# Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models*

Ulrich K. Müller
Princeton University
Department of Economics
Princeton, NJ, 08544

First Version: May 2010
This Version: September 2010

**Abstract**

The paper derives measures of prior sensitivity and prior informativeness for posterior results in large Bayesian models 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. An application to Smets and Wouters' (2007) dynamic stochastic general equilibrium model shows that for many structural parameters, the prior is very informative, and posterior means are quite sensitive to changes in prior means. In contrast, the prior plays a much less important role for key impulse responses and variance decompositions.

**JEL classification:** C11, C52, E3, E5

**Keywords:** Local sensitivity analysis, weak identification, dynamic stochastic general equilibrium model

*I thank Andriy Norets, Chris Sims, Stefan Hoderlein, Arthur Lewbel and participants at a seminar at the New York Federal Reserve Bank for helpful discussions.

# 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 (Sims and Zha (2006), Smets and Wouters (2003, 2007), Fernandez-Villaverde and Rubio-Ramirez (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), among others). These models typically contain a moderate or large number of unknown parameters, requiring the specification of a corresponding prior. The empirical conclusions are based on the center and spread of the resulting posterior, which are estimated using Markov Chain Monte Carlo (MCMC) methods.

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 sometimes even necessary in large scale macroeconomic applications (cf. An and Schorfheide (2007)). 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 more implicit effects of the typical product prior for the whole parameter vector are 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 subjective prior beliefs.

Current standard practice is two provide to 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. These statistics are not necessarily very informative about the relative importance of the prior and likelihood.

*Example 1:* We observe data from the model $Y \sim \mathcal{N}(\theta, \Sigma)$ with $\Sigma$ known, and are interested in the first element of the $k$ dimensional vector $\theta$. Suppose $Y_i = \theta_i + \frac{1}{6}\varepsilon_i + 6\varepsilon_0$,

$i = 1, \cdots, k$ where $\varepsilon_i \sim iid\mathcal{N}(0,1)$ and $k = 40$. With a $\mathcal{N}(0, I_k)$ prior for $\theta$, and conditional on the realization $Y = Y^0 = (Y_1^0, \cdots, Y_k^0)'$ arising from $\varepsilon_i = (-1)^i$ for $i = 0, \cdots, k$, the posterior of $\theta_1$ is $\theta_1 | Y = Y^0 \sim \mathcal{N}(-0.158, 0.051)$. With the prior changed to $\theta \sim \mathcal{N}(0, 2I_k)$, we obtain instead $\theta_1 | Y = Y^0 \sim \mathcal{N}(-0.156, 0.077)$.

For both priors, the posterior distribution of $\theta_1$ is very different from the $\mathcal{N}(0,1)$ prior, and especially the posterior mean only varies moderately across the two priors. Superficially, this suggests that the prior only plays a modest role in the posterior of $\theta_1$. But without knowledge of the last $k-1$ elements of $\theta$, only $Y_1 \sim \mathcal{N}(\theta_1, 36.03)$ contains useful informative about $\theta_1$ (note that the MLE is $\hat{\theta}_1 = Y_1$). The data thus contains only a very limited amount of information about $\theta_1$, and the sharp posterior $\theta_1 | Y = Y^0 \sim \mathcal{N}(-0.158, 0.051)$ merely reflects interactions between $\Sigma$ and the prior on the remaining elements of $\theta$ (for comparison, $\theta_1 | Y_1 = Y_1^0 \sim \mathcal{N}(0.158, 0.973)$ under the standard normal prior for $\theta_1$). ▲

The goal of this paper is 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 prior for individual parameters?

We approach both questions 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. Also, 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 original prior distribution in an exponential family. This choice has a certain theoretical appeal, as discussed below. But what is more, this embedding leads to an simple expression for the derivative matrix as a function of the prior and posterior variance-covariance matrices. The derived prior sensitivity and prior informativeness measures thus become trivial to compute in practice.

More concretely, with a scalar unknown parameter, the suggested prior sensitivity measure and prior informativeness measure are proportional to the derivative of the posterior mean with respect to the prior mean. The derivative usually takes on values between zero and one, with values close to one indicating both high sensitivity and strong informativeness of the prior. What is more, the derivative also has an interpretation as the approximative fraction of prior information to total (posterior) information.

In the multivariate parameter case it makes sense to normalize the parameters to have identity prior covariance matrix to ensure comparability of their scales. A natural measure of the prior sensitivity for a particular parameter is then given by the Euclidian norm of its derivative vector—this approximates the largest change of the posterior mean that can be induced by changing the prior mean by a vector of unit length. In contrast to the scalar case, however, this is no longer an appropriate prior informativeness measure. To see why, suppose the parameter of interest is the sum of two parameters that are independent in both the prior and the likelihood. If for one of them, the prior is not at all informative (a derivative close to zero), and for the other one, the prior is very informative (a derivative close to one), then the prior is also very informative for the sum. Alternatively, if the prior is moderately informative for both, then this is also the case for the sum. Yet the prior sensitivity measure—the Euclidian norm of the two derivatives—takes on similar values in both cases.

To obtain a more suitable prior informativeness measure, we identify the unique function of the derivative matrix that satisfies a set of reasonable axiomatic requirements. One of the requirements is that the informativeness measure is compatible with the fraction of information interpretation mentioned above. The end result are two easily computed statistics: the prior sensitivity PS (approximately) measures the maximal change of the posterior mean when the prior means are varied by the multivariate analogue of one standard deviation. The prior informativeness measure PI $\in [0,1]$ summarizes the relative amount of prior information in the posterior, and thus quantifies to which degree the posterior results are driven by data information. The measure PI can thus 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.

*Example 1, ctd:* Under the $\theta \sim \mathcal{N}(0, I_k)$ prior, the prior sensitivity measure for $\theta_1$ equals PS $= 0.160$, and the prior informativeness measure equals PI $= 0.973$. Both indicate an important role for the prior in the posterior $\theta_1|Y \sim \mathcal{N}(-0.158, 0.051)$: a change of the prior mean of unit length can induce a change in the posterior mean of $0.160/\sqrt{0.051} = 0.706$ posterior standard deviations, and the prior informativeness measure PI is close to unity, indicating a dominant role of the prior for the posterior of $\theta_1$. ▲

An application to the Smets and Wouters (2007) DSGE model shows that the prior is very informative for many of the structural parameters, and posterior results are quite sensitive to prior mean changes. In contrast, the parameters describing the shock processes are much more pinned down by the likelihood. Interestingly, the prior also plays only a moderate role for the posterior results of key impulse responses and variance decompositions, at least when an additional two structural parameters are fixed compared to the Smets and Wouters (2007) specification. The reason is that to a large degree, the shock parameters determine the value of both the impulse responses and variance decompositions, while the structural parameters have a comparatively minor influence. Some of the important implications of this DSGE model are thus not driven by the prior.

Although the measures PS and PI are based on the same derivative matrix and may thus 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, Jammalamadaka, and Liu (1996) and Perez, Martin, and Rufo (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. The prior sensitivity measure PS thus essentially becomes a special case of their analyses. The contribution of the present paper with regard to the measure PS merely consists of the suggestion of the exponential family embedding, and of the normalization by the prior derivative matrix.

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 very recent literature that, initiated by Canova and Sala (2009), discusses identification of modern DSGE models, such as Iskrev (2010b) 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.

The remainder of the paper is organized as follows. Section 2 derives PS and PI for both the scalar and vector parameter case. Section 3 discusses inequalities for PS and PI, analogue measures for functions of the original parameters, implementation issues, a detailed comparison with Rothenberg's (1971) definition of identifiability, and the relationship to McCulloch's (1989) study of prior robustness. Section 4 contains the empirical results for the Smets and Wouters (2007) model, and Section 5 concludes.

# 2 Derivation of Measures

## 2.1 Scalar Parameter

Denote the parameter of interest by $\theta$, which is a scalar in this subsection. Let the prior density be $p$, with mean $\mu_p = E_p[\theta]$ and variance $\sigma_p^2 = V_p[\theta]$, where here and below subscripts of the expectation and variance indicate the measure of integration. The posterior density $\pi$ is derived from $p$ and the likelihood function $l$ via $\pi(\theta) = p(\theta)l(\theta)/\int p(h)l(h)dh$.

Now embed the prior density in a family $p_\alpha$ with $p_0 = p$, score function $s_\alpha(\theta) = d\ln p_\alpha(\theta)/d\alpha$ and prior mean $\int \theta p_\alpha(\theta)d\theta = \mu_p + \alpha$. The posterior mean as a function of $\alpha$ 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, Martin, and Rufo (2006) for details),

$$\frac{d\mu_\pi(\alpha)}{d\alpha}|_{\alpha=0} = E_\pi[(\theta - E_\pi[\theta])s_0(\theta)] \qquad (1)$$

which is recognized as the posterior covariance between $\theta$ 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.

With a large sample size $n$ and the true value of $\theta$ equal to $\theta_0$, the Bernstein-von Mises Theorem yields that the posterior is approximately Gaussian, centered at the MLE $\hat{\theta}$ and with variance of approximately $\sigma^2(\theta_0)/n$, the inverse of second derivative of the likelihood evaluated at $\theta_0$. From the first order Taylor expansion $s_0(\theta) \approx s_0(\hat{\theta}) + (\theta - \hat{\theta})\kappa(\hat{\theta})$ with $\kappa(\theta) = ds_0(\theta)/d\theta$, (1) will usually be close to $\sigma^2(\theta_0)\kappa(\hat{\theta})/n$ in large samples. As expected,

(1) is of the order $O(n^{-1})$ and thus small in large samples, but the exact value also depends on $\kappa(\hat{\theta}) \approx \kappa(\theta_0)$. So even if $\sigma^2(\theta_{01}) = \sigma^2(\theta_{02})$ for some $\theta_{01} \neq \theta_{02}$, so that the amount of likelihood information is identical for $\theta_0 = \theta_{01}$ and $\theta_0 = \theta_{02}$, the same prior can lead to very different values of the derivative (1). This might be considered undesirable for a measure that seeks to disentangle prior and likelihood information, as the amount of prior and likelihood information are arguably the same for both values of $\theta_0$. A corresponding equality of (1) can only be ensured if $\kappa(\theta)$ does not depend on $\theta$. For this to happen, the function $s_0(\theta)$ has to be linear in $\theta$, as in the exponential family embedding

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

with cumulant function $C(\alpha) = \log \int p(\theta) \exp[\alpha\theta/\sigma_p^2] d\theta$ 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] For Gaussian $p$, (2) simply corresponds to a normal with the same variance and mean shifted by $\alpha$. More generally, the derivative of the variance of $p_\alpha$ at $\alpha = 0$ under (2) equals $E_p[(\theta - \mu_p)^3]/\sigma_p^2$, which is small compared to $\sigma_p^2$ as long as $p$ is not too skewed. Also, for $\alpha > 0$, $p_\alpha(\theta) > p(\theta)$ for all $\theta > \mu_p$, and $p_\alpha(\theta) < p(\theta)$ for all $\theta < \mu_p$ (and vice versa for $\alpha < 0$). The pivot of the exponential tilting in (2) is thus always the original mean $\mu_p$.

Under (2), a calculation yields

$$\frac{d\mu_\pi(\alpha)}{d\alpha}|_{\alpha=0} = J = \frac{V_\pi[\theta]}{\sigma_p^2} \tag{3}$$

so that the derivative simply becomes the ratio of posterior to prior variance. A natural measure for the prior sensitivity might be the linear approximation to the change of the posterior mean that can be induced by a prior mean of one prior standard deviation. With (3), this results in

$$\text{PS} = \sigma_p J = \frac{V_\pi[\theta]}{\sigma_p}.$$

Typically, the derivative $J$ takes on values between zero and one, in which case it usefully measures the relative prior informativeness, PI $= J$: When the data perfectly pins down $\theta$,

[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 $p_\alpha(\theta) = 2c(\alpha)p(\theta)/(1 + \exp\left[-2\alpha\theta/\sigma_p^2\right])$ where $c(\alpha) > 0$ ensures that $\int p_\alpha(\theta)d\theta = 1$ for all $\alpha$. 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 $d\mu_\pi(\alpha)/d\alpha|_{\alpha=0}$.

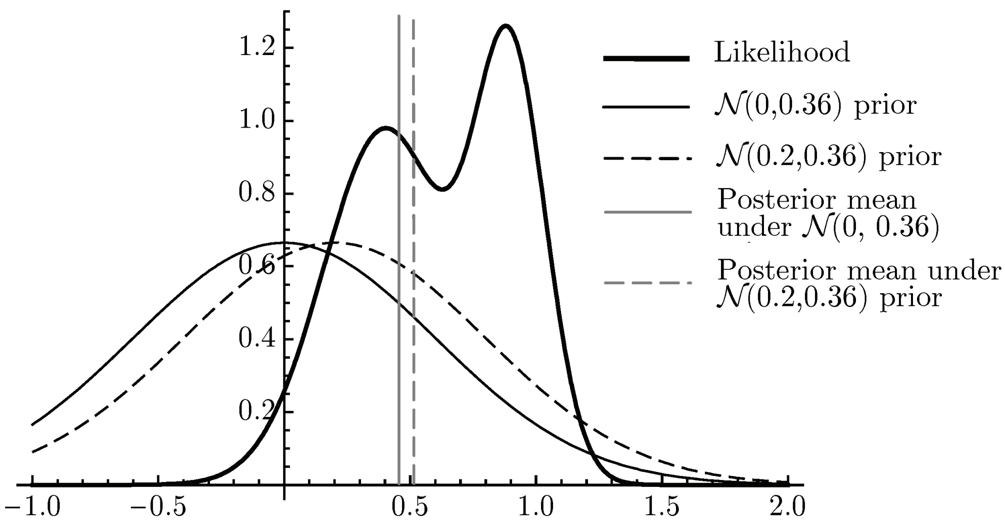


Figure 1: Example of posterior mean as function of prior mean

we have $V_\pi[\theta] = 0$, and correspondingly, prior informativeness is zero, $\mathrm{PI} = 0$. In the other extreme, with a perfectly flat likelihood, the posterior is equal to the prior, and $\mathrm{PI} = 1$. 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 we define

$$\mathrm{PI} = \min(J, 1). \tag{4}$$

If $\mathrm{PI} = 1/3$, say, then local changes of the prior mean are pushed through to the posterior mean at the rate of $1/3$. So in a loose sense, one might say that $1/3$ of the location information in the posterior stems from the location information in the prior, and the remaining $2/3$ is likelihood information. Somewhat more precisely, suppose that both the log-density and the log-likelihood are quadratic in $\theta$, i.e. $l(\theta) \propto \exp[-\frac{1}{2}(\theta - \mu_l)/\sigma_l^2]$ (as arising from observing $\theta$ with Gaussian noise of variance $\sigma_l^2$), and $p_0 \sim \mathcal{N}(\mu_p, \sigma_p^2)$, so that $p_\alpha \sim \mathcal{N}(\mu_p + \alpha, \sigma_p^2)$ under (2). By a standard calculation, the posterior mean then satisfies

$$\mu_\pi(\alpha) = w(\mu_p + \alpha) + (1 - w)\mu_l \qquad \text{with } w = \frac{\sigma_p^{-2}}{\sigma_p^{-2} + \sigma_l^{-2}}. \tag{5}$$

With the precision $\sigma_p^{-2}$ and $\sigma_l^{-2}$ measuring the amount of information in the prior and likelihood, 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 log-prior and log-likelihood are approximately quadratic, then this interpretation will typically remain a useful approximation. Figure 1 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 $E[Y] = \theta$. Here $w$ with $\sigma_l^2 = V[Y]$ evaluates to $w = 0.224$, and PI $= V_\pi[\theta]/0.36 = 0.249$ (with a range of PI $\in [0.169, 0.252]$ for $-1.5 \leq \mu_p \leq 1.5$), even though the log-likelihood is far from quadratic. 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, instead merely summarize its local characteristics, which would be quite misleading in the example of Figure 1. It is also worth noting that the value of PI is fully determined by the likelihood and prior, so that it adheres to the likelihood principle. This desirable feature is not shared by the Fisher Information, considered by Iskrev (2008, 2010a, 2010b), Traum and Yang (2010) and Andrle (2010), which averages over the amount of sample information in samples that did not realize.[2]

## 2.2 Vector Parameter

Let $\theta$ be the $k \times 1$ vector of unknown parameters, denote by $V_p[\theta]$ the full-rank variance-covariance matrix of the prior $p$ on $\theta$, and let $\pi$ be the posterior density computed from the likelihood $l$. It is useful to initially determine the prior sensitivity and prior informativeness of the normalized parameter $\theta^* = G\theta$, where the $k \times k$ matrix $G$ satisfies $GV_p[\theta]G' = I_k$. Denote the implied prior and posterior of $\theta^*$ by $p^*(\theta^*) = |G|^{-1}p(G^{-1}\theta^*)$ and $\pi^*$, respectively. The first two moments of the prior on $\theta^*$ then are $E_{p^*}[\theta^*] = GE_p[\theta]$ and $V_{p^*}[\theta^*] = I_k$, so that the prior scale of the elements in $\theta^*$ is identical. In analogy to (2), consider the exponential tilting

$$p^*_\alpha(\theta^*) = \exp[\alpha^{*\prime}\theta^* - C(\alpha^*)]p^*(\theta^*) \tag{6}$$

with cumulant function $C(\alpha^*) = \ln \int \exp[\alpha^{*\prime}\theta^*]p(\theta^*)d\theta^*$, which exists for small enough $||\alpha^*||$ whenever the moment generating function of $p$ exists, at least in a neighborhood of zero. Note that the mean of $\theta = G^{-1}\theta^*$ under (6) equals $E_p[\theta] + G^{-1}\alpha^*$. One might worry that

[2] In the example of Figure 1, suppose that in addition to $Y$, we observe 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 we obtained about $\theta$. In contrast, the Fisher Information in this experiment is the probability weighted average of these two values for $\sigma_l^{-2}$.

the implied density for $\theta^*$ depends on the exact choice of the matrix $G$, because in general, changing the means of a random vector and rotating it or changing the means of the rotated vector does not amount to the same transformation of the density. A further advantage of (6) is that these operations commute in the exponential family embedding: If two matrices $G_1$ and $G_2$ both satisfy $G_i V_p[\theta] G_i' = I_k$, $i = 1, 2$, then $G_2 = QG_1$ for an orthogonal matrix $Q$, and the density of $\theta$ implied by (6) with $G = G_1$ and $\alpha^* = \alpha_1^*$ is the same as that implied by $G = G_2$ and $\alpha^* = Q\alpha_1^*$.[3]

Let $\mu_\pi^*(\alpha^*)$ be the posterior mean of $\theta^*$ under the prior (6). The $k \times k$ derivative matrix then is the symmetric matrix

$$J^* = \frac{\partial \mu_\pi^*(\alpha^*)}{\partial \alpha^{*\prime}}|_{\alpha^*=0} = V_{\pi^*}[\theta^*] = GV_\pi[\theta]G'.$$

Consider first the prior sensitivity for a particular linear combination of the normalized parameters $v^{*\prime}\theta^*$, with posterior mean derivative relative to $\alpha^{*\prime}$ equal to $v^{*\prime}J^*$. The largest rate of change of the posterior mean is (cf. Corollary 1 of Basu, Jammalamadaka, and Liu (1996))

$$\text{PS} = \max_{||\alpha^*||=1} v^{*\prime}J^*\alpha^* = \sqrt{v^{*\prime}J^{*2}v^*}$$

with the worst-case direction $\alpha^* = J^*v^*/||J^*v^*||$. So if the interest is in the linear combination $v'\theta$ in the original parametrization, then $v'\theta = (G^{-1\prime}v)'\theta^*$ yields

$$\text{PS} = \sqrt{v'G^{-1}J^{*2}G^{-1\prime}v} = \sqrt{v'V_\pi[\theta]V_p[\theta]^{-1}V_\pi[\theta]v}. \tag{7}$$

This is alternatively recognized as the local approximation to the largest change in the posterior mean of $\theta$ that can be induced by varying the prior mean $\alpha = G^{-1}\alpha^*$ by one unit in the Mahalanobis metric $\sqrt{\alpha' V_p[\theta]^{-1}\alpha}$, the multivariate analogue of "one prior standard deviation", with worst-case direction

$$\alpha = \frac{G^{-1}J^*v^*}{||J^*v^*||} = \frac{V_\pi[\theta]v}{\sqrt{v'V_\pi[\theta]V_p[\theta]^{-1}V_\pi[\theta]v}}. \tag{8}$$

Now turn to measuring prior informativeness in the vector case. Let's first consider the straightforward case where the linear combination $v^{*\prime}\theta^*$ of interest is independent of all other parameters $w^{*\prime}\theta^*$ with $w^{*\prime}v^* = 0$ in both the the likelihood and the prior $p^*$. The

[3] If the moment generating function does not exist, one can again define an alternative embedding analogously to the scalar case, with identical $\partial \ln p_\alpha^*(\theta)/\partial\alpha^*|_{\alpha^*=0}$, and rotation and mean change still commute.

independence then also holds in the prior (6) with $\alpha^* = a \cdot v^*$, and the derivative of the posterior mean of $\theta^*$ with respect to the scalar $a$ satisfies

$$J^* v^* = \lambda v^* \tag{9}$$

because prior mean changes in the direction of $v^*$ only affects the posterior of $v^{*\prime}\theta^*$ (but not the posterior of $w^{*\prime}\theta^*$, for any $w^{*\prime}v^* = 0$). Given the independence, the prior informativeness measure of $v^{*\prime}\theta^*$ should not depend on properties of the prior and likelihood of $w^{*\prime}\theta^*$. The problem is thus effectively one-dimensional, and $\min(\lambda, 1)$ is the natural measure for the prior informativeness of $v^{*\prime}\theta^*$, in accordance with the discussion in Section 2.1. More generally, whenever $v^*$ is an eigenvector of $J^*$ as in (9), a local prior mean change yields a local posterior mean change in the same direction, and it makes sense to use the eigenvalue (the "push through" rate) as the measure of prior informativeness of $v^{*\prime}\theta^*$.

Since $J^*$ is symmetric, it has a spectral decomposition $J^* = \sum_{i=1}^k \lambda_i q_i q_i'$, where the eigenvectors $q_i$ are linearly independent and of unit length. Any vector $v^*$ can be re-expressed in the coordinate system defined by the $q_i$, $v^* = \sum_{i=1}^k \omega_i q_i$. By the argument above, $\min(\lambda_i, 1)$ is the prior informativeness for the $k$ axes, but it remains to determine an appropriate value for PI for non-degenerate directions $\omega = (\omega_1, \cdots, \omega_k)'$. We approach this issue by imposing conditions on potential functions $\overline{\mathrm{PI}}_k$ that map $\{\omega_i, \lambda_i, q_i\}_{i=1}^k$ to a measure of prior informativeness.

**Condition 1** *For any integers $k$ and $m < k$, and any values of $\{\{\omega_i, \lambda_i, q_i\}_{i=1}^k\}$:*

*(a)* $\overline{\mathrm{PI}}_k((\omega_1, \lambda_1, q_1), \cdots, (\omega_k, \lambda_k, q_k)) = \mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2 \\ \lambda_1\end{pmatrix}, \cdots, \begin{pmatrix}\omega_k^2 \\ \lambda_k\end{pmatrix}\right)$;

*(b)* $\mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2 \\ \lambda_1\end{pmatrix}, \cdots, \begin{pmatrix}\omega_k^2 \\ \lambda_k\end{pmatrix}\right) = \mathrm{PI}_k\left(\begin{pmatrix}\omega_{p(1)}^2 \\ \lambda_{p(1)}\end{pmatrix}, \cdots, \begin{pmatrix}\omega_{p(k)}^2 \\ \lambda_{p(k)}\end{pmatrix}\right)$ *for any permutation $p$ of the first $k$ natural numbers;*

*(c)* $\mathrm{PI}_1\left(\begin{pmatrix}\omega_1^2 \\ \lambda_1\end{pmatrix}\right) = \min(\lambda_1, 1)$;

*(d)* $\mathrm{PI}_{k+1}\left(\begin{pmatrix}\omega_1^2 \\ \lambda_1\end{pmatrix}, \cdots, \begin{pmatrix}\omega_k^2 \\ \lambda_k\end{pmatrix}, \begin{pmatrix}0 \\ \lambda_{k+1}\end{pmatrix}\right) = \mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2 \\ \lambda_1\end{pmatrix}, \cdots, \begin{pmatrix}\omega_k^2 \\ \lambda_k\end{pmatrix}\right)$;

*(e)* $\mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2 \\ \lambda_1\end{pmatrix}, \cdots, \begin{pmatrix}\omega_k^2 \\ \lambda_k\end{pmatrix}\right)$ *has range $[0, 1]$, is weakly increasing in $\lambda_1$, and, for $\omega_1^2 > 0$ and $\max_{i \leq k} \lambda_i < 1$, is continuous in $(\omega_1^2, \lambda_1)$ and strictly increasing and differentiable in $\lambda_1$;*

*(f)* $\mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2 \\ \lambda_1\end{pmatrix}, \cdots, \begin{pmatrix}\omega_{k-2}^2 \\ \lambda_{k-2}\end{pmatrix}, \begin{pmatrix}\omega_{k-1}^2 \\ \lambda_k\end{pmatrix}, \begin{pmatrix}\omega_k^2 \\ \lambda_k\end{pmatrix}\right)$

$$= \mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2 \\ \lambda_1\end{pmatrix}, \cdots, \begin{pmatrix}\omega_{k-2}^2 \\ \lambda_{k-2}\end{pmatrix}, \begin{pmatrix}\omega_{k-1}^2 + \omega_k^2 \\ \lambda_k\end{pmatrix}, \begin{pmatrix}0 \\ \lambda_k\end{pmatrix}\right);$$

*(g)* $\mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2\\ \lambda_1\end{pmatrix},\cdots,\begin{pmatrix}\omega_k^2\\ \lambda_k\end{pmatrix}\right)=\mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2\\ \bar\lambda_m\end{pmatrix},\cdots,\begin{pmatrix}\omega_m^2\\ \bar\lambda_m\end{pmatrix},\begin{pmatrix}\omega_{m+1}^2\\ \lambda_{m+1}\end{pmatrix},\cdots,\begin{pmatrix}\omega_k^2\\ \lambda_k\end{pmatrix}\right)$ *for* $\bar\lambda_m = \mathrm{PI}_m\left(\begin{pmatrix}\omega_1^2\\ \lambda_1\end{pmatrix},\cdots,\begin{pmatrix}\omega_m^2\\ \lambda_m\end{pmatrix}\right)$.

Parts (a)-(e) are probably uncontroversial: part (a) impose that $\overline{\mathrm{PI}}_k$ does not vary with either the sign of $\omega_i$ or the eigenvectors $q_i$; part (b) imposes permutation invariance relative to the order of the eigenvalue/eigenvector pairs; part (c) says that $\min(\lambda_1, 1)$ is the prior informativeness for the first axis; part (d) asserts that zero weight on some axis amounts to a dimension reduction; part (e) imposes continuity, differentiability and monotonicity in the prior informativeness measures of the $k$ axes. For part (f), if the prior is equally informative in two orthogonal directions, then the relative loading on the two should not matter for the overall prior informativeness. Finally, part (g) is an internal consistency requirement: if $\bar\lambda_m$ accurately summarizes prior informativeness of the direction $(\omega_1,\cdots,\omega_m,0,\cdots,0)'$ (which may equivalently be computed via $\mathrm{PI}_m$, using repeatedly the second part of condition (d)), then the prior informativeness of the direction $(\omega_1,\cdots,\omega_m,\omega_{m+1},\cdots,\omega_k)'$ should remain unchanged when $\lambda_1,\cdots,\lambda_m$ are all replaced by $\bar\lambda_m$.

As an additional constraint, one might find it desirable if $\overline{\mathrm{PI}}_k$ was compatible with the fraction of information interpretation of PI discussed at the end of Section 2.1. Specifically, suppose $k=2$ with $\theta^*=(\theta_1^*,\theta_2^*)'$, $\theta_1^*$ and $\theta_2^*$ are independent in both the prior and likelihood, and the log-prior density and log-likelihood of $\theta_1^*$ are (at least approximately) quadratic. The prior informativeness $\lambda_1$ of $\theta_1^*$ then has the interpretation of the fraction of prior information for the posterior mean, $\lambda_1 = 1/(1+\sigma_{1l}^{-2}) < 1$, where $\sigma_{1l}^{-2}$ is the (approximate) leading coefficient in the log-likelihood of $\theta_1^*$, as in the example of Figure 1. Further, assume that $\theta_2^*$ is perfectly identified through the likelihood, $\lambda_2 = 0$. The likelihood information about the parameter $\bar\theta_{12}^* = \theta_1^* + \theta_2^*$ then is as good as that about $\theta_1^*$ (the log-likelihood of $\bar\theta_{12}^*$ is identical to that of $\theta_1^*$, which is approximately quadratic with coefficient $\sigma_{1l}^{-2}$). At the same time, with the prior variances of $\theta_1^*$ and $\theta_2^*$ normalized to unity, the prior information (=precision) on $\bar\theta_{12}^*$ is equal to 1/2. The ratio of prior information and posterior information thus becomes

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

leading to the following condition on $\mathrm{PI}_2$.

**Condition 2** *For* $\lambda_1 < 1$, $\mathrm{PI}_2\left(\begin{pmatrix}1\\ \lambda_1\end{pmatrix},\begin{pmatrix}1\\ 0\end{pmatrix}\right) = \frac{\lambda_1}{2-\lambda_1}$.

The following Theorem identifies the unique function that satisfies Conditions 1 and 2.

**Theorem 1** *Under Conditions 1 and 2,*

$$\overline{\mathrm{PI}}_k((\omega_1,\lambda_1,q_1),\cdots,(\omega_k,\lambda_k,q_k)) = \begin{cases} 1 & \textit{if } (\max_{i\leq k}\omega_i^2\mathbf{1}[\lambda_i\geq 1]) > 0 \\ 1-\frac{\sum_{i=1}^k \omega_i^2}{\sum_{i=1}^k \frac{\omega_i^2}{1-\lambda_i}} = 1-\frac{v^{*\prime}v^*}{v^{*\prime}(I_k-J^*)^{-1}v^*} & \textit{otherwise} \end{cases}$$

Theorem 1 essentially follows from combining the implication of Condition 1 as derived in Kitagawa (1934), whose result builds on the classic results of Kolmogorov (1930) and Nagumo (1930) on axiomatic foundations for quasi-arithmetic means, with Condition 2 and the differentiability assumption in Condition 1 (e). See the appendix for details.

From now on, we use PI to denote the prior informativeness measure derived in Theorem 1. Note that $\mathrm{PI}=1$ whenever $v^*=\sum_{i=1}^k\omega_i q_i$ has positive weight on any $q_i$ with eigenvalue of one or larger. This makes intuitive sense: if posterior results for $q_i'\theta^*$ are wholly determined by the prior for some $i$, then so are the posterior results for any linear combination $v^{*\prime}\theta^*$ that involves $q_i'\theta^*$. This motivates the following definition.

**Definition 2** *A prior is of* limited overall informativeness *if* $\lambda_{\max}=\max_{i\leq k}\lambda_i<1$*, that is if the largest eigenvalue* $\lambda_{\max}$ *of* $V_p[\theta]^{-1}V_\pi[\theta]$ *is smaller than unity.*

Under overall limited prior informativeness in the sense of Definition 2, PI can alternatively be written as

$$\mathrm{PI}=\frac{v^{*\prime}J(I_k-J^*)^{-1}v^*}{v^{*\prime}(I_k-J^*)^{-1}v^*}=\frac{v^{*\prime}J^*\alpha^*}{v^{*\prime}\alpha^*} \tag{10}$$

where $\alpha^*=(I_k-J^*)^{-1}v^*$. It can therefore also be viewed as a ratio of the derivative of the posterior mean of $v^{*\prime}\theta^*$ and the derivative of the prior mean of $v^{*\prime}\theta^*$. In this ratio, the direction $\alpha^*=(I_k-J^*)^{-1}v^*$ of the change of the prior mean of $\theta^*$ is such that for all nuisance parameters $w^{*\prime}\theta^*$ with $w^{*\prime}v^*=0$, the ratio is equal to one, $w^{*\prime}J^*\alpha^*/w^{*\prime}\alpha^*=1$. So PI also has the interpretation of the push-through rate of local changes of the prior mean of $v^{*\prime}\theta^*$ to changes in the posterior mean of $v^{*\prime}\theta^*$, where the prior mean is changed in a direction that ensures that all "resistance" exclusively stems from the direction of interest $v^*$.

If the parameter of interest is expressed in the original parameterization $v'\theta$, we obtain

$$\mathrm{PI}=1-\frac{v'V_p[\theta]v}{v'V_p[\theta](V_p[\theta]-V_\pi[\theta])^{-1}V_p[\theta]v} \tag{11}$$

via $v'\theta = (G^{-1\prime}v)'\theta^*$ under limited overall prior informativeness. Also, a linear combination $v^{*\prime}\theta^*$ corresponds to $v'\theta$ in the original parameterization, where $v = G'\theta^*$. Thus, if $v^*$ is an eigenvector of $J^*$, as in (9), then $J^* = GV_\pi[\theta]G'$ implies

$$V_p[\theta]^{-1}V_\pi[\theta]v = \lambda v \tag{12}$$

for this $v$. The eigenvectors $v^*$ of $J^*$ thus correspond to the eigenvectors of $V_p[\theta]^{-1}V_\pi[\theta]$ in the original parameterization. In particular, under overall limited prior informativeness, the prior plays the relatively most dominating role for the linear combination $v'\theta$ with $v$ the eigenvector corresponding to the largest eigenvalue of $V_p[\theta]^{-1}V_\pi[\theta]$. Finally, the direction $\alpha^* = (I_k - J^*)^{-1}v^*$ in (10) corresponds to

$$\alpha = (I_k - V_p[\theta]^{-1}V_\pi[\theta]V_p[\theta]^{-1})^{-1}v \tag{13}$$

in the original parameterization.

*Example 2:* Suppose $Y \sim \mathcal{N}(\theta, \Sigma)$ with $\Sigma$ positive definite and known, and the prior on $\theta$ is Gaussian $\theta \sim \mathcal{N}(\mu_p, V_p[\theta])$. Then $V_\pi[\theta] = V_p[\theta] - V_p[\theta](\Sigma + V_p[\theta])^{-1}V_p[\theta] = (V_p[\theta]^{-1} + \Sigma^{-1})^{-1}$, and the prior has overall limited informativeness, since $(V_p[\theta]^{-1} + \Sigma^{-1})^{-1}V_p[\theta]^{-1} = (I_k + V_p[\theta]^{1/2}\Sigma^{-1}V_p[\theta]^{1/2})^{-1}$ has all eigenvalues smaller than one. Further, the prior informativeness of $v'\theta$ is

$$\text{PI} = \frac{v'\Sigma v}{v'(\Sigma + V_p[\theta])v}.$$

This is recognized as the prior informativeness in the scalar experiment with data $Y_v \sim \mathcal{N}(\theta_v, v'\Sigma v)$ and prior $\theta_v \sim \mathcal{N}(v'\mu_p, v'V_p[\theta]v)$. Intuitively, in absence of knowledge of $\theta$, the likelihood information about $v'\theta$ is given by $Y_v = \theta'Y \sim \mathcal{N}(v'\theta, v'\Sigma v)$, and the prior on $v'\theta$ is $\mathcal{N}(v'\mu_p, v'V_p[\theta]v)$. ▲

# 3 Discussion and Extensions

## 3.1 Inequalities

The following 7 inequalities follow from the definitions of PS and PI in Section 2. See the Appendix for details. Recall that $\lambda_{\max}$ of Definition 2 is the largest eigenvalue of $V_p[\theta]^{-1}V_\pi[\theta]$.

1. For any $v$,

$$\text{PS} \leq \sqrt{\lambda_{\max}}\sqrt{v'V_\pi[\theta]v},$$

that is the sensitivity measure of $v'\theta$ is at most $\sqrt{\lambda_{\max}}$ posterior standard deviations. In other words, the (linear approximation of) the largest change in the posterior mean that can be induced by unit change of the the prior mean in the metric $\sqrt{\alpha' V_p[\theta]^{-1}\alpha}$, is never larger than $\sqrt{\lambda_{\max}}$ posterior standard deviations.

2. For any $v$,

$$\text{PS} \leq \lambda_{\max}\sqrt{v' V_p[\theta] v},$$

that is the sensitivity measure of $v'\theta$ is at most $\lambda_{\max}$ prior standard deviations. If all eigenvalues of $V_p[\theta]^{-1}V_\pi[\theta]$ are small, then the (linear approximation of) the largest change in the posterior mean that can be induced by unit change of the prior mean in the metric $\sqrt{\alpha' V_p[\theta]^{-1}\alpha}$, is small relative to the prior standard deviation.

3. For any $v$,

$$\frac{\text{PS}}{\sqrt{v' V_\pi[\theta] v}} \geq \frac{\sqrt{v' V_\pi[\theta] v}}{\sqrt{v' V_p[\theta] v}}$$

that is the sensitivity measure of $v'\theta$, expressed in multiples of posterior standard deviations, is never smaller than the ratio of posterior and prior deviations.

4. For any $v$,

$$\text{PI} \geq \min(\frac{v' V_\pi[\theta] v}{v' V_p[\theta] v}, 1)$$

that is the ratio of posterior and prior variance provides a lower bound for the prior informativeness of $v'\theta$, provided it is smaller than one. A large posterior variance relative to the prior variance is sufficient for a dominating role of the prior as measured by PI.

5. For any $v$,

$$\text{PI} \leq \lambda_{\max}$$

that is if all eigenvalues of $V_p[\theta]^{-1}V_\pi[\theta]$ are small, then the prior is not dominant as measured by PI for all parameters $v'\theta$.

6. For any $v$, if $\lambda_{\max} \leq 1/3$, then

$$\text{PI} \leq \frac{\text{PS}}{\sqrt{v' V_p[\theta] v}}$$

that is if the prior is at most moderately informative all parameters, then a small prior sensitivity PS relative to the prior standard deviation is sufficient for a small prior informativeness of a particular parameter.

7. For any $v$, if $\lambda_{\max} \leq 1$, then
$$\frac{\text{PS}}{\sqrt{v'V_p[\theta]v}} \leq \sqrt{\tfrac{3}{2}\,\text{PI}}$$
that is small prior informativeness of a particular parameter is sufficient for a small prior sensitivity PS relative to the prior standard deviation.

The most remarkable of these relations might be Inequality 1: Under overall limited prior informativeness, 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 never larger 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 $\alpha$ in the prior mean that is small in terms of the $\sqrt{\alpha' V_p[\theta]^{-1}\alpha}$ metric (at least under the linear approximation based on the derivative).

## 3.2 Functions of Parameters

In many applications, there is interest not only in the unknown $k \times 1$ parameters $\theta$, 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 $\gamma$ with respect to the prior mean of $\theta$ in (6) is, under weak regularity conditions, the $1 \times k$ vector

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

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

In analogy to PS, define $\text{PS}_\gamma$ as the largest (linear approximation to the) change of the posterior mean of $\gamma$ that can be induced by a unit change $\alpha$ of the prior mean in the metric $\sqrt{\alpha' V_p[\theta]^{-1}\alpha}$,

$$\begin{aligned} \text{PS}_\gamma &= \max_{\alpha' V_p[\theta]^{-1}\alpha = 1} J_\gamma \alpha \\ &= \sqrt{J_\gamma' V_p[\theta] J_\gamma} \\ &= \sqrt{E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])']V_p[\theta]^{-1}E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])]}. \end{aligned} \tag{14}$$

The measure $\text{PS}_\gamma$ is alternatively recognized as the sensitivity measure PS of the linear combination $\upsilon'\theta$ with

$$\upsilon = \upsilon_\gamma = V_\pi[\theta]^{-1}E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])]. \tag{15}$$

This ensures that whenever $\Gamma$ is linear, $\Gamma(\theta) = c_\gamma + \upsilon'\theta$, $\text{PS}_\gamma = \text{PS}$. Also, since the posterior covariance matrix of $(\theta', \gamma)'$ is non-negative definite, and $E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])]$ is the posterior covariance between $\gamma$ and $\theta$, $V_\pi[\Gamma(\theta)] \geq \upsilon'_\gamma V_\pi[\theta]\upsilon_\gamma$. The analogue of Inequality 1 of the last subsection, $\text{PS}_\gamma \leq \sqrt{\lambda_{\max}}\sqrt{V_\pi[\Gamma(\theta)]}$ thus still holds for any $\Gamma$ with finite posterior variance.

Similarly, define the prior informativeness $\text{PI}_\gamma$ of $\gamma$ as the prior informativeness measure PI of the linear combination $\upsilon'\theta$ with the same derivative of the posterior mean as $\gamma$, that is with $\upsilon = \upsilon_\gamma$. Under overall limited prior informativeness, we obtain

$$\text{PI}_\gamma = 1 - \frac{\upsilon'_\gamma V_p[\theta]\upsilon_\gamma}{\upsilon'_\gamma V_p[\theta](V_p[\theta] - V_\pi[\theta])^{-1}V_p[\theta]\upsilon_\gamma}. \tag{16}$$

This definition again ensures agreement with PI for linear $\Gamma$, and also Inequality 4 of the last section holds for $\text{PI}_\gamma$, $\text{PI}_\gamma \leq \lambda_{\max}$. What is more, even if $\Gamma(\theta)$ is non-linear, but can be written as a function of $\upsilon'\theta$ alone, then $\text{PI}_\gamma$ is still equal to the prior informativeness measure PI of $\upsilon'\theta$ as long as $E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])] \neq 0$ and $\upsilon'\theta$ is independent of all other parameters $w'\theta$ with $w'\upsilon = 0$ in the posterior.

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

$$R^2_\gamma = \frac{\upsilon'_\gamma V_\pi[\theta]\upsilon_\gamma}{V_\pi[\Gamma(\theta)]} = \frac{E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])']V_\pi[\theta]^{-1}E_\pi[\Gamma(\theta)(\theta - E_\pi[\theta])]}{V_\pi[\Gamma(\theta)]}. \tag{17}$$

Values of $R^2_\gamma$ close to one indicate a very similar posterior behavior of $\gamma$ and $\upsilon'_\gamma\theta$, so that $\text{PI}_\gamma$ becomes a more compelling measure for the prior informativeness of $\gamma$. In large samples, the Bernstein-von Mises Theorem ensures convergence of the posterior of $\theta$ to a Gaussian with vanishing variance, so that a Delta-method type argument applied to $\gamma = \Gamma(\theta)$ yields $R^2_\gamma \to 1$ with probability converging to one for differentiable, sample size independent $\Gamma$.

It is not necessary that the function $\gamma$ is a function of $\theta$ alone, but it may also depend on the realized data (so that formally, $\Gamma$ is indexed by the data). For example, $\text{PS}_\gamma$ and $\text{PI}_\gamma$ might be applied to learn about the role of the prior for a forecast, which is a function of both the model parameters $\theta$ and the realized data. As an illustration, consider a one-step ahead forecast in a AR(1) model $y_t - \mu = \rho(y_{t-1} - \mu)$, 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 far away from the sample mean $\bar{y}$ (and thus the approximate posterior mean of $\mu$), then $\rho$ is relatively more influential relative to $\mu$, which is properly reflected in the measures $\mathrm{PS}_\gamma$ and $\mathrm{PI}_\gamma$.

### 3.3 Implementation and Interpretation

It is entirely straightforward to implement the suggested statistics PS, PI, $\mathrm{PS}_\gamma$, $\mathrm{PI}_\gamma$ and $R^2_\gamma$ from standard MCMC output by replacing the expectation $E_\pi$ and variance $V_\pi$ by the MCMC sample mean $\hat{E}_\pi$ and variance $\hat{V}_\pi$. The prior mean $E_p$ and variances $V_p$ are typically known in closed form, but could also be calculated by Monte Carlo simulation if necessary.

Monte Carlo estimates always contain some degree of estimation error. This means that even if $v'\theta$ is exactly independent of $w'\theta$ in the posterior, their (unconstrained) estimated posterior covariance is never exactly zero. Whenever the prior has unlimited informativeness in the sense of Definition 2, PI of $v'\theta$ computed with $(E_\pi, V_\pi)$ replaced by $(\hat{E}_\pi, \hat{V}_\pi)$ thus equals unity even if $v$ is orthogonal to all true eigenvectors corresponding to eigenvalues larger or equal to unity. As a practical matter it therefore makes sense to restrict computation of PI (and $\mathrm{PI}_\gamma$) to cases where the prior has overall limited informativeness. What is more, it probably also does not makes sense to compute PI if the largest eigenvalue of $V_p[\theta]^{-1}\hat{V}_\pi[\theta]$ is only marginally below unity. One the one hand, the appearance of limited overall informativeness might simply be due to estimation error in $\hat{V}_\pi[\theta]$. On the other hand, inspection of the formula in Theorem 1 shows that PI becomes very sensitive to the exact value of the eigenvalues $\lambda_i$ once they are close to unity. The difference between, say, $\lambda = 0.99$ and $\lambda = 0.96$ for a particular eigenvector might well reflect details of the choice of prior shape, etc., rather than a fourfold difference in the relative amount of likelihood information as implied by the formula in Theorem 1. In practice, it therefore seems sensible to adopt a rather conservative view of overall limited prior informativeness and to demand that the largest eigenvalue of $V_p[\theta]^{-1}\hat{V}_\pi[\theta]$ is no larger than, say, 0.90 (and to conclude that otherwise, prior information plays a potentially dominant role for all parameters).

Typically, the elements $\theta_1, \cdots, \theta_k$ of $\theta$ are of direct interest. Letting $v$ equal to the $i$th column of $I_k$, one obtains via (7) and (11) a prior robustness measure PS and an prior informativeness measure PI for each $\theta_i$. As discussed above, the interpretation of PS is that the interval with endpoints $E_\pi[\theta_i] \pm a\,\mathrm{PS}$ is an approximation to the set of posterior mean values that can be obtained through changes of the prior mean by the multivariate analogue of $a$ prior standard deviations. The natural comparison is a purely marginal prior sensitivity analysis where only the prior mean of $\theta_i$ (but not of $\theta_j$ with $j \neq i$) is varied by $a$ prior standard deviations, which leads to the interval with endpoints $E_\pi[\theta_i] \pm aV_\pi[\theta_i]/\sqrt{V_p[\theta_i]}$ (cf. (3) of the univariate analysis). By inequality 3 of Section 3.1 above, the former interval always contains the latter. Further, PI can at least loosely be interpreted as the fraction of prior information in the posterior mean $E_\pi[\theta_i]$. Here the comparison is the derivative $V_\pi[\theta_i]/V_p[\theta_i]$ (cf. (4) of the univariate analysis) of a purely marginal analysis that only considers the effect of changes in the prior mean of $\theta_i$. For both PS and PI, the difference to such a marginal analysis becomes potentially large if the correlation pattern in the posterior is substantially different from the correlation pattern in the prior.

This "joint" multivariate analysis by PS and PI requires that changes of parameter values, standardized by prior standard deviations, are comparable across different parameters that are uncorrelated in the prior. In particular, the choice of a very vague prior for some, but not all parameters (maybe based on the rationale that the likelihood is likely to be very informative for those parameters) make PS and PI difficult to interpret. To illustrate the point, suppose $\theta = (\theta_1, \theta_2)'$ with data $Y \sim \mathcal{N}(\theta, I_2)$. Under the prior $\theta \sim \mathcal{N}(0, \text{diag}(1, 100))$ an analysis using PI indicates a much smaller role for the prior for $\theta_2$ compared to $\theta_1$, and thus a correspondingly more dominant role of the likelihood. This only makes sense if the actual prior uncertainty about $\theta_2$ is much larger compared to $\theta_1$, so that PI correctly reflects the relatively more pronounced reduction of prior uncertainty about $\theta_2$ through the likelihood information (and accordingly, in terms of the the normalize parameter $\theta^* = (\theta_1, \theta_2/10)'$, the likelihood $Y^* = (Y_1, Y_2/10)' \sim \mathcal{N}(\theta^*, \text{diag}(1, 1/00))$ is more informative about $\theta_2^*$).

Note that neither PS nor PI are invariant to nonlinear transformations of $\theta$ in general. Also from this perspective it is therefore important that prior standard deviations form a meaningful yardstick for prior uncertainty. Under an absolutely continuous product prior, one could in principle ensure invariance at least to arbitrary one-to-one transformations of the scalar individual parameters by always letting $\theta$ of Section 2.1 to be the unique element-by-element monotone increasing transformation of the original parameter vector that leads to a $\mathcal{N}(0, I_k)$ prior, and by computing PS and PI of the original parameter by the formulas developed in Section 3.2 above. A change of the prior mean of $\theta_i$ could then be interpreted via the implied quantile shifts of the prior on the $i$th original parameter. In practice, though, there will often exist a parameterization in which the elements of $\theta$ have directly meaningful units, and it might well be easier to think about changes in the prior mean in those units, standardized by the prior standard deviation.

One can also go beyond the scalar measures PS and PI and examine key directions in the $k$ dimensional parameter space. Two approaches come to mind. First, for a given parameter of interest $\theta_i$, it is instructive to compute the directions of prior mean changes that underlie the values of PS and PI. By (8), the $i$th column of $V_\pi[\theta]$ is the direction that induces the largest change of the posterior mean of $\theta_i$. Similarly, by (13), the $i$th column of $(I_k - V_p[\theta]^{-1}V_\pi[\theta]V_p[\theta]^{-1})^{-1}$ induces the push-through rate PI of prior mean changes of $\theta_i$ to posterior mean changes of $\theta_i$ (and a push-through rate of one for $\theta_j$, $j \neq i$). Second, one can ask for what linear combination of the parameters is the prior most dominant, or which direction is most sensitive to prior mean changes. Both of these questions lead to the same direction: the eigenvector of $V_p[\theta]^{-1}V_\pi[\theta]$ that corresponds to the largest eigenvalue.

## 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 $\theta = \theta_0$.[4] The likelihood function $l$ is, of course, simply given by $l(\theta) = f(y;\theta)$ after observing $Y = y$. Thus, if $\theta_0$ is such that $l(\theta) = l(\theta_0)$ implies $\theta = \theta_0$, then $\theta_0$ 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 $f(y_0;\theta) \neq f(y_0;\theta_0)$. In particular, 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 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 well 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 would have contained information does not mitigate this fact.

[4] Rothenberg (1971) further defines a parameter point $\theta = \theta_0$ to be *locally identifiable* if it there exists an open neighborhood $\Theta_0$ of $\theta_0$ so that $\theta_0$ is identifiable with the parameter space restricted to $\Theta = \Theta_0$.

Rothenberg's (1971) definition is useful for the more theoretical question of whether model parameters could *in principle* be told apart by empirical studies. But it also implies something about the shape of the likelihood: if $\theta_0$ is not identifiable because $f(y;\theta) = f(y;\theta_0)$ for all $y \in \mathcal{Y}$ for some $\theta \neq \theta_0$, then also $l(\theta) = l(\theta_0)$, for all possible observations.[5] In particular, if there exists a hyperplane $\Theta_{hp}$ so that $f(y;\theta) = f(y;\theta_0)$ for all $\theta, \theta_0 \in \Theta_{hp}$ and $y \in \mathcal{Y}$, then $l$ is constant on that hyperplane. The prior informativeness measure PI will then equal unity for any linear combination of $\theta$ that is not orthogonal to this hyperplane, and $\lambda_{\max} \geq 1$. An empirical finding of limited overall prior informativeness in the sense of Definition 2 (i.e. that $\lambda_{\max} < 1$) thus rules out at least this hyperplane form of the lack of identifiability. At the same time, if $f(y;\theta) = f(y;\theta_0)$ for all $\theta, \theta_0 \in \Theta'$, but $\Theta'$ does not contain a hyperplane, then $\lambda_{\max}$ might still be smaller than unity. Whether or not this is the "right" answer depends on $\Theta'$—if $\Theta'$ is "small", then lack of identifiability does not imply that nothing useful can be learned about $\theta$. 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 7 digits. Then no value of $\theta$ 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 both models, as the global shape of the likelihood is almost identical.

An important practical advantage 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)).

In summary, the concept and appeal of the prior informativeness measure PI is quite distinct from the standard frequentist definition of identification. The approach pursued here is thus largely complementary to the recent results on identification in DSGE models by Iskrev (2010a, 2010b), Komunjer and Ng (2009) and Andrle (2010) that build on Rothenberg's (1971) definition.

[5] Lack of local identifiability at the MLE $\theta = \hat{\theta}$ implies a singular hessian matrix at $\hat{\theta}$ (assuming $l$ admits two derivatives). As a consequence, a numerical finding of a non-singluar hessian at $\hat{\theta}$ implies local identifiability of $\theta = \hat{\theta}$ in the sense of Rothenberg (1971).

### 3.5 Relationship to McCulloch (1989)

McCulloch (1989) investigates the local influence of a particular prior distribution (or model) on posterior results. In particular, he studies local changes $\delta$ of a general prior hyperparameter, and studies its impact on the prior and posterior distributions relative to the baseline specification, with distributional changes measured by the Kullback-Leibler divergence. The hyperparameter is deemed influential if small distributional changes of the prior lead to a large distributional change of the posterior. By a second-order Taylor expansion, McCulloch (1989) shows that the change of the Kullback-Leibler divergence is approximated by quadratic forms in $\delta$ around the Fisher-Information matrices $\mathfrak{I}_p$ and $\mathfrak{I}_\pi$ of the prior and posterior distributions, respectively. This implies that the overall influence of $\delta$ is usefully measured by the largest eigenvalue $\lambda_{\mathfrak{I}}$ of $\mathfrak{I}_p^{-1}\mathfrak{I}_\pi$.

The parameter $\alpha$ in the exponential family embedding (6) can, of course, be viewed as a hyperparameter, with the baseline specification $\alpha = 0$. A straightforward calculation further shows that the Fisher information of $\alpha$ in the prior and posteriors are given by $\mathfrak{I}_p = V_p[\theta]^{-1}$ and $\mathfrak{I}_\pi = V_p[\theta]^{-1}V_\pi[\theta]V_p[\theta]^{-1}$ at $\alpha = 0$. Thus, in the exponential family embedding (6), $\mathfrak{I}_\pi\mathfrak{I}_p^{-1} = V_\pi[\theta]V_p[\theta]^{-1}$, and $\lambda_{\mathfrak{I}} = \lambda_{\max}$ of Definition 2 in Section 2.2. The upper bound $\lambda_{\max}$ on prior informativeness PI (cf. Inequality 4) therefore also has the interpretation as the largest Kulback-Leibler difference between the posterior distribution from its baseline, relative to the Kullback-Leibler difference between the prior and its baseline, that can be induced by local variation of the prior mean.

## 4 Application to Smets and Wouters (2007)

In an influential paper, Smets and Wouters (2007) use Bayesian MCMC techniques to estimate a log-linearized DSGE model on US postwar data. The underlying economic model features sticky prices and wages, habit formation in consumption, variable capital utilization and investment adjustment costs. Table 1 summarizes the dynamics of the seven structural shocks $\varepsilon_t$, which are driven by independent Gaussian innovations $\eta_t$. 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 with Dynare using quarterly data on output growth, consumption growth, investment growth, real wage growth, inflation, hours worked and nominal interest rate. We refer to Smets and Wouters (2007) for further

Table 1: Dynamic Specification of Structural Shocks

| | | |
|---|---|---|
| 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 | $\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \eta_t^w - \mu_w \eta_{t-1}^w$ | $\eta_t^w \sim iid\mathcal{N}(0, \omega_w^2)$ |

details on the model and the log-linearization.

In their estimation, Smets and Wouters fix 5 of the 41 parameters: the depreciation rate (0.025 on a quarterly basis), the GDP ratio of steady state exogenous spending (18 percent), the steady-state markup in the labor market (1.5) and the two Kimball aggregators in the goods and labor markets (both 10). The shock processes of Table 1 are parametrized by a total of 17 parameters (the "shock" parameters), and the remaining 19 parameters are listed in Table 2 (the "structural" parameters). We adopt the same independent prior on these 36 parameters as Smets and Wouters (2007), except for the 7 standard deviations $\omega$ of Table 1. There we choose the Gamma distribution on the precision $1/\omega^2$ so that the implied mean and standard deviation of $\omega$ 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 with the degree of prior uncertainty for the other estimated parameters,[6] which facilitates the interpretation of the prior sensitivity and prior informativeness measures, as discussed in Section 3.3.

Estimation of the model with Dynare version 4.0.4 using 4 independent chains with 200,000 draws each essentially reproduces the posterior results in Smets and Wouters (2007). The three largest eigenvalues of $V_p[\theta]^{-1}V_\pi[\theta]$ are 1.25, 0.90 and 0.68, so that the prior has overall unlimited informativeness in the sense of Definition 2. Inspection of the eigenvector (normalized to unit length) associated with the largest eigenvalue shows a loading of 0.94 on the steady state inflation rate $\bar{\pi}$. What is more, the ratio of posterior to prior standard deviation of $\bar{\pi}$ is equal to 1.07. The likelihood thus seems entirely uninformative about $\bar{\pi}$. To proceed further, we set $\bar{\pi}$ equal to its prior mean 0.625, and exclude it from estimation.

[6] 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 $\omega$ on page 592 of Smets and Wouters (2007) ($\sigma$ in their notation) does not match their actual choice reported in their Table 1B.

Table 2: Structural Parameters of Smets and Wouters (2007)

| | |
|---|---|
| $\varphi$ | elasticity of capital adjustment cost function |
| $\sigma_c$ | elasticity of intertemporal substitution |
| $h$ | external habit formation |
| $\xi_w$ | calvo probability in labor market |
| $\sigma_l$ | elasticity of labor supply with respect to real wage |
| $\xi_p$ | calvo probability in goods market |
| $\iota_w$ | degree of wage indexation |
| $\iota_p$ | degree of price indexation |
| $\psi$ | normalized elasticity of capital utilization adjustment cost function |
| $\Phi$ | fixed cost of intermediate good producers |
| $r_\pi$ | inflation coefficient in monetary policy reaction function |
| $\rho$ | 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 |
| $\bar{\pi}$ | steady state inflation rate |
| $100(\beta^{-1}-1)$ | normalized household discount factor |
| $\bar{l}$ | steady state hours worked |
| $\bar{\gamma}$ | steady state quarterly growth rate |
| $\alpha$ | capital share in production |

Table 3: Prior and Posterior of Selected Parameters with $\bar{\pi}$ fixed

| | | | Prior | | Posterior | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ev | | $\mu_p$ | $\sigma_p$ | $\mu_\pi$ | $\sigma_\pi$ | $\sigma_\pi^2/\sigma_p^2$ | PS | PI |
| $\xi_w$ | 0.54 | $\mathcal{B}$ | 0.50 | 0.10 | 0.70 | 0.07 | 0.45 | 0.06 | 0.94 |
| $\sigma_l$ | 0.55 | $\mathcal{N}$ | 2.00 | 0.75 | 1.82 | 0.57 | 0.58 | 0.50 | 0.94 |
| $\xi_p$ | 0.28 | $\mathcal{B}$ | 0.50 | 0.10 | 0.65 | 0.06 | 0.35 | 0.04 | 0.82 |
| $\iota_w$ | -0.39 | $\mathcal{B}$ | 0.50 | 0.15 | 0.56 | 0.13 | 0.70 | 0.11 | 0.91 |
| $\psi$ | -0.29 | $\mathcal{B}$ | 0.50 | 0.15 | 0.55 | 0.11 | 0.56 | 0.09 | 0.85 |

Notes: $\mathcal{B}$ and $\mathcal{N}$ are Beta and Normal prior distributions with mean and variance $\mu_p$ and $\sigma_p^2$. The "ev" column reports the elements of the eigenvector corresponding to the largest eigenvalue of $V_p[\theta]^{-1}V_\pi[\theta]$.

Re-estimation with $\bar{\pi}$ fixed at 0.625 yields the three largest eigenvalues equal to 0.98, 0.71 and 0.69. Table 3 shows the prior and posterior results for the 5 parameters with a loading of the eigenvector corresponding to the largest eigenvalue of more than 0.15 in absolute value (these are very close to the loadings of the eigenvector corresponding to the second largest eigenvalue in the estimation including $\bar{\pi}$). A comparison of marginal prior and posterior distribution might suggest that the likelihood determines these 5 parameters at least to some degree, with a reduction of the prior variance of at least 30% for all parameters. But the values of PI show that this is misleading and instead point to an essentially dominating role of the prior, qualitatively similar to (although much less extreme than) Example 1 of the introduction. To make further progress, we also fix $\sigma_l$ at its prior mean $\sigma_l = 2.00$.

With $\bar{\pi}$ and $\sigma_l$ fixed, the three largest eigenvalues of $V_p[\theta]^{-1}V_\pi[\theta]$ become 0.78, 0.68 and 0.60. Table 4 reports the full set of prior and posterior results for the remaining 34 estimated parameters. For all shock parameters, the role of the prior is quite limited, with PI of at most 0.21, and often much below. In contrast, 10 of the 17 structural parameters have PI $\geq 1/3$, indicating that to a substantial degree, posteriors reflect prior information. Similarly, if one were uncertain about the appropriate prior means, then this would lead to substantially more additional variability in the posterior mean results for the structural parameters compared to the shock parameters, as indicated by PS. On the whole the entries for $\sigma_\pi^2/\sigma_p^2$ are quite close to those for PI. In this application the likelihood information about the different parameters does not seem to be highly correlated, approximately matching the product prior specification. Notable exceptions are $\rho_p$, $\rho$ and $r_y$, where the relatively smaller ratios $\sigma_\pi^2/\sigma_p^2$ substantially understate the role of the prior.

Tables 5 and 6 contain prior and posterior results for key impulse responses and variance decompositions of output forecasts. For some of the impulse responses with long horizon, the ratio of posterior to prior variance is much larger than $\text{PI}_\gamma$ (which is impossible for PI by Inequality 4 of Section 3.1), and sometimes substantially above one. Intuitively, in an AR(1) model with coefficient $\rho$, the 16 period ahead impulse response is $\rho^{16}$. With, say, a uniform prior on $[0, .95]$ for $\rho$, $\rho^{16}$ is very small with high prior probability. If the likelihood strongly favors values of $\rho$ close to 0.90, say, but without being very informative, then larger values of $\rho^{16}$ are more likely under the posterior compared to the prior, yielding a relatively larger posterior variance. In this example, the highly nonlinear transformation results in a very fat-tailed prior distribution on $\rho^{16}$, and one prior standard deviation of $\rho^{16}$ describes a very different degree of prior uncertainty than one prior standard deviation of $\rho$. It is

Table 4: Prior and Posterior of Parameters with $\bar{\pi}$ and $\sigma_l$ fixed

| | | Prior | | Posterior | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $\mu_p$ | $\sigma_p$ | $\mu_\pi$ | $\sigma_\pi$ | $\sigma_\pi^2/\sigma_p^2$ | PS | PI |
| $\omega_a$ | $\mathcal{IG}$ | 0.30 | 0.20 | 0.46 | 0.03 | 0.02 | 0.01 | 0.02 |
| $\omega_b$ | $\mathcal{IG}$ | 0.30 | 0.20 | 0.24 | 0.02 | 0.01 | 0.01 | 0.02 |
| $\omega_g$ | $\mathcal{IG}$ | 0.30 | 0.20 | 0.53 | 0.03 | 0.02 | 0.01 | 0.02 |
| $\omega_i$ | $\mathcal{IG}$ | 0.30 | 0.20 | 0.46 | 0.05 | 0.06 | 0.02 | 0.07 |
| $\omega_r$ | $\mathcal{IG}$ | 0.30 | 0.20 | 0.25 | 0.02 | 0.01 | 0.00 | 0.01 |
| $\omega_p$ | $\mathcal{IG}$ | 0.30 | 0.20 | 0.15 | 0.02 | 0.01 | 0.01 | 0.01 |
| $\omega_w$ | $\mathcal{IG}$ | 0.30 | 0.20 | 0.25 | 0.02 | 0.01 | 0.01 | 0.02 |
| $\rho_a$ | $\mathcal{B}$ | 0.50 | 0.20 | 0.96 | 0.01 | 0.00 | 0.00 | 0.00 |
| $\rho_b$ | $\mathcal{B}$ | 0.50 | 0.20 | 0.21 | 0.08 | 0.18 | 0.04 | 0.19 |
| $\rho_g$ | $\mathcal{B}$ | 0.50 | 0.20 | 0.97 | 0.01 | 0.00 | 0.00 | 0.00 |
| $\rho_i$ | $\mathcal{B}$ | 0.50 | 0.20 | 0.70 | 0.06 | 0.09 | 0.03 | 0.11 |
| $\rho_r$ | $\mathcal{B}$ | 0.50 | 0.20 | 0.15 | 0.06 | 0.10 | 0.02 | 0.11 |
| $\rho_p$ | $\mathcal{B}$ | 0.50 | 0.20 | 0.90 | 0.05 | 0.06 | 0.03 | 0.10 |
| $\rho_w$ | $\mathcal{B}$ | 0.50 | 0.20 | 0.97 | 0.01 | 0.00 | 0.00 | 0.00 |
| $\mu_p$ | $\mathcal{B}$ | 0.50 | 0.20 | 0.74 | 0.09 | 0.18 | 0.05 | 0.21 |
| $\mu_w$ | $\mathcal{B}$ | 0.50 | 0.20 | 0.86 | 0.05 | 0.07 | 0.03 | 0.10 |
| $\rho_{ga}$ | $\mathcal{N}$ | 0.50 | 0.25 | 0.52 | 0.09 | 0.13 | 0.03 | 0.13 |
| $\varphi$ | $\mathcal{N}$ | 4.00 | 1.50 | 5.73 | 1.04 | 0.48 | 0.75 | 0.51 |
| $\sigma_c$ | $\mathcal{N}$ | 1.50 | 0.38 | 1.39 | 0.13 | 0.12 | 0.07 | 0.16 |
| $h$ | $\mathcal{B}$ | 0.70 | 0.10 | 0.72 | 0.04 | 0.17 | 0.02 | 0.22 |
| $\xi_w$ | $\mathcal{B}$ | 0.50 | 0.10 | 0.72 | 0.05 | 0.29 | 0.04 | 0.38 |
| $\xi_p$ | $\mathcal{B}$ | 0.50 | 0.10 | 0.65 | 0.06 | 0.34 | 0.04 | 0.47 |
| $\iota_w$ | $\mathcal{B}$ | 0.50 | 0.15 | 0.56 | 0.13 | 0.70 | 0.11 | 0.73 |
| $\iota_p$ | $\mathcal{B}$ | 0.50 | 0.15 | 0.26 | 0.09 | 0.35 | 0.06 | 0.38 |
| $\psi$ | $\mathcal{B}$ | 0.50 | 0.15 | 0.54 | 0.11 | 0.53 | 0.08 | 0.59 |
| $\Phi$ | $\mathcal{N}$ | 1.25 | 0.13 | 1.61 | 0.08 | 0.38 | 0.05 | 0.41 |
| $r_\pi$ | $\mathcal{N}$ | 1.50 | 0.25 | 2.04 | 0.17 | 0.48 | 0.14 | 0.58 |
| $\rho$ | $\mathcal{B}$ | 0.75 | 0.10 | 0.81 | 0.02 | 0.06 | 0.01 | 0.09 |
| $r_y$ | $\mathcal{N}$ | 0.13 | 0.05 | 0.10 | 0.02 | 0.20 | 0.02 | 0.31 |
| $r_{\Delta y}$ | $\mathcal{N}$ | 0.13 | 0.05 | 0.22 | 0.03 | 0.30 | 0.02 | 0.34 |
| $100(\beta^{-1}-1)$ | $\mathcal{G}$ | 0.25 | 0.10 | 0.17 | 0.06 | 0.33 | 0.03 | 0.34 |
| $\bar{l}$ | $\mathcal{N}$ | 0.00 | 2.00 | 1.36 | 0.90 | 0.20 | 0.44 | 0.22 |
| $\bar{\gamma}$ | $\mathcal{N}$ | 0.40 | 0.10 | 0.44 | 0.01 | 0.02 | 0.00 | 0.02 |
| $\alpha$ | $\mathcal{N}$ | 0.30 | 0.05 | 0.19 | 0.02 | 0.12 | 0.01 | 0.14 |

Notes: $\mathcal{N}$, $\mathcal{B}$ and $\mathcal{G}$ are Normal, Beta and Gamma prior distributions with mean and variance $\mu_p$ and $\sigma_p^2$, and $\mathcal{IG}$ is a Gamma prior distribution on $1/\omega^2$ that implies a mean and variance of $\mu_p$ and $\sigma_p^2$ on $\omega$.

Table 5: Prior and Posterior Description of Impulse Responses with $\bar{\pi}$ and $\sigma_l$ fixed

| | | Prior | | Posterior | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| series | hor. | $\mu_p$ | $\sigma_p$ | $\mu_\pi$ | $\sigma_\pi$ | $\sigma_\pi^2/\sigma_p^2$ | $PS_\gamma$ | $PI_\gamma$ | $R_\gamma^2$ | $r_{sh}$ |
| Responses to Productivity Shock | | | | | | | | | | |
| output | 1 | 0.45 | 0.36 | 0.73 | 0.10 | 0.07 | 0.04 | 0.06 | 1.00 | 0.97 |
| output | 4 | 0.35 | 0.25 | 1.27 | 0.12 | 0.23 | 0.07 | 0.02 | 0.99 | 0.98 |
| output | 16 | 0.03 | 0.10 | 1.20 | 0.15 | 2.46 | 0.04 | 0.00 | 0.99 | 1.00 |
| hours | 1 | -0.76 | 0.34 | -0.62 | 0.07 | 0.05 | 0.03 | 0.08 | 0.99 | 0.96 |
| hours | 4 | 0.07 | 0.12 | -0.15 | 0.07 | 0.35 | 0.04 | 0.19 | 0.98 | 0.76 |
| hours | 16 | -0.01 | 0.01 | 0.08 | 0.04 | 8.80 | 0.02 | 0.01 | 0.92 | 0.99 |
| inflation | 1 | -0.25 | 0.14 | -0.12 | 0.03 | 0.04 | 0.02 | 0.18 | 0.98 | 0.75 |
| inflation | 4 | 0.00 | 0.07 | -0.09 | 0.02 | 0.05 | 0.01 | 0.08 | 0.93 | 0.89 |
| inflation | 16 | -0.01 | 0.01 | 0.00 | 0.00 | 0.31 | 0.00 | 0.15 | 0.86 | 0.83 |
| interest rate | 1 | -0.20 | 0.09 | -0.14 | 0.02 | 0.03 | 0.01 | 0.02 | 0.97 | 0.96 |
| interest rate | 4 | -0.09 | 0.09 | -0.16 | 0.02 | 0.05 | 0.01 | 0.01 | 0.94 | 0.97 |
| interest rate | 16 | 0.00 | 0.01 | -0.02 | 0.01 | 0.65 | 0.00 | 0.18 | 0.88 | 0.82 |
| Responses to Monetary Policy Shock | | | | | | | | | | |
| output | 1 | -0.14 | 0.15 | -0.11 | 0.02 | 0.01 | 0.01 | 0.01 | 0.91 | 0.93 |
| output | 4 | -0.20 | 0.39 | -0.34 | 0.05 | 0.01 | 0.02 | 0.01 | 0.83 | 0.96 |
| output | 16 | -0.01 | 0.11 | -0.24 | 0.08 | 0.58 | 0.03 | 0.01 | 0.82 | 0.98 |
| hours | 1 | -0.07 | 0.08 | -0.05 | 0.01 | 0.02 | 0.01 | 0.04 | 0.95 | 0.87 |
| hours | 4 | -0.13 | 0.24 | -0.19 | 0.03 | 0.02 | 0.02 | 0.02 | 0.88 | 0.96 |
| hours | 16 | 0.00 | 0.05 | -0.10 | 0.04 | 0.67 | 0.01 | 0.02 | 0.82 | 0.98 |
| inflation | 1 | 0.30 | 0.24 | 0.25 | 0.02 | 0.00 | 0.00 | 0.00 | 0.84 | 0.99 |
| inflation | 4 | -0.03 | 0.10 | 0.06 | 0.02 | 0.03 | 0.01 | 0.01 | 0.82 | 0.98 |
| inflation | 16 | 0.00 | 0.02 | 0.00 | 0.01 | 0.28 | 0.00 | 0.12 | 0.71 | 0.87 |
| interest rate | 1 | 0.09 | 0.14 | 0.07 | 0.01 | 0.00 | 0.00 | 0.01 | 0.95 | 0.90 |
| interest rate | 4 | 0.02 | 0.10 | 0.07 | 0.01 | 0.02 | 0.00 | 0.01 | 0.85 | 0.94 |
| interest rate | 16 | 0.00 | 0.03 | 0.00 | 0.01 | 0.07 | 0.00 | 0.09 | 0.72 | 0.87 |

Notes: The column "hor." is the forecast horizon in quarters. $r_{sh}$ measures the relative importance of the shock parameters for the posterior results. The prior mean and variances are estimated from 10,000 independet draws from the prior (diregarding the approximately 3% of the prior draws for which the log-linearized model does not admit a unique solution).

Table 6: Prior and Posterior Description of Variance Deomposition of Output Forecasts with $\bar{\pi}$ and $\sigma_l$ fixed

| | Prior | | Posterior | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\mu_p$ | $\sigma_p$ | $\mu_\pi$ | $\sigma_\pi$ | $\sigma_\pi^2/\sigma_p^2$ | $PS_\gamma$ | $PI_\gamma$ | $R_\gamma^2$ | $r_{sh}$ |
| one quarter ahead forecast | | | | | | | | | |
| productivity | 0.03 | 0.06 | 0.16 | 0.04 | 0.43 | 0.02 | 0.04 | 0.99 | 0.97 |
| risk premium | 0.60 | 0.27 | 0.27 | 0.03 | 0.02 | 0.01 | 0.01 | 0.98 | 0.98 |
| exogenous spending | 0.11 | 0.13 | 0.36 | 0.04 | 0.09 | 0.01 | 0.02 | 0.99 | 0.99 |
| investment | 0.05 | 0.07 | 0.13 | 0.03 | 0.15 | 0.01 | 0.05 | 0.96 | 0.91 |
| monetary policy | 0.19 | 0.23 | 0.05 | 0.01 | 0.00 | 0.01 | 0.02 | 0.97 | 0.88 |
| price markup | 0.02 | 0.06 | 0.02 | 0.00 | 0.01 | 0.00 | 0.01 | 0.89 | 0.93 |
| wage markup | 0.00 | 0.01 | 0.00 | 0.00 | 0.08 | 0.00 | 0.04 | 0.79 | 0.92 |
| four quarter ahead forecast | | | | | | | | | |
| productivity | 0.02 | 0.05 | 0.24 | 0.05 | 0.70 | 0.02 | 0.03 | 0.98 | 0.97 |
| risk premium | 0.55 | 0.31 | 0.14 | 0.03 | 0.01 | 0.01 | 0.02 | 0.96 | 0.95 |
| exogenous spending | 0.05 | 0.07 | 0.18 | 0.03 | 0.15 | 0.01 | 0.04 | 0.97 | 0.97 |
| investment | 0.06 | 0.10 | 0.23 | 0.05 | 0.21 | 0.02 | 0.04 | 0.95 | 0.94 |
| monetary policy | 0.26 | 0.27 | 0.09 | 0.02 | 0.01 | 0.01 | 0.03 | 0.96 | 0.87 |
| price markup | 0.05 | 0.11 | 0.06 | 0.02 | 0.02 | 0.01 | 0.01 | 0.82 | 0.96 |
| wage markup | 0.02 | 0.06 | 0.05 | 0.03 | 0.22 | 0.02 | 0.03 | 0.90 | 0.94 |
| sixteen quarter ahead forecast | | | | | | | | | |
| productivity | 0.03 | 0.07 | 0.32 | 0.05 | 0.68 | 0.02 | 0.01 | 0.90 | 0.99 |
| risk premium | 0.50 | 0.31 | 0.03 | 0.01 | 0.00 | 0.00 | 0.02 | 0.90 | 0.96 |
| exogenous spending | 0.04 | 0.07 | 0.06 | 0.02 | 0.05 | 0.01 | 0.02 | 0.92 | 0.99 |
| investment | 0.07 | 0.12 | 0.13 | 0.05 | 0.15 | 0.02 | 0.04 | 0.91 | 0.98 |
| monetary policy | 0.26 | 0.27 | 0.04 | 0.02 | 0.01 | 0.01 | 0.06 | 0.89 | 0.86 |
| price markup | 0.06 | 0.13 | 0.10 | 0.03 | 0.05 | 0.01 | 0.01 | 0.77 | 0.98 |
| wage markup | 0.04 | 0.12 | 0.31 | 0.08 | 0.48 | 0.04 | 0.01 | 0.81 | 0.98 |

Notes: $r_{sh}$ measures the relative importance of the shock parameters for the posterior results. The prior mean and variances are estimated from 10,000 independet draws from the prior (diregarding the approximately 3% of the prior draws for which the log-linearized model does not admit a unique solution).

thus quite reasonable to measure prior informativeness of $\rho$ with its prior standard deviation as a yardstick, and to then assign the same prior informativeness to $\rho^{16}$, disregarding the prior (and posterior) standard deviation of $\rho^{16}$. Correspondingly, $\mathrm{PI}_\gamma$ in Tables 5 and 6 is not a function of either prior or posterior standard deviations of the impulse responses and variance decompositions, but rather equals the prior informativeness PI of the corresponding linear function of $\theta$, as explained in Section 3.2 above.

The consistently high $R^2_\gamma$ shows that these linear functions are good approximations of the posterior relationship between $\theta$ and the impulse responses and variance decompositions. It is striking how informative the likelihood is for the impulse responses and variance decompositions compared to the structural parameters in Table 4. The reason is that they are largely determined by the relatively well pinned down shock parameters: Let $c_\gamma + v'_\gamma\theta$ be the linear approximation in the posterior of a given impulse response or variance decomposition (cf. (15)), and partition $v_\gamma$ and $\theta$ into shock and structural parameters, respectively, $v_\gamma = (v'_{\gamma\,\mathrm{sh}}, v'_{\gamma\,\mathrm{st}})'$ and $\theta = (\theta'_{\mathrm{sh}}, \theta'_{\mathrm{st}})'$. If the range of plausible values for the parameters is proportional to the respective prior standard deviations, then the relative importance of the shock and structural parameters is usefully measured by $r_{\mathrm{sh}} = ||V_p[\theta_{\mathrm{sh}}]^{1/2} v_{\gamma\,\mathrm{sh}}|| / ||V_p[\theta]^{1/2} v_\gamma||$ and $r_{\mathrm{st}} = ||V_p[\theta_{\mathrm{st}}]^{1/2} v_{\gamma\,\mathrm{st}}|| / ||V_p[\theta]^{1/2} v_\gamma||$, respectively, where $r_{\mathrm{sh}} + r_{\mathrm{st}} = 1$. The last column of Tables 5 and 6 reports $r_{\mathrm{sh}}$, which is never below 0.75 and mostly very close to one.[7] It thus seems that the DSGE model of Smets and Wouters (2007) determines in which rotation and magnitude the structural shocks enter the reduced form VAR largely independently of the structural parameters, and also the dynamics are mostly driven by the autocorrelations of the shock processes.[8] Of course, one could easily imagine that the structural parameters enter other functions of $\theta$ 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.

[7] With the Smets and Wouters (2007) prior on the standard deviations $\omega$ 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 $V_p[\theta_{\mathrm{sh}}]$ become larger, further increasing the value of $r_{\mathrm{sh}}$, and thus decreasing $\mathrm{PI}_\gamma$ (cf. equation (16) in Section 3.1).

[8] Alternatively, one might replace $v_\gamma$ by the derivative of $\gamma = \Gamma(\theta)$ with respect to $\theta$ at, say, the posterior mode of $\theta$, and unreported results show that this leads to almost identical results for the analogue of $r_{\mathrm{sh}}$.

# 5 Conclusion

The paper develops prior sensitivity and prior informativeness measures that shed some light on the role of the prior and likelihood for posterior results in large Bayesian models. Both measures 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 two statistics from MCMC output.

The two suggested measures are scalar summary statistics. By definition, they cannot reflect all features of the high-dimensional likelihood and its interaction with the prior, and one can imagine other useful summary statistics that highlight different aspects. At the same time, with a starting point of the derivative of the posterior mean with respect to the prior mean, it is shown that reasonable restrictions on the properties of such summary statistics naturally lead to the suggested measures.

# A Appendix

**Proof of Theorem 1:**
By Condition 1 (b) and (d), we can restrict attention to the case where $\omega_i^2 > 0$ for all $i = 1, \cdots, k$. Consider first the case of overall limited prior informativeness in the sense of Definition 2. We start by showing that Condition 1 implies the four Axioms of Kitagawa (1934), with $\mathrm{PI}_k$, $k$ and $(\omega_i^2, \lambda_i)$ playing the role of $M_n$, $n$ and $(w_i, x_i)$ in Kitagawa's notation. Condition 1 (b) implies Axiom 1. Condition 1 (e) implies Axiom 2. Repeated application of Condition 1 (f) yields $\mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2\\ \lambda_k\end{pmatrix}, \begin{pmatrix}\omega_2^2\\ \lambda_k\end{pmatrix}, \cdots, \begin{pmatrix}\omega_k^2\\ \lambda_k\end{pmatrix}\right) = \mathrm{PI}_1\left(\begin{pmatrix}\sum_{i=1}^k \omega_i^2\\ \lambda_k\end{pmatrix}\right)$, which equals $\min(\lambda_k, 1) = \lambda_k$ by Condition 1 (c). This shows that Axiom 3 is satisfied. Finally, for Axiom 4, note that applying Condition 1 (g), (b), (f) (repeatedly) yields $\mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2\\ \lambda_1\end{pmatrix}, \cdots, \begin{pmatrix}\omega_k^2\\ \lambda_k\end{pmatrix}\right) = \mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2\\ \bar\lambda_m\end{pmatrix}, \cdots, \begin{pmatrix}\omega_m^2\\ \bar\lambda_m\end{pmatrix}, \begin{pmatrix}\omega_{m+1}^2\\ \lambda_{m+1}\end{pmatrix}, \cdots \begin{pmatrix}\omega_k\\ \lambda_k\end{pmatrix}\right) = \mathrm{PI}_{k-m+1}\left(\begin{pmatrix}\omega_{m+1}^2\\ \lambda_{m+1}\end{pmatrix}, \cdots, \begin{pmatrix}\omega_k^2\\ \lambda_k\end{pmatrix}, \begin{pmatrix}\sum_{i=1}^m \omega_m^2\\ \bar\lambda_m\end{pmatrix}\right)$, so that Axiom 4 follows from another application of Condition 1 (b), with Kitagawa's $w_r^*$ equal to $w_r^* = \sum_{i=1}^r w_i$. Thus, Kitagawa's results are applicable and imply that $\mathrm{PI}_k$ is of the form

$$\mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2\\ \lambda_1\end{pmatrix}, \cdots, \begin{pmatrix}\omega_k^2\\ \lambda_k\end{pmatrix}\right) = g^{-1}\left(\frac{\sum_{i=1}^k \omega_i^2 g(\lambda_i)}{\sum_{i=1}^k \omega_i^2}\right)$$

where $g : [0, 1) \mapsto \mathbb{R}$ is a strictly monotone increasing, continuous function with strictly monotone increasing and continuous inverse $g^{-1}$ (the continuity is not asserted by Kitagawa, but follows from Kolmogorov's (1930) Theorem invoked in Kitagawa's proof). Without loss of generality, normalize $g(0) = 1$.

We now show that $g$ is differentiable at 0. This is obvious if $g$ is constant, so suppose it is not. Recall that every strictly monotone function is Lebesgue almost everywhere differentiable. Thus, the two $[0, 1) \mapsto \mathbb{R}$ functions $f(\lambda) = \frac{1}{2}g(0) + \frac{1}{2}g(\lambda)$ and $g^{-1}(f(\lambda))$ are almost everywhere differentiable. Pick $\lambda_0 > 0$ such that both are differentiable at $\lambda = \lambda_0$. We first argue that this implies that $g^{-1}$ is differentiable at $x_0 = f(\lambda_0)$. Let $h_n$ be arbitrary nonzero reals converging to zero as $n \to \infty$. By continuity and monotonicity of $f$, there exist, for all large enough $n$, $h_n' \neq 0$ such that $h_n = f(\lambda_0 + h_n') - f(\lambda_0)$. Thus

$$\Delta = \lim_{n\to\infty} \frac{g^{-1}(x_0 + h_n) - g^{-1}(x_0)}{h_n}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \frac{g^{-1}(f(\lambda_0+h_n')) - g^{-1}(f(\lambda_0))}{f(\lambda_0+h_n') - f(\lambda_0)} \\
&= \lim_{n\to\infty} \frac{g^{-1}(f(\lambda_0+h_n')) - g^{-1}(f(\lambda_0))}{h_n'} \cdot \frac{h_n'}{f(\lambda_0+h_n') - f(\lambda_0)} \\
&= \frac{dg^{-1}(f(\lambda))}{d\lambda}|_{\lambda=\lambda_0} / \frac{df(\lambda)}{d\lambda}|_{\lambda=\lambda_0}
\end{aligned}
$$

by the product rule for limits, and by the continuity of $g$ at 0,

$$
\frac{g^{-1}\left(\frac{1}{2}g(h_n)+\frac{1}{2}g(\lambda_0)\right) - g^{-1}(\frac{1}{2}g(0)+\frac{1}{2}g(\lambda_0))}{h_n} = \tfrac{1}{2}\Delta \frac{g(h_n)-g(0)}{h_n} + \xi(h_n)
$$

where $\xi(h_n) \to 0$. Now by Condition 1 (e), $\mathrm{PI}_2\left(\begin{pmatrix}1\\ \lambda_1\end{pmatrix}, \begin{pmatrix}1\\ \lambda_0\end{pmatrix}\right) = g^{-1}\left(\frac{1}{2}g(\lambda_1)+\frac{1}{2}g(\lambda_0)\right)$ is differentiable in $\lambda_1$ at $\lambda_1 = 0$. Thus,

$$
\lim_{n\to\infty} \tfrac{1}{2}\Delta \frac{g(h_n)-g(0)}{h_n} + \xi(h_n)
$$

exists and doesn't depend on $h_n$, which implies differentiability of $g$ at 0.

Now by Condition 2,

$$
\mathrm{PI}_2\left(\begin{pmatrix}1\\ \lambda_1\end{pmatrix}, \begin{pmatrix}1\\ 0\end{pmatrix}\right) = g^{-1}\left(\tfrac{1}{2}g(\lambda_1)+\tfrac{1}{2}g(0)\right) = \frac{\lambda_1}{2-\lambda_1}
$$

so that $g(\lambda_1) + g(0) = 2g(\lambda_1/(2-\lambda_1))$ for all $\lambda_1 \in [0,1)$. Define the continuous and strictly monotone increasing function $\varphi : [0,1) \mapsto \mathbb{R}$ as $\varphi(\lambda) = 1/(1-\lambda)$, and let $h : \mathbb{R} \mapsto \mathbb{R}$ be the monotone increasing function such that $g(\lambda) = h(\varphi(\lambda))$. Then $h(x)$ has a positive derivative at $x = 1$, $h(1) = 1$, and

$$
h\left(\frac{1}{1-\lambda_1}\right) + h(1) = 2h\left(\frac{2-\lambda_1}{2-2\lambda_1}\right).
$$

With $\lambda_1 = 1 - 1/x$, we obtain $h(x) + 1 = 2h((x+1)/2)$ for all $x \in [1,\infty)$. Repeated substitution yields $h(x) - 1 = 2^j h(2^{-j}x + (1-2^{-j})1) - 2^j$ for all integer $j$, so that for $x_1, x_2 \in (1,\infty)$

$$
\frac{h(x_1)-1}{h(x_2)-1} = \frac{h(1+2^{-j}(x_1-1))-1}{h(1+2^{-j}(x_2-1))-1} = \frac{dh(x)/dx|_{x=1}}{dh(x)/dx|_{x=1}} = 1.
$$

Thus $h$ is linear function, and the result follows.

Finally, consider the case where $\lambda_i \geq 1$ for some $i$. Let $\tilde{\lambda}_i(n) = 1 - h_n$ if $\lambda_i \geq 1$ and $\tilde{\lambda}_i(n) = \lambda_i$ otherwise, where $h_n$ is a positive sequence converging to zero. Applying the result for the overall identified case, we obtain $\lim_{n\to\infty} \mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2\\ \tilde\lambda_1(n)\end{pmatrix},\cdots,\begin{pmatrix}\omega_k^2\\ \tilde\lambda_k(n)\end{pmatrix}\right)=1$. Furthermore, by Condition 1 (b) and (e), $\mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2\\ \lambda_1\end{pmatrix},\cdots,\begin{pmatrix}\omega_k^2\\ \lambda_k\end{pmatrix}\right)\geq \mathrm{PI}_k\left(\begin{pmatrix}\omega_1^2\\ \tilde\lambda_1(n)\end{pmatrix},\cdots,\begin{pmatrix}\omega_k^2\\ \tilde\lambda_k(n)\end{pmatrix}\right)$ for all $n$, so that the result follows from the range upper bound in Condition 1 (e).

**Inequalities:** Note that for any vector $v^*=G^{-1\prime}v=\sum_{i=1}^k \omega_i q_i$, we have $v'V_p[\theta]v=v^{*\prime}v^*=\sum_{i=1}^k\omega_i^2$, $v'V_\pi[\theta]v=v^{*\prime}J^*v^*=\sum_{i=1}^k\omega_i^2\lambda_i$, $v'V_\pi[\theta]V_p[\theta]^{-1}V_\pi[\theta]v=v^{*\prime}J^{*2}v^*=\sum_{i=1}^k\omega_i^2\lambda_i^2$ and $\mathrm{PI}=\varphi^{-1}(\sum_{i=1}^k\omega_i^2\varphi(\lambda_i)/\sum_{i=1}^k\omega_i^2)$ with $\varphi(\lambda)=1/(1-\lambda)$.

1. Follows from $\sum_{i=1}^k\omega_i^2\lambda_i^2\leq\lambda_{\max}\sum_{i=1}^k\omega_i^2\lambda_i$.
2. Follows from $\sum_{i=1}^k\omega_i^2\lambda_i^2\leq\lambda_{\max}^2\sum_{i=1}^k\omega_i^2$.
3. Follows from $\sum_{i=1}^k\omega_i^2\lambda_i^2/\sum_{i=1}^k\omega_i^2\leq(\sum_{i=1}^k\omega_i^2\lambda_i/\sum_{i=1}^k\omega_i^2)^2$ by convexity.
4. Follows from $\sum_{i=1}^k\omega_i^2\varphi(\lambda_i)/\sum_{i=1}^k\omega_i^2\geq\varphi(\sum_{i=1}^k\omega_i^2\lambda_i/\sum_{i=1}^k\omega_i^2)$ by convexity of $\varphi$.
5. Follows from $\sum_{i=1}^k\omega_i^2\varphi(\lambda_i)\leq\sum_{i=1}^k\omega_i^2\varphi(\lambda_{\max})$ for $\lambda_{\max}\leq 1$, and inequality is trivial otherwise.
6. Note that $\mathrm{PS}/\sqrt{v'V_p[\theta]v}=\varphi_{\mathrm{PS}}^{-1}(\sum_{i=1}^k\omega_i^2\varphi_{\mathrm{PS}}(\lambda_i)/\sum_{i=1}^k\omega_i^2)$ with $\varphi_{\mathrm{PS}}(x)=x^2$. Both PI and $\mathrm{PS}/\sqrt{v'V_p[\theta]v}$ can thus be considered the certainty equivalence of an expected utility maximizer with utility function $\varphi$ and $\varphi_{\mathrm{PS}}$, respectively, facing a lottery with payoff's $\{\lambda_i\}_{i=1}^k$ with probabilities $\{\omega_i^2/\sum_{j=1}^k\omega_j^2\}_{i=1}^k$. The result now follows from Pratt's (1964) Theorem 1, since a calculation shows that $\varphi$ has a weakly larger (negative) coefficient of absolute risk aversion than $\varphi_{\mathrm{PS}}$ on the interval $[0,1/3]$.
7. Follows from $v^{*\prime}(I_k-J^*)^{-1}v^*=v^{*\prime}\sum_{i=1}^\infty(J^*)^iv^*\geq v^{*\prime}(I+J^*+J^{*2})v^*$, so that $\mathrm{PI}=1-v^{*\prime}v^*/v^{*\prime}(I_k-J^*)^{-1}v^*\geq v^{*\prime}(J^*+J^{*2})v^*/v^{*\prime}(I+J^*+J^{*2})v^*\geq\frac{2}{3}v^{*\prime}J^{*2}v^*/v^{*\prime}v^*$.

# References


An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26, 113–172.

Andrle, M. (2010): "A Note on Identificiation Patterns in DSGE Models," *ECB Working Paper Series 1235*.

Basu, S., S. R. Jammalamadaka, and W. Liu (1996): "Local Posterior Robustness with Parametric Priors: Maximum and Average Sensitivity," in *Maximum Entropy and Bayesian Methods: Santa Barbara, California, U.S.A., 1993*, ed. by G. R. Heidbreder, pp. 97–106, Dordrecht, The Netherlands. Kluwer.

Berger, J. O. (1994): "An Overview of Robust Bayesian Analysis," *Test*, 3, 5–124.

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

Cogley, T., and T. J. Sargent (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US," *Review of Economic Dynamics*, 8, 262–302.

Fernandez-Villaverde, J., and J. F. Rubio-Ramirez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74, 1059–1087.

Gustafson, P. (2000): "Local Robustness in Bayesian Analysis," in *Robust Bayesian Analysis*, ed. by D. R. Insua, and F. Ruggeri, pp. 71–88. Springer-Verlag, New York.

Iskrev, N. (2008): "Evaluating the Information Matrix in Linearized DSGE Models," *Economics Letters*, 99, 607–610.

——— (2010a): "Evaluating the Strength of Identification in DSGE Models: An A Priori Approach," *Working paper*.

——— (2010b): "Local Identification in DSGE Models," *Journal of Monetary Economics*, 57, 189–202.

Justiniano, A., and G. E. Primiceri (2008): "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98, 604–641.

Kim, C.-J., and C. R. Nelson (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., and S. Ng (2009): "Dynamic Identification of DSGE Models," *Discussion paper, U.C. San Diego.*

McCulloch, R. E. (1989): "Local Model Influence," *Journal of the American Statistical Association*, 84, 473–478.

Nagumo, M. (1930): "Über eine Klasse der Mittelwerte," *Japan Journal of Mathematics*, 7, 71–79.

Perez, C. J., J. Martin, and M. J. Rufo (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., and T. Zha (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 *Robust Bayesian Analysis*, ed. by D. R. Insua, and F. Ruggeri, pp. 59–108, New York. Springer-Verlag.

Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–1175.

——— (2007): "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

Traum, N., and S. S. Yang (2010): "Does Government Debt Crowd Out Investment? A Bayesian DSGE Approach," *CBO Working Paper Series*.