thus computationally trivial to obtain the derivative matrix J from standard MCMC output.

More concretely, suppose the scalar parameter of interest is $\nu'\theta$. The derivative matrix J can then straightforwardly be used to compute a prior sensitivity measure PS that approximates the largest change of the posterior mean that can be induced by changing the prior mean by the multivariate analogue of one prior standard deviation:

$$\mathrm{PS}=\max_{\sqrt{\alpha'\Sigma_p^{-1}\alpha}=1}\nu'J\alpha=\sqrt{\nu'\Sigma_\pi\Sigma_p^{-1}\Sigma_\pi\nu}. \tag{2}$$

Furthermore, as argued above, the derivative matrix is also a useful starting point for the construction of a prior informativeness measure PI $\in[0,1]$ that summarizes the relative amount of prior information in the posterior. For models with a single scalar parameter, the suggested measure PI is simply equal to PI $=\min(J,1)$ (note that it is possible, even though somewhat special for the posterior variance to be larger than the prior variance). In models with a vector parameter and scalar parameter of interest $\nu'\theta$, we impose axiomatic requirements on potential mappings from J (and Σ_p) to the unit interval to determine a measure that provides a sense for the fraction of prior information in the posterior results. Specifically, as long as the largest eigenvalue of J is smaller than unity, PI $\in[0,1]$ is given by

$$\mathrm{PI}=1-\frac{\nu'\Sigma_p\nu}{\nu'\Sigma_p(\Sigma_p-\Sigma_\pi)^{-1}\Sigma_p\nu}. \tag{3}$$

The measure PI can also usefully be thought of as measuring ‘‘identification strength’’ (with large values of PI indicating weak identification), although ‘‘relative informativeness’’ of prior and likelihood seems a more accurate designation. Both measures can also be applied to functions of parameters, such as impulse responses.

As an illustration, consider posterior results about the calvo probability in the labor market ξ_w in Smets and Wouters' (2007) DSGE model with 36 estimated parameters (a detailed discussion of this application is in Section 4.2). The prior on ξ_w is Beta with mean 0.50 and standard deviation 0.10, and the posterior has mean 0.70 and standard deviation 0.066. The derivative of the posterior mean of ξ_w relative to the prior mean of ξ_w is 0.43, so that a one prior standard deviation change of the prior mean leads to a change in the posterior mean of about 0.043. But the prior on the other parameters also has a substantial influence on the posterior results for ξ_w: for instance, the partial derivative of the posterior mean of ξ_w with respect to the prior means of the elasticity of labor supply with respect to real wage σ_l, the calvo probability in the goods market ξ_p and the MA(1) parameter μ_w in the wage markup shock are 0.036, 0.12 and 0.067, respectively, implying that a one prior standard deviation change of the prior mean of these parameters leads to a change of the posterior mean of ξ_w of approximately 0.027, 0.012 and 0.013, respectively. These cross effects are taken into account by the measure PS, which equals 0.055 for ξ_w. Thus, varying the prior on the 36 parameters by the multivariate analogue of one prior standard deviation can induce a change in the posterior mean of ξ_w that is nearly as big as the posterior standard deviation of 0.066. Furthermore, the ratio of posterior to prior variance of ξ_w yields 0.43, which may suggest that the prior contributes less than half of the posterior information. But taking again the cross effects of the multivariate prior into account, one obtains instead $\text{PI}=0.75$ for ξ_w, pointing to an even more prominent role of the multivariate prior information for the posterior results on ξ_w.

Section 4 contains a more detailed application of the two measures to a small and larger scale DSGE estimation. Substantively, the analysis shows that in the three equation DSGE model of Lubik and Schorfheide (2004), U.S. postwar data contains little information about the coefficient of risk aversion and the slope of the Phillips curve. Posterior results about other model parameters, as well as impulse responses and variance decompositions are highly sensitive to the prior about these two parameters. In the Smets and Wouters (2007) model, the prior is very informative for many of the structural parameters, while the parameters describing the shock processes are much better pinned down by data information, at least conditional on the prior information on the steady state inflation rate. Interestingly, key impulse responses and variance decompositions inherit the relatively moderate role of prior information from the shock parameters. At least in this data set, impulse responses and variance decompositions are not mainly driven by the prior.

Although the measures PS and PI are based on the same derivative matrix and may thus be considered a natural pair, the two statistics are related to quite distinct literatures. On the one hand, the prior sensitivity measure PS belongs to the large Bayesian robustness literature that considers the effect of local changes of the prior distribution. Berger (1994), Gustafson (2000) and Sivaganesan (2000) provide overviews and references. More specifically, Basu et al. (1996), Geweke (1999) and Perez et al. (2006) study the local sensitivity of the posterior mean in a parametric class of priors, which amounts to the computation of the posterior mean derivative with respect to the prior hyperparameter. Millar (2004) observes that if the scalar marginal prior distribution is in the exponential family, then the derivative with respect to the prior mean is simply given by the ratio of the posterior to prior variance. The measure PS thus merely amounts to an (mathematically straightforward) extension and specialization of these previous results. The derivation of the statistic PS is still useful, though, as it provides a computationally trivial and easily interpreted default scalar measure for the local prior sensitivity of a particular scalar parameter of interest (or real valued function of the underlying parameter vector).

The prior informativeness measure PI, on the other hand, does not seem to have a close counterpart in the literature. Poirier (1998) observes that lack of identification of some parameters entails that their conditional posterior distribution is always the same as in the prior, but not necessarily their marginal posterior distribution. The measure PI, however, does not take identification or lack thereof as a given, but summarizes the *amount* of likelihood information about a specific parameter in a high dimensional model, relative to the prior information. This property also distinguishes it from the recent literature that, initiated by Canova and Sala (2009), analyzes identification of DSGE models, such as Iskrev (2010) or Komunjer and Ng (2009). The differences to this literature go further, though, as the frequentist notion of identification (or identifiability) as defined by Rothenberg (1971) is neither necessary nor sufficient for low prior informativeness as measured by PI. Roughly speaking, identifiability is most useful for assessing whether data can *potentially* provide information about model parameters, whereas the likelihood based measure PI describes how much information is contained *in a given data set* relative to the prior information—see Section 3.4 for further details. Leamer (1973) discusses interpretational challenges when multivariate prior and sample information do not align in the context of the normal linear regression model. Lindley (1955) defined entropy based measures of data informativeness, which were further investigated by, e.g., Goel and DeGroot (1981), Goel (1983), Soofi (1990) and Ebrahimi et al. (1999). The appeal of PI derived here relative to this existing literature is its applicability to general models and priors, its computational simplicity and its tight connection to the readily interpretable derivative matrix.

The remainder of the paper is organized as follows. Section 2 derives the measures PS and PI. Section 3 discusses inequalities for PS and PI, analogue measures for functions of the original parameters, conditioning on a subset of the prior information, and a detailed comparison with Rothenberg's (1971) definition of frequentist identifiability. Section 4 contains the empirical results for the Lubik and Schorfheide (2004) and the Smets and Wouters (2007) DSGE models. Section 5 concludes.

2. Derivation of measures

This section contains the derivation of the prior sensitivity and prior informativeness measures. The first subsection is concerned with models with a scalar parameter θ, and the following two subsections discuss models with a vector valued parameter θ.

2.1. Scalar parameter

Denote the baseline prior density for the scalar model parameter θ by p, with mean $\mu_p = E_p[\theta]$ and variance $\sigma_p^2 = E_p[(\theta-\mu_p)^2]$. Here and below subscripts of the expectation indicate the measure of integration. The posterior density π is derived from p and the likelihood function l via $\pi(\theta) = p(\theta)l(\theta)/\int p(h)l(h)\,dh$. Assume that the posterior distribution has finite variance σ_π^2.

Now embed the baseline prior density in a family p_α with $p_0=p$, score function $s_\alpha(\theta)=d\log p_\alpha(\theta)/d\alpha$ and $d(\int\theta p_\alpha(\theta)\,d\theta)/d\alpha=1$, so that at least for values of α close to zero, the prior mean is approximately $\int\theta p_\alpha(\theta)\,d\theta\approx\mu_p+\alpha$. The posterior mean as a function of α then equals $\mu_\pi(\alpha)=\int\theta p_\alpha(\theta)l(\theta)\,d\theta/\int p_\alpha(\theta)l(\theta)\,d\theta$, and under weak regularity conditions that justify differentiation under the integral (see, for instance, Perez et al., 2006 for details),

$$\begin{aligned}\frac{d\mu_\pi(\alpha)}{d\alpha}\Big|_{\alpha=0}&=\frac{\int\theta s_0(\theta)p_0(\theta)l(\theta)\,d\theta}{\int p_0(\theta)l(\theta)\,d\theta}-\frac{\left(\int\theta p_0(\theta)l(\theta)\,d\theta\right)\left(\int s_0(\theta)p_0(\theta)l(\theta)\,d\theta\right)}{\left(\int p_0(\theta)l(\theta)\,d\theta\right)^2}\\&=E_\pi[(\theta-E_\pi[\theta])s_0(\theta)]\end{aligned}\tag{4}$$

which is recognized as the posterior covariance between θ and $s_0(\theta)$. As explained in the introduction, the idea is to use this derivative as a basis for measuring both prior sensitivity and prior informativeness.

In general, $s_0(\theta)$ is not necessarily a monotone function of θ. For instance, a family of Gamma priors with mean $\mu_p+\alpha$ and fixed variance always has $ds_0(\theta)/d\theta<0$ for large θ. If the likelihood concentrates on such values, then increasing the prior mean leads to a *smaller* posterior mean. Thus, even though one might think of the embedding of a baseline Gamma prior into a family of Gamma priors with different means but equal variances as the most natural one, it leads to the counterintuitive result that increasing the prior mean leads to a *decrease* of the posterior mean whenever the likelihood strongly favors large values of θ.

More generally, also nonlinearities in $s_0(\theta)$ are potentially unattractive: Suppose a model with a given prior is estimated on two different data sets, and the two posteriors happen to be equally informative as measured by σ_π^2. One might well demand that a measure of prior importance should correspondingly come to a similar conclusion about the relative importance of prior and likelihood in either estimation exercise. With $s_0(\theta)$ nonlinear, however, the derivative in (4) can take on very different values. A generic equality can only be guaranteed if $s_0(\theta)$ is linear in θ, as in the exponential family embedding

$$p_\alpha(\theta)=p(\theta)\exp[\alpha(\theta-\mu_p)/\sigma_p^2-C(\alpha)]\tag{5}$$

with cumulant function $C(\alpha)=\log\int p(\theta)\exp[\alpha(\theta-\mu_p)/\sigma_p^2]\,d\theta$ and $s_0(\theta)=(\theta-\mu_p)/\sigma_p^2$. This is a well defined family of densities for small enough $|\alpha|$ whenever the moment generating function of p exists, at least in an open interval containing zero.[1] By construction, the derivative of the prior mean relative to α in the family (5) is equal to unity at $\alpha=0$, so that locally, α has the interpretation of a mean shift of the baseline prior. This holds also globally for a Gaussian baseline prior, as p_α is then simply the density of $\mathcal{N}(\mu_p+\alpha,\sigma_p^2)$. More generally, the derivative of the variance of p_α at $\alpha=0$ under (5) equals $E_p[(\theta-\mu_p)^3]/\sigma_p^2$, so that the percentage change in the variance that is induced by changing the mean by a fraction of the prior standard deviation is small as long as the baseline prior is not too skewed. Fig. 1 shows p_α for two quite non-Gaussian baseline Beta(3/2,3) and Gamma(4,1/2) priors and $\alpha\in\{-\sigma_p,-\frac{1}{2}\sigma_p,\frac{1}{2}\sigma_p,\sigma_p\}$. These values of α induce prior mean shifts of $\{-0.73\sigma_p,-0.43\sigma_p,0.55\sigma_p,1.12\sigma_p\}$ and $\{-0.67\sigma_p,-0.40\sigma_p,0.67\sigma_p,2\sigma_p\}$, respectively. At least for moderate values of α, the mean shift interpretation thus remains a reasonable approximation.

Since (5) implies $s_0(\theta)=(\theta-\mu_p)/\sigma_p^2$, (4) reduces to

$$\frac{d\mu_\pi(\alpha)}{d\alpha}\Big|_{\alpha=0}=J=\frac{\sigma_\pi^2}{\sigma_p^2}\tag{6}$$

so that the derivative of the posterior mean relative to the prior mean simply becomes the ratio of posterior and prior variance (cf. (11) in Millar, 2004).

The prior sensitivity can now usefully be measured by the linear approximation to the change of the posterior mean that can be induced by increasing the prior mean by one prior standard deviation. With (6), this results in

$$\mathrm{PS}=\sigma_pJ=\frac{\sigma_\pi^2}{\sigma_p}.\tag{7}$$

Additionally, the derivative J can also be used directly to measure the prior informativeness: When the likelihood perfectly pins down θ and the relative prior informativeness is zero, then prior mean changes leave the posterior mean unchanged, and $J=0$. In the other extreme, with a completely flat likelihood all information stems from the prior, the posterior mean is identical to the prior mean, and $J=1$. Values of J above unity are possible, though, as the posterior variance can be larger than the prior variance. This poses no problem for the derivative interpretation of PS, but "more than 100% prior importance" is much less compelling for a prior informativeness measure, so define

$$\mathrm{PI}=\min(J,1).\tag{8}$$

Values of PI between zero and one may be thought of as a numerical measure for the relative importance of prior information for the posterior results. More precisely, suppose that both the prior log-density and the log-likelihood are quadratic in θ, i.e. $l(\theta)\propto\exp[-\frac{1}{2}(\theta-\mu_l)^2/\sigma_l^2]$ (as arising from observing θ with Gaussian noise of variance σ_l^2), and

[1] If p is such that the moment generating function does not exist (as, for instance, for the inverse Gamma distribution), an alternative, less familiar embedding is given by $p_\alpha(\theta)=2c(\alpha)p(\theta)/(1+\exp[-2\alpha(\theta-\mu_p)/\sigma_p^2])$ where $c(\alpha)>0$ ensures that $\int p_\alpha(\theta)\,d\theta=1$ for all α. This embedding always exists as long as p has two moments, and also leads to $s_0(\theta)=(\theta-\mu_p)/\sigma_p^2$, and therefore to an identical expression for $d\mu_\pi(\alpha)/d\alpha|_{\alpha=0}$.

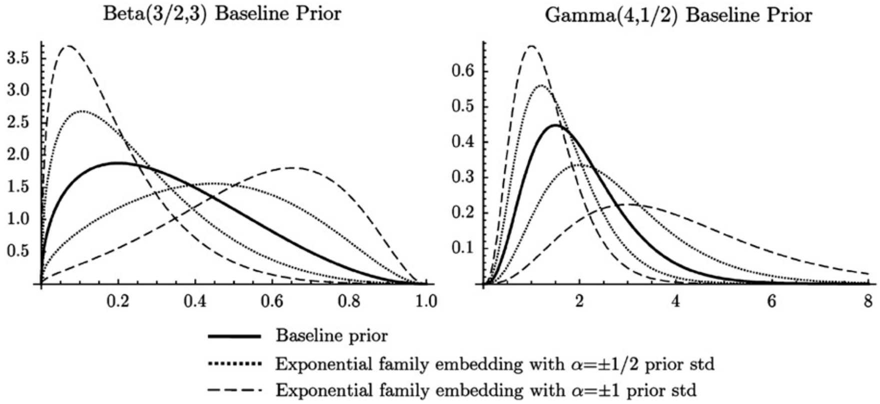

Fig. 1. Exponential family embedding of a Beta and Gamma baseline prior.
Notes: This figure illustrates how a baseline prior is transformed in the exponential family embedding. The exponential family is indexed by α, with $\alpha=0$ recovering the baseline prior, and positive and negative values of α induce positive and negative mean shifts of the prior distribution, respectively.

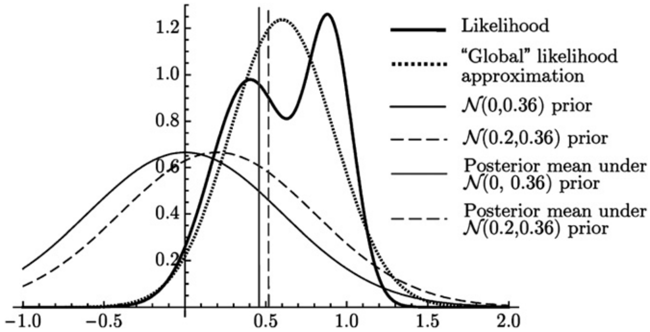

Fig. 2. Numerical example of a prior mean shift and its impact on the posterior mean.
Notes: This figure shows how the prior mean shift from 0.0 to 0.2 affects the posterior mean for an example likelihood. A very similar mean shift would be obtained for the "global" log-quadratic approximation to the likelihood. This demonstrates that the mean shift measure is not strongly affected by local features of the likelihood.

$p\sim\mathcal{N}(\mu_p,\sigma_p^2)$, so that $p_\alpha\sim\mathcal{N}(\mu_p+\alpha,\sigma_p^2)$ under (5). By a standard calculation, the posterior mean then satisfies

$$\mu_\pi(\alpha)=w(\mu_p+\alpha)+(1-w)\mu_l \quad \text{with } w=\frac{\sigma_p^{-2}}{\sigma_p^{-2}+\sigma_l^{-2}}. \tag{9}$$

With the precisions σ_p^{-2} and σ_l^{-2} measuring the amount of information in the prior and likelihood, respectively, we thus obtain a more explicit interpretation of $\mathrm{PI}=d\mu_\pi(\alpha)/d\alpha=w$ as the fraction of prior information for the posterior mean.

If the prior log-density and log-likelihood are only approximately quadratic, then this interpretation will typically remain a useful approximation. Fig. 2 provides an illustration with $p_\alpha\sim\mathcal{N}(\mu_p+\alpha,0.36)$ with $\mu_p=0$ and a likelihood arising from observing $Y=0.6$, where $Y\sim\mathcal{N}(\theta-0.3,0.02)$ with probability 0.4 and $Y\sim\mathcal{N}(\theta+0.2,0.06)$ with probability 0.6, so that $E[Y]=\theta$. The overall information content about θ for this draw of Y is reasonably well approximated by the quadratic log-likelihood with mean μ_l and variance σ_l^2 computed from the likelihood normalized to integrate to one, so that μ_l and σ_l^2 are the posterior mean and variance that one would obtain from a completely flat prior. This is depicted as the "global" likelihood approximation in Fig. 2. Now w in (9) with this value of σ_l^2 evaluates to $w=0.224$, and $\mathrm{PI}=\sigma_\pi^2/0.36=0.249$ (with a range of $\mathrm{PI}\in[0.193,0.252]$ for $-1\le\mu_p\le 1$). Thus, even though the log-likelihood is far from quadratic, PI gives a good sense of its overall informativeness in this example. Intuitively, the posterior mean $\mu_\pi(\alpha)$ is a weighted average of the likelihood, and thus reflects its global shape. Other plausible measures for the informativeness of the data, such as the curvature of the likelihood at its peak, merely summarize its local characteristics, which would be quite misleading in the example of Fig. 2.

2.2. Prior sensitivity with a vector parameter

Now let $\theta=(\theta_1,\ldots,\theta_k)'$ be $k\times 1$, and embed the prior density p with mean μ_p and covariance matrix Σ_p in the exponential family

$$p_\alpha(\theta)=\exp[\alpha'\Sigma_p^{-1}(\theta-\mu_p)-C(\alpha)]p(\theta) \tag{10}$$

indexed by the $k \times 1$ vector α and cumulant function $C(\alpha) = \log \int \exp[\alpha' \Sigma_p^{-1}(\theta - \mu_p)]p(\theta)\,d\theta$, which exists for small enough $\|\alpha\|$ whenever the moment generating function of p exists, at least in a neighborhood of zero.[2] By construction, $d(\int \theta p_\alpha(\theta)\,d\theta)/d\alpha'|_{\alpha=0} = I_k$, so that at least local to $\alpha = 0$, p_α is a family of priors with prior mean equal to $\mu_p + \alpha$.

Let $\mu_\pi(\alpha)$ be the posterior mean of θ under the prior (10). The $k \times k$ derivative matrix then is

$$J = \frac{\partial \mu_\pi(\alpha)}{\partial \alpha'}\Big|_{\alpha=0} = \Sigma_\pi \Sigma_p^{-1}. \tag{11}$$

The jth column of J contains the partial derivatives of the posterior mean of θ with respect to the prior mean of θ_j.

In Sections 2.2 and 2.3, the scalar parameter of interest is always the linear combination $\nu'\theta$ (which reduces to θ_j with ν the jth column of I_k). The derivative vector of the posterior mean of $\nu'\theta$ is $\nu' J$. Thus, if the magnitude of a prior mean change α is measured in the Mahalanobis metric $\sqrt{\alpha' \Sigma_p^{-1} \alpha}$, then the local approximation to the largest change of the posterior mean of θ that can be induced by a unit change of the prior mean is given by (cf. Corollary 1 of Basu et al., 1996)

$$\mathrm{PS} = \max_{\sqrt{\alpha' \Sigma_p^{-1} \alpha} = 1} \nu' J \alpha = \sqrt{\nu' \Sigma_\pi \Sigma_p^{-1} \Sigma_\pi \nu}. \tag{12}$$

The interval with endpoints $E_\pi[\nu'\theta] \pm a\mathrm{PS}$ is thus a local approximation to the set of posterior mean values that can be induced by changing the prior mean by the multivariate analogue of a prior standard deviations. Note that the direction α that induces the largest change is proportional to $\Sigma_\pi \nu$. Thus, if the parameter of interest is $\theta_j = \nu'\theta$, then the largest (local) change of the posterior mean is induced by shifting the prior mean by a multiple of the jth column of the posterior covariance matrix Σ_π.

Alternatively, one might ask which linear combination of the parameters is most sensitive to local changes in the prior mean. If the magnitude of posterior mean changes is measured relative to the posterior standard deviation, then the most sensitive directions ν_S are given by

$$\nu_S \in \mathrm{argmax}_\nu \frac{\sqrt{\nu' \Sigma_\pi \Sigma_p^{-1} \Sigma_\pi \nu}}{\sqrt{\nu' \Sigma_\pi \nu}}. \tag{13}$$

This is solved by the vectors ν_S that are proportional to the eigenvector of $\Sigma_p^{-1}\Sigma_\pi$ associated with the largest eigenvalue.

2.3. Prior informativeness with a vector parameter

Now turn to measuring prior informativeness in the vector case. As an initial motivation, suppose the likelihood arises from a Gaussian shift experiment $Y \sim \mathcal{N}(\theta, \Sigma_l)$ with Σ_l full rank and known, and also the prior is Gaussian, $\theta \sim \mathcal{N}(\mu_p, \Sigma_p)$, where θ is $k \times 1$ and Σ_p is full rank. The scalar parameter of interest is $\nu'\theta$. Now $\nu'Y$ is both the maximum likelihood and uniformly minimum variance unbiased estimator of $\nu'\theta$. The likelihood information about $\nu'\theta$ is thus arguably summarized by the scalar random variable $Y_\nu = \nu'Y \sim \mathcal{N}(\nu'\theta, \nu'\Sigma_l\nu)$. The multivariate prior $\theta \sim \mathcal{N}(\mu_p, \Sigma_p)$ implies the scalar Gaussian prior $p_\nu \sim \mathcal{N}(\nu'\mu_p, \nu'\Sigma_p\nu)$ on $\nu'\theta$. In this scalar problem of observing Y_ν with prior p_ν on $\nu'\theta$, the analysis of Section 2.1 implies that the fraction of prior information for the posterior mean of $\nu'\theta$ is $(\nu'\Sigma_p\nu)^{-1}/((\nu'\Sigma_p\nu)^{-1} + (\nu'\Sigma_l\nu)^{-1})$, so that

$$\mathrm{PI}_G = 1 - \frac{\nu'\Sigma_p\nu}{\nu'\Sigma_p(\Sigma_p - \Sigma_\pi)^{-1}\Sigma_p\nu}, \tag{14}$$

where $\Sigma_\pi = \Sigma_p - \Sigma_p(\Sigma_l + \Sigma_p)^{-1}\Sigma_p = (\Sigma_p^{-1} + \Sigma_l^{-1})^{-1}$ is the posterior covariance matrix in the $k \times 1$ problem. Thus, at least in this Gaussian setup, $\mathrm{PI}_G \in (0,1)$ is the natural generalization of the scalar prior informativeness measure PI to the vector parameter case.

At the same time, the Gaussian framework of this derivation of PI_G is obviously quite special, and it does not apply as such to most applied problems.[3] Abstractly, the question is how to suitably generalize PI in (8) of the scalar case to a reasonable measure of prior informativeness for $\nu'\theta$ when the derivative J is the $k \times k$ matrix (11). One way of proceeding is to impose constraints on potential mappings from J (and Σ_p and ν) to the unit interval. Specifically, we will argue that a number of reasonable axiomatic requirements on such mappings lead to the measure $\mathrm{PI} = \mathrm{PI}_G$, at least as long as the largest eigenvalue of J is smaller than unity (if not, PI is typically equal to one). Thus, to the extent that these requirements are compelling, (14) emerges as the unique scalar function of J that summarizes the fraction of prior information for $\nu'\theta$, also for non-Gaussian models and priors.

This result is formally derived in the online appendix. More informally, the three sets of requirements involve invariance to reparameterizations; conditions on dimension reductions, coherency and continuity; and consistency with the fraction of information interpretation in a particular bivariate context. A brief summary of each follows.

Invariance to linear reparametrizations. Computing PI after reparametrizing the problem in terms of $\theta^* = H\theta$ leads to the same informativeness measure, for all parameters of interest $\nu'\theta$ and full rank matrices H.

[2] Otherwise, one can always define an alternative embedding analogously to the scalar case, with identical s_0 and $d(\int \theta p_\alpha(\theta)\,d\theta)/d\alpha'|_{\alpha=0} = I_k$.

[3] Inference in a linear regression with Gaussian errors and the usual conjugate priors on the regression coefficient and error variance almost fits this framework, though, as the marginal likelihood of the regression coefficients has a multivariate student-t kernel, which is very close to log-quadratic, unless the sample size is very small.

With an appropriate choice of H, one can reduce the problem to the case where the prior covariance matrix is the identity matrix, and the posterior covariance matrix is diagonal. The derivative matrix J is then given by the diagonal matrix $J=\mathrm{diag}(\lambda_1,\ldots,\lambda_k)$, where the λ_i's correspond to the eigenvalues of J in the original parameterization.

Dimension reductions, coherence and continuity when $\Sigma_p=I_k$ and $J=\mathrm{diag}(\lambda_1,\ldots,\lambda_k)$. (a) If $\nu'\theta=\theta_1$, then $\mathrm{PI}=\min(\lambda_1,1)$, in accordance with (8). (b) Zeros in ν are equivalent to facing a lower dimensional problem with the corresponding rows and columns of J and Σ_p deleted. (c) PI has range [0,1], is weakly increasing in all λ_i and satisfies some continuity and differentiability conditions. (d) If $\lambda_1=\lambda_2$, then any ν of the same length and with the same last $k-2$ elements leads to the same PI. (e) Replacing $\lambda_1,\lambda_2,\ldots,\lambda_m$, $m<k$, with PI^+ computed from $\nu^+=(\nu_1,\ldots,\nu_m,0,\ldots,0)'$ leaves PI of $\nu=(\nu_1,\ldots,\nu_m,\nu_{m+1},\ldots,\nu_k)'$ unchanged, that is $\overline{\lambda}_m=\mathrm{PI}^+$ is the coherent average value of $\lambda_1,\ldots,\lambda_m$.

This second set of requirements ensures that PI of $\nu'\theta$ in the reparameterization with $\Sigma_p=I_k$ and $J=\mathrm{diag}(\lambda_1,\ldots,\lambda_k)$ is given by a generalized weighted average of the individual information strengths λ_i of the parameters θ_i. Specifically, the results of Kitagawa (1934), who builds on the classic results of Kolmogorov (1930) and Nagumo (1930) on axiomatic foundations for quasi-arithmetic means, imply that PI is then of the form

$$\mathrm{PI}=\phi^{-1}\left(\frac{\sum_{i=1}^k \nu_i^2\phi(\lambda_i)}{\sum_{i=1}^k \nu_i^2}\right) \tag{15}$$

for some increasing and continuous function ϕ, at least as long as $\max_i\lambda_i<1$.

Consistency with fraction of information interpretation when $\Sigma_p=I_2$, $J=\mathrm{diag}(\lambda_1,0)$ and $\nu=(1,1)'$. A derivative matrix of $J=\mathrm{diag}(\lambda_1,0)$ arises when θ_2 is perfectly pinned down by the likelihood. In that case, the shape of the likelihood for the parameter of interest $\nu'\theta=\theta_1+\theta_2$ is the same as the shape of the likelihood of θ_1. As discussed in Section 2.1, λ_1 has an interpretation as the fraction of prior information in the posterior information for θ_1, that is $\lambda_1=1/(1+\sigma_{1l}^{-2})$, where σ_{1l}^2 is the approximate global curvature of the log-likelihood of θ_1, and thus also of $\nu'\theta$. The prior information (=precision) on $\nu'\theta$ equals $1/\nu'\Sigma_p\nu=1/2$. The fraction of information interpretation thus requires that PI of $\nu'\theta=\theta_1+\theta_2$ equals

$$\frac{1/2}{1/2+\sigma_{1l}^{-2}}=\frac{\lambda_1}{2-\lambda_1}. \tag{16}$$

With θ_2 known, the likelihood forms a ridge, and the third requirement specifies a suitable quantification of the importance of the bivariate prior in this special case. In conjunction with the differentiability constraint in part (c) of the second requirement, this pins down ϕ in (15) to be equal to $\phi(\lambda)=1/(1-\lambda)$.

To state the precise relationship between PI and PI_G, it is useful to introduce the following definition.

Definition 1. A prior is of *limited overall informativeness* if the largest eigenvalue $\lambda_{\max}$ of J is smaller than one.

In words, limited overall prior informativeness means that for all possible parameters of interest $\nu'\theta$, the posterior variance is smaller than the prior variance. If this is the case, we obtain that the only informativeness measure that is compatible with the three sets of requirements above is given by

$$\mathrm{PI}=\mathrm{PI}_G=1-\frac{\nu'\Sigma_p\nu}{\nu'\Sigma_p(\Sigma_p-\Sigma_\pi)^{-1}\Sigma_p\nu}. \tag{17}$$

Without limited overall informativeness, the value of PI is equal to one, unless ν is exactly orthogonal to all eigenvectors of J whose eigenvalue is larger than one. Intuitively, if $\nu'\theta$ is partially a function of a linear combination of parameters for which the prior information completely dominates, then the prior also completely dominates for $\nu'\theta$.

3. Discussion and extensions

This section discusses some general inequalities for the measures PS and PI, their generalization to functions of parameters, a modification of the measures that conditions on some subset of prior information, and the relationship of the analysis here with the frequentist concept of identification.

3.1. Inequalities

In the exponential family embedding (10), the prior and posterior covariance matrices play a dual role as measures of spread of the respective distribution, and as inputs to the derivative matrix J of posterior means relative to prior means. This leads to a number of interesting relations involving PS and PI (see the online appendix for details)

$$\mathrm{PS}\le\sqrt{\lambda_{\max}}\sqrt{\nu'\Sigma_\pi\nu} \tag{18}$$

$$\mathrm{PS}\ge\frac{\nu'\Sigma_\pi\nu}{\sqrt{\nu'\Sigma_p\nu}} \tag{19}$$

$$\mathrm{PI}\ge\min\left(\frac{\nu'\Sigma_\pi\nu}{\nu'\Sigma_p\nu},1\right) \tag{20}$$

$$\mathrm{PI} \le \min(1,\lambda_{\max}) \tag{21}$$

$$\mathrm{PI} \le \frac{\mathrm{PS}}{\sqrt{\nu'\Sigma_p\nu}} \quad \text{if } \lambda_{\max} \le 1/3 \tag{22}$$

$$\frac{\mathrm{PS}}{\sqrt{\nu'\Sigma_p\nu}} \le \sqrt{\frac{3}{2}\mathrm{PI}} \quad \text{if } \lambda_{\max} < 1 \tag{23}$$

The most remarkable of these relations might be (19): Under overall limited prior informativeness (so that $\lambda_{\max} < 1$), the maximal variation of the posterior mean that can be induced by varying the prior mean by the multivariate analogue of a prior standard deviations is always smaller than a posterior standard deviations. A highly significant posterior result, that is a posterior mean that is several posterior standard deviations different from zero, can never be overturned by a variation α in the prior mean that is small in terms of the $\sqrt{\alpha'\Sigma_p^{-1}\alpha}$ metric (at least under the linear approximation based on the derivative). Inequalities (19) and (20) formally show that PS and PI are at least as large as what would be obtained by a "marginal" analysis based on the analogous expression using only the prior and posterior variances of $\nu'\theta$. The largest possible value for PI is $\min(1,\lambda_{\max})$, as demonstrated by inequality (22), and this value is obtained with ν proportional to the eigenvector of $\Sigma_p^{-1}\Sigma_\pi$ that corresponds to its largest eigenvalue. Note that this is the same ν that also maximizes $\mathrm{PS}/\sqrt{\nu'\Sigma_\pi\nu}$, as discussed at the end of Section 2.2 above. Finally, the last two inequalities (22) and (23), which rely in part on a result in Pratt (1964), show that once PS is normalized by the prior standard deviation, the two measures cannot take on very different values, at least as long as $\lambda_{\max}$ is small.

3.2. *Functions of parameters*

In many applications, there is interest not only in the unknown $k \times 1$ parameters θ, but also in particular functions of them. Let $\gamma = \Gamma(\theta)$, where $\Gamma : \mathbb{R}^k \mapsto \mathbb{R}$. In the notation of Section 2.2, the derivative of the posterior mean of γ with respect to the prior mean of θ in (10) is, under weak regularity conditions, the $1 \times k$ vector

$$J_\gamma = E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])']\Sigma_p^{-1} \tag{24}$$

where $E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])']$ is recognized as the posterior covariance between γ and θ'.

In analogy to PS, define PS_γ as the largest change of the posterior mean of γ that can be induced by a unit change α of the prior mean in the metric $\sqrt{\alpha'\Sigma_p^{-1}\alpha}$,

$$\begin{aligned}\mathrm{PS}_\gamma &= \max_{\sqrt{\alpha'\Sigma_p^{-1}\alpha}=1} J_\gamma\alpha = \sqrt{J_\gamma\Sigma_p J_\gamma{}'} \\ &= \sqrt{E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])']\Sigma_p^{-1}E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])]}.\end{aligned} \tag{25}$$

The measure PS_γ is alternatively recognized as the sensitivity measure PS of the linear combination $\nu'\theta$ with

$$\nu = \nu_\gamma = \Sigma_\pi^{-1}E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])]. \tag{26}$$

This ensures that whenever Γ is linear, $\Gamma(\theta) = c_\gamma + \nu'\theta$, $\mathrm{PS}_\gamma = \mathrm{PS}$. Also, since the posterior covariance matrix of $(\theta',\gamma)'$ is positive semi-definite, and $E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])]$ is the posterior covariance between γ and θ, the posterior variance $\sigma_\gamma^2 = E_\pi[(\Gamma(\theta) - E_\pi[\Gamma(\theta)])^2]$ satisfies $\sigma_\gamma^2 \ge \nu_\gamma'\Sigma_\pi\nu_\gamma$. The analogue of inequality (19) of the last subsection, $\mathrm{PS}_\gamma \le \sqrt{\lambda_{\max}}\sigma_\gamma$ thus still holds for any Γ with finite posterior variance.

Similarly, define the prior informativeness PI_γ of γ as the prior informativeness measure PI of the linear combination $\nu'\theta$ with the same derivative of the posterior mean as γ, that is with $\nu = \nu_\gamma$. Thus, under overall limited prior informativeness,

$$\mathrm{PI}_\gamma = 1 - \frac{\nu_\gamma'\Sigma_p\nu_\gamma}{\nu_\gamma'\Sigma_p(\Sigma_p - \Sigma_\pi)^{-1}\Sigma_p\nu_\gamma}. \tag{27}$$

This definition again ensures agreement with PI for linear Γ, and also inequality (22) for PI_γ, $\mathrm{PI}_\gamma \le \lambda_{\max}$.

For highly nonlinear Γ one might worry about the general appropriateness of equating the prior informativeness of γ with that of $\nu_\gamma'\theta$. A useful statistic in that regard is the R^2 of a linear regression of $\gamma = \Gamma(\theta)$ on θ in the posterior,

$$R_\gamma^2 = \frac{\nu_\gamma'\Sigma_\pi\nu_\gamma}{\sigma_\gamma^2} = \frac{E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])']\Sigma_\pi^{-1}E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])]}{\sigma_\gamma^2}. \tag{28}$$

Values of R_γ^2 close to one indicate a very similar posterior behavior of γ and $\nu_\gamma'\theta$, so that PI_γ becomes a more compelling measure for the prior informativeness of γ. In large samples, the Bernstein–von Mises theorem implies convergence of the posterior of θ to a Gaussian with vanishing variance, so that a delta-method type argument applied to $\gamma = \Gamma(\theta)$ yields $R_\gamma^2 \to 1$ with probability converging to one for differentiable, sample size independent Γ.

It is not necessary that γ is a function of θ alone, but it may also depend on the realized data (so that formally, Γ is indexed by the data). For example, PS_γ and PI_γ might be applied to learn about the role of the prior for a forecast, which is a function of both the model parameters θ and the realized data. As an illustration, consider a one-step ahead forecast in an AR(1) model $y_t-\mu=\rho(y_{t-1}-\mu)+\varepsilon_t$, where the last observation is y_T and $\theta=(\mu,\rho)'$. Here $\gamma=\Gamma(\theta)=\mu+\rho(y_T-\mu)$. If y_T takes on a value very different from the sample mean $\overline{y}$ (and thus the approximate posterior mean of μ), then ρ is relatively more influential than μ, which is properly reflected in the measures PS_γ and PI_γ.

One can also set up functions Γ to learn about the sensitivity of posterior results beyond the posterior mean. For instance, with $\gamma=\Gamma(\theta)=\mathbf{1}[\theta_j>0]$, the posterior mean of γ is the posterior probability that θ_j is positive, and J_γ contains the derivatives of this posterior probability with respect to the prior mean of θ.

3.3. Conditional analysis

The measure PI is designed to reveal the importance of the whole multivariate prior distribution for the posterior results about the parameter of interest θ_j. In some applications, though – possibly due to an analysis using PI – one might be fully aware that the data are not informative about a particular parameter θ_i, $i\neq j$, and it is clear that its prior is important for the posterior results. The interesting question then is whether *conditional* on the prior information about θ_i, other parts of the prior are particularly informative. In practice, such a conditional analysis can be performed by dropping the ith row and column of Σ_p and Σ_π in (17) when computing PI (and possibly also PS), which always leads to (weakly) smaller value of PS and PI as long as the prior on θ_i is independent of the remaining prior.[4]

Note that there is nothing wrong with continuing to include such θ_i in the estimation: Bayes rule ensures that the posterior is the coherent update of prior beliefs from data information, even if there is little or no information regarding θ_i. The alternative of fixing θ_i to a particular value has the disadvantage that the posterior then fails to reflect uncertainty about θ_i. What is more, the derivative matrix J cannot be computed for a degenerate prior on θ_i, so that the impact of the fixed value of θ_i on the posterior results on other parameters cannot be assessed in this straightforward manner.

A prior informativeness analysis conditional on the prior information about θ_i can be formally motivated by a two stage view of information acquisition about θ_i: At time 0, almost nothing is known about θ_i, corresponding to beliefs $p_{i,0}$ with large variance $\sigma^2_{i,0}$. At time 1, a data set A is analyzed, leading to a much tighter posterior $\pi_{i,1}$ about θ_i with variance $\sigma^2_{i,1}$. The data set currently under investigation is obtained at time 2, and the investigator specifies the prior on θ_i that corresponds to the posterior at time 1, that is $p_{i,2}=\pi_{i,1}$. Assuming that there is no other link between the parameters of the current study and the stage A data, then the posterior $\pi_{i,2}$ that results from this analysis is also the posterior of jointly observing the stage A and the current data set, with prior $p_{0,i}$ on θ_i. But in this view, the appropriate value for the prior variance is $\sigma^2_{i,0}$, and as $\sigma^2_{i,0}\to\infty$, a calculation shows that PI for θ_j, $j\neq i$ converges to the value that one obtains in the conditional analysis that simply drops the ith row and column of Σ_p and Σ_π. Thus, if the prior for a poorly identified nuisance parameter can be reasonably viewed as representing the posterior from previous data analyses with an originally very vague prior, then a conditional analysis approximates the prior informativeness of the remaining parameters relative to the entire data information.

An alternative motivation for a conditional analysis arises from interpreting prior distributions as part of the stochastic specification of the model: In the Bayesian framework, there is no difference between an unknown (but nonstochastic) nuisance parameter equipped with some prior distribution, and a stochastic specification of this unknown parameter with probability distribution equal to the prior. For example, in a panel model, one can reasonably view unit specific intercepts as unknown parameters equipped with a prior, or as realizations of a random process. In either case one obtains the same posterior, and draws the same conclusions. At the same time, PI (and PS) depend on this classification, and treating nuisance parameters as stochastic in this sense again amounts to dropping the corresponding rows and columns of Σ_p and Σ_π.

Similarly, for functions of parameters $\gamma=\Gamma(\theta)$, one can condition on prior information about θ_i by dropping the ith column and row in the computation of PI_γ in (27). The measure PI_γ then ignores variation in γ that is induced by θ_i, and focusses exclusively on the relative importance of the prior on the other parameters.

3.4. Relationship to frequentist identification

Denote the density of the observables $Y\in\mathcal{Y}$ by $f(y;\theta)$, where the parameter is $\theta\in\Theta$. Rothenberg (1971) defines $\theta_0\in\Theta$ to be *identifiable* if $f(y;\theta)=f(y;\theta_0)$ for all $y\in\mathcal{Y}$ implies $\theta=\theta_0$. The likelihood function l is, of course, simply given by $l(\theta)=f(y;\theta)$ after observing $Y=y$. Thus, if θ_0 is such that $l(\theta)=l(\theta_0)$ implies $\theta=\theta_0$, then θ_0 is identifiable. In particular, the existence of a *unique* maximizer $\hat{\theta}$ of l is sufficient for identifiability of the parameter value $\theta=\hat{\theta}$.

The converse is not true, though: even if $l(\theta)=l(\theta_0)$ for $\theta\neq\theta_0$ and $l(\theta)=f(y;\theta)$, there might well exist $y_0\neq y$ for which $f(y_0;\theta)\neq f(y_0;\theta_0)$. Even an entirely flat likelihood l does not imply lack of identifiability—it could be that for some other draw of the data, the likelihood does contain information. For instance, think of a state dependent model with observed

[4] This follows immediately from the definition of PS in (12), and also holds for PI, since the formula for partitioned inverses implies that any $(k-1)\times(k-1)$ submatrix of $(\Sigma_p-\Sigma_\pi)^{-1}$ is (weakly) larger than the inverse of the corresponding $(k-1)\times(k-1)$ submatrix of $(\Sigma_p-\Sigma_\pi)$.

states. If one of the states never occurs in the observed data, then the likelihood of the model parameters in that state is completely flat, yet all parameters of the model could well be identifiable in the sense of Rothenberg. An entirely flat likelihood would always lead to a prior informativeness measure PI of unity, as discussed in Section 2.1. One might argue that in this example, this is the "right" answer for communicating empirical results—the data that was observed does not contain information about the parameter of interest, and the possibility that other potential data from the same model could have contained information does not mitigate this fact. In other words, the value of PI is fully determined by the likelihood and prior, and thus adheres to the likelihood principle. In contrast, measures based on the Fisher Information, as considered by Iskrev (2010), Traum and Yang (2010) and Andrle (2010), average over the amount of sample information in samples that did not realize.[5]

Rothenberg's (1971) definition is useful for the more theoretical question of whether model parameters could *in principle* be told apart by empirical studies, that is whether hypothetical knowledge of the population distribution of Y would pin down the model parameters. The relationship between this concept and the measure PI is as follows. Assume first that θ is not identifiable because all values of θ in some hyperplane Θ_{hp} lead to the same density. As a leading example, suppose some element of θ does not affect the density of Y. Then surely also the likelihood is constant on this hyperplane, for all possible realizations of y. Thus, prior mean shifts along the hyperplane will lead to one-to-one shifts of the posterior mean, and $\lambda_{\max} \geq 1$. An empirical finding of limited overall prior informativeness in the sense of Definition 1 (i.e. that $\lambda_{\max} < 1$) therefore rules out at least this hyperplane form of lack of identifiability.

Second, assume that θ_0 is not identifiable because $f(y;\theta) = f(y;\theta_0)$ for all $\theta, \theta_0 \in \Theta'$, but Θ' does not contain a hyperplane. In that case, there is no direction in which the likelihood is necessarily flat, and $\lambda_{\max}$ might well be smaller than unity, despite the lack of identifiability. Whether or this is the "right" answer depends on Θ'—if Θ' is "small", then lack of identifiability does not imply that nothing useful can be learned about θ. For an extreme illustration, suppose $Y \sim \mathcal{N}(r(\theta), 1)$, $\theta \in \mathbb{R}$, and $r : \mathbb{R} \mapsto \mathbb{R}$ rounds its inputs to 10 digits. Then no value of θ is identifiable, and the likelihood is a step function. But almost no information is lost relative to the experiment $Y \sim \mathcal{N}(\theta, 1)$. Accordingly, PI will behave almost the same way in the model with and without rounding, as the global shape of the likelihood is almost identical.

An important practical appeal of PI is that it *quantifies* prior and likelihood informativeness, in contrast to the binary "identifiable or not" of Rothenberg's definition. Many DSGE models, for instance, may well have identifiable parameter values in the sense of Rothenberg, although the information in the data about parameters of interest might be very limited (cf. the discussion of weak identification in Canova and Sala, 2009). Also, Rothenberg's definition of identifiability concerns a specific parameter *value* θ_0. But in practice the parameter is unknown, leading to the difficult question for which value(s) of θ_0 identifiability should be analyzed, and what conclusion is to be drawn if different identifiability results are obtained for different plausible θ_0. In contrast, PI is a single statistic that summarizes the relative prior and data informativeness for any scalar parameter $\nu'\theta$. Finally, the approach here is in no way tied to an underlying linear or Gaussian model. For instance, in a DSGE context, PI can easily be computed also for posterior results from a likelihood that is based on higher order approximations of the decision rules around the steady state, such as those developed in Fernández-Villaverde et al. (2010).

The concept and appeal of the prior informativeness measure PI is thus quite distinct from the standard frequentist definition of identification, so that the approach pursued here is largely complementary to the recent results on identification in log-linearized DSGE models by Iskrev (2010), Komunjer and Ng (2009) and Andrle (2010).

4. Applications

The measures PS and PI are illustrated in two applications: Lubik and Schorfheide's (2004) small scale New Keynesian monetary DSGE model, and Smets and Wouters' (2007) larger scale DSGE model.

4.1. *Lubik and Schorfheide (2004)*

After log-linearization, the model analyzed in Lubik and Schorfheide (2004) (henceforth LS) is given by the three equations

$$x_t = E_t[x_{t+1}] - \tau(R_t - E_t[\pi_{t+1}]) + g_t \tag{29}$$

$$\pi_t = \beta E_t[\pi_{t+1}] + \kappa(x_t - z_t) \tag{30}$$

$$R_t = \rho_R R_{t-1} + (1-\rho_R)(\psi_1 \pi_t + \psi_2(x_t - z_t)) + \varepsilon_{R,t} \tag{31}$$

where x_t, π_t and R_t are percentage deviations from steady state output, inflation and interest rate, respectively. The discount factor β is approximated by $\beta = (1 + r^*/100)^{-1/4}$ with r^* the steady state annual real interest rate, and the annual steady state inflation rate is denoted by π^*. In addition to the i.i.d. monetary policy shock $\varepsilon_{R,t}$, the two additional shock

[5] In the example of Fig. 2 of Section 2.1, suppose that in addition to Y, it is also observed from which of the two Gaussians Y was drawn. The log-likelihood then is quadratic in either case, and PI$=w$ exactly with $\sigma_l^2 = 0.02$ or $\sigma_l^2 = 0.06$. The statistic PI thus reflects the actual amount of information about θ in the realized sample. In contrast, the Fisher Information in this experiment is the probability weighted average of these two values for σ_l^{-2}.

Table 1
Parameter prior and posterior results in Lubik and Schorfheide (2004).

Parameter	Prior			Posterior		σ_π^2/σ_p^2	PI	PS	$\widehat{\text{PS}}_{1/2}$	$\widehat{\text{PS}}_1$
	Shape	μ_p	σ_p	μ_π	σ_π					
ψ_1	$\mathcal{G}$	1.50	0.25	1.31	0.13	0.27	0.28	0.09	0.09	0.10
ψ_2	$\mathcal{G}$	0.25	0.15	0.12	0.08	0.28	0.28	0.05	0.06	0.08
ρ_R	$\mathcal{B}$	0.50	0.20	0.73	0.03	0.03	0.03	0.02	0.03	0.03
π^*	$\mathcal{G}$	4.00	2.00	4.23	0.52	0.07	0.07	0.15	0.16	0.15
r^*	$\mathcal{G}$	2.00	1.00	2.06	0.36	0.13	0.13	0.13	0.13	0.13
κ	$\mathcal{G}$	0.50	0.20	0.48	0.18	0.84	NA	0.20	0.21	0.23
τ^{-1}	$\mathcal{G}$	2.00	0.50	3.19	0.64	1.64	NA	0.86	0.95	1.07
ρ_g	$\mathcal{B}$	0.70	0.10	0.87	0.03	0.07	0.08	0.02	0.02	0.03
ρ_z	$\mathcal{B}$	0.70	0.10	0.80	0.03	0.08	0.09	0.01	0.01	0.02
ρ_{gz}	$\mathcal{N}_{[-1,1]}$	0.00	0.40	0.79	0.12	0.09	0.09	0.08	0.11	0.14
ω_R	$\mathcal{IG}$	0.31	0.16	0.28	0.02	0.02	0.02	0.01	0.02	0.02
ω_g	$\mathcal{IG}$	0.38	0.20	0.16	0.02	0.01	0.01	0.01	0.01	0.02
ω_z	$\mathcal{IG}$	1.00	0.52	0.97	0.08	0.02	0.03	0.03	0.03	0.04

Notes: $\mathcal{B}$, $\mathcal{G}$ and $\mathcal{N}_{[-1,1]}$, are Beta, Gamma and Normal (restricted to the $[-1,1]$ interval) prior distributions with mean and variance μ_p and σ_p^2, and $\mathcal{IG}$ is a Gamma prior distribution on $1/\omega^2$ that implies a mean and variance of μ_p and σ_p^2 on ω. The values of PI are conditional (in the sense of Section 3.3) on the prior information on τ^{-1} and κ. $\widehat{\text{PS}}_{1/2}$ and $\widehat{\text{PS}}_1$ are defined in Eq. (33) and are nonderivative based measures of prior sensitivity analogous to PS.

processes are the demand shock g_t and the productivity shock z_t

$$g_t = \rho_g g_{t-1} + \varepsilon_{g,t}, \quad z_t = \rho_z z_{t-1} + \varepsilon_{z,t} \tag{32}$$

where the correlation between the i.i.d. shocks $\varepsilon_{g,t}$ and $\varepsilon_{z,t}$ is equal to ρ_{zg}.

Using LS's data and code, the model is estimated on HP-filtered U.S. postwar data. In their analysis, LS focus on the possibility of indeterminacy of the system (29), (30) and (31) due to a tepid response of monetary policy (31) to inflation (i.e., small value of ψ_1) in the 60s and 70s. For simplicity and comparability to other studies, we instead impose a time invariant, determinate monetary policy rule for the whole post-war period 1960:I–1997:IV. The prior on the 13 parameters is as in LS, except that we adopt del Negro and Schorfheide's (2004) prior on ψ_1 with little mass on the indeterminacy region.

Parameter prior and posterior results are given in Table 1. For the risk aversion τ^{-1} (or intertemporal substitution elasticity), the prior variance is smaller than the posterior variance, suggesting that the prior plays a dominant role.[6] Thus without conditioning on any prior information, the prior is of unlimited informativeness in the sense of Definition 1. To obtain nontrivial results for PI, we thus condition on the prior for τ^{-1}. Moreover, even conditional on the prior information about τ^{-1}, the largest eigenvalue $\lambda_{\max}$ of $\Sigma_p^{-1}\Sigma_\pi$ is $\lambda_{\max} = 0.97$, and the corresponding eigenvector (normalized to unit length) has a loading of 0.91 on κ. Even though $\lambda_{\max}$ is now smaller than unity, as a practical matter it makes sense to conclude that also the prior for κ is of dominating importance: On the one hand, the appearance of limited overall informativeness might simply be due to estimation error in Σ_π. On the other hand, PI becomes very sensitive to the exact value of the eigenvalues λ_i once they are close to unity.

With the conditioning on prior information about τ^{-1} and κ, the remaining entries for PI are very close to σ_π^2/σ_p^2. In general, the marginal analysis based on the derivative σ_π^2/σ_p^2 (cf. (6) of Section 2.1) always leads to smaller values than the joint prior informativeness measure PI by inequality (20). The difference between PI and σ_π^2/σ_p^2 becomes potentially large if the correlation pattern in the posterior is substantially different from the correlation pattern in the prior. In this application, at least after integrating out τ^{-1} and κ, the likelihood information about the parameters does not seem to be highly correlated, approximately matching the independent prior specification. Overall the values of PI in this application show overwhelming prior importance of κ and τ^{-1}, but conditional on this information, the prior contributes less than 30% to the posterior results of the other parameters.

Now turn to the prior sensitivity measure PS, computed without conditioning on the prior for κ and τ^{-1}. The values of PS are usefully compared to the posterior standard deviation σ_π, as uncertainty about the appropriate prior mean of (the multivariate analogue of) a prior standard deviations leads to changes of the posterior mean within $\pm a$PS, at least under the linear approximation based on the derivative. The results for the steady-state inflation rate π^* are the least sensitive relative to its posterior standard deviation, with a posterior mean between 4 and 4.5 for $a = 1$. In the other extreme, the additional uncertainty about τ^{-1} induced by $a = 1$ is larger than the baseline posterior uncertainty.

To get some sense for the quality of this approximation, consider two empirical measures of posterior mean sensitivity. Specifically, embed the baseline prior in the exponential family (10) (except for the three inverse Gamma priors, which are

[6] The posterior mean of τ^{-1} is substantially different from the prior mean, so that the likelihood does contain information about τ^{-1}. But the qualitative conclusion of an overwhelming prior importance for τ^{-1} seems warranted: changing the variance of the Gamma prior to 2 without changing the mean, for instance, yields $\mu_\pi = 13.2$ and $\sigma_\pi = 4.1$.

Table 2
Standardized matrix of derivatives of posterior means with respect to prior means G in Lubik and Schorfheide (2004).

Parameter	ψ_1	ψ_2	ρ_R	π^*	r^*	κ	τ^{-1}	ρ_g	ρ_z	ρ_{gz}	ω_R	ω_g	ω_z
ψ_1	0.52					0.32	−0.30	0.13		0.05			
ψ_2		0.52				0.22	−0.09		0.07				
ρ_R	0.11		0.16			−0.48	0.54			−0.14	−0.06		
π^*				0.26	0.13								
r^*				0.09	0.36								
κ	0.18	0.12	−0.08			0.92	−0.47	0.13		0.19	0.06		−0.06
τ^{-1}	−0.12		0.07			−0.34	1.28	−0.12					
ρ_g	0.25					0.43	−0.57	0.26		0.16		−0.08	
ρ_z	−0.05	0.12				−0.08	0.11		0.29	−0.10			−0.07
ρ_{gz}	0.09		−0.08			0.61	−0.13	0.14	−0.09	0.29		−0.05	
ω_R	0.14	0.11	−0.07			0.41	−0.46	0.08		0.07	0.13		
ω_g	−0.07					−0.27		−0.20		−0.15		0.11	
ω_z		−0.13				−0.30	0.11		−0.14				0.16

Notes: Each column contains derivative based approximations to the change in the posterior mean of the various parameters, measured in posterior standard deviations, that results from shifting the prior mean of the parameter in the column heading by one prior standard deviation. Entries of absolute value smaller than 0.05 are left blank.

embedded in the family described in 1).[7] For each parameter, one can then numerically determine[8]

$$\widehat{\mathrm{PS}}_a = \frac{1}{a} \max_{\sqrt{\alpha' \Sigma_p^{-1} \alpha} = a} |\mu_\pi(\alpha) - \mu_\pi(0)| \tag{33}$$

the actual largest change of the posterior mean, expressed in units of a (and by construction, $\lim_{a \searrow 0} \widehat{\mathrm{PS}}_a = \mathrm{PS}$). As can be seen in Table 1, the linear approximation using PS is very accurate for $a = 1/2$ and still quite good for $a = 1$.

A more detailed picture of the sensitivity of the posterior results emerges by direct inspection of the derivative matrix J. A useful standardized version of J is the $k \times k$ matrix

$$G = \mathrm{diag}(\sigma_{\pi,1}, \ldots, \sigma_{\pi,k})^{-1} J \, \mathrm{diag}(\sigma_{p,1}, \ldots, \sigma_{p,k}) \tag{34}$$

where $\sigma_{p,j}$ and $\sigma_{\pi,j}$ are the prior and posterior standard deviations of θ_j, respectively. The jth column of G has the interpretation of the (approximate) change of the posterior mean of θ, measured in units of posterior standard deviations, that results from increasing the prior mean of θ_j by one prior standard deviation. Table 2 reports G for the LS example. The off-diagonal elements show that changing the prior mean on any given parameter often has substantial consequences also for the posterior mean of other parameters. This is especially true for τ^{-1} and κ, whose prior means tend to push the posterior means of parameters in opposite directions by substantial amounts. Looking across the rows of ρ_R, ρ_g, ρ_{gz}, ω_R, ω_g and ω_z, the off-diagonal elements in the τ^{-1} and κ columns are even larger than the diagonal elements, so that changing the prior mean of τ^{-1} or κ by one prior standard deviation has a larger effect on the posterior mean of ρ_R, ρ_g, ρ_{gz}, ω_R, ω_g and ω_z than changing their own prior mean by one prior standard deviation. These cross effects are incorporated in the corresponding values of PS, which are given by the length of each row of G multiplied by the posterior standard deviation. Substantively, prior beliefs of a steeper Phillips curve (higher κ) lead to posterior beliefs of a more aggressive (higher ψ_1) and less smoothing (lower ρ_R) monetary policy rule, while a priori beliefs of higher risk aversion (higher τ^{-1}) have the opposite effect.

Table 3 extends the analysis to impulse responses (IRs) and variance decompositions (VDs), based on the definitions in Section 3.2 of the prior sensitivity measure PS_γ and prior informativeness measures PI_γ for these specific functions of the primitive parameter θ. The $1 \times k$ vector G_γ of these functions $\gamma = \Gamma(\theta)$ is the standardized version of the derivative vector J_γ in (24),

$$G_\gamma = \frac{1}{\sigma_\gamma} J_\gamma \, \mathrm{diag}(\sigma_{p,1}, \ldots, \sigma_{p,k}) \tag{35}$$

with the interpretation that the elements in G_γ denote the (linear approximation to the) change of the posterior mean of γ, measured in posterior standard deviations, that arises by increasing the prior mean of θ_j by one prior standard deviation. The large entries in the columns for τ^{-1} and κ in Table 3 demonstrate that the prior sensitivity of the IRs and VDs posterior means is mostly driven by these two parameters. For instance, the variance decomposition shows that with the baseline prior, 83% of the variation in output is driven by the demand shock g. This fraction is seen to further increase substantially

[7] Alternatively, one could reparametrize ω in terms of ω^{-1} or ω^{-2}, and apply the exponential family embedding for all 13 parameters. This yields almost identical results for the other 10 parameters.

[8] The largest posterior mean shift is determined by numerical maximization, with the posterior mean computed by importance sampling: The posterior mean with a prior indexed by α is the weighted average of the posterior draws obtained from the baseline prior, with weights equal to the ratio of the exponential family density and the baseline prior density.

Table 3
Posterior impulse response and variance decomposition results in Lubik and Schorfheide (2004).

	μ_π	σ_π	PS_γ	PI_γ	Selected elements of G_γ							
					ψ_1	ψ_2	ρ_R	κ	τ^{-1}	ρ_g	ρ_z	ρ_{gz}
One quarter impulse responses												
$R\to x$	−0.21	0.03	0.03	0.01				0.36	0.66			0.14
$R\to \pi$	−0.82	0.19	0.21	0.02	−0.22	−0.10	0.05	−0.74	0.76	−0.15		−0.16
$R\to r$	0.81	0.08	0.06	0.01	−0.15	−0.06		−0.56	0.58	−0.11		−0.12
$g\to x$	0.55	0.11	0.11	0.07	−0.26	−0.12	0.08	−0.76	0.63	−0.17		−0.17
$g\to \pi$	2.94	0.56	0.38	0.05			−0.06	0.61		0.11	−0.05	0.24
$g\to r$	1.11	0.27	0.23	0.06	0.11	0.06	−0.10	0.77	−0.28	0.12		0.23
$z\to x$	0.42	0.09	0.10	0.03	0.27	0.14	−0.07	0.68	−0.81	0.18		0.14
$z\to \pi$	−2.44	0.47	0.29	0.04	0.06		0.06	−0.47	−0.33	−0.05	0.06	−0.22
$z\to r$	−0.93	0.22	0.16	0.04	−0.05		0.10	−0.69	0.06	−0.09	0.06	−0.23
Four quarter impulse responses												
$R\to x$	−0.03	0.02	0.01	0.05	0.14	0.07	−0.11	0.78	−0.48	0.13		0.23
$R\to \pi$	−0.12	0.03	0.02	0.02		0.06	−0.13	0.53	−0.61	0.06		0.09
$R\to r$	0.13	0.06	0.07	0.03	−0.17	−0.08	0.10	−0.76	0.82	−0.15		−0.20
$g\to x$	0.14	0.06	0.06	0.05	−0.25	−0.11	0.08	−0.77	0.74	−0.15		−0.18
$g\to \pi$	1.05	0.23	0.11	0.07	−0.06	−0.06	−0.06	0.33	0.18	0.14		0.21
$g\to r$	1.57	0.30	0.21	0.09	0.13		−0.07	0.61		0.16		0.24
$z\to x$	0.38	0.05	0.05	0.03	0.19	0.13		0.47	−0.60	0.17	0.13	0.09
$z\to \pi$	−0.58	0.17	0.13	0.02	0.28			0.15	−0.70	0.05	−0.13	
$z\to r$	−1.10	0.19	0.13	0.01				−0.35	−0.52			−0.17
Variance decompositions												
$R\to x$	0.03	0.01	0.01	0.02		−0.06	0.05	−0.47	−0.34	−0.06		−0.19
$g\to x$	0.83	0.06	0.04	0.04	−0.09	−0.09		0.26	0.45		−0.12	0.22
$z\to x$	0.14	0.06	0.03	0.05	0.10	0.11		−0.21	−0.43		0.14	−0.21
$R\to \pi$	0.12	0.04	0.04	0.02	0.23	0.12		0.61	−0.63	0.08		0.11
$g\to \pi$	0.47	0.12	0.09	0.02	0.19			0.42	−0.56	0.21	−0.07	0.15
$z\to \pi$	0.41	0.13	0.13	0.02	−0.24			−0.58	0.70	−0.21	0.07	−0.17
$R\to r$	0.12	0.05	0.04	0.02	−0.25	−0.06		−0.60	0.62	−0.21		−0.15
$g\to r$	0.56	0.17	0.15	0.03	0.25			0.55	−0.62	0.21	−0.09	0.18
$z\to r$	0.33	0.13	0.11	0.03	−0.23			−0.50	0.58	−0.20	0.11	−0.18

Notes: x, π, r are output, inflation and interest rates, respectively, and R, g and z are the monetary policy, demand and supply shocks (orthogonalized such that the supply shock affects $\varepsilon_{z,t}$ in (32) only). The values of PI_γ are conditional (in the sense of Section 3.3) on the prior information about κ and τ^{-1}. The elements of G_γ are derivative based approximations to the change in the posterior mean of the impulse responses and variance decompositions, measured in posterior standard deviations, that results from shifting the prior mean of the parameter in the column heading by one prior standard deviation. Entries of G_γ left blank are smaller than 0.05 in absolute value, and unreported elements of G_γ are uniformly smaller than 0.07 in absolute value.

under prior beliefs of a steeper Phillips curve and higher risk aversion. At the same time, the posterior mean of most impulse responses changes in opposite directions as a function of prior mean increases of κ and τ^{-1}. Thus, a prior mean increase for κ (a steeper Phillips curve) accompanied by a prior mean decrease for τ^{-1} (less risk aversion) leads to substantially larger impulse responses $g\to\pi$, $g\to r$ and $z\to x$ while dampening the impulse responses $R\to r$ and $g\to x$. Interestingly, the effect of such a prior change on the response of inflation to a monetary policy shock is of opposite signs at the one and four quarter horizons.

4.2. Smets and Wouters (2007)

Smets and Wouters (2007) (henceforth, SW) estimate a larger scale log-linearized DSGE model on U.S. postwar data. Their model features sticky prices and wages, habit formation in consumption, variable capital utilization and investment adjustment costs. Table 4 summarizes the dynamics of the seven structural shocks ε_t, which are driven by independent Gaussian innovations η_t. In total, the model has 14 endogenous variables (output, consumption, investment, utilized and installed capital, capacity utilization, hours worked, real wage, rental rate of capital, inflation, nominal interest rate, Tobin's q, and price and wage markups) and is estimated using quarterly data on output growth, consumption growth, investment growth, real wage growth, inflation, hours worked and a nominal interest rate. See SW for further details on the model and the data.

The seven structural shock processes in Table 4 are parametrized by a total of 17 parameters (the "shock" parameters), and the remaining 19 estimated parameters are listed in Table 5 (the "structural" parameters). We adopt the same independent prior on these 36 parameters as Smets and Wouters (2007), except for the seven standard deviations ω of Table 4. There we choose the Gamma distribution on the precision $1/\omega^2$ so that the implied mean and standard deviation of ω is 0.3 and 0.2, respectively, compared to 0.1 and 2.0 of Smets and Wouters (2007). Our tighter prior seems more in line

Table 4
Dynamic specification of structural shocks in Smets and Wouters (2007).

Productivity	$\varepsilon_t^a=\rho_a\varepsilon_{t-1}^a+\eta_t^a$	$\eta_t^a \sim iid\mathcal{N}(0,\omega_a^2)$
Risk premium	$\varepsilon_t^b=\rho_b\varepsilon_{t-1}^b+\eta_t^b$	$\eta_t^b \sim iid\mathcal{N}(0,\omega_b^2)$
Exogenous spending	$\varepsilon_t^g=\rho_g\varepsilon_{t-1}^g+\eta_t^g+\rho_{ga}\eta_t^a$	$\eta_t^g \sim iid\mathcal{N}(0,\omega_g^2)$
Investment	$\varepsilon_t^i=\rho_i\varepsilon_{t-1}^i+\eta_t^i$	$\eta_t^i \sim iid\mathcal{N}(0,\omega_i^2)$
Monetary policy	$\varepsilon_t^r=\rho_r\varepsilon_{t-1}^r+\eta_t^r$	$\eta_t^r \sim iid\mathcal{N}(0,\omega_r^2)$
Price markup	$\varepsilon_t^p=\rho_p\varepsilon_{t-1}^p+\eta_t^p-\mu_p\eta_{t-1}^p$	$\eta_t^p \sim iid\mathcal{N}(0,\omega_p^2)$
Wage markup	$\varepsilon_t^w=\rho_w\varepsilon_{t-1}^w+\eta_t^w-\mu_w\eta_{t-1}^w$	$\eta_t^w \sim iid\mathcal{N}(0,\omega_w^2)$

Notes: The table describes the time series specification of the seven structural shocks in the Smets and Wouters (2007) model.

Table 5
Estimated structural parameters in Smets and Wouters (2007).

φ	Elasticity of capital adjustment cost function
σ_c	Elasticity of intertemporal substitution
h	External habit formation
ξ_w	Calvo probability in labor market
σ_l	Elasticity of labor supply with respect to real wage
ξ_p	Calvo probability in goods market
ι_w	Degree of wage indexation
ι_p	Degree of price indexation
ψ	Normalized elasticity of capital utilization adjustment cost function
Φ	Fixed cost of intermediate good producers
r_π	Inflation coefficient in monetary policy reaction function
ρ	Interest rate smoothing in monetary policy reaction function
r_y	Output gap coefficient in monetary policy reaction function
$r_{\Delta y}$	Short-run feedback of change in output gap in monetary policy function
$\overline{\pi}$	Steady state inflation rate
$100(\beta^{-1}-1)$	Normalized household discount factor
$\overline{l}$	Steady state hours worked
$\overline{\gamma}$	Steady state quarterly growth rate
α	Capital share in production

Notes: The table defines the additional estimated parameters in the Smets and Wouters (2007) model beyond the parameters describing the seven shock processes in Table 4.

with the degree of prior uncertainty for the other estimated parameters,[9] which facilitates the interpretation of the prior sensitivity and prior informativeness measures.

Estimation of the model with Dynare essentially reproduces the posterior results in SW. The three largest eigenvalues of $\Sigma_p^{-1}\Sigma_\pi$ are 1.25, 0.90 and 0.68. Inspection of the eigenvector associated with the largest eigenvalue shows a loading of 0.95 on the steady state inflation rate $\overline{\pi}$, whose posterior variance is also larger than the prior variance. The data thus seems to contain very little information about $\overline{\pi}$ relative to the fairly tight prior, and in the sequel, we condition on the prior information about $\overline{\pi}$ in the computation of PI and PI_γ.[10]

Table 6 reports parameter prior and posterior results. Conditional on the prior information about $\overline{\pi}$, the prior does not seem to play a very important role for the shock parameters, with values of PI of at most 0.22, and often much below. In contrast, 12 of the 17 structural parameters have $\mathrm{PI}\geq 1/3$, indicating that to a substantial degree, posteriors reflect prior information. Especially for ξ_w, σ_l, ξ_p, ρ, r_y and μ_w, a marginal analysis based on the ratio σ_π^2/σ_p^2 substantially understates the role of the prior compared to the joint analysis using PI (cf. inequality (20)).

Tables 7 and 8 report posterior results for key IRs and the one-step ahead VD of output forecasts. It is striking how relatively unimportant the prior seems to be for these functions compared to the structural parameters in Table 6, as indicated by the uniformly small values of PI_γ. To understand why, let $c_\gamma+\nu_\gamma'\theta$ be the linear approximation in the posterior of a given IR or VD (cf. (26)). Since the values of R_γ^2 are quite close to one, IRs and VDs are well approximated by this linear function in the posterior distribution. Now partition ν_γ and θ into shock and structural parameters, respectively, $\nu_\gamma=(\nu_{\gamma\mathrm{sh}}',\nu_{\gamma\mathrm{st}}')'$ and $\theta=(\theta_{\mathrm{sh}}',\theta_{\mathrm{st}}')'$. The relative magnitudes of $\nu_{\gamma\mathrm{sh}}$ and $\nu_{\gamma\mathrm{st}}$ may then serve as indicators for the relative

[9] There might have been some confusion about Dynare's interpretation of the "inverse Gamma distribution" parameters for standard deviations, as the verbal description of the prior on ω on page 592 of Smets and Wouters (2007) (σ in their notation) does not match their actual choice reported in their Table 1B.

[10] Conditionally, $\lambda_{\max}=0.90$. An additional conditioning on the prior information about σ_l (the parameter with largest value of PI in Table 6) reduces $\lambda_{\max}$ further to 0.78, and leaves the values of PI and PI_γ in Tables 6–8 largely unchanged.

Table 6
Parameter prior and posterior results in Smets and Wouters (2007).

	Prior			Posterior		PS	σ_π^2/σ_p^2	PI
Parameter	Shape	μ_p	σ_p	μ_π	σ_π			
φ	$\mathcal{N}$	4.00	1.50	5.74	1.03	0.75	0.48	0.53
σ_c	$\mathcal{N}$	1.50	0.38	1.38	0.13	0.07	0.12	0.16
h	$\mathcal{B}$	0.70	0.10	0.72	0.04	0.02	0.17	0.24
ξ_w	$\mathcal{B}$	0.50	0.10	0.70	0.07	0.06	0.43	0.75
σ_l	$\mathcal{N}$	2.00	0.75	1.83	0.56	0.47	0.56	0.78
ξ_p	$\mathcal{B}$	0.50	0.10	0.65	0.06	0.04	0.31	0.50
ι_w	$\mathcal{B}$	0.50	0.15	0.56	0.12	0.10	0.68	0.73
ι_p	$\mathcal{B}$	0.50	0.15	0.25	0.09	0.06	0.35	0.37
ψ	$\mathcal{B}$	0.50	0.15	0.55	0.11	0.09	0.54	0.64
Φ	$\mathcal{N}$	1.25	0.13	1.61	0.08	0.05	0.39	0.48
r_π	$\mathcal{N}$	1.50	0.25	2.05	0.18	0.14	0.50	0.60
ρ	$\mathcal{B}$	0.75	0.10	0.81	0.02	0.02	0.06	0.15
r_y	$\mathcal{N}$	0.13	0.05	0.09	0.02	0.02	0.21	0.35
$r_{\Delta y}$	$\mathcal{N}$	0.13	0.05	0.22	0.03	0.02	0.31	0.34
$\overline{\pi}$	$\mathcal{G}$	0.63	0.10	0.79	0.11	0.12	1.14	NA
$100(\beta^{-1}-1)$	$\mathcal{G}$	0.25	0.10	0.17	0.06	0.04	0.35	0.37
$\overline{l}$	$\mathcal{N}$	0.00	2.00	0.52	1.09	0.84	0.30	0.32
$\overline{\gamma}$	$\mathcal{N}$	0.40	0.10	0.43	0.01	0.01	0.02	0.03
α	$\mathcal{N}$	0.30	0.05	0.19	0.02	0.01	0.13	0.15
ω_a	$\mathcal{IG}$	0.30	0.20	0.46	0.03	0.01	0.02	0.02
ω_b	$\mathcal{IG}$	0.30	0.20	0.24	0.02	0.01	0.01	0.02
ω_g	$\mathcal{IG}$	0.30	0.20	0.53	0.03	0.01	0.02	0.02
ω_i	$\mathcal{IG}$	0.30	0.20	0.45	0.05	0.02	0.06	0.07
ω_r	$\mathcal{IG}$	0.30	0.20	0.25	0.02	0.00	0.01	0.01
ω_p	$\mathcal{IG}$	0.30	0.20	0.15	0.02	0.01	0.01	0.01
ω_w	$\mathcal{IG}$	0.30	0.20	0.24	0.02	0.01	0.01	0.02
ρ_a	$\mathcal{B}$	0.50	0.20	0.96	0.01	0.00	0.00	0.00
ρ_b	$\mathcal{B}$	0.50	0.20	0.21	0.08	0.04	0.18	0.19
ρ_g	$\mathcal{B}$	0.50	0.20	0.98	0.01	0.00	0.00	0.00
ρ_i	$\mathcal{B}$	0.50	0.20	0.71	0.06	0.03	0.09	0.12
ρ_r	$\mathcal{B}$	0.50	0.20	0.15	0.06	0.03	0.10	0.14
ρ_p	$\mathcal{B}$	0.50	0.20	0.89	0.05	0.03	0.06	0.10
ρ_w	$\mathcal{B}$	0.50	0.20	0.97	0.02	0.01	0.01	0.01
μ_p	$\mathcal{B}$	0.50	0.20	0.73	0.09	0.05	0.18	0.22
μ_w	$\mathcal{B}$	0.50	0.20	0.84	0.06	0.04	0.09	0.22
ρ_{ga}	$\mathcal{N}$	0.50	0.25	0.52	0.09	0.03	0.13	0.13

Notes: $\mathcal{N}$, $\mathcal{B}$ and $\mathcal{G}$ are Normal, Beta and Gamma prior distributions with mean and variance μ_p and σ_p^2, and $\mathcal{IG}$ is a Gamma prior distribution on $1/\omega^2$ that implies a mean and variance of μ_p and σ_p^2 on ω. The entries for PI are conditional on the prior information about $\overline{\pi}$ in the sense of Section 3.3.

importance of θ_{sh} and θ_{st} in the determination of the given IR or VD. In particular, one may measure the magnitude of $\nu_{\gamma\text{sh}}$ and $\nu_{\gamma\text{st}}$ by the prior variance that they imply for $c_\gamma+\nu_\gamma'\theta$, i.e. $\nu_{\gamma\text{sh}}'\Sigma_{p\text{sh}}\nu_{\gamma\text{sh}}$ and $\nu_{\gamma\text{st}}'\Sigma_{p\text{st}}\nu_{\gamma\text{st}}$, respectively, where $\Sigma_{p\text{sh}}$ and $\Sigma_{p\text{st}}$ are the prior variances of θ_{sh} and θ_{st}. The relative importance of $\nu_{\gamma\text{sh}}$ is then given by the relative contribution $r_{\text{sh}}=\nu_{\gamma\text{sh}}'\Sigma_{p\text{sh}}\nu_{\gamma\text{sh}}/\nu_\gamma'\Sigma_p\nu_\gamma$ to the overall prior variance $\nu_\gamma'\Sigma_p\nu_\gamma=\nu_{\gamma\text{sh}}'\Sigma_{p\text{sh}}\nu_{\gamma\text{sh}}+\nu_{\gamma\text{st}}'\Sigma_{p\text{st}}\nu_{\gamma\text{st}}$. The large values of r_{sh} reported in Tables 7 and 8 indicate a relatively dominant role of the shock parameters.[11]

One might conclude from these numbers that the structural parameters θ_{st} simply do not matter much for the value of key IRs and VDs in the SW model. Note, however, that r_{sh} was computed using a linear approximation for the given IR or VD in the *posterior*, which, of course, incorporates likelihood information on θ. A data independent, purely *a priori* measure of the relative importance of the shock parameters for IRs and VDs in the SW model is $r_{\text{sh}}^p=\nu_{\gamma\text{sh}}^{p\prime}\Sigma_{p\text{sh}}\nu_{\gamma\text{sh}}^p/\nu_\gamma^{p\prime}\Sigma_p\nu_\gamma^p$, where $\nu_\gamma^p=(\nu_{\gamma\text{sh}}^{p\prime},\nu_{\gamma\text{st}}^{p\prime})'$ are the coefficients in a linear regression of γ on θ in the prior, with an overall coefficient of determination $R_{\gamma,p}^2=\nu_\gamma^{p\prime}\Sigma_p\nu_\gamma^p$. The lower values of $R_{\gamma,p}^2$ show that some IRs and VDs are highly nonlinear functions of the underlying parameters, making the interpretation of r_{sh}^p more difficult. Nevertheless, r_{sh}^p is typically lower than r_{sh}, and sometimes substantially so. Thus, it is not that IRs and VDs never change substantially as the structural parameters are varied over their prior support. Rather, in the application of the SW model to US postwar data, the likelihood favors (so that the

[11] With the Smets and Wouters (2007) prior on the standard deviations ω of the shock processes, these results become even more pronounced: First, the increase in prior variance directly leads to smaller prior informativeness. Second, the corresponding elements in $\Sigma_{p\text{sh}}$ become larger, further increasing the value of r_{sh}, and thus decreasing PI_γ (cf. Eq. (27) in Section 3.2).

Table 7
Impulse response posterior results in Smets and Wouters (2007).

Series	μ_π	σ_π	PS_γ	PI_γ	R^2_γ	r_{sh}	$R^2_{p,\gamma}$	$r_{p,\mathrm{sh}}$
One quarter responses to productivity shock								
Output	0.33	0.05	0.02	0.05	0.99	0.96	0.72	0.93
Hours	−0.29	0.03	0.01	0.04	0.99	0.96	0.79	0.96
Inflation	−0.06	0.01	0.01	0.11	0.98	0.69	0.68	0.64
Interest rate	−0.06	0.01	0.00	0.03	0.96	0.93	0.06	0.74
Four quarter responses to productivity shock								
Output	0.58	0.06	0.03	0.02	0.98	0.97	0.66	0.90
Hours	−0.07	0.03	0.02	0.14	0.97	0.66	0.47	0.50
Inflation	−0.04	0.01	0.00	0.07	0.93	0.84	0.24	0.43
Interest rate	−0.07	0.01	0.00	0.01	0.94	0.94	0.04	0.80
One quarter responses to monetary policy shock								
output	−0.19	0.02	0.01	0.03	0.98	0.78	0.51	0.67
Hours	−0.13	0.02	0.01	0.03	0.98	0.77	0.51	0.66
Inflation	−0.04	0.01	0.01	0.13	0.96	0.53	0.49	0.69
Interest rate	0.18	0.01	0.00	0.01	0.99	0.98	0.04	0.78
Four quarter responses to monetary policy shock								
Output	−0.34	0.05	0.03	0.04	0.97	0.73	0.51	0.64
Hours	−0.22	0.04	0.02	0.05	0.97	0.73	0.53	0.64
Inflation	−0.05	0.01	0.01	0.09	0.96	0.52	0.47	0.75
Interest rate	0.04	0.01	0.01	0.12	0.97	0.84	0.12	0.82

Notes: Entries for PI_γ are conditional on prior information about $\overline{\pi}$ in the sense of Section 3.3. R^2_γ and $R^2_{p,\gamma}$ are R^2s in a linear regression of the impulse response value on the 36 underlying parameters in the posterior and prior, respectively. Based on these linear approximations for the value of the impulse response, r_{sh} and $r_{p,\mathrm{sh}}$ measure the relative contribution of the "shock" parameters (those that appear in Table 4) in the overall prior uncertainty about the impulse response.

Table 8
Posterior results for decomposition of output variance in Smets and Wouters (2007).

	μ_π	σ_π	PS_γ	PI_γ	R^2_γ	r_{sh}	$R^2_{p,\gamma}$	$r_{p,\mathrm{sh}}$
Productivity	0.16	0.04	0.02	0.04	0.98	0.94	0.47	0.88
Risk premium	0.27	0.03	0.01	0.02	0.98	0.95	0.76	0.72
Exogenous spending	0.36	0.04	0.01	0.03	0.99	0.98	0.70	0.91
Investment	0.13	0.03	0.01	0.05	0.96	0.83	0.65	0.90
Monetary policy	0.05	0.01	0.01	0.02	0.97	0.78	0.70	0.62
Price markup	0.02	0.01	0.00	0.02	0.89	0.88	0.51	0.85
Wage markup	0.00	0.00	0.00	0.06	0.80	0.84	0.09	0.76

Notes: See Table 7.

posterior concentrates on) values of θ where prior uncertainty about the structural parameters is relatively unimportant for the determination of IRs and VDs.

One can easily imagine that the structural parameters enter other functions of θ of interest, such as the welfare effects of alternative monetary policy regimes, in a more prominent way, and the important role of the prior for the structural parameters would then translate into a correspondingly important role for the posterior of such functions.

5. Conclusion

This paper develops measures that shed some light on the role of the prior and likelihood for posterior results in large Bayesian models. The two suggested statistics are based on the derivative matrix of the posterior mean relative to a specific parametric variation in the prior distribution, which turns out to be a simple function of the posterior and prior covariance matrices. It is thus entirely straightforward to compute the measures from the output of standard posterior samplers.

The suggested prior informativeness and prior sensitivity measures are scalar summary statistics. They cannot reflect all features of the high-dimensional likelihood and its interaction with the prior, and one can imagine other useful statistics that highlight different aspects. At the same time, the exponential family embedding of the baseline prior is arguably an attractive starting point for studying the role of the prior information: It leads to a tight link between local prior sensitivity and prior and posterior spread as measured by the second moment, which facilitates computation and interpretation. In addition, it is shown that reasonable axiomatic restrictions on scalar summary statistics about overall prior informativeness based on this embedding lead to the suggested measure.

Acknowledgments

I thank the editor, the associate editor and an anonymous referee for helpful suggestions and comments. I am indebted to Marco Del Negro for his very useful discussion at the 2010 DSGE workshop in Atlanta. I also would like to thank Fabio Canova, Diogo Guillen, Stefan Hoderlein, Arthur Lewbel, Serena Ng, Andriy Norets, Giorgio Primiceri, Chris Sims, Sharon Traiberman and participants at seminars at the New York Federal Reserve Bank, Harvard/MIT, and at the Atlanta workshop for helpful comments.

Appendix A. Supplementary derivations

Supplementary derivations associated with this article can be found in the online version at http://dx.doi.org.10.1016/j.jmoneco.2012.09.003.

References

An, S., Schorfheide, F., 2007. Bayesian analysis of DSGE models. Econometric Reviews 26, 113–172.
Andrle, M., 2010. A Note on Identification Patterns in DSGE Models. ECB Working Paper Series 1235.
Basu, S., Jammalamadaka, S.R., Liu, W., 1996. Local posterior robustness with parametric priors: maximum and average sensitivity. In: Heidbreder, G.R. (Ed.), Maximum Entropy and Bayesian Methods: Santa Barbara, California, U.S.A., 1993. Kluwer Dordrecht, The Netherlands, pp. 97–106.
Berger, J.O., 1994. An overview of robust Bayesian analysis. Test 3, 5–124.
Canova, F., Sala, L., 2009. Back to square one: identification issues in DSGE models. Journal of Monetary Economics 56, 431–449.
Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII US. Review of Economic Dynamics 8, 262–302.
del Negro, M., Schorfheide, F., 2004. Priors from general equilibrium models for VARs. International Economic Review 45, 643–673.
Ebrahimi, N., Maasoumi, E., Soofi, E.S., 1999. Measuring informativeness of data by entropy and variance. In: Slottje, D. (Ed.), Advances in Econometrics: Income Distribution and Methodology of Science, Essays in Honor of Camilo Dagumi, Springer Verlag.
Fernández-Villaverde, J., Rubio-Ramírez, J.F., 2007. Estimating macroeconomic models: a likelihood approach. Review of Economic Studies 74, 1059–1087.
Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramírez, J.F., 2010. Fortune or Virtue: Time-Variant Volatilities Versus Parameter Drifting In U.S. Data. NBER Working Paper No. 15928.
Geweke, J., 1999. Simulation methods for model criticism and robustness analysis In: Berger, J.O., Bernardo, J.M., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 6. Oxford University Press, Oxford, pp. 275–299.
Goel, P.K., 1983. Information measures and Bayesian hierarchical models. Journal of the American Statistical Association 78, 408–410.
Goel, P.K., DeGroot, M.H., 1981. Information about hyperparameters in hierarchical models. Journal of the American Statistical Association 76, 140–147.
Gustafson, P., 2000. Local robustness in Bayesian analysis. In: Insua, D.R., Ruggeri, F. (Eds.), Robust Bayesian Analysis, Springer-Verlag, New York, pp. 71–88.
Iskrev, N., 2010. Local identification in DSGE models. Journal of Monetary Economics 57, 189–202.
Justiniano, A., Primiceri, G.E., 2008. The time-varying volatility of macroeconomic fluctuations. American Economic Review 98, 604–641.
Kim, C.-J., Nelson, C.R., 1999. Has the economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. The Review of Economics and Statistics 81, 608–616.
Kitagawa, T., 1934. On some class of weighted means. Proceedings of the Physico-Mathematical Society of Japan 16, 311–320.
Kolmogorov, A., 1930. Sur la notion de la moyenne. Atti della Reale Accademia dei Lincei, Rendiconti 6, 388–391.
Komunjer, I., Ng, S., 2009. Dynamic Identification of DSGE Models. Discussion paper, U.C. San Diego.
Leamer, E.E., 1973. Multicollinearity: a Bayesian interpretation. The Review of Economics and Statistics 55, 371–380.
Lindley, D.V., 1955. On a measure of information provided by an experiment. Annals of Mathematical Statistics 27, 986–1005.
Lubik, T., Schorfheide, F., 2004. Testing for indeterminacy: an application to U.S. monetary policy. American Economic Review 94, 190–217.
Millar, R.B., 2004. Sensitivity of bayes estimators to hyper-parameters with an application to maximum yield from fisheries. Biometrics 60, 536–542.
Nagumo, M., 1930. über eine klasse der mittelwerte. Japan Journal of Mathematics 7, 71–79.
Perez, C.J., Martin, J., Rufo, M.J., 2006. MCMC-based local parametric sensitivity estimations. Computational Statistics & Data Analysis 51, 823–835.
Poirier, D.J., 1998. Revising beliefs in nonidentified models. Econometric Theory 14, 483–509.
Pratt, J.W., 1964. Risk aversion in the small and in the large. Econometrica 32, 122–136.
Primiceri, G.E., 2005. Time varying structural vector autoregressions and monetary policy. The Review of Economic Studies 72, 821–852.
Rothenberg, T.J., 1971. Identification in parametric models. Econometrica 39, 577–591.
Sims, C.A., Zha, T., 2006. Were there regime switches in U.S. monetary policy. American Economic Review 96, 54–81.
Sivaganesan, S., 2000. Global and local robustness approaches: uses and limitations. In: Insua, D.R., Ruggeri, F. (Eds.), Robust Bayesian Analysis, Springer-Verlag, New York, pp. 59–108.
Smets, F., Wouters, R., 2003. An estimated dynamic stochastic general equilibrium model of the euro area. Journal of the European Economic Association 1, 1123–1175.
Smets, F., Wouters, R., 2007. Shocks and frictions in U.S. business cycles: a Bayesian DSGE approach. American Economic Review 97, 586–606.
Soofi, E.S., 1990. Effects of collinearity on information about regression coefficients. Journal of Econometrics 43, 255–274.
Traum, N., Yang, S.S., 2010. Does Government Debt Crowd Out Investment? A Bayesian DSGE Approach. CBO Working Paper Series.