Global Robust Bayesian Analysis in Large Models[*]

Paul Ho[†]

Princeton University

pho@princeton.edu

October 31, 2018

Abstract

This paper develops tools for global prior sensitivity analysis in large Bayesian models. Without imposing parametric restrictions, the framework provides bounds for a wide range of posterior statistics given any prior that is close to the original in relative entropy. The methodology also reveals parts of the prior that are important for the posterior statistics of interest. To implement these calculations in large models, we develop a sequential Monte Carlo algorithm and use approximations to the likelihood and statistic of interest. We use the framework to study error bands for the impulse response of output to a monetary policy shock in the New Keynesian model of Smets and Wouters (2007). The error bands depend asymmetrically on the prior through features of the likelihood that are hard to detect without this formal prior sensitivity analysis.

1 Introduction

In Bayesian estimation, we are confronted with the questions of how much our posterior estimates depend on our prior, and which parts of the prior are most important. Tackling these questions analytically is difficult because both the likelihood and the statistics of interest may be complicated functions of the parameters over which we define the prior. On the other hand, it is infeasible to repeat the estimation for all possible priors. Given the complicated dependence of the posterior statistic on the prior, one may wish to consider nonparametric changes in the joint prior of multiple parameters, rather than restricting oneself to a particular class of priors that relies on assumptions such as independence or distributional form. Existing approaches to prior sensitivity either restrict the researcher to infinitesimal parametric changes to the prior or become infeasible outside simple low-dimensional settings.

This paper develops a method to investigate the sensitivity of conclusions across a nonparametric set of priors but remains feasible in large models. We shall refer to this method as *relative entropy prior sensitivity* (REPS). The calculation searches over a set of priors that are close to the original prior, then finds the *worst-case* prior that leads to the largest change in the reported posterior estimates. The set of priors to be considered is a relative entropy ball around the original prior, following the robust control literature (Petersen et al. (2000); Hansen and Sargent (2001)), which allows us to search across a flexible nonparametric set of priors. The solution for the worst-case prior and posterior requires solving for only one scalar regardless of the dimensionality of the parameter, then reweighting draws from the original prior and posterior. To overcome cases where direct reweighting results in poor approximations for the worst-case distributions, we develop a sequential Monte Carlo (SMC) algorithm to obtain draws from these distributions.

The prior sensitivity analysis informs an econometrician of how sensitive her posterior results are to the prior, and identifies particular features of the prior that are important for these results. For example, if the econometrician reports the posterior mean of an elasticity, she would search for the priors that respectively minimize and maximize this mean for a given relative entropy, thus obtaining bounds on the posterior mean. The worst-case prior will differ most from the original prior in dimensions that are most important for the posterior mean. These are parts of parameter space that are not well-identified by the likelihood but matter for the posterior mean.

To generate draws from the worst-case prior and posterior in complicated and high-dimensional settings, we adapt the SMC algorithm of Herbst and Schorfheide (2014) and use approximations of the likelihood and statistic of interest. In principle, one could take a sufficiently large number of draws from the original prior and posterior, solve for the worst-case prior and posterior, then use importance sampling to reweight the draws from the original distributions. However, importance sampling performs poorly when the distribution of these weights has fat tails. SMC overcomes the challenge by solving for a sequence of intermediate priors and posteriors, and recursively obtaining draws from each of these intermediate distributions. However, SMC can become computationally infeasible due to the need to repeatedly compute the likelihood and statistic of interest. To reduce the computational burden, we use approximations of the inputs of the REPS calculations in a procedure we refer to as *approximate relative entropy prior sensitivity* (AREPS). AREPS yields draws from an approximate worst-case prior and posterior, which can then be reweighted to obtain draws from the the exact worst-case distributions.

To gauge the sensitivity of one's posterior results to the prior, we provide a rule of thumb for what a large or small relative entropy is. In particular, we derive asymptotics for the relative entropy needed to shift the posterior mean in the Gaussian location model by a factor of the posterior standard deviation. Using a formula based on these asymptotics, we offer a rule of thumb to interpret the level of relative entropy in more general settings. This allows a practitioner to quantify the sensitivity of her posterior to the prior by measuring how different the original and worst-case priors need to be to generate a given change in the posterior estimate.

Our main application is the impulse response of output to a monetary policy shock in the large-scale New Keynesian model of Smets and Wouters (2007). We use REPS to construct bounds that contain the pointwise 68% error bands arising from any prior in a relative entropy ball around the original prior, and compare the bounds when we distort the prior for the Taylor rule parameters only to when we distort the nominal frictions parameters only. The impulse response is more sensitive to the prior of the nominal frictions parameters because the likelihood for these parameters is less concentrated. For both sets of parameters, the upper bound of the error bands is more sensitive to changes to the prior and the impulse response is more dependent on the prior at long horizons. The worst-case prior adds mass to certain regions of the tail of the posterior where the likelihood is large. In addition, the worst-case distortions show that joint changes in the prior can result in

larger changes in the posterior than if one were to distort the marginals only. This dependence on the prior may be hard to discern without the systematic framework provided by REPS.

1.1 Related literature

REPS overcomes key limitations of existing approaches to prior sensitivity analysis. Local methods (e.g. Gustafson (2000); Müller (2012)) consider derivatives of specific posterior quantities with respect to the prior. Although these methods can be used for high-dimensional problems, they not only focus on infinitesimal changes in the prior and posterior, but are also restricted to parametric changes in the prior. As we show in both our main application and stylized examples, these restrictions may result in misleading conclusions about prior sensitivity. REPS does not impose such restrictions, allowing for joint nonparameteric distortions across parameters. Global methods (e.g. Berger and Berliner (1986); Moreno (2000)) allow for a larger class of priors and consider potentially large changes in the posterior, but lose tractability once we move beyond simple applications. REPS is a global method that is feasible in settings with high-dimensional and complicated likelihoods such as large-scale New Keynesian models. Moreover, it can be applied to a range of statistics such as means, cumulative distribution functions, and credible sets of a wide class of functionals of the model parameters.

The key difference between REPS and the existing robust control literature is that the worst-case prior from REPS depends on the likelihood. Hansen and Sargent (2007) also solve for a worst-case prior that is constrained to be close to an economic agent's original prior in relative entropy. However, they consider an ex-ante problem that does not condition on the observed data. In contrast, here we consider an econometrician analyzing the prior after observing data, and therefore condition our worst-case prior on the data. Conditioning on the data allows REPS to account for characteristics of the likelihood that are important for our posterior results. In related work, Giacomini et al. (2016) construct a relative entropy ball around the prior for a partially identified parameter. Because the parameter is not identified without further restrictions, this reduces to the usual robust control problem of Hansen and Sargent (2001) conditional on the identified parameters in the model. The minimax problem presented in Giacomini et al. (2016) under a quadratic loss function is a special case of REPS. This paper develops a more general methodology that applies to any Bayesian estimation problem, in contrast to Giacomini et al.

([2016](#)) who focus on partial identification problems.

The importance of prior sensitivity analysis is especially salient in Bayesian dynamic stochastic general equilibrium (DSGE) models like our main application (see Herbst and Schorfheide ([2015](#)); Fernández-Villaverde et al. ([2016](#)) for overviews). These models have many parameters connected by numerous equilibrium conditions. As a result, priors typically rely on simplifying assumptions such as independence or conjugacy, which potentially matter for the posterior. For example, Del Negro and Schorfheide ([2008](#)) show that the joint prior matters for the posterior estimates of the role of nominal rigidities. Moreover, some of the parameters do not have a tight range of values that is widely accepted, which may give rise to an audience with heterogeneous priors. Systematic prior sensitivity analysis provides diagnostics for whether such an audience should be concerned about one's posterior estimates. The need for prior sensitivity analysis is further motivated the identification problems in DSGE models described by Canova and Sala ([2009](#)). In contrast to the subsequent literature on identification in DSGE models (e.g. Iskrev ([2010](#)); Komunjer and Ng ([2011](#)); Koop et al. ([2013](#))) that primarily focuses on asymptotic identification, the framework here takes the Bayesian approach of conditioning on current observed data.

While our main application is a DSGE model, REPS can be applied to any Bayesian estimation. For instance, Bayesian methods are widely used in the estimation of vector autoregressions (VARs). Del Negro and Schorfheide ([2004](#)) and Giannone et al. ([2017](#)) show that the priors in VARs play an important role for forecasting. Baumeister and Hamilton ([2015](#)) and Giacomini and Kitagawa ([2015](#)) show that the role of priors is especially prominent when structural VARs are partially identified. Bayesian methods have also been used by Abdulkadiroğlu et al. ([2015](#)) and Avery et al. ([2016](#)) to estimate matching models.

1.2 Outline

The organization of the paper is as follows. I introduce the REPS framework in Section [2](#) and demonstrate the methodology using two stylized examples in Section [3](#). Section [4](#) derives the asymptotics from the Gaussian location model, which motivates a rule of thumb to quantify the difference between priors. I discuss the numerical implementation in Section [5](#). In Section [6](#), we turn to the DSGE model of Smets and Wouters ([2007](#)) as our main application. Section [7](#) concludes.

2 Relative entropy prior sensitivity

2.1 Setting

Consider the Bayesian estimation of a parameter $\theta \in \Theta$ given data X. Bayes rule states that the prior $\pi(\theta)$ and likelihood $L(\theta|X)$ imply the posterior:

$$p(\theta|X) \propto \pi(\theta) L(\theta|X).$$ (2.1)

Suppose we are interested in the posterior of a function $\psi : \Theta \to \Psi$ of the parameter θ. We do not impose restrictions on Ψ, allowing the function ψ to be multidimensional. For example, $\psi(\theta)$ could be an elasticity, a variance decomposition, or an impulse response function at an arbitrary range of horizons. Denote the expectation under an arbitrary probability measure f by $\mathbb{E}_f[\cdot]$. We define the objective function $\gamma_\psi : \Psi \to \mathbb{R}$ so that $\mathbb{E}_p[\gamma_\psi(\psi)]$ captures the property of the posterior distribution of ψ that we are interested in. For instance, if we set $\gamma_\psi(\psi) = \psi$, then $\mathbb{E}_p[\gamma_\psi(\psi)]$ is the posterior mean of ψ. Altenatively, if we take $\gamma_\psi(\psi) = \mathbf{1}\{\psi \leq \psi^*\}$, then $\mathbb{E}_p[\gamma_\psi(\psi)]$ is the cumulative distribution function of ψ evaluated at ψ^*. We denote $\gamma(\theta) \equiv \gamma_\psi(\psi(\theta))$, and study how $\mathbb{E}_p[\gamma(\theta)]$ depends on π.

2.2 Alternative priors and implied posteriors

To study the dependence of the posterior p on the prior π, we shall search across a set of alternative priors. For each alternative prior, we need to describe the implied distorted posterior.

Consider an alternative prior $\tilde{\pi}$ that is absolutely continuous with respect to π. We can write:

$$\tilde{\pi}(\theta) \equiv M(\theta) \pi(\theta),$$ (2.2)

where M is the Radon-Nikodym derivative of $\tilde{\pi}$ with respect to π. Since $\tilde{\pi}$ is a probability distribution, we have $M > 1$ and $\mathbb{E}_\pi[M] = 1$.

Given the likelihood L, the prior $\tilde{\pi}$ implies the distorted posterior:

$$\tilde{p}(\theta|X) = \frac{M(\theta)}{\mathbb{E}_p[M]} p(\theta|X).$$ (2.3)

To see this, notice that Bayes rule implies:

$$\tilde{p}(\theta|X) \propto \tilde{\pi}(\theta) L(\theta|X)$$

$$= M(\theta) \pi(\theta) L(\theta|X) = M(\theta) p(\theta|X). \tag{2.4}$$

The normalization by $\mathbb{E}_p[M]$ ensures that $\tilde{p}$ integrates to one.

For any function $g : \Theta \to G$, the prior and posterior expectations arising from the alternative prior $\tilde{\pi}$ can be written $\mathbb{E}_\pi[M(\theta) g(\theta)]$ and $\mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]} g(\theta)\right]$ respectively. Given draws $\{\theta_{\pi,j}\}_{j=1}^J$ from π and $\{\theta_{p,j}\}_{j=1}^J$ from p respectively, we can evaluate these expectations by importance sampling, using the weighted sample averages:

$$\mathbb{E}_{\tilde{\pi}}[g(\theta)] \approx \frac{1}{J} \sum_{j=1}^J M(\theta_{\pi,j}) g(\theta_{\pi,j}) \tag{2.5}$$

$$\mathbb{E}_{\tilde{p}}[g(\theta)] \approx \frac{\sum_{j=1}^J M(\theta_{p,j}) g(\theta_{p,j})}{\sum_{j=1}^J M(\theta_{p,j})} \tag{2.6}$$

As is standard with importance sampling, the accuracy of these approximations depends on the tail behavior of the distribution of M. If M has fat tails under the prior or posterior, the respective approximations will perform poorly.

2.3 Setup

I now describe the optimization problem that we use to analyze the sensitivity of the posterior estimates to the prior. We search across a set of alternative priors, seeking the worst-case prior that yields the largest change in the posterior mean of the objective function γ. Since the original prior typically contains useful information about the distribution of θ, we restrict the set of priors to be close to the original prior in relative entropy. We shall refer to this framework as *relative entropy prior sensitivity* (REPS).

Comparing the change in the prior to the change in the posterior mean of γ tells us of how much the posterior mean of γ depends on the prior. Comparing the worst-case prior to the original prior indicates parts of the prior that are important for determining the posterior mean of γ.

2.3.1 Primal problem

Formally, we solve an extension of the constraint problem of Hansen and Sargent (2001):

$$\min_{M(\theta):\mathbb{E}_\pi[M]=1} \mathbb{E}_p \left[\frac{M(\theta)}{\mathbb{E}_p[M]} \gamma(\theta) \right] \tag{2.7}$$

$$\text{s.t. } \mathbb{E}_\pi \left[M(\theta) \log M(\theta) \right] \leq \mathcal{R} \tag{2.8}$$

The minimization over M satisfying $\mathbb{E}_\pi[M] = 1$ is equivalent to minimizing over alternative priors, as the random variable M indexes the possible priors. We choose the prior that minimizes $\mathbb{E}_p \left[\frac{M(\theta)}{\mathbb{E}_p[M]} \gamma(\theta) \right]$, which is the distorted posterior mean of γ. Replacing the minimization operator with a maximization operator gives the upper bound for the posterior mean. The left-hand side of the constraint (2.8) is the relative entropy or Kullback-Leibler divergence of the alternative prior relative to the original prior. $\mathcal{R}$ is a positive constant that provides an upper bound on the relative entropy, limiting us to priors that are statistically difficult to distinguish from the original prior. As $\mathcal{R} \to 0$, the worst-case and original priors converge as we are restricted to choosing $M = 1$. Section 4 gives benchmarks for large and small values of $\mathcal{R}$.

Relative entropy measures the information that a Bayesian with the original prior π needs to gather to change her beliefs to the alternative prior $M\pi$ and is invariant to the parameterization of θ. These properties provide theoretical justification for the use of relative entropy. Moreover, we will show that the functional form for relative entropy delivers an analytic solution to (2.7)-(2.8) that allows REPS to maintain tractability in large models.

REPS differs from the usual constraint problem of Hansen and Sargent (2001) and the prior robustness problem from Hansen and Sargent (2007) because the objective function in (2.7) and the relative entropy in (2.8) are taken under different probability measures. This difference in probability measures arises because we are interested in how ex-ante beliefs affect ex-post estimates. Since we wish to report the *posterior* estimate of $\mathbb{E}_p[\gamma(\theta)]$, our objective function is the distorted *posterior* mean of γ, which conditions on the observed data. However, we wish to consider small changes in the *prior*, and are thus led to restrict the relative entropy of the alternative *priors* with respect to the original prior.

2.3.2 Dual problem

Instead of specifying the bound $\mathcal{R}$ on relative entropy, it is convenient to specify the worst-case posterior mean $\tilde{\gamma}$ and solve the dual problem:

$$\min_{M(\theta):\mathbb{E}_\pi[M]=1} \mathbb{E}_\pi\left[M(\theta)\log M(\theta)\right] \tag{2.9}$$

$$\text{s.t. } \mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\gamma(\theta)\right] = \tilde{\gamma} \tag{2.10}$$

We are now searching across priors that imply that γ has posterior mean $\tilde{\gamma}$, and picking the one that is closest to the original prior in terms of relative entropy. We justify the formulation (2.9)-(2.10) below and argue that it simplifies the solution. Robertson et al. (2005) minimize relative entropy subject to moment constraints, in order to find the forecasting model satisfying the posterior moment constraints that is closest to some benchmark model. Unlike them, we take the relative entropy of the prior instead of the posterior.

2.4 Solution

The solution to (2.9)-(2.10) has the form:

$$M(\theta) \propto \exp\left[\lambda L(\theta|X)(\gamma(\theta) - \tilde{\gamma})\right], \tag{2.11}$$

where $\lambda \in \mathbb{R}$ is a constant to be solved for from the constraint (2.10). We have therefore reduced the minimization over a nonparametric set of priors to a problem with one equation and one unknown, regardless of the dimensionality of θ. This is the key feature that makes REPS feasible in large models.

The distortion M depends on the parameter θ through the objective function γ and likelihood L. The worst-case distortion M reweights according to γ because the statistic of interest is the posterior mean of γ. The direction and degree of reweighting depends on λ, which is the Lagrange multiplier on (2.10) scaled by $\mathbb{E}_p[M]$. λ is analogous to the penalty parameter in the multiplier problem of Hansen and Sargent (2001). If $\tilde{\gamma} < \mathbb{E}_p[\gamma(\theta)]$, then in order to reduce the posterior mean of γ, we require $\lambda < 0$ so that the worst-case prior places more weight on smaller values of γ.

The difference between the solution (2.11) and the worst-case distortion in Hansen and Sargent

(2001) is that the distortion in (2.11) is scaled by the likelihood L, which captures the role of the data in (2.9)-(2.10). The worst-case distortion depends on the likelihood because the expectations in (2.9) and (2.10) are taken under different probability measures. Since the posterior is proportional to the product of the prior and the likelihood, concentrating distortions in the high-likelihood regions generates large changes in the posterior from small distortions of the prior. If the likelihood is flat, we return to the standard exponential tilt of Hansen and Sargent (2001).

To solve for λ, substitute (2.11) into (2.10):

$$\tilde{\gamma} = \mathbb{E}_p \left[\underbrace{\frac{\exp\left[\lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right) - \hat{\gamma}\right)\right]}{\mathbb{E}_p\left[\exp\left[\lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right) - \hat{\gamma}\right)\right]\right]}}_{\text{change of measure}} \gamma\left(\theta\right) \right]. \tag{2.12}$$

The right-hand side is the posterior expectation of γ after a change of measure that depends on λ. As λ increases, the change of measure places more weight on large values of γ, thus increasing the right-hand side. Since the left-hand side is a constant and the right-hand side is increasing in λ, (2.12) implies a unique solution for λ that is straightforward to solve for numerically.

Appendix A shows that (2.7)-(2.8) also produces a solution of the form (2.11), but requires us to solve for both λ and $\tilde{\gamma}$. The multiplier representation of (2.7)-(2.8), which is the Lagrangian problem, specifies λ and requires us to solve for $\tilde{\gamma}$. We favor the dual representation (2.9)-(2.10) because $\tilde{\gamma}$ has a straightforward interpretation, while λ has no clear economic interpretation and is difficult to specify ex-ante. It is therefore more convenient to specify a reasonable change in the posterior mean of γ, then check how much the prior needs to be distorted in order to generate this change. Section 4 suggests ways to quantify the change in the prior for a sequence of values for $\tilde{\gamma}$, in order to establish the relationship between the set of priors and worst-case posterior mean $\tilde{\gamma}$.

2.5 Extensions

The REPS framework allows for flexibility in application. We now consider several prior sensitivity problems that can be incorporated into (2.9)-(2.10).

2.5.1 Credible intervals

One can adapt the constraint (2.10) to analyze the prior sensitivity of quantiles of the function of interest $\psi(\theta)$. This allows us to produce credible intervals for ψ that are robust to deviations in the prior. In particular, suppose we are interested in the dependence of the qth quantile of ψ on the prior. Then we can solve (2.9) subject to:

$$\mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\mathbf{1}\left\{\psi(\theta) < \tilde{\psi}\right\}\right] = q, \tag{2.13}$$

where $\tilde{\psi}$ is the worst-case quantile of ψ. Since (2.13) is (2.10) with $\tilde{\gamma} = q$ and $\gamma(\theta) = \mathbf{1}\left\{\psi(\theta) < \tilde{\psi}\right\}$, the solution has the form (2.11), with the appropriate substitution for $\tilde{\gamma}$ and γ.

With these substitutions, we have $\gamma(\theta) - \tilde{\gamma} \in \{1 - \tilde{\gamma}, -\tilde{\gamma}\}$. M therefore only takes on extreme values if θ has high likelihood. If the likelihood is flat, then

$$M(\theta) = \begin{cases} \exp\left[\lambda(1-\tilde{\gamma})\right] & \psi(\theta) < \tilde{\psi} \\ \exp\left[-\lambda\tilde{\gamma}\right] & \psi(\theta) \geq \tilde{\psi} \end{cases} \tag{2.14}$$

In contrast, if γ were not bounded, then M would take on extreme values either when θ has high likelihood or when θ implies an extreme value of γ.

2.5.2 Subspaces

Taking the expectations in (2.9)-(2.10) over the marginal prior and posterior of a subspace Θ^* of Θ allows us to study the dependence of the posterior on the marginal prior over Θ^* instead of the entire space Θ. Such an exercise can be useful if there is a natural partition for θ. For example, in a New Keynesian model, one may be especially concerned about a subset of parameters whose priors are hard to calibrate from existing data. The reduced dimensionality of Θ^* can simplify the analysis of the worst-case distortions.

More formally, consider

$$\min_{M(\theta^*):\mathbb{E}_{\pi^*}[M]=1} \mathbb{E}_{\pi^*}\left[M(\theta^*)\log M(\theta^*)\right] \tag{2.15}$$

$$\text{s.t. } \mathbb{E}_{p^*}\left[\frac{M(\theta^*)}{\mathbb{E}_{p^*}[M]}\mathbb{E}_p\left[\gamma(\theta)\,|\theta^*\right]\right] = \tilde{\gamma} \tag{2.16}$$

where π^* and p^* are the marginal prior and posterior over Θ^*. The constraint (2.16) arises by applying the law of iterated expectations to (2.10) and noting that M now depends on θ^* only. Define $\gamma^*(\theta^*) \equiv \mathbb{E}_p\left[\gamma(\theta)\,|\theta^*\right]$. The solution to (2.15)-(2.16) is:

$$M(\theta^*) \propto \exp\left[\lambda L^*(\theta^*|X)(\gamma^*(\theta^*)-\tilde{\gamma})\right], \tag{2.17}$$

where $L^*(\theta^*|X) \equiv p^*(\theta^*|X)/\pi^*(\theta^*)$ is the marginal likelihood of θ^*. The solution (2.17) to the subspace problem differs from the original solution (2.11) in two aspects. Firstly, the likelihood L is replaced with the marginal likelihood L^*. Secondly, the objective function $\gamma(\theta)$ is replaced by its expectation conditional on θ^*. Section 5 proposes using approximations for both of these objects when implementing these calculations in practice.

(2.15)-(2.16) nests the problem in Giacomini et al. (2016) under a quadratic loss function. Since the posterior expectation minimizes expected quadratic loss, equation (2.16) is equivalent to:

$$\min_{\gamma'}\left(\gamma' - \mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\gamma(\theta)\right]\right)^2 = \tilde{\gamma} \tag{2.18}$$

The primal problem is thus identical to that of Giacomini et al. (2016) if we take θ^* to be the unidentified parameter. However, in the problem of Giacomini et al. (2016), θ^* is not identified, and there is hence no distinction between prior and posterior once we condition on the identified parameters.

2.5.3 Additional constraints

We can further restrict the set of permissible priors by including prior or posterior moment restrictions to (2.9)-(2.10), as in the "tilted robustness" problem of Bidder et al. (2016). Each additional restriction produces one additional multiplier to solve for, while the moment restriction provides
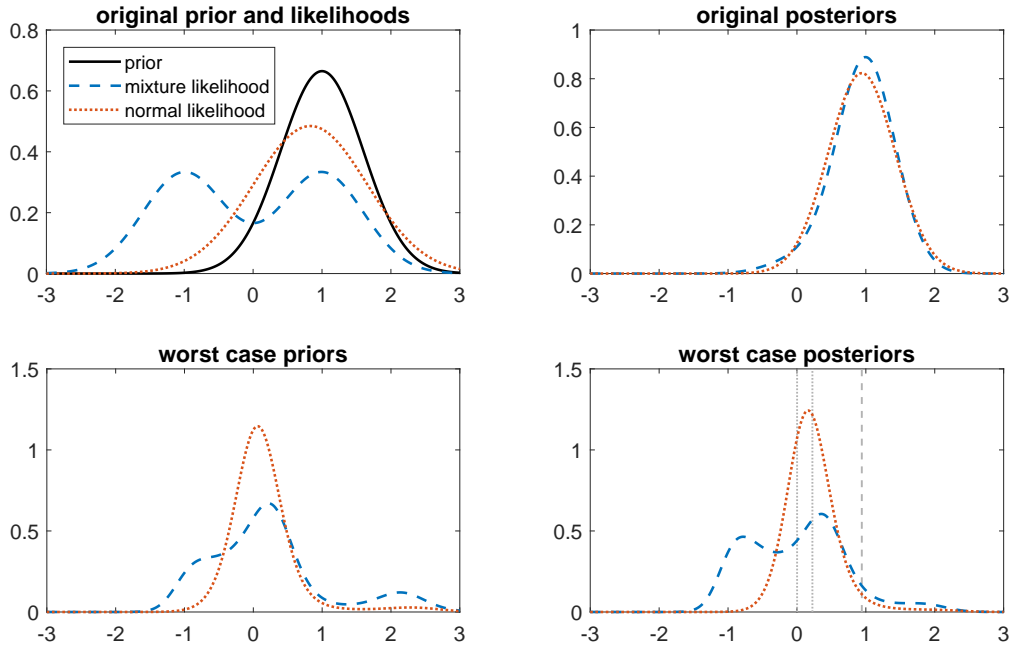
Figure 3.1: Likelihoods, priors and posteriors for mixture and normal likelihoods. Blue dashed lines correspond to mixture likelihood, red dotted lines correspond to normal likelihood. **Top left:** Original prior and likelihoods; **Top right:** Original posteriors; **Bottom left:** Worst case priors; **Bottom right:** Worst case posteriors, with original and worst case means in gray.

the additional equation with which to solve for the new unknown. See Appendix A for details.

3   Two illustrative examples

We now use two simple examples to illustrate how REPS can diagnose dependence on the prior that may be hard to detect otherwise. The first shows that REPS accounts for fat tails in the posterior, while the second demonstrates how prior sensitivity can depend on the mapping between the primitive parameters and the function of interest ψ. Such features are prevalent in many applications, and can be difficult to detect in high-dimensional problems, even with existing prior sensitivity tools.

3.1   Example 1: multiple modes in the likelihood

The first example shows that REPS accounts for tail behavior of the likelihood that may be hard to distinguish from visual inspection of the posterior and prior.

Suppose $\theta \in \mathbb{R}$, and we have the prior $\theta \sim \mathcal{N}\left(1, 0.6^2\right)$. We consider two alternative likelihoods — a mixture model:

$$X \sim \begin{cases} \mathcal{N}\left(-\theta, 0.6^2\right) & \text{w.p. } 0.5 \\ \mathcal{N}\left(\theta, 0.6^2\right) & \text{w.p. } 0.5 \end{cases} \tag{3.1}$$

with data $X = 1$, and a Gaussian model:

$$X \sim \mathcal{N}\left(\theta, 0.678^2\right) \tag{3.2}$$

with data $X = 0.831$. The parameters of the Gaussian model are picked so that both models produce posteriors with mean 0.942 and standard deviation 0.485. The top right panel of Figure 3.1 shows that the two posteriors are indeed hard to distinguish visually. On the other hand, the top left panel of Figure 3.1 shows that the mixture likelihood has modes at -1 and 1, while the normal likelihood has only one mode at 0.831.

REPS shows that the posterior mean of θ has greater sensitivity to changes in the prior under the mixture likelihood. In particular, fixing $\gamma\left(\theta\right) = \theta$ and $\mathcal{R} = 1.25$, we solve (2.7)-(2.8) for both models. Since the original prior and $\mathcal{R}$ are identical across the models, we are choosing from the same set of priors in both cases. However, the mixture model's worst-case posterior mean of 0.002 is substantially lower than the normal model's worst-case posterior mean of 0.224. Without prior sensitivity analysis, the similarity of the posteriors in the two models might mislead one to believe that the models are equally sensitive to changes in the prior. Local prior sensitivity methods may also fail to detect the difference in the prior sensitivity. For instance, the derivative of the posterior mean with respect to the prior mean, as considered in Müller (2012), is 0.653 for the mixture likelihood and 0.654 for the normal likelihood, suggesting that the posterior mean in the two cases is roughly equally sensitive to the prior.

The worst-case priors and posteriors, plotted in the bottom row of Figure 3.1, indicate the importance of the wide range of alternative priors considered by REPS. With the normal likelihood, the worst-case prior for θ is approximately normal, centered around 0. The worst-case posterior is also approximately normal, also centered around the new mean of 0.2. In contrast, with the mixture likelihood, the worst-case prior is flatter, and places a relatively large mass around $\theta = -1$. The worst-case posterior now bimodal, with a second mode around $\theta = -1$ corresponding to the

second mode in the likelihood, which was not visible from the original posterior. The worst-case distortions are informative about parts of the prior the one should be concerned about even if one does not regard the exact shape of the worst-case prior as being plausible.

This example illustrates how the robustness of a posterior estimate to changes in the prior can depend on peaks in the likelihood that are dampened by the original prior. In such cases, visual inspection of the prior and posterior can mislead one to believe a result is more robust than it actually is. Herbst and Schorfheide (2014) show that under more diffuse priors, the DSGE models of Smets and Wouters (2007) and Schmitt-Grohé and Uribe (2012) produce multimodal posteriors that can alter inference relative to a tighter prior. We find that these features matter for posterior inference in our main application to the DSGE model of Smets and Wouters (2007) in Section 6. In general, such multimodality is hard to detect without reestimating the model for different priors. Flatter marginal priors may not reveal these modes, since the parameterization and independence assumptions matter for how flattening the marginals impacts the posterior of the object of the interest. REPS provides a systematic approach to prior sensitivity analysis that accounts for features of the likelihood in the tail of the posterior, such as multiple modes, if they are important for one's posterior results.

3.2 Example 2: log-normal distribution

We now show that even if the prior and likelihood are Gaussian, REPS detects that the sensitivity to the prior depends on the function ψ of the parameters that we are interested in.

Suppose $\theta \in \mathbb{R}$ and we have prior $\theta \sim \mathcal{N}(0,1)$, $X \sim \mathcal{N}(\theta,1)$, and observe data $X = 0$. The posterior is $\theta \sim \mathcal{N}\left(0, \frac{1}{2}\right)$. Suppose we wish to do REPS analysis on $\psi(\theta) = \exp(\theta)$, so that ψ is log-normal with mean 1.28 and standard deviation 1.03. Figure 3.2 shows that the posterior of ψ is skewed even though the posterior for θ is symmetric.

REPS shows that this asymmetry matters for the sensitivity of the posterior mean of ψ to changes in the prior. In particular, fix $\gamma(\theta) = \psi(\theta)$, and take $\mathcal{R} = 0.57$, which corresponds to a one standard deviation change in the posterior mean of θ. For this value of $\mathcal{R}$, the posterior mean of ψ has a maximum value of 2.50 (an increase of 1.18 standard deviations) and minimum value of 0.59 (a decrease of 0.67 standard deviations). The asymmetry in sensitivity arises because the ψ is bounded below by zero, but has a posterior with a fat right tail.
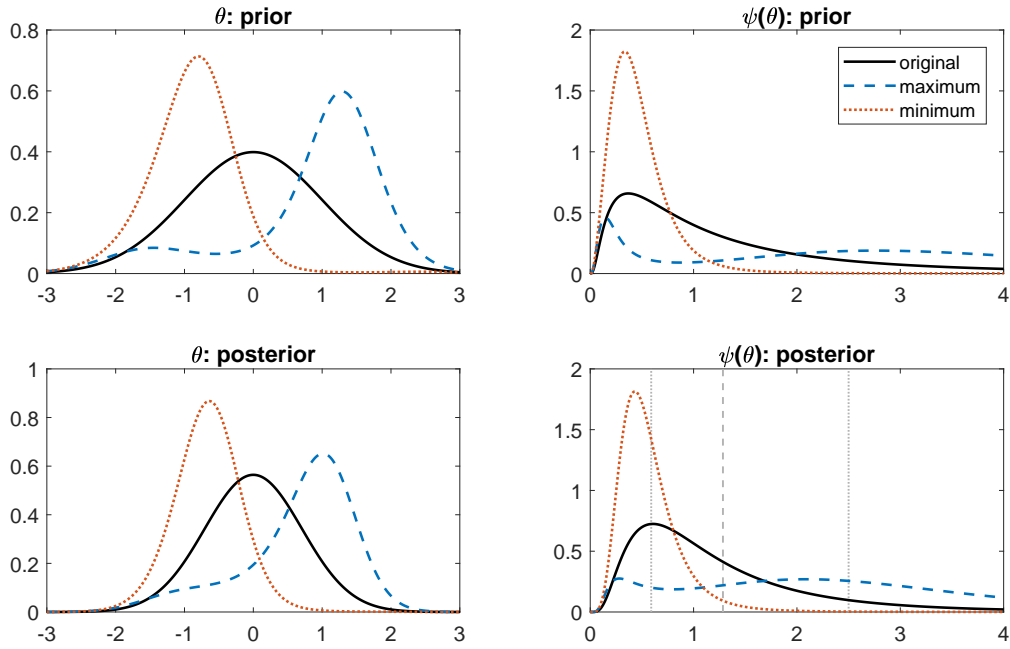
Figure 3.2: Prior and posterior of θ and $\psi(\theta)$. Black solid lines correspond to original distribution, blue dashed lines correspond to distributions that maximize mean, red dotted lines correspond to distributions that minimize mean. **Top left:** Prior of θ; **Top right:** Prior of $\psi(\theta)$ **Bottom left:** Posterior of θ; **Bottom right:** Posterior of $\psi(\theta)$.

The worst-case distortions are also asymmetric. The prior that maximizes the mean distorts the tails more relative to the prior that minimizes the mean, because the convexity of the exponential function amplifies (dampens) the effect of distortions on the right (left) tail of θ on the mean of ψ. The asymmetry arises despite the symmetry of the Gaussian prior and likelihood. Since relative entropy is invariant to transformations of θ, the set of priors does not depend on the parameterization of the problem.

These insights generalize to more complex settings, where ψ may be a complicated function whose sensitivity to the prior may be hard to analyze. For example, if ψ is an impulse response in a DSGE model, one would need to solve the model and then evaluate the impulse response. The challenge is compounded when θ is high-dimensional. Even if visual inspection of the posterior of ψ reveals that it could be sensitive to the prior, it is not clear without further analysis which parts of the prior of θ are important. REPS accounts for the function of interest through the γ term in the solution (2.11) while checking across a wide range of alternative priors.

4 Quantifying the change in prior

In order to quantify prior sensitivity, we need to quantify how much the prior has changed in order to produce the specified change in posterior mean. In this section, I show how $\mathcal{R}$ in the Gaussian location model depends on features such as the number of dimensions, prior variance, sample size, and sample mean. I then provide a formula that summarizes these relationships, and give practitioners a rule of thumb for what a large or small value of $\mathcal{R}$ is in more general settings.

4.1 Gaussian location model

We first study the properties of the Gaussian location model.

4.1.1 Setup

We estimate a d-dimensional vector $\theta = (\theta_1, ..., \theta_d)'$ whose true value is θ_0. We have prior $\theta \sim \mathcal{N}(0, \Sigma_\pi)$ and observe T iid realizations of $X \sim \mathcal{N}(\theta, \Omega)$ with sample mean $\overline{X}_T$. Assume Ω is full rank. This yields a posterior $\theta \sim \mathcal{N}(\theta_{p,T}, \Sigma_{p,T})$ where

$$\theta_{p,T} = T \Sigma_{p,T} \Omega^{-1} \overline{X}_T \tag{4.1}$$

$$\Sigma_{p,T} = \left(\Sigma_\pi^{-1} + T \Omega^{-1} \right)^{-1} \tag{4.2}$$

Denote the posterior standard deviation of θ_i by $\sigma_{i,p,T}$.

4.1.2 Asymptotics

Lemma 1. *Suppose $\gamma(\theta) = \theta_1$ and $\tilde{\gamma}_T = \theta_{1,p,T} - c\sigma_{1,p,T}$ where $c \in \mathbb{R}_+$. Then as $T \to \infty$, the solution to (2.9)-(2.10) satisfies $T^{\frac{d}{2}} \mathcal{R}_T \to \overline{R}\pi(\theta_0) |\Omega|^{-\frac{1}{2}}$ for some $\overline{R}$.*

Lemma 1 states that the relative entropy needed to shift the posterior mean by c posterior standard deviations declines at rate $T^{\frac{d}{2}}$, and depends on the variance of the data and the prior at θ_0. We have not specified the scaling factor $\overline{R}$, which could vary depending on context, including the number of dimensions d and amount of distortion c. When we use Lemma 1 to provide a rule of thumb for $\mathcal{R}$, we shall also discuss values for $\overline{R}$. We now discuss the intuition of Lemma 1 here and provide the formal proof in Appendix A.
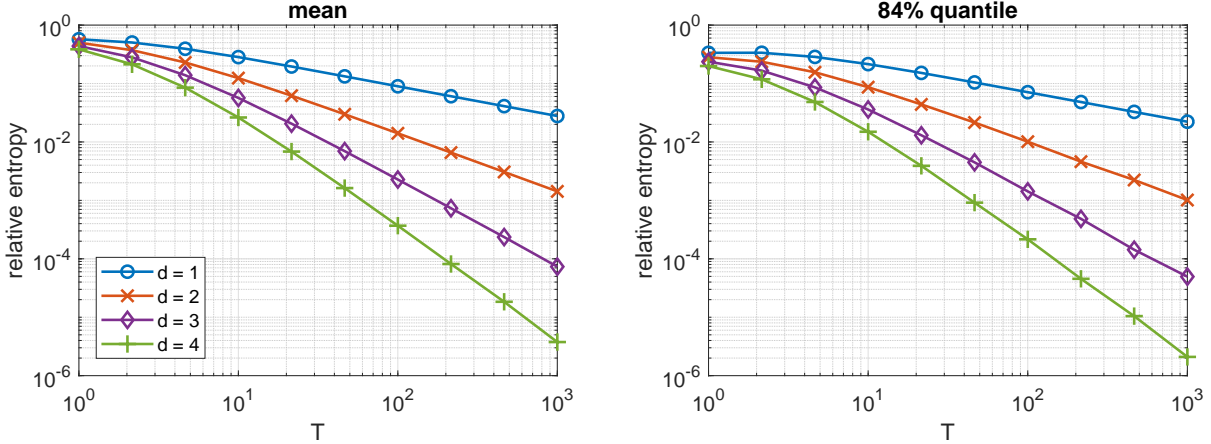
Figure 4.1: Relative entropy for given number of observations in Gaussian location model with $\theta \sim \mathcal{N}(0, I)$, $\overline{X}_T = 0$, $d \in \{1, ..., 4\}$. **Left:** increase in posterior mean of θ_1 by one posterior standard deviation; **Right:** increase in 84% quantile of θ_1 by one posterior standard deviation.

As the likelihood becomes more concentrated, the distortions asymptotically concentrate around a small region whose volume shrinks at rate $T^{\frac{d}{2}}$. The $|\Omega|$ term accounts for the dispersion of the likelihood for a given T. Since the likelihood concentrates around θ_0, the asymptotic relative entropy is scaled by $\pi(\theta_0)$. We can think of regions of high prior probability as being close to pre-sample observations of X. The relative entropy increases with $\pi(\theta_0)$ because it is harder to change our beliefs in a region that we have knowledge about.

For other models that satisfy the Bernstein-von Mises theorem, we will have the same asymptotic result, with $|\Omega|$ replaced by the Fisher information matrix. In addition, the same asymptotics apply for the mean or quantile of any linear combination of θ.

We now show that Lemma 1 provides a good approximation for the relative entropy in the Gaussian location model even for relatively small values of T. In each case, we set $\Sigma_\pi = \Omega = I$ and show that the relative entropy needed to:

1. increase the posterior mean of θ_1 by one posterior standard deviation; or

2. increase the 84% quantile of θ_1 by one posterior standard deviation,

approximately scales with $T^{-\frac{d}{2}}$ and $\pi(\theta_0)$, as predicted by Lemma 1.

Sample size and dimension. To show the dependence of $\mathcal{R}$ on $T^{\frac{d}{2}}$, we first fix $\overline{X}_T = 0$ and vary $d \in \{1, ..., 4\}$ and $T \in \{1, ..., 10^3\}$. Figure 4.1 shows the relative entropy for different values of
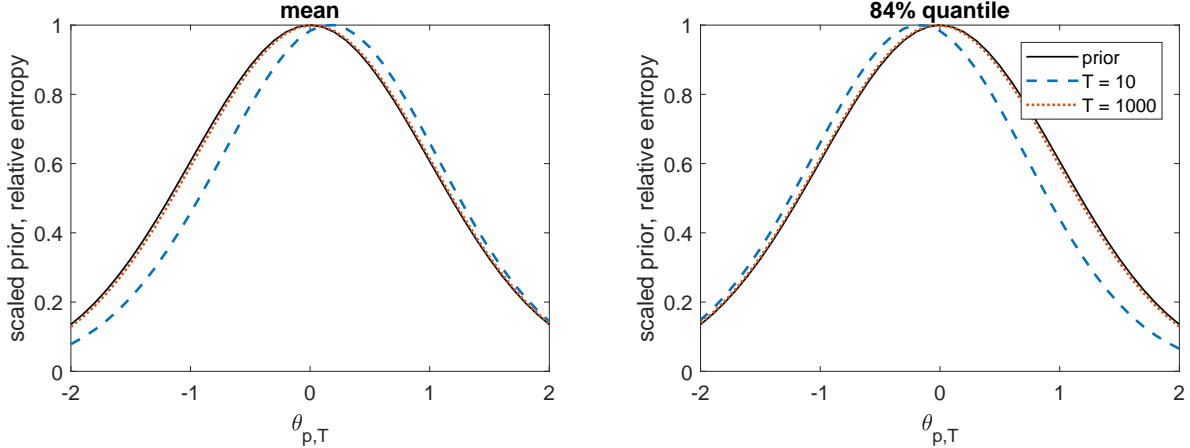
17

Figure 4.2: Relative entropy for given posterior mean with $\theta \sim \mathcal{N}(0,1)$, $\theta_{p,T} \in [-2,2]$. **Left:** increase in posterior mean of θ_1 by one posterior standard deviation; **Right:** increase in 84% quantile of θ_1 by one posterior standard deviation.

T and d. Firstly, notice that Lemma 1 gives an accurate approximation for the behavior of relative entropy as T increases for $T \geq 10$, with a gradient of $-\frac{d}{2}$ when we plot $\log \mathcal{R}$ against $\log T$. When T is small, the data has not yet swamped the prior. The resulting greater sensitivity to the prior is reflected in the relative entropy for $T < 10$ being small compared to when $T \geq 10$, relative to what is predicted by Lemma 1.

Sample mean. To show the dependence of $\mathcal{R}$ on $\pi(\theta_0)$, we now fix $d = 1$ and vary $\overline{X}_T$ such that $\theta_{p,T} \in [-2,2]$. We do this for $T \in \{10, 1000\}$. Figure 4.2 compares the relative entropy for different values of $\theta_{p,T}$ to the prior $\pi(\theta_{p,T})$ at that point, normalizing the values so all plots have a maximum of one. Even with $T = 10$, the relative entropy is almost proportional to the prior $\pi(\theta_{p,T})$ at the posterior mean. With $T = 1000$, the scaled prior and relative entropy are visually indistinguishable. We consider the prior at $\theta_{p,T}$ instead of $\overline{X}_T$ for two reasons. Firstly, locating the maximum likelihood may be computationally involved in settings other than the Gaussian location model, while evaluating the posterior mean is trivial given Monte Carlo draws. Secondly, since $\theta_{p,T} \to \theta_0$, using the posterior mean is asympotically equivalent to using $\overline{X}_T$.

4.2 Rule of thumb for $\mathcal{R}$

More generally, we calibrate $\mathcal{R}$ by taking a Gaussian approximation of the prior and posterior, then picking the relative entropy to match the corresponding Gaussian location model. In particular, for
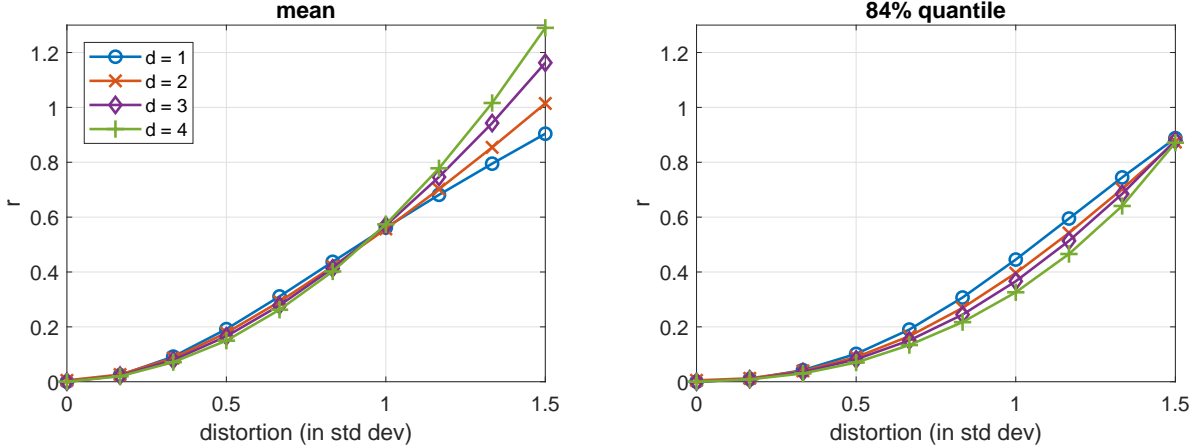
Figure 4.3: Scaled relative entropy r and distortion in Gaussian location model with $\theta \sim \mathcal{N}(0, I)$, $\overline{X}_T = 0$, $T = 100$, $d \in \{1, ..., 4\}$. All distortions are scaled by posterior standard deviation. **Left:** increase in posterior mean of θ_1 by one posterior standard deviation; **Right:** increase in 84% quantile of θ_1 by one posterior standard deviation.

prior and posterior means (μ_π, μ_p) and variances (Σ_π, Σ_p), consider the approximation $\mathcal{N}(\mu_\pi, \Sigma_\pi)$ and $\mathcal{N}(\mu_p, \Sigma_p)$ for the prior and posterior respectively. Define the dispersion of the likelihood:

$$\Sigma_\ell = \left(\Sigma_p^{-1} - \Sigma_\pi^{-1}\right)^{-1} \tag{4.3}$$

In a Gaussian location model with T observations of $X \sim \mathcal{N}(\theta, \Omega)$, we have $\Sigma_\ell = \frac{1}{T}\Omega$.

We parameterize the relative entropy as:

$$\mathcal{R} = 1.6^\delta r \frac{\pi(\mu_p)}{\pi(\mu_\pi)} \sqrt{\frac{|\Sigma_\ell|}{|\Sigma_\pi|}} \quad \text{where} \quad \delta = \begin{cases} 0 & d = 1 \\ d & d > 1 \end{cases} \tag{4.4}$$

The last two terms arise from asymptotics in Lemma 1, and the 1.6^δ scaling accounts for further differences across dimensions. Figure 4.3 shows the values of r for different changes in the mean of 84% quantile in the Gaussian location model with $\Sigma_\pi = \Omega = I$, and $T = 100$. As a rule of thumb, $r < 0.05$ is small and $r > 0.50$ is large. Figure 4.3 shows that for both the mean and 84% quantile of the Gaussian model, $r = 0.05$ corresponds to an approximately one quarter standard deviation change, while $r = 0.50$ corresponds to a roughly one standard deviation change.

One can also obtain a Gaussian approximations using the Hessian of the prior and posterior at their modes. However, the local nature of this approximation is less consistent with the intuition

in Lemma 1 that we should scale relative entropy according to the dispersion of the likelihood. For example, if a model is set identified, the likelihood may concentrate around a given region of the parameter space. However, if the likelihood is flat in the idenfied set, then the Hessian of the posterior and prior at the modes may be close, leading to $|\Sigma_\ell|$ to be large even if the identified set is small.

5   Implementation

We now discuss the numerical implementation of the calculations in Section 2. Throughout, we will assume that we have a set of Monte Carlo draws from the prior and posterior respectively.

5.1   Importance sampling

If the distribution of the worst-case distortion M does not have fat tails, then we can solve for M and evaluate the worst-case prior and posterior using importance sampling. In particular, for any λ, we can evaluate the right-hand side of (2.12), approximating the expectations with Monte Carlo sample averages:

$$\mathbb{E}_p\left[\frac{\exp\left[\lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right)-\tilde{\gamma}\right)\right]}{\mathbb{E}_p\left[\exp\left[\lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right)-\tilde{\gamma}\right)\right]\right]}\gamma\left(\theta\right)\right] \approx \frac{\sum_{j=1}^{J}\exp\left[\lambda L\left(\theta_{p,j}|X\right)\left(\gamma\left(\theta_{p,j}\right)-\tilde{\gamma}\right)\right]\gamma\left(\theta_{p,j}\right)}{\sum_{j=1}^{J}\exp\left[\lambda L\left(\theta_{p,j}|X\right)\left(\gamma\left(\theta_{p,j}\right)-\tilde{\gamma}\right)\right]}. \quad (5.1)$$

We can then solve (2.12) using this Monte Carlo approximation. Given the solution for λ, we can now compute M for any value of θ. Reweighting the original Monte Carlo draws by M then gives us the worst-case prior and posterior.

In addition, Lemma 4 in Appendix A shows that we can evaluate relative entropy using the expression:

$$\mathbb{E}_\pi\left[M\left(\theta\right)\log M\left(\theta\right)\right] = -\log\mathbb{E}_\pi\left[\exp\left[\lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right)-\tilde{\gamma}\right)\right]\right] \quad (5.2)$$

The right hand side is negative log of the normalizing constant that ensures $M\pi$ integrates to one. It is therefore straightforward to evaluate the right-hand side of (5.2) using draws from the prior. This produces more precise estimates than averaging $\log M$ across the draws from the worst-case prior because such a calculation would require the normalizing constant as well.

Importance sampling performs poorly when the distribution of M has fat tails, which occurs

when the likelihood is sharply peaked. This problem becomes more severe as we increase the dimensionality of Θ or the number of observations. The solution (2.11) shows that when the likelihood is sharply peaked, the distortions are concentrated in a small region of the parameter space, but the distortions in that region are large. As a result, in order to accurately approximate the prior and posterior distortions, we need an increasingly large number of Monte Carlo draws for the high likelihood region to be sufficiently well sampled.

5.2 Sequential Monte Carlo

When importance sampling fails, we can use sequential Monte Carlo (SMC) to generate draws from the worst-case prior and posterior. Rather that using importance sampling to move directly from the original to the worst-case distributions, SMC introduces a sequence of *bridge distributions* that serve as intermediate steps between the original and worst-case distributions. Beginning with draws from the original prior and posterior, referred to as *particles*, we iteratively construct particle approximations of the bridge distributions, before arriving at a particle approximation of the worst-case prior and posterior.

5.2.1 Algorithm

Our algorithm is based on Herbst and Schorfheide (2014), who begin with draws from a prior and transition to the posterior of a DSGE model by constructing bridge distributions that are proportional to the product of the prior and the likelihood raised to some exponent. Adapting their algorithm requires constructing bridge distributions for our setting, and computing the corresponding weights between consecutive bridge distributions.

We take the bridge distributions to be the worst-case priors $\{\pi_i\}_{i=1}^{N_{\mathrm{SMC}}}$ and posteriors $\{p_i\}_{i=1}^{N_{\mathrm{SMC}}}$ arising from the solution of (2.9)-(2.10) for a sequence of intermediate worst-case posterior means $\tilde{\gamma}_0 > ... > \tilde{\gamma}_{N_{\mathrm{SMC}}}$, with $\tilde{\gamma}_0 = \mathbb{E}_p\left[\gamma\left(\theta\right)\right]$ and $\tilde{\gamma}_{N_{\mathrm{SMC}}} = \tilde{\gamma}$. When studying quantiles, we can fix the quantile of interest q and construct a sequence of intermediate worst-case quantiles $\tilde{\psi}_0 > ... > \tilde{\psi}_{N_{\mathrm{SMC}}}$, where $\tilde{\psi}_0$ is the quantile under the original posterior, and $\tilde{\psi}_{N_{\mathrm{SMC}}} = \tilde{\psi}$ is the worst-case quantile. With the bridge distributions in hand, we sketch out the SMC procedure below in Algorithm 1. We leave the details to Appendix B.

Algorithm 1: Sequential Monte Carlo for REPS

    Input: Draws $\{\theta_{\pi,j}\}_{j=1}^{N_P}$ and $\{\theta_{p,j}\}_{j=1}^{N_P}$ from original prior and posterior.

    Output: Draws $\{\theta_{\tilde{\pi},j}\}_{j=1}^{N_P}$ and $\{\theta_{\tilde{p},j}\}_{j=1}^{N_P}$ from worst-case prior and posterior.

    Initialize: Set $\{\theta_{\pi_0,j}\}_{j=1}^{N_P} = \{\theta_{\pi,j}\}_{j=1}^{N_P}$ and $\{\theta_{p_0,j}\}_{j=1}^{N_P} = \{\theta_{p,j}\}_{j=1}^{N_P}$.

    for $i = 1$ **to** N_{SMC} **do**

        compute weights: Solve for λ_i and compute $m_i \equiv \pi_i / \pi_{i-1}$ for each draw.

        selection: Draw from $\{\theta_{\pi_{i-1},j}\}_{j=1}^{N_P}$ and $\{\theta_{p_{i-1},j}\}_{j=1}^{N_P}$ using a multinomial distribution

        with probability weights proportional to $m_i(\theta_{\pi_{i-1},j})$ and $m_i(\theta_{p_{i-1},j})$ respectively.

        mutation: For each draw, take N_{MH} Metropolis-Hastings steps.

    end

    return $\{\theta_{\tilde{\pi},j}\}_{j=1}^{N_P} = \{\theta_{\pi_{N_{\mathrm{SMC}}},j}\}_{j=1}^{N_P}$ and $\{\theta_{\tilde{p},j}\}_{j=1}^{N_P} = \{\theta_{p_{N_{\mathrm{SMC}}},j}\}_{j=1}^{N_P}$.

5.2.2 Computational constraints

In general, the main computational bottleneck in Algorithm 1 is the mutation step. In particular, let the number of particles and Metropolis-Hastings mutation steps be N_P and N_{MH} respectively. Obtaining particle approximations of the worst-case prior and posterior each require us to compute L and γ for $N_P \times N_{\mathrm{MH}} \times N_{\mathrm{SMC}}$ different parameter values. Both of these may be computationally expensive to compute. For example, if γ is an impulse response function in a DSGE model, then one needs to solve the model for each draw of θ in order to compute the impulse response. To compute the likelihood L, one needs to run a Kalman filter using the solved model for the state equation. However, since the particles are independent of each other, the mutation step can be parallelized, as in Herbst and Schorfheide (2014). In addition, if we only wish to solve for λ, we can execute Algorithm 1 for the posterior only, halving the number of likelihood evaluations.

    Algorithm 1 raises two computational challenges. Firstly, if we are interested in more than one statistic, we need to repeat the SMC algorithm for each objective function we are interested in. For instance, if one were interested in the error bands for a set of impulse response functions in a DSGE model, one would need to repeat Algorithm 1 for each impulse response function at multiple horizons. For large-scale DSGE models, this can be computationally demanding. Secondly, to apply Algorithm 1 to problem (2.15)-(2.16), we require the marginal likelihood L^* and conditional expectation γ^*, both of which may be difficult to compute.

Algorithm 2: Approximate REPS

    Input: Draws $\{\theta_{\pi,j}\}_{j=1}^{N_{\mathrm{P}}}$ and $\{\theta_{p,j}\}_{j=1}^{N_{\mathrm{P}}}$ from original prior and posterior.

    Output: Draws $\{\theta_{\tilde{\pi},j}\}_{j=1}^{N_{\mathrm{P}}}$ and $\{\theta_{\tilde{p},j}\}_{j=1}^{N_{\mathrm{P}}}$ from approximate worst-case prior and posterior.

    1. Approximate π and p by $\hat{\pi}$ and $\hat{p}$.
    2. Use $\hat{\pi}$ and $\hat{p}$ to obtain an approximation $\widehat{L}(\theta^*|X) \approx cL^*(\theta^*|X)$, where c is a constant.
    3. Obtain an estimate for $\hat{\gamma}(\theta) \approx \mathbb{E}_p[\gamma(\theta)|\theta^*]$.
    4. Do Algorithm 1, replacing (π, p, L^*, γ) with $(\hat{\pi}, \hat{p}, \widehat{L}, \hat{\gamma})$.

    return $\{\theta_{\tilde{\pi},j}\}_{j=1}^{N_P} = \{\theta_{\pi_{N_{\mathrm{SMC}}},j}\}_{j=1}^{N_P}$ and $\{\theta_{\tilde{p},j}\}_{j=1}^{N_P} = \{\theta_{p_{N_{\mathrm{SMC}}},j}\}_{j=1}^{N_P}$.

5.3 Approximate REPS

To overcome the limitations of Algorithm 1, we consider an approximation to the REPS calculations, which we shall refer to as *approximate relative entropy prior sensitivity* (AREPS).

5.3.1 Algorithm

The main idea of AREPS is to replace π, p, L^* and γ^* with approximations $\hat{\pi}$, $\hat{p}$, $\widehat{L}$ and $\hat{\gamma}$. In particular, we fit a set of basis functions to the Monte Carlo draws that we obtained from the original estimation as observations, to obtain approximations of $\hat{\pi}$, $\hat{p}$ and $\hat{\gamma}$ respectively for π, p and γ.[1] If we are doing the REPS analysis over the entire Θ, then $\widehat{L}$ and $\hat{\gamma}$ correspond to approximations for L and γ respectively. Analogously to semiparametric methods, we use basis functions to allow for flexibility while mitigating the curse of dimensionality that arises in fully nonparametric estimation.

    With the appropriate approximations, the approximate likelihood $\widehat{L}$ for a set of particles can be computed in vectorized form. As a result, each SMC run no longer requires parallelization, and one can parallelize across SMC algorithms rather than over particles within the SMC algorithm. In addition, since we no longer need to compute the true L^* or γ^*, output from packages such as Dynare can be directly fed into the algorithm. The steps are summarized in Algorithm 2. I present the algorithm as if we are interested in REPS for the subspace Θ^*.

---

[1]The motivation for the approximations is similar in spirit to variational Bayesian inference. While variational Bayesian methods seek the approximating distribution that is closest to the true posterior in relative entropy, here we make use of the fact that we have existing Monte Carlo draws that we can use to directly approximate the distributions and functions.

5.3.2 Computational details

For computational efficiency, the approximations $\hat{\pi}$, $\hat{p}$ and $\hat{\gamma}$ should be fast to evaluate. For example, in our application in Section 6, we use a Gaussian mixture model to approximate p, and a quadratic logit to approximate γ^*.

For numerical accuracy, L^* and γ^* need to be well approximated in regions with the largest distortions. Since the approximations typically perform more poorly in the tails of the distributions, AREPS would provide misleading results if the distortions M take on extreme values in the tails of π^*, p^* or γ^*. This problem tends to be less severe when γ^* is bounded. For example, AREPS would generally produce more accurate results when we study quantiles, since $\gamma^* \in [0, 1]$ and M takes on extreme values in the high likelihood regions, which would tend to be near the posterior mode.

If closed form expressions for L^* and γ^* are available, we can transform the draws from the approximate worst-case prior and posterior to obtain draws from the true worst-case prior and posterior after executing Algorithm 2. See Appendix B for details.

6 Application: Smets and Wouters (2007)

For our main application, we consider the workhorse DSGE model from Smets and Wouters (2007). We show that even in such a large model, REPS is not only feasible but also accounts for features of the likelihood that are especially hard to diagnose in such high-dimensional settings.

6.1 Model and estimation

Model. Smets and Wouters (2007) presents a large-scale New Keynesian model with sticky wages and prices, wage and price indexation, habit formation, investment adjustment costs, variable capital utilization, and fixed costs in production. The model includes total factor productivity, risk premium, investment-specific, wage mark-up, price mark-up, government spending, and monetary policy shocks. This gives a total of thirty-six parameters.

Original estimation. We use quarterly data from 1984Q1 to 2007Q4 for seven series from the Federal Reserve Economic Data (FRED): GDP growth, consumption growth, investment growth,

24

wage growth, hours, inflation, and the federal funds rate. The series are updated vintages of those used in Smets and Wouters (2007), for the period after the start of the Great Moderation. For our benchmark estimation, we take our prior from Smets and Wouters (2007), and make 1.5 million Markov Chain Monte Carlo draws after discarding 40,000 burn-in draws from the posterior using a standard Metropolis-Hastings algorithm[2].

6.2   Prior sensitivity

Object of interest.   We construct robust 68% error bands for the impulse response of output to a one percentage point decrease in interest rates up to five years from impact. In particular, we solve the REPS problem separately for each horizon and posterior quantile determining the error band. This provides conservative error bands that uniformly contains the set of all error bands arising from the chosen set of priors.

Parameters of interest.   We consider the sensitivity of the error bands to changes in the prior of two groups of structural parameters. The first set of parameters is $\{\rho, r_\pi, r_y, r_{\Delta y}\}$ from the monetary policy rule:

$$\hat{r}_t = \rho \hat{r}_{t-1} + (1 - \rho) \left[r_\pi \hat{\pi}_t + r_y \left(\hat{y}_t - \hat{y}_t^* \right) + r_{\Delta y} \left[(\hat{y}_t - \hat{y}_t^*) - (\hat{y}_{t-1} - \hat{y}_{t-1}^*) \right] \right] + \varepsilon_t^r, \qquad (6.1)$$

where $\hat{r}_t$ is the interest rate, $\hat{\pi}_t$ is the inflation rate, $\hat{y}_t - \hat{y}_t^*$ is the output gap, and ε_t^r is an exogenous AR(1) shock process. The second set of parameters is $\{\xi_w, \xi_p, \iota_w, \iota_p\}$, which determine the level of wage rigidity, price rigidity, wage indexation, and price indexation respectively. Each of these parameters have range $[0, 1]$, with 0 corresponding to the flexible wage and price benchmarks. We compare the resulting robust error bands and analyze the worst-case priors for each set of parameters.

Economic theory suggests that both may be important for determining the response of output to monetary policy. The Taylor rule indicates how persistent the monetary policy shock will be, and how the monetary authority will respond to dampen future deviations in inflation and the

---

[2]We estimate the equations as presented in the text of Smets and Wouters (2007). In their estimation, Smets and Wouters (2007) use an alternative scaling for their risk premium shock. The scaling does not change the estimation results materially.
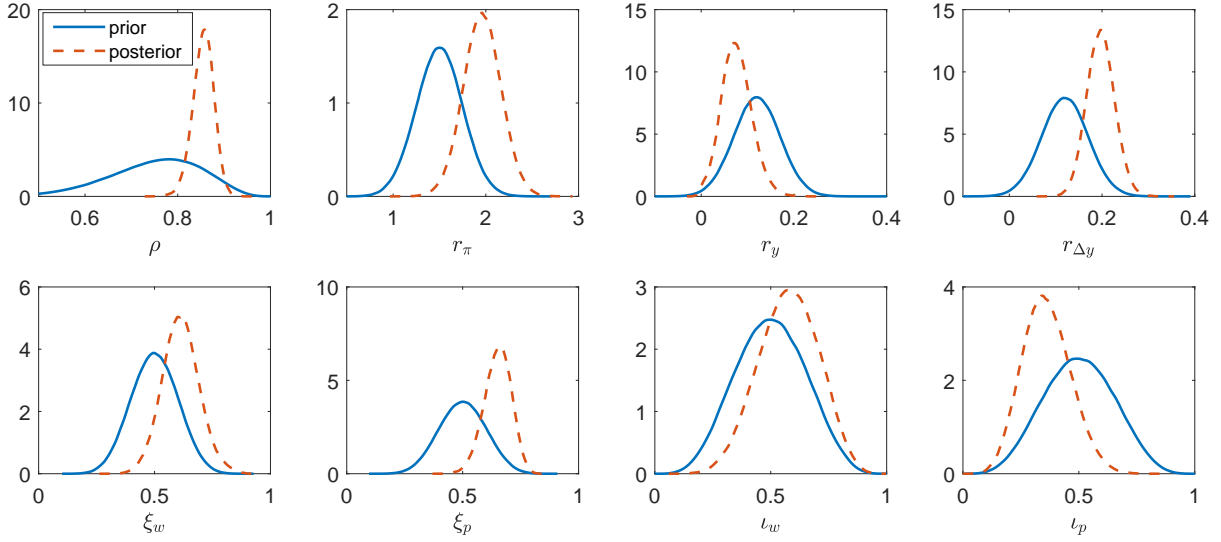
Figure 6.1: Original prior and posterior of Taylor rule parameters $\{\rho, r_\pi, r_y, r_{\Delta y}\}$ and nominal frictions parameters $\{\xi_w, \xi_p, \iota_w, \iota_p\}$.

output gap that arise from the initial monetary policy shock. In addition, households have rational expectations about this future path of interest rates, and make decisions that determine the response of output in equilibrium. Similarly, the nominal frictions determine how much prices can adjust in response to monetary policy, and thus how much output has to move in equilibrium. Nevertheless, because the impulse response function is determined in equilibrium, it is hard to make precise analytical statements about the effect of changing these parameters on the impulse response.

Figure 6.1 suggests that most of these parameters are not sharply identified, as their marginal posteriors are not much more concentrated than their priors. However, since these parameters are jointly distributed with each other and with the remaining 28 parameters, it is not clear ex ante how changing the marginal prior of these subsets of parameters will affect the impulse response. Not only do we need to consider the joint distribution within each subset, but we also need to account for how the conditional likelihood of the remaining parameters depends on the subset of parameters whose prior we distort.

Computational details. Since the prior is fast to evaluate, we use the exact prior. The posterior is approximated using a Gaussian mixture with 40 components. The impulse response at each horizon is approximated using a quadratic regression, yielding R^2s of between 0.97 to 0.98. Finally, the conditional probability that the impulse response is less than a given cutoff is estimated with
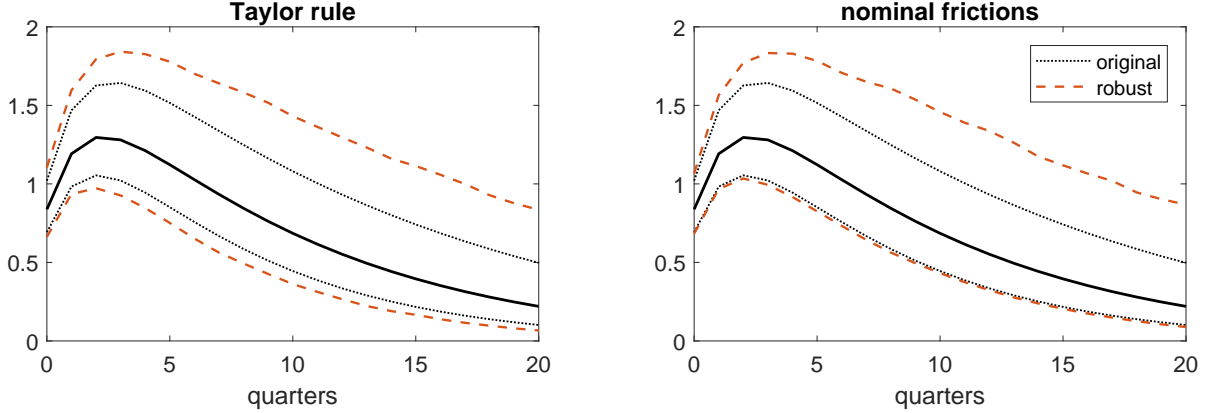
26

Figure 6.2: Robust 68% error bands. Black lines show original median (solid) and error bands (dotted); red dashed lines show robust error bands. **Left:** Taylor rule prior, with $r = 5 \times 10^{-3}$; **Right:** nominal frictions prior, with $r = 2.5 \times 10^{-4}$.

a quadratic logit. For the SMC, we use 5×10^4 particles, 250 SMC stages, and 10 Metropolis-Hastings draws at each stage. We average the results across ten runs of the SMC. Appendix D provides further details, including checks for the accuracy of the approximations.

6.3 Robust error bands

Figure 6.2 plots the original 68% error bands for the impulse response, together with the robust error bands. Appendix D provides details on the construction of the error bands.

Sensitivity of error bands. Figure 6.2 plots the resulting robust error bands. For each set of parameters, we fix the relative entropy so that the maximum distance between the original and robust error bands is approximately one posterior standard deviation. Using the rule of thumb in equation (4.4), this requires $r = 5 \times 10^{-3}$ for the Taylor rule and $r = 2.5 \times 10^{-4}$ for the nominal frictions prior. These values of r are one and two orders of magnitude smaller than the $r = 0.05$ benchmark respectively, which shows that the posterior error bands for the impulse response are very sensitive to the priors of the two sets of parameters, especially the nominal frictions prior. There are two reasons for the high degree of sensitivity.

Firstly, the likelihood for both the Taylor rule and the nominal frictions parameters is dispersed, allowing a small change in the prior to generate a large change in the posterior of the parameters. The statistic $\sqrt{|\Sigma_\ell| / |\Sigma_\pi|}$, where Σ_ℓ is defined in (4.3), measures the dispersion of the likelihood

relative to the prior. This statistic is 0.23 for the Taylor rule and 1.53 for the nominal frictions parameters. In the Gaussian location model, these values correspond to having a standard normal prior and observing just one observation of $X \sim \mathcal{N}\left(\theta, \omega^2 I\right)$ with $\omega = 0.69$ and $\omega = 1.11$ respectively. With such a dispersed likelihood, the asymptotics in Lemma 1 overpredicts the relative entropy needed to change the posterior estimates. In particular, in Figure 4.1, the relative entropy with $T = 1$ is an order of magnitude smaller than what is predicted by the log-log trend for $T > 10$. In addition, the likelihood for the nominal frictions parameters is less informative, resulting in a higher degree of sensitivity to the prior.

Secondly, both the Taylor rule and nominal frictions parameters are good predictors of the impulse response function. For example, using the Monte Carlo posterior draws as observations and running a quadratic regression of the impulse response 20 quarters after impact on the Taylor rule parameters yields an R^2 of 0.82. Running the regression on the nominal frictions parameters instead yields an R^2 of 0.48. As a result, a given change in the posterior for the Taylor rule or nominal frictions parameters shifts the posterior for the impulse response substantially. Even though the nominal frictions parameters predict the impulse response less well, their prior is more important because of the dispersion of their marginal likelihood.

Asymmetry.   However, the sensitivity to the prior is neither uniform across horizons nor between bounds. The robust error bands diverge from the original error bands as we increase the horizon, indicating that the impulse response depends more on the prior at longer horizons. Intuitively, since low frequency fluctuations are estimated less precisely than high frequency fluctuations, the marginal likelihood for the impulse response is more concentrated around short horizons. Moreover, the impulse response at short horizons are less well-predicted by the parameters of interest. For example, running a quadratic regression of the impulse response on impact on the Taylor rule parameters yields an R^2 of 0.27, and running the regression on the nominal frictions parameters instead yields an R^2 of 0.05.

In addition, the upper bound is more sensitive to changes in the prior than the lower bound for both sets of parameters. This indicates that given the model and data, a policymaker should be more concerned about underestimating than overestimating the expansionary effects of a surprise decrease in interest rates. Like the log-normal example shown in Figure 3.2, the original posterior

of the impulse response is skewed to the right, which partially explains the greater sensitivity of the upper bound of the error bands. However, the shape of the posterior for the impulse response does not fully capture the asymmetry in sensitivity. In particular, the asymmetry is more pronounced for the nominal frictions parameters, reflecting differences in the mapping from the two sets of parameters to the impulse response.

As a simple example, consider a Gaussian location model with two independent components, and suppose we are interested in the sensitivity of the mean of $\psi(\theta) = \theta_1 + \exp(\theta_2)$ to the prior. The log-normal component from $\exp(\theta_2)$ will introduce skew into the posterior for $\psi(\theta)$, as in the example in Figure 3.2. However, the effect of θ_1 on the mean of $\psi(\theta)$ is symmetric. As a result, if we only distort the marginal prior of θ_1, we would have symmetric bounds on the posterior mean of $\psi(\theta)$. The asymmetry in sensitivity will only arise if we distort the prior of θ_2 as well. It is much more complicated to analytically establish such properties for the impulse response function here. Nevertheless, the robust error bands in Figure 6.2 account for how the impulse response depends on the Taylor rule and nominal frictions parameters.

The wide robust error bands add to the literature documenting the sensitivity of Bayesian DSGE estimates to the prior (e.g. Canova and Sala (2009)). Even though a model may be weakly identified, the direction that lacks identification may not be important for the statistics one is most interested in. For the impulse response here, we see different degrees of sensitivity depending on the horizon, quantile, or subset of parameters considered.

6.4   Worst-case distributions

As an illustration, we now study the worst-case priors and posteriors for the impulse response one year after impact. For the Taylor rule prior, we consider the worst-case distortions that decrease the lower bound by $1/3$ standard deviation and increase the upper bound by one standard deviation respectively. For the nominal frictions prior, we consider the worst-case distortions that decrease the lower bound by $1/20$ standard deviation and increase the upper bound by one standard deviation respectively. These deviations correspond approximately to the robust error bands in Figure 6.2, with values of $r = 5 \times 10^{-3}$ for the Taylor rule and $r = 2.5 \times 10^{-4}$ for the nominal frictions prior respectively. We show that REPS is jointly accounting for the effect the parameters on the impulse response and the joint likelihood. Moreover, we show that the worst-case distortions

Parameter		Lower bound		Upper bound	
		$\frac{\mathbb{E}_{\tilde{\pi}}[\cdot]-\mathbb{E}_{\pi}[\cdot]}{\sigma_{\pi}[\cdot]}$	$\frac{\mathbb{E}_{\tilde{p}}[\cdot]-\mathbb{E}_{p}[\cdot]}{\sigma_{p}[\cdot]}$	$\frac{\mathbb{E}_{\tilde{\pi}}[\cdot]-\mathbb{E}_{\pi}[\cdot]}{\sigma_{\pi}[\cdot]}$	$\frac{\mathbb{E}_{\tilde{p}}[\cdot]-\mathbb{E}_{p}[\cdot]}{\sigma_{p}[\cdot]}$
Taylor rule					
ρ	persistence	0.007	−0.165	−0.003	0.125
r_{π}	inflation coef.	0.000	0.743	−0.004	−0.234
r_y	output gap coef.	0.005	0.192	−0.001	−0.282
$r_{\Delta y}$	output gap growth coef.	−0.002	−0.072	−0.005	0.204
Nominal frictions					
ξ_w	wage rigidity	−0.009	−0.162	0.006	0.517
ξ_p	price rigidity	−0.040	−0.076	−0.000	−0.041
ι_w	wage indexation	−0.001	−0.048	−0.002	0.037
ι_p	price indexation	0.004	0.108	0.008	−0.212

Table 6.1: Difference between worst-case and original prior and posterior means, normalized by standard deviations. Worst-case distributions correspond to impulse response four quarters from impact. **Taylor rule:** Lower bound decreased by 1/3 standard deviations and upper bound increased by one standard deviation under worst case; **Nominal frictions:** Lower bound decreased by 1/20 standard deviations and upper bound increased by one standard deviation under worst case.

capture important features of the likelihood that may be neglected with a more naive analysis.

We summarize the distortions by the changes in the prior and posterior means in Table 6.1. The prior means change less than the posterior mean because the distortions are nonparametric and applied jointly to the various parameters. For example, changes in the skew of a distribution can leave both first and second moments unchanged, but lead to substantial changes in the posterior results if the likelihood is high at one tail of the prior distribution. In addition, because the changes in the prior are concentrated in the high likelihood regions, they may appear small when integrated out, but still lead to relatively large changes in the posterior means. The distortions are asymmetric across the upper and lower bounds, emphasizing that the upper and lower bounds of the error bands depend on the priors in different ways.

In what follows, we study the worst-case distortions for each set of parameters in greater detail. The distortions reveal useful information about the model and likelihood that would be difficult to uncover by simple inspection of the posterior. Much of the analysis conditions on the size of the impulse response. In order to shift the lower (upper) bound of the error band, we need to shift

Parameter		Lower bound	Upper bound
Taylor rule			
ρ	persistence	0.088	0.232
r_π	inflation coef.	-0.057	-0.170
r_y	output gap coef.	-0.036	-0.233
$r_{\Delta y}$	output gap growth coef.	0.024	0.112
Nominal frictions			
ξ_w	wage rigidity	0.049	0.260
ξ_p	price rigidity	0.019	-0.060
ι_w	wage indexation	0.016	0.053
ι_p	price indexation	-0.000	0.015

Table 6.2:   Regression of impulse response four quarters from impact on parameters. **Lower bound:** Conditional on impulse response being at least one standard deviation below its mean; **Upper bound:** Conditional on impulse response being at least one standard deviation above its mean.

mass to the left (right) tail of the distribution of the impulse response, and therefore focus on draws for which the impulse response in one standard deviation below (above) the mean. Comparing the behavior of the left and right tails of the impulse response reveals the reasons for the asymmetry in distortions.

To understand the worst-case distortions, we use the fact that the distortion (2.17) depends on the parameters through the objective function $\gamma^* (\theta^*)$ and the marginal likelihood $L^* (\theta^*|X)$. For each set of parameters, we run regressions of the impulse response on the parameters using the Monte Carlo draws from the original posterior, in order to analyze the relationship between the parameters and the impulse response. Since $\gamma^* (\theta^*)$ in (2.17) is a conditional expectation, this regression captures both the direct effect of the parameters on the impulse response, as well as the indirect effect from the conditional distribution of the remaining parameters under the posterior. The impulse response and parameters are normalized to mean zero and standard deviation one, so the coefficients can be interpreted as the number of standard deviations the impulse response increases by in response to a one standard deviation increase in the corresponding parameter on average. The results are reported in Table 6.2. We then analyze the original and worst-case distributions in order to understand the shape of the likelihood.

6.4.1 Taylor rule distortions

Dependence of impulse response on parameters. The regression results for the Taylor rule parameters depend on the size of the impulse response. In particular, when the impulse response is small, the coefficients and R^2 are smaller than when the impulse response is large. The Taylor rule parameters thus have a small effect on average and explain less of the variation in left tail of the impulse response function, which leads to the lower bound being less sensitive to changes in the Taylor rule prior than the upper bound.

The coefficients indicate that the impulse response decreases in response to an increase in r_π or r_y, and increases in response to an increase in ρ or $r_{\Delta y}$. These results are consistent with both economic intuition and the signs of the distortions. The response to a monetary policy shock will be stronger if the Taylor rule is more persistent or less responsive to changes in the output gap and inflation. $r_{\Delta y}$ smooths variation in the output gap, and can move the impulse response in either direction, which is reflected in the small magnitudes of the coefficients on $r_{\Delta y}$.

Likelihood. The likelihood offers further insights on three features of the worst-case posterior. Firstly, the distortion for r_π is especially large for the lower bound relative to the magnitude of the regression coefficient. Secondly, the worst-case prior distorts ρ minimally even though the regression suggests that ρ has a relatively large effect on the impulse response. Finally, the relative distortions for r_π and $r_{\Delta y}$ are larger for the upper bound.

We begin by comparing the original and worst-case marginal posteriors for r_π, shown in Figure 6.3. The worst-case posterior for the lower bound is bimodal, with an additional peak around $r_\pi = 2.75$, revealing a high likelihood in the right tail of r_π. The observation corroborates the results of Herbst and Schorfheide (2014), who find that the posterior mean of r_π moves from 2.04 to 2.78 when one replaces the prior from Smets and Wouters (2007) with a more diffuse one. This mode in the likelihood arises from fitting the data to the Taylor rule. In particular, using data for the fed funds rate, inflation, and output gap from FRED and estimating the Taylor rule using a linear regression, we find a large inflation coefficient of 2.59[3]. This large value arises due to the low frequency variation in the data. As a heuristic exercise, we run the regression of the trend and cyclical components of HP filtered data, and find coefficients of 2.66 and 0.92 respectively. Sala

---

[3]The regression ignores the autocorrelation of 0.26 in the monetary policy shock.
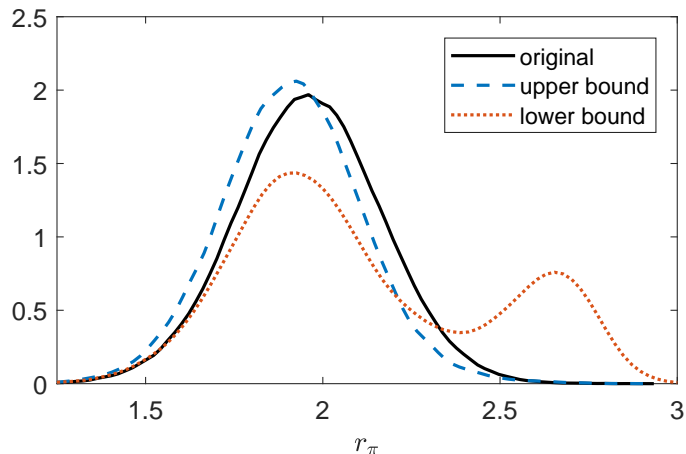
Figure 6.3: Original and worst-case posteriors of r_π. **Black solid line:** original posterior; **Blue dashed line:** worst-case posterior for upper bound; **Red dotted line:** worst-case posterior for lower bound.

(2015) formally estimates a similar DSGE model in the frequency domain, and finds a posterior estimate of 1.81 for the low frequency component, which is significantly larger than the estimate of 1.12 for the high frequency component[4]. The prior from Smets and Wouters (2007) shrinks more towards smaller values of r_π, making it difficult to detect the possibility of an additional mode without reestimating the model with a new prior. However, one may not know ex ante which priors might result in these changes in the posterior, or whether the changes are consequential to the impulse response. REPS flexibly accounts for such features in the likelihood if they are important for the posterior outcomes of interest.

The likelihood reveals two reasons for the small distortions in ρ. Firstly, Figure 6.1 shows that ρ is relatively sharply identified by the likelihood. In particular, the ratio of posterior to prior standard deviation for ρ is smaller than that of r_π, r_y and $r_{\Delta y}$, with a value of 0.23 as compared to 0.82, 0.66 and 0.60 respectively. As a result, larger changes in the prior would be needed to produce the same change in the posterior or ρ, which would result in a greater relative entropy cost. In addition, the correlations of ρ with r_π, r_y and $r_{\Delta y}$ are in conflict with their effect on the impulse response. In particular, ρ has positive correlations with r_π and r_y of 0.18 and 0.33 respectively, and a negative correlation of -0.10 with $r_{\Delta y}$. However, the regression results suggest that ρ should be distorted in the opposite direction from r_π and r_y, but in the same direction as $r_{\Delta y}$. Such a

---

[4]The posterior estimates are smaller than our estimate of $r_\pi = 2.04$ because restricting the estimation to a particular set of frequencies reduces the information available to form the likelihood, thus amplifying the effect of the prior.

distortion is costly in terms of relative entropy as it runs against the likelihood, which makes it optimal to distort ρ less. Since distorting ρ is costly, the worst-case prior focuses its distortions on the remaining parameters. More broadly, these results suggest that one may understate the dependence of the posterior results on the prior if one changes the marginal priors purely based on their effects on the object of interest.

The likelihood also supports the large distortions in r_y and $r_{\Delta y}$ for the upper bound relative to the lower bound. Figure 6.1 shows that the posterior for r_y is centered around small parameter values relative to the prior, while the posterior for $r_{\Delta y}$ is centered around large parameter values relative to the prior. Therefore, decreasing r_y and increasing $r_{\Delta y}$ on average imply distortions around higher likelihood regions than increasing r_y and decreasing $r_{\Delta y}$. Since the impulse response is on average decreasing in r_y and increasing in $r_{\Delta y}$, it is optimal to distort r_y and $r_{\Delta y}$ more when increasing the upper bound of the error bands. The joint distortion is reinforced by the negative correlation of -0.26 between r_y and $r_{\Delta y}$, which is consistent with the two parameters being distorted in opposite directions. The asymmetry further emphasizes the need to do the REPS computations separately for each bound. Even though both worst-case priors correspond to the same impulse response, the optimal distortions for the upper and lower bound can be very different due to asymmetry in the likelihood and the mapping from parameters to impulse response.

6.4.2 Nominal frictions distortions

We now analyze the worst-case distortions for the nominal frictions prior to understand several features of the worst-case posterior means in Table 6.1. Firstly, the wage rigidity parameter ξ_w is distorted relatively more, especially for the upper bound. Next, the posterior means for price rigidity ξ_p and wage indexation ι_w move in contradictory directions when we go from the lower bound to the upper bound. Finally, the worst-case posterior mean for price indexation ι_p increases for the lower bound and decreases for the upper bound, contradicting the standard intuition that reducing nominal frictions should dampen the impulse response.

Dependence of impulse response on parameters. The largest coefficient from the regression reported in Table 6.2 is the one on wage rigidity ξ_w, which partly explains why the large distortion for ξ_w. Moreover, the regression coefficient on ξ_p is negative in the regression for the upper bound,
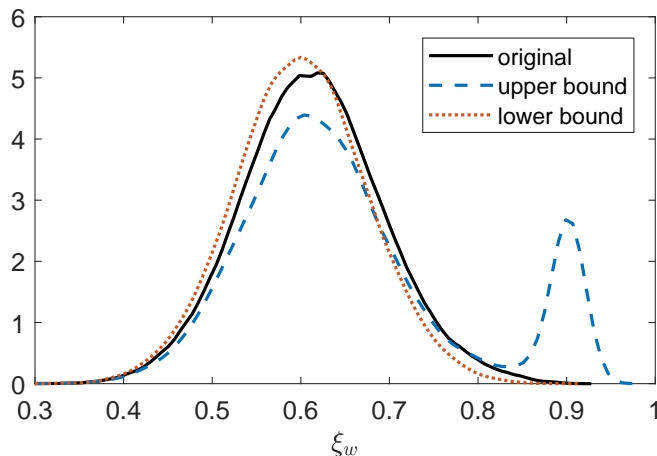
Figure 6.4: Original and worst-case posteriors for ξ_w. **Black solid line:** original posterior; **Blue dashed line:** worst-case posterior for upper bound; **Red dotted line:** worst-case posterior for lower bound.

rationalizing the counterintuitive direction of the distortion of the prior on ξ_p. The negative coefficient arises because of an omitted variable bias — ξ_p is correlated with other parameters that also affect the impulse response, biasing the regression coefficient relative to what we would have found if we controlled for all the parameters in the model. The coefficient obtained without controlling for the remaining parameters is the relevant one here because we are leaving the prior of all other parameters unchanged. REPS accounts for the fact that changing the prior for ξ_p changes the posterior of the impulse response through both the marginal effect of ξ_p and the effect of any other correlated parameters.

On the other hand, the coefficient of wage indexation ι_w for the upper bound regression contradicts the shift of the posterior distribution towards smaller values. Moreover, the small regression coefficients on price indexation ι_p are inconsistent with both the direction and magnitude of the worst-case distortions.

Likelihood. The worst-case posteriors for ξ_w, shown in Figure 6.4, provide an explanation for the especially large change in the posterior mean of ξ_w for the upper bound. In particular, the worst-case posterior for the upper bound is bimodal, with a new mode around $\xi_w = 0.90$. As with the lower bound worst-case posterior of r_π, this is in line with the diffuse prior estimates of Herbst and Schorfheide (2014). In particular, the posterior mean of ξ_w shifts from 0.70 under the Smets and Wouters (2007) prior to 0.93 under the diffuse prior. This is a larger change than that of ξ_p,

	Lower bound					Upper bound			
	ξ_w	ξ_p	ι_w	ι_p		ξ_w	ξ_p	ι_w	ι_p
ξ_w	1				ξ_w	1			
ξ_p	0.19	1			ξ_p	−0.20	1		
ι_w	−0.13	−0.23	1		ι_w	−0.05	−0.31	1	
ι_p	−0.07	−0.23	−0.15	1	ι_p	−0.08	−0.15	−0.14	1

Table 6.3: Posterior correlation of nominal frictions parameters. **Left:** Conditional on impulse response being at least one standard deviation below its mean; **Right:** Conditional on impulse response being at least one standard deviation above its mean.

ι_w, and ι_p, whose posterior means move from 0.66, 0.59, and 0.22 to 0.72, 0.73 and 0.11 respectively under the diffuse prior. Again, we find that REPS accounts for peaks in the likelihood that are hard to detect without reestimating the model under the appropriate prior. In addition, the asymmetry between the upper and lower robust error bands suggests that this new mode in the posterior for ξ_w is more consequential for the impulse response than the new mode in the lower bound worst-case posterior of r_π. Indeed, the regression coefficient for ξ_w in Table 6.2 is larger in magnitude than that for r_π.

The posterior correlations, reported in Table 6.3, help to account for the counterintuitive distortions in price indexation ι_p. The negative correlation of ι_p with ξ_w, ξ_p and ι_w imply that increases in these parameters correspond on average to a decrease in ι_p. Hence the likelihood favors moving ι_p in the opposite direction from the other three parameters. In addition, Figure 6.5 shows that under the worst-case posterior for the upper bound, the new mode for ξ_w corresponds to small values of ι_p, decreasing the posterior mean of ι_p. More generally, these distortions indicate that if we check the robustness of the posterior by changing the prior of a set of parameters in the direction suggested by economic intuition, our resulting estimates may understate the sensitivity of the posterior because we have failed to account for correlations across these parameters.

The posterior correlations for ξ_p further emphasize this point and provide an additional explanation for the inconsistent direction of distortions for the two worst-case posteriors of ξ_p. The rigidity parameters ξ_w and ξ_p have a positive correlation of 0.19 conditional on the impulse response being small, and a negative correlation of -0.20 conditional on the impulse response being large. This drives ξ_p in the same direction as ξ_w for the lower bound, and in the opposite direction from ξ_w for the upper bound. Given the distortions for ξ_w, it is therefore optimal to decrease ξ_p for both the upper and lower bound. In addition, ξ_p is negatively correlated with ι_w for both the lower
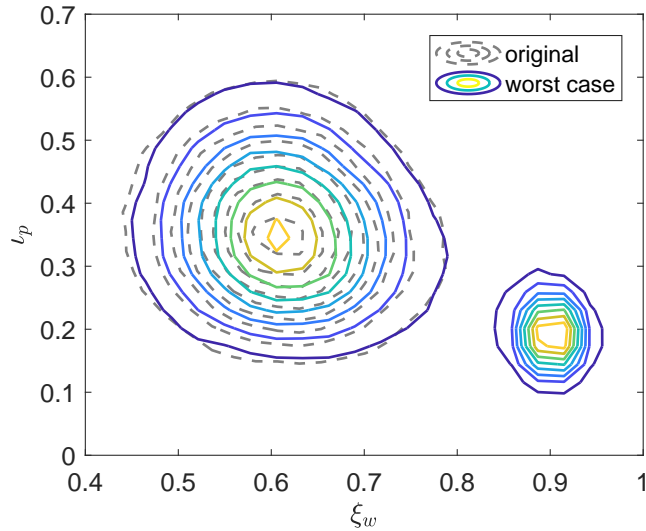
Figure 6.5: Original and lower bound worst-case posteriors for (ξ_w, ι_p). **Gray dashed lines:** original posterior; **Colored solid lines:** worst-case posterior.

and upper bounds. This implies that distorting their priors in the same direction would concentrate distortions in low likelihood regions of the parameter space, which explains why the posterior distortions are small and, for the upper bound, the prior distortions for ξ_p and ι_w are in opposite directions.

6.5 Beyond Smets and Wouters (2007)

We now summarize the general insights from the application above, and discuss several broader issues for REPS.

The REPS framework jointly accounts for the shape of the likelihood and the correlation of the parameters with the object of interest, which we sought to disentangle by studying the distributions and the regression results. These are useful but heuristic tools. The regression analysis helps uncover the average relationship between the parameters and the impulse response, but does not account for nonlinearities. By conditioning on the level of the impulse response, we showed that these nonlinearities are nontrivial and can affect the worst-case distributions. We also simplified the analysis by studying the joint posteriors of the parameters pairwise at most. However, the entire joint distribution of both the subvectors of interest and the undistorted parameters matters for the dependence of the posterior on the prior. REPS accounts for these comovements when providing bounds on the posterior outcomes, even in cases where it is hard to uncover these comovements in

high-dimensional settings.

Nevertheless, even these blunt tools reveal that one may incorrectly infer the dependence of a posterior result on the prior without formal prior sensitivity analysis. Firstly, economic intuition alone does not fully determine how a change in a given parameter changes a function of the full vector of parameters. When fitted to the data, the model induces dependence across parameters in the posterior. The effect of changing any given parameter depends on this joint distribution. Secondly, posterior inference can be affected by the behavior of the likelihood in the tail of both the prior and posterior. As shown in the first example of Section 3, both visual inspection of the prior and posterior and local prior sensitivity analysis may not detect this. Thirdly, the asymmetry in the robust error bands show that prior sensitivity depends on both the function of interest ψ and the property of this function's posterior distribution represented by γ. Even if we were to use REPS to place bounds on the mean, the bounds may not be symmetric around the original mean. Capturing these nonlinearities requires either a global approach to prior sensitivity analysis or higher order local prior sensitivity analysis. All these issues reiterate the importance of conditioning on the data in the REPS problem (2.7)-(2.8) to account for these properties in the likelihood.

Finally, while we have focused on the uniquely determined worst-case distributions as a diagnostic, one may find an alternative prior in the same set that implies change in the posterior statistic of interest that is only marginally smaller. Thus the worst-case distributions provide sufficient but not necessary distortions to generate large changes in the posterior results.

7 Conclusion

Prior sensitivity analysis should be a part of any rigorous Bayesian analysis. However, such analysis is often either absent or adhoc. Reestimating the model for many alternative priors can be computationally demanding, and it is infeasible to reestimate the model for the full set of possible priors. On the other hand, few prior sensitivity tools have been developed for economic applications.

Relative entropy prior sensitivity allows for global prior sensitivity analysis even in large models. The global nature of REPS accounts for features of the likelihood that may be neglected with local methods or simple inspection of the prior and posterior. The framework allows us to study the robustness of credible intervals to the prior, or to focus on subvectors of the parameter that may

be of special interest.

The New Keynesian model of Smets and Wouters (2007) provides a laboratory that shows how REPS can reveal which properties of the posterior are most sensitive to changes in the prior, and which parts of the prior are most important for these properties. Moreover, the worst-case distortions uncover features of the likelihood that are important for posterior inference yet hard to detect. These are useful diagnostics for any Bayesian estimation, pointing to future work to be done applying REPS to a wider range of applications.

References

Abdulkadiroğlu, A., N. Agarwal, and P. A. Pathak (2015). The welfare effects of coordinated assignment: Evidence from the NYC HS match. *NBER Working Paper*.

Avery, C. N., M. E. Glickman, C. M. Hoxby, and A. Metrick (2016). A revealed preference ranking of U.S. colleges and universities. *Quarterly Journal of Economics 128*(1), 425–467.

Baumeister, C. and J. D. Hamilton (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica 83*(5), 1963–1999.

Berger, J. and L. M. Berliner (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *The Annals of Statistics*, 461–486.

Bidder, R., R. Giacomini, and A. McKenna (2016). Stress testing with misspecified models. *Working Paper*.

Canova, F. and L. Sala (2009). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics 56*(4), 431–449.

Del Negro, M. and F. Schorfheide (2004). Priors from general equilibrium models for VARs. *International Economic Review 45*(2), 643–673.

Del Negro, M. and F. Schorfheide (2008). Forming priors for DSGE models (and how it affects the assessment of nominal rigidities). *Journal of Monetary Economics 55*(7), 1191–1208.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016). Solution and estimation methods for DSGE models. *Handbook of Macroeconomics 2*, 527–724.

Giacomini, R. and T. Kitagawa (2015). Robust inference about partially identified SVARs. *Working Paper*.

Giacomini, R., T. Kitagawa, and H. Uhlig (2016). Estimation under ambiguity. *Working Paper*.

Giannone, D., M. Lenza, and G. Primiceri (2017). Priors for the long run. *Working Paper*.

Gustafson, P. (2000). Local robustness in Bayesian analysis. In *Robust Bayesian Analysis*, pp. 71–88. Springer.

Hansen, L. P. and T. J. Sargent (2001). Robust control and model uncertainty. *American Economic Review 91*(2), 60–66.

Hansen, L. P. and T. J. Sargent (2007). Recursive robust estimation and control without commitment. *Journal of Economic Theory 136*(1), 1–27.

Herbst, E. and F. Schorfheide (2014). Sequential Monte Carlo sampling for DSGE models. *Journal of Applied Econometrics 29*(7), 1073–1098.

Herbst, E. P. and F. Schorfheide (2015). *Bayesian Estimation of DSGE Models.* Princeton University Press.

Iskrev, N. (2010). Local identification in DSGE models. *Journal of Monetary Economics 57*(2), 189–202.

Komunjer, I. and S. Ng (2011). Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica 79*(6), 1995–2032.

Koop, G., M. H. Pesaran, and R. P. Smith (2013). On identification of Bayesian DSGE models. *Journal of Business & Economic Statistics 31*(3), 300–314.

Moreno, E. (2000). Global Bayesian robustness for some classes of prior distributions. In *Robust Bayesian Analysis*, pp. 45–70. Springer.

Müller, U. K. (2012). Measuring prior sensitivity and prior informativeness in large Bayesian models. *Journal of Monetary Economics 59*(6), 581–597.

Petersen, I. R., M. R. James, and P. Dupuis (2000). Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Transactions on Automatic Control 45*(3), 398–412.

Robertson, J. C., E. W. Tallman, and C. H. Whiteman (2005). Forecasting using relative entropy. *Journal of Money, Credit, and Banking 37*(3), 383–401.

Sala, L. (2015). DSGE models in the frequency domain. *Journal of Applied Econometrics 30*, 219–240.

Schmitt-Grohé, S. and M. Uribe (2012). What's news in business cycles. *Econometrica 80*(6), 2733–2764.

Smets, F. and R. Wouters (2007). Shocks and frictions in us business cycles: A Bayesian DSGE approach. *The American Economic Review 97*(3), 586–606.

Appendix

A   Proofs

A.1   Solution to primal and dual problems

Lemma 2. *The solutions for M in problems (2.7)-(2.8) and (2.9)-(2.10) both have the form (2.11).*

Proof. Denote the marginal data density under prior π and likelihood L by $\zeta \equiv \int \pi(\theta) L(\theta|X) d\theta$. First recall that

$$p(\theta|X) = \frac{\pi(\theta) L(\theta)}{\zeta} \tag{A.1}$$

Consider the dual problem (2.9)-(2.10). Attaching the multiplier $\lambda \mathbb{E}_p[M]\zeta$ to (2.10) and the multiplier μ to the constraint $\mathbb{E}_\pi[M] = 1$, we have the first-order condition:

$$0 = \mu + 1 + \log M(\theta) - \lambda L(\theta|X)(\gamma(\theta) - \tilde{\gamma}), \tag{A.2}$$

which we can rearrange to obtain (2.11).

Now consider the primal problem (2.7)-(2.8). Attaching the multiplier $1/(\lambda \mathbb{E}_p[M]\zeta)$ to (2.8) and the multiplier $\mu/(\lambda \mathbb{E}_p[M]\zeta)$ to the constraint $\mathbb{E}_\pi[M] = 1$, we have the first-order condition:

$$
\begin{aligned}
0 &= \mu + 1 + \log M(\theta) - \lambda \mathbb{E}_p[M] L(\theta|X) \left(\frac{\gamma(\theta)}{\mathbb{E}_p[M]} - \frac{\mathbb{E}_p[M(\theta)\gamma(\theta)]}{\mathbb{E}_p[M]^2} \right) \\
&= \mu + 1 + \log M(\theta) - \lambda L(\theta|X) \left(\gamma(\theta) - \mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\gamma(\theta) \right] \right).
\end{aligned} \tag{A.3}
$$

Rearranging (A.3), we have:

$$M(\theta) \propto \exp\left[\lambda L(\theta|X) \left(\gamma(\theta) - \mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\gamma(\theta) \right] \right) \right], \tag{A.4}$$

which has the same form as (2.11) once we replace $\tilde{\gamma}$ with the full expression for the worst-case posterior mean of γ.   □

A.2   Subspaces

Lemma 3. *The solution for M is problem (2.15)-(2.16) is (2.17).*

Proof. Notice that

$$\mathbb{E}_{p^*}\left[M\left(\theta^*\right)\gamma\left(\theta\right)\right] = \mathbb{E}_{p^*}\left[M\left(\theta^*\right)\mathbb{E}_p\left[\gamma\left(\theta\right)|\theta^*\right]\right] \tag{A.5}$$

It immediately follows that the first-order condition of (2.15)-(2.16) is:

$$0 = \mu + 1 + \log M\left(\theta\right) - \lambda L\left(\theta|X\right)\left(\mathbb{E}_p\left[\gamma\left(\theta\right)|\theta^*\right] - \tilde{\gamma}\right), \tag{A.6}$$

which simplifies to (2.17). $\qquad\square$

A.3 Additional constraints

Consider the constrained optimization problem:

$$\min_{M(\theta):\mathbb{E}_\pi[M]=1} \mathbb{E}_\pi\left[M\left(\theta\right)\log M\left(\theta\right)\right] \tag{A.7}$$

$$\text{s.t. } \mathbb{E}_p\left[\frac{M\left(\theta\right)}{\mathbb{E}_p\left[M\right]}\gamma\left(\theta\right)\right] = \tilde{\gamma} \tag{A.8}$$

$$\mathbb{E}_p\left[\frac{M\left(\theta\right)}{\mathbb{E}_p\left[M\right]}g_{p,k}\left(\theta\right)\right] = \tilde{g}_{p,k} \text{ for } k = 1, ..., K \tag{A.9}$$

$$\mathbb{E}_\pi\left[M\left(\theta\right)g_{\pi,l}\left(\theta\right)\right] = \tilde{g}_{\pi,l} \text{ for } l = 1, ..., L \tag{A.10}$$

This is the problem (2.9)-(2.10) augmented by the additional moment conditions (A.9)-(A.10).

Attaching multipliers $\lambda\mathbb{E}_p\left[M\right]\zeta$, $\mu_{p,k}\mathbb{E}_p\left[M\right]\zeta$ and $\mu_{\pi,l}$ to constraints (A.8), (A.9) and (A.10) respectively, we obtain the first-order condition:

$$\begin{aligned} 0 =& \mu + 1 + \log M\left(\theta\right) - \lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right) - \tilde{\gamma}\right) \\ & - \sum_{k=1}^K \mu_{p,k}L\left(\theta|X\right)\left(g_{p,k}\left(\theta\right) - \tilde{g}_{p,k}\right) - \sum_{l=1}^L \mu_{\pi,l}\left(g_{\pi,l}\left(\theta\right) - \tilde{g}_{\pi,l}\right), \end{aligned} \tag{A.11}$$

where μ is the multiplier on the constraint $\mathbb{E}_\pi\left[M\right] = 1$. We rearrange (A.11) to obtain the solution:

$$\begin{aligned} M\left(\theta\right) \propto & \exp\left[\lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right) - \tilde{\gamma}\right)\right] \\ & \times \exp\left[L\left(\theta|X\right)\sum_{k=1}^K \mu_{p,k}\left(g_{p,k}\left(\theta\right) - \tilde{g}_{p,k}\right) + \sum_{l=1}^L \mu_{\pi,l}\left(g_{\pi,l}\left(\theta\right) - \tilde{g}_{\pi,l}\right)\right], \end{aligned} \tag{A.12}$$

where the second term introduces $K + L$ additional unknowns arising from the moment conditions (A.9)-(A.10).

A.4 Evaluating relative entropy

Lemma 4. *The solution for the minimum relative entropy in (2.9)-(2.10) is:*

$$\mathbb{E}_\pi \left[M\left(\theta\right) \log M\left(\theta\right) \right] = -\log \mathbb{E}_\pi \left[\exp\left[\lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right) - \tilde{\gamma}\right) \right] \right] \tag{A.13}$$

Proof. Define $\kappa \equiv 1/\mathbb{E}_\pi \left[\exp\left[\lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right) - \tilde{\gamma}\right) \right] \right]$. Taking logs of the solution (2.11) yields:

$$\log M\left(\theta\right) = \log \kappa + \lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right) - \tilde{\gamma}\right). \tag{A.14}$$

Denote the marginal data density by $\zeta \equiv \int \pi\left(\theta\right) L\left(\theta|X\right) d\theta$ and denote the worst-case prior and posterior by $\tilde{\pi}$ and $\tilde{p}$ respectively. Expand the expression for relative entropy:

$$\begin{aligned}
\mathbb{E}_\pi \left[M\left(\theta\right) \log M\left(\theta\right) \right] &= \int M\left(\theta\right) \log M\left(\theta\right) \pi\left(\theta\right) d\theta \\
&= \int \left[\log \kappa + \lambda L\left(\theta|X\right)\left(\gamma\left(\theta\right) - \tilde{\gamma}\right) \right] \tilde{\pi}\left(\theta\right) d\theta \\
&= \log \kappa + \int \lambda \zeta\left(\gamma\left(\theta\right) - \tilde{\gamma}\right) \frac{\tilde{\pi}\left(\theta\right) L\left(\theta|X\right)}{\zeta} d\theta \\
&= \log \kappa + \lambda \zeta \left(\int \gamma\left(\theta\right) \tilde{p}\left(\theta|X\right) d\theta - \tilde{\gamma} \right) = \log \kappa \tag{A.15}
\end{aligned}$$

The third equality uses the fact that $\tilde{\pi}$ integrates to one. The fourth equality uses the fact that $\tilde{p} = \tilde{\pi} L/\zeta$. The last equality uses (2.10) and the fact that $\mathbb{E}_p \left[\frac{M(\theta)}{\mathbb{E}_p[M]} \gamma\left(\theta\right) \right] = \int \gamma\left(\theta\right) \tilde{p}\left(\theta|X\right) d\theta$. $\qquad \square$

A.5 Asymptotics

Define $\Sigma_{\ell,T} \equiv \frac{1}{T}\Omega$ to be the variance of the likelihood.

Proof. (Lemma 1) Assume $\overline{X}_T = \theta_0$, and first consider the case with Ω diagonal, which implies $\Sigma_{\ell,T}$ is diagonal. Define $\Delta\left(\theta\right) \equiv \theta - \theta_0$. Abusing notation, we can write the likelihood as a function of Δ:

$$L\left(\Delta; \Sigma_\ell\right) = \left| 2\pi \Sigma_{\ell,T} \right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \Delta' \Sigma_{\ell,T}^{-1} \Delta \right] = \left| \Sigma_\ell^T \right|^{-\frac{1}{2}} L\left(\Sigma_{\ell,T}^{-\frac{1}{2}} \Delta; I \right) \tag{A.16}$$

Abusing notation again, write (2.11) as a function of Δ:

$$M\left(\Delta;\Sigma_\ell\right) \propto \exp\left[\lambda\left(\Sigma_\ell^T\right) L\left(\Delta;\Sigma_{\ell,T}\right)\left(\Delta_1 + \left(1 - \frac{T\sigma_{p,T}^1}{\omega_1}\right)\theta_0 + c\sigma_{p,T}^1\right)\right] \tag{A.17}$$

As $T \to \infty$, since $\frac{T\sigma_{p,T}^1}{\omega_1} \to 1$, there exists λ^*, M^* such that:

$$|\Sigma_{\ell,T}|^{-\frac{1}{2}} \lambda\left(\Sigma_{\ell,T}\right) \to \lambda^* \tag{A.18}$$

$$M\left(\Sigma_{\ell,T}^{\frac{1}{2}}\Delta;\Sigma_{\ell,T}\right) \to M^*\left(\Delta\right) \tag{A.19}$$

for (2.10) and $\mathbb{E}_\pi\left[M\left(\Delta\right)\right] = 1$ to be satisfied. In particular, we have

$$M^*\left(\Delta\right) \propto \exp\left[\lambda^* L\left(\Delta;I\right)\left(\Delta_1 + c\omega_1\right)\right] \tag{A.20}$$

Denoting $\hat{\pi}\left(\Delta\right) \equiv \pi\left(\Delta + \theta_0\right)$, the relative entropy is

$$T^{\frac{d}{2}}\mathcal{R}_T = T^{\frac{d}{2}} \int M\left(\Delta;\Sigma_{\ell,T}\right) \log\left[M\left(\Delta;\Sigma_{\ell,T}\right)\right] \hat{\pi}\left(\Delta\right) d\Delta \tag{A.21}$$

$$\approx T^{\frac{d}{2}} \int M^*\left(\Sigma_{\ell,T}^{-\frac{1}{2}}\Delta\right) \log\left[M^*\left(\Sigma_{\ell,T}^{-\frac{1}{2}}\Delta\right)\right] \hat{\pi}\left(\Delta\right) d\Delta \tag{A.22}$$

$$\approx \overline{R}\pi\left(\theta_0\right) T^{\frac{d}{2}} |\Sigma_{\ell,T}|^{-\frac{1}{2}} = \overline{R}\pi\left(\theta_0\right) |\Omega|^{-\frac{1}{2}} \tag{A.23}$$

fror some constant r. The second line follows because we can find, for any ε, some neighborhood N_ε around zero such that $M^*\left(\Delta\right) \log\left[M^*\left(\Delta\right)\right] < \varepsilon$ for all $\Delta \notin N_\varepsilon$.

When Ω is not diagonal, we first note that we can decompose the likelihood of θ into the marginal likelihood of θ_1 and the conditional likelihood of $\theta_{2:d}|\theta_1$, both of which remain Gaussian. An eigendecomposition of $\theta_{2:d}|\theta_1$ reparameterizes the likelihood in terms of orthogonal components, after which we can apply the proof for diagonal Ω. Finally, the proof follows through with general $\overline{X}_T$ since $\overline{X}_T \to \theta_0$. $\qquad\square$

B Sequential Monte Carlo

B.1 Implementation details

Constructing bridge distributions. To define the sequence of worst-case means, one can take:

$$\gamma_i = \mathbb{E}_p\left[\gamma(\theta)\right] + \left(\tilde{\gamma} - \mathbb{E}_p\left[\gamma(\theta)\right]\right)\left(\frac{i}{N_{\text{SMC}}}\right)^\nu. \tag{B.1}$$

A smaller value of ν corresponds to larger initial steps, and smaller steps toward the end of the SMC algorithm. Substituting $\tilde{\gamma}_i$ into the right-hand side of (2.10) for each i yields a sequence of distortions:

$$M_i(\theta) \propto \exp\left[\lambda_i L\left(\theta|X\right)\left(\gamma(\theta) - \tilde{\gamma}_i\right)\right], \tag{B.2}$$

which in turn imply a sequence of intermediate worst-case priors $\{\pi_i\}_{i=0}^{N_{\text{SMC}}}$ and posteriors $\{p_i\}_{i=0}^{N_{\text{SMC}}}$.

Transition between bridge distributions. To transition iteratively through these bridge distributions, we use transition weights π_i/π_{i-1} and p_i/p_{i-1}, both of which are proportional to:

$$m_i(\theta) \equiv \frac{M_i(\theta)}{M_{i-1}(\theta)} \propto \exp\left[L\left(\theta|X\right)\left[\lambda_i\left(\gamma(\theta) - \tilde{\gamma}_i\right) - \lambda_{i-1}\left(\gamma(\theta) - \tilde{\gamma}_{i-1}\right)\right]\right]. \tag{B.3}$$

Given λ_{i-1} and draws from π_{i-1} and p_{i-1}, the only unknown remaining is λ_i, which we solve for from (2.11), which we rewrite as:

$$\tilde{\gamma} = \mathbb{E}_{p_{i-1}}\left[\frac{\exp\left[L\left(\theta|X\right)\left[\lambda_i\left(\gamma(\theta) - \tilde{\gamma}_i\right) - \lambda_{i-1}\left(\gamma(\theta) - \tilde{\gamma}_{i-1}\right)\right]\right]}{\mathbb{E}_{p_{i-1}}\left[\exp\left[L\left(\theta|X\right)\left[\lambda_i\left(\gamma(\theta) - \tilde{\gamma}_i\right) - \lambda_{i-1}\left(\gamma(\theta) - \tilde{\gamma}_{i-1}\right)\right]\right]\right]}\gamma(\theta)\right]. \tag{B.4}$$

With a sufficiently large N_{SMC}, importance sampling of p_i from p_{i-1} is feasible. We can then solve for λ_i in (B.4) by using the particle approximation of p_{i-1} to evaluate the expectation on the right-hand side, analogously to (5.1).

Number of particles, mutation steps, and SMC steps. Given the sequence $\{\tilde{\gamma}_i\}_{i=1}^{N_{\text{SMC}}}$, three parameters need to be chosen for Algorithm 1: the number of particles N_{P}, the number of Metropolis-Hastings mutation steps N_{MH}, and the number of SMC steps N_{SMC}. Relative to Herbst and Schorfheide (2014), it is more important here to have a large number of particles, so that the

expectations in equation (B.4) are evaluated accurately when solving for λ_i. Similarly, N_{MH} must be sufficiently large in order to solve for λ_i accurately. If λ_i is computed accurately, the posterior mean of γ evaluated from the particles before and after the mutation step should be identical up to sampling error. We can check if N_{SMC} is sufficiently large by ensuring that at each stage, the distribution of m_i is well-behaved in the tails.

Moving from approximate to true worst-case distributions. Algorithm 2 provides draws from an approximate worst-case prior and posterior, with distortions

$$\widetilde{M}\left(\theta^*\right) \propto \exp\left[\tilde{\lambda}\widehat{L}\left(\theta^*|X\right)\left(\hat{\gamma}\left(\theta^*\right) - \tilde{\gamma}\right)\right] \tag{B.5}$$

instead of (2.17). To transform these draws into draws from the true worst-case distribution, notice that the Radon-Nikodym derivative between the true and approximate worst-case distributions is

$$\frac{M\left(\theta^*\right)}{\widetilde{M}\left(\theta^*\right)} \propto \frac{\exp\left[\lambda^* L^*\left(\theta^*|X\right)\left(\gamma^*\left(\theta^*\right) - \tilde{\gamma}\right)\right]}{\exp\left[\tilde{\lambda}\widehat{L}\left(\theta^*|X\right)\left(\hat{\gamma}\left(\theta^*\right) - \tilde{\gamma}\right)\right]} \tag{B.6}$$

where we can solve for λ^* using the constraint (2.16). Once we solve for λ^*, we can begin with the approximate worst-case draws, then use the selection and mutation steps from Algorithm 1 to obtain draws from the true worst-case distributions.

B.2 Evaluating relative entropy

We now use the output from the SMC in Algorithm 1 together with Lemma 4 to evaluate the relative entropy of the worst-case prior relative to the original prior.

To use Lemma 4, recall that the intermediate weights in Algorithm 1 have the form:

$$m_i\left(\theta\right) \equiv \frac{M_i\left(\theta\right)}{M_{i-1}\left(\theta\right)} = \frac{\kappa_i}{\kappa_{i-1}} \exp\left[L\left(\theta|X\right)\left[\lambda_i\left(\gamma\left(\theta\right) - \tilde{\gamma}_i\right) - \lambda_{i-1}\left(\gamma\left(\theta\right) - \tilde{\gamma}_{i-1}\right)\right]\right]. \tag{B.7}$$

Since $\kappa = \kappa_{N_{\mathrm{SMC}}} = \prod_{i=1}^{N_{\mathrm{SMC}}} \frac{\kappa_i}{\kappa_{i-1}}$, at each stage we evaluate:

$$\frac{\kappa_i}{\kappa_{i-1}} = \mathbb{E}_{\pi_{i-1}}\left[\exp\left[L\left(\theta|X\right)\left[\lambda_i\left(\gamma\left(\theta\right) - \tilde{\gamma}_i\right) - \lambda_{i-1}\left(\gamma\left(\theta\right) - \tilde{\gamma}_{i-1}\right)\right]\right]\right], \tag{B.8}$$

from which we obtain:

$$\mathbb{E}_\pi\left[M_i\left(\theta\right)\log M_i\left(\theta\right)\right] = \sum_{\iota=1}^{i}\log\frac{\kappa_\iota}{\kappa_{\iota-1}} \tag{B.9}$$

for $i = 1, ..., N_{\text{SMC}}$. Directly evaluating the relative entropies from the particle approximations would itself require solving for κ_i, leading to greater numerical error.

C Gaussian location model

All calculations for the Gaussian location model in Section 4 for dimensionality $d = 1$ and $d = 2$ are done using grids. For $d = 1$, the grid has range $[-8, 8]$ and $10^5 + 1$ uniformly spaced grid points. For $d = 2$, the grid has range $[-5, 5] \times [-5, 5]$ and has $10^3 + 1$ uniformly spaced grid points in each direction. For $d > 2$, I use the SMC Algorithm 1 with $(N_{\text{P}}, N_{\text{MH}}, N_{\text{SMC}}) = (d \times 10^5, 10, 100)$, and average across 25 runs.

D Smets and Wouters (2007)

D.1 Gaussian mixture approximation of posterior

To approximate p, I take a Gaussian mixture approximation of $p\left(\hat{\theta}|X\right)$, where $\hat{\theta}$ is the following transformation of θ:

$$\hat{\theta}_i = \begin{cases} \theta_i & \theta_i \in (-\infty, \infty) \\ \log\left(\theta_i\right) & \theta_i \in (0, \infty) \\ \log\left(\frac{1}{\theta_i} - 1\right) & \theta_i \in (0, 1) \end{cases} \tag{D.1}$$

which is chosen so that all the components of $\hat{\theta}$ are bounded neither above nor below. This transformation improves the quality of the approximation, especially around the tails, because the marginals of the transformed parameters are closer to being Gaussian. There is suggestive evidence that the Gaussian mixture approximates the posterior well. Figure D.1 plots the marginals of each parameter under the posterior and the Gaussian mixture approximation. We see that the marginals are visually indistinguishable. The first and second moments are also very similar. A (one component) Gaussian approximation would match these moments perfectly except for sampling error.
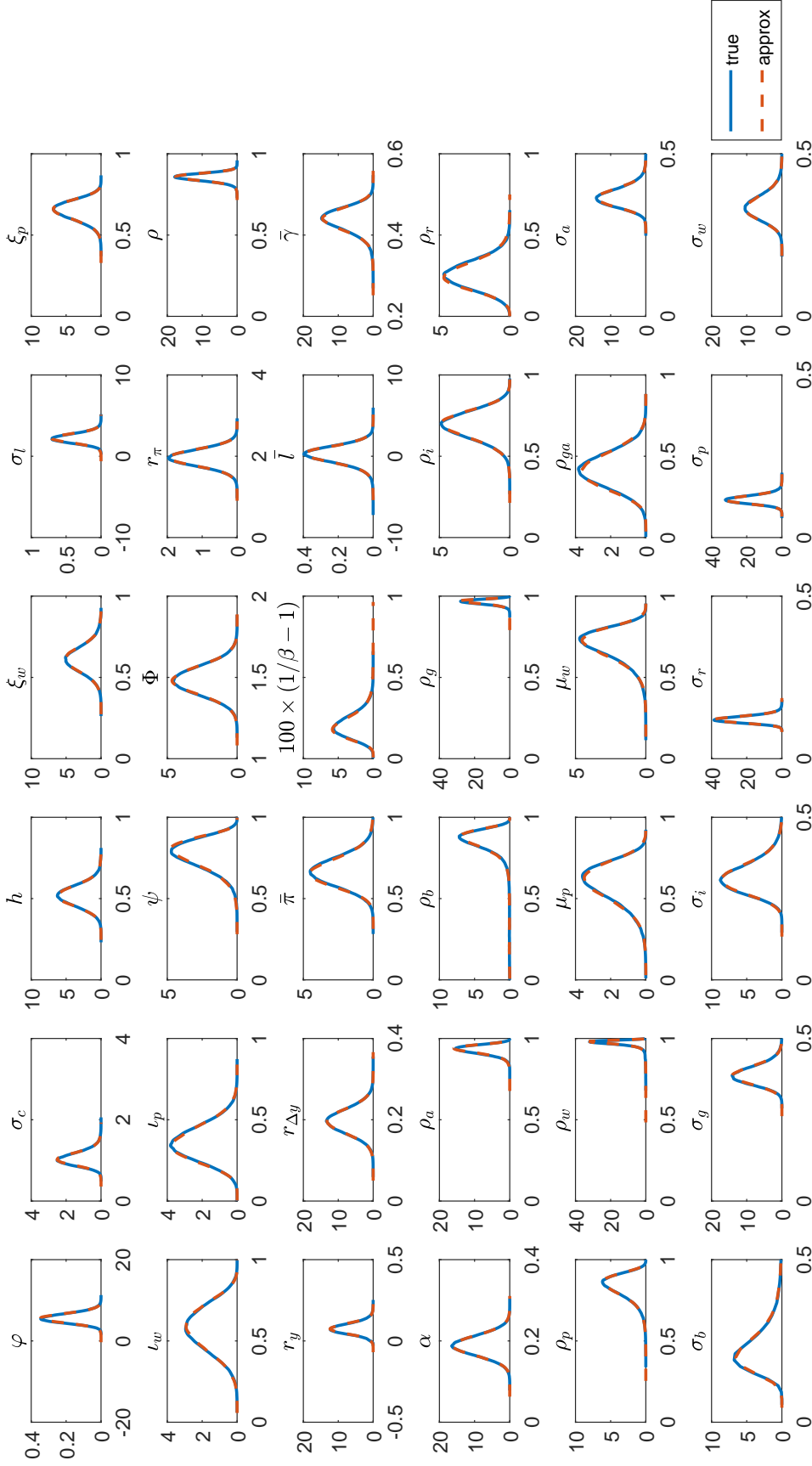
Figure D.1: Original and approximate marginal posteriors. Blue solid line corresponds to true posterior; red dashed line corresponds to Gaussian mixture approximation.

D.2 Sequential Monte Carlo

Worst-case quantiles. Let $\psi_h(\theta)$ be the impulse response of output to a 100 basis point decrease in interest rates at horizon h. Let $Q_f[\psi_h; q]$ be the qth quantile of ψ_h under the distribution f, and let $\sigma_f[\cdot]$ denote the standard deviation under distribution f. We set the worst-case qth quantile $\tilde{\psi}_h^q$ for horizon h as follows. For the Taylor rule prior, for the lower bound we choose:

$$
\tilde{\psi}_h^{0.16} = \begin{cases} Q_p[\psi_h; 0.16] - \frac{1}{8}\sigma_p[\psi_h] & h \le 2 \text{ or } h \ge 16 \\[2mm] Q_p[\psi_h; 0.16] - \frac{1}{5}\sigma_p[\psi_h] & 13 \le h \le 15 \\[2mm] Q_p[\psi_h; 0.16] - \frac{1}{4}\sigma_p[\psi_h] & h = 3 \text{ or } 10 \le h \le 12 \\[2mm] Q_p[\psi_h; 0.16] - \frac{1}{3}\sigma_p[\psi_h] & \text{otherwise} \end{cases} \tag{D.2}
$$

and for the upper bound we choose

$$
\tilde{\psi}_h^{0.84} = \begin{cases} Q_p[\psi_h; 0.84] + \frac{1}{2}\sigma_p[\psi_h] & h \le 4 \\[2mm] Q_p[\psi_h; 0.84] + \frac{3}{4}\sigma_p[\psi_h] & 5 \le h \le 8 \\[2mm] Q_p[\psi_h; 0.84] + \sigma_p[\psi_h] & \text{otherwise} \end{cases} \tag{D.3}
$$

For the nominal frictions prior, for the lower bound we choose:

$$
\tilde{\psi}_h^{0.16} = \begin{cases} Q_p[\psi_h; 0.16] - \frac{1}{240}\sigma_p[\psi_h] & h \le 2 \\[2mm] Q_p[\psi_h; 0.16] - \frac{1}{80}\sigma_p[\psi_h] & h = 3 \\[2mm] Q_p[\psi_h; 0.16] - \frac{1}{40}\sigma_p[\psi_h] & 10 \le h \le 21 \\[2mm] Q_p[\psi_h; 0.16] - \frac{1}{20}\sigma_p[\psi_h] & \text{otherwise} \end{cases} \tag{D.4}
$$

and for the upper bound we choose

$$
\tilde{\psi}_h^{0.84} = \begin{cases} Q_p[\psi_h; 0.84] + \frac{1}{2}\sigma_p[\psi_h] & h \le 4 \\[2mm] Q_p[\psi_h; 0.84] + \frac{3}{4}\sigma_p[\psi_h] & 5 \le h \le 8 \\[2mm] Q_p[\psi_h; 0.84] + \sigma_p[\psi_h] & \text{otherwise} \end{cases} \tag{D.5}
$$

The worst-case quantiles are chosen so that they imply similar sized distortions in terms of relative entropy.

Bridge distributions. To construct the bridge distributions, we consider the sequence of quantiles analogously to (B.1):

$$\tilde{\psi}_{h,i}^{q} = Q_p\left[\psi_h; q\right] + \left(\tilde{\psi}_h^q - Q_p\left[\psi_h; q\right]\right)\left(\frac{i}{N_{\text{SMC}}}\right)^{\nu} \tag{D.6}$$

For the Taylor rule prior, we set $\nu = \frac{1}{2}$ and $\nu = \frac{3}{4}$ for the lower and upper bound respectively. For the nominal frictions prior, we set $\nu = \frac{1}{3}$ and $\nu = \frac{3}{4}$ for the lower and upper bound respectively.

Computing time. A single run of the SMC in MATLAB takes approximately three hours. For each bound and set of parameters, we parallelize the computations by horizon across 21 cores.

D.3   Robust error bands

To generate the robust error bands, we use a quadratic regression for each bound and horizon to predict the worst-case quantile for a given relative entropy. In particular, at each stage of each SMC run, we evaluate the quantile and relative entropy. Aggregating across the 10 SMC runs, we obtain $250 \times 10 = 2500$ draws, to which we fit a quadratic regression of the quantile on the relative entropy. The robust error bands are constructed from the fitted values for a given level of distortion. For robustness, I also fit mixture regression to account more flexibly for nonlinearities, but find substantive differences.