Working Paper Series

# Global Robust Bayesian Analysis in Large Models

WP 20-07

Paul Ho
Federal Reserve Bank of Richmond



FEDERAL RESERVE BANK OF RICHMOND®
Richmond ▪ Baltimore ▪ Charlotte

# Global Robust Bayesian Analysis in Large Models*

Paul Ho†

Federal Reserve Bank of Richmond‡

paul.ho@rich.frb.org

June 30, 2020

**Abstract**

This paper develops a tool for global prior sensitivity analysis in large Bayesian models. Without imposing parametric restrictions, the methodology provides bounds for posterior means or quantiles given any prior close to the original in relative entropy, and reveals features of the prior that are important for the posterior statistics of interest. We develop a sequential Monte Carlo algorithm and use approximations to the likelihood and statistic of interest to implement the calculations. Applying the methodology to the error bands for the impulse response of output to a monetary policy shock in the New Keynesian model of Smets and Wouters (2007), we show that the upper bound of the error bands is very sensitive to the prior but the lower bound is not, with the prior on wage rigidity playing a particularly important role.

*Download the latest version of the paper here.

†I am indebted to Jaroslav Borovička, Chris Sims and Mark Watson for their guidance. I thank Timothy Christensen, Liyu Dou, Ulrich Müller, Mikkel Plagborg-Møller, Frank Schorfheide, Denis Tkachenko, and numerous seminar and conference participants for comments and suggestions. I am also grateful for financial support from the Alfred P. Sloan Foundation, the CME Group Foundation, Fidelity Management & Research, the Macro Financial Modeling Initiative, and the International Association for Applied Econometrics I received for this project when I was working on it as a student at Princeton University.

‡The views expressed herein are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

# 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. In particular, given the complicated dependence of the posterior statistic on the prior, one may wish to consider a vast range of nonparametric changes in the joint prior of multiple parameters, rather than limiting oneself to a restrictive class of priors that relies on assumptions such as independence or distributional form. Existing prior sensitivity tools either restrict one to infinitesimal parametric changes to the prior or are infeasible outside relatively simple or low-dimensional settings.

This paper develops a method to investigate the sensitivity of conclusions across a nonparametric set of priors, while remaining feasible in large models. We refer to this method as *relative entropy prior sensitivity* (REPS). The calculation searches across priors that are close to the original prior in relative entropy, then finds the *worst-case* prior that leads to the largest change in the reported posterior estimates. Even though the set of alternative priors is nonparametric, the solution for the worst-case prior and posterior requires solving for only one scalar, regardless of the number of parameters, 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 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. One should be particularly concerned if the posterior is sensitive to features of the prior arising from ad hoc assumptions such as independence or distributional forms that were used solely for convenience.

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

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. Nevertheless, 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 exact worst-case distributions. In existing applications, the computational time for AREPS is of the same order of magnitude as that of estimating the model once.

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. The rule of thumb consists of a formula that has one free parameter, which we calibrate using asymptotic behavior the Gaussian location model. This allows a practitioner to quantify the sensitivity of her posterior to the prior by measuring how much a given change in the prior can affect the posterior estimate, and comparing the sensitivity to a Gaussian location model with a large number of observations. The comparison indicates the prior sensitivity of the estimation relative to what one would have concluded by looking only at prior and posterior variances.

Our main application is the impulse response of output to a monetary policy shock in the New Keynesian model of Smets and Wouters (2007). We use REPS to construct bounds that contain pointwise 68% error bands arising from any prior in a relative entropy ball around the original prior and to compare the bounds when we distort the prior for the Taylor rule parameters to when we distort the prior for the nominal frictions parameters. In contrast to Müller (2012), who finds that the impulse responses are relatively insensitive to the priors for the structural parameters, REPS reveals that the upper bound of the error bands is very sensitive to changes to the prior but the lower bound is not. The impulse response is more dependent on the prior at long horizons. The impulse response is more sensitive to the prior of the nominal frictions parameters than the Taylor rule parameters, and is especially sensitive to the prior on the wage rigidity parameter.

REPS detects dependence on the prior in the Smets and Wouters (2007) estimation that would be hard to discern with other approaches to prior sensitivity such as that of Müller (2012). 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. Without having to reestimate the model for different priors, we find results that are consistent with the extensive literature studying the New Keynesian model of Smets and Wouters (2007), providing support for the validity of the methodology and the numerical implementation.

**Related literature.** REPS overcomes key limitations of existing approaches to prior sensitivity analysis.[1] Local methods (e.g., Gustafson (2000); Müller (2012)) consider derivatives of specific posterior quantities with respect to the prior. These methods focus on only infinitesimal changes in the prior and posterior and are often restricted to parametric changes in the prior. Our main application and stylized examples show that these restrictions can 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 wider class of priors and consider potentially large changes in the posterior, but are infeasible outside a limited range of applications. REPS is a global method that is feasible in settings with high-dimensional and complicated likelihoods such as dynamic stochastic general equilibrium (DSGE) models. Moreover, it can be applied to a range of statistics such as means and credible intervals of a wide class of functionals of the model parameters. REPS is also a more general methodology that applies to any Bayesian estimation problem, in contrast to the literature focusing on prior sensitivity in partial identification problems (e.g., Giacomini and Kitagawa (2018); Giacomini, Kitagawa, and Uhlig (2019)).

The use of relative entropy follows the robust control literature (Petersen et al. (2000); Hansen and Sargent (2001)). 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 the posterior results. In related work, Giacomini, Kitagawa, and Uhlig (2019) construct a relative entropy ball around the prior for set-identified parameters conditional on the identified subvector of parameters. In contrast, when measuring the sensitivity of the posterior to the prior of a subvector of the parameter, REPS focuses on the marginal prior rather than the conditional prior.

[1] The econometrics literature on prior sensitivity analysis dates back to Chamberlain and Leamer (1976) and Leamer (1982). The early statistics literature on the topic is reviewed in Berger et al. (1994).

The importance of prior sensitivity analysis is especially salient in Bayesian 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. Systematic prior sensitivity analysis provides diagnostics for whether an audience with heterogenous prior should be concerned about one's posterior estimates. The need for prior sensitivity analysis is further motivated by 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. Ho (2020) applies the REPS methodology to an overlapping generations model, showing the importance of capital and lifecycle consumption data for identifying the effects of an aging population on interest rates. Bayesian methods are also widely used in the estimation of vector autoregressions (VARs). Del Negro and Schorfheide (2004) and Giannone et al. (2018) show that the priors in VARs play an important role for forecasting. Baumeister and Hamilton (2015) and Giacomini, Kitagawa, and Read (2019) show that priors are important when structural VARs are partially identified. Finally, Bayesian methods have been used by Abdulkadiroglu et al. (2017) and Avery et al. (2013) to estimate matching models.

**Outline.** I introduce the REPS framework in Section 2 and demonstrate the methodology using two stylized examples in Section 3. Section 4 derives the rule of thumb to quantify the difference between priors. I discuss implementation in Section 5. In Section 6, I apply the methodology to the DSGE model of Smets and Wouters (2007). Section 7 concludes.

# 2 Relative entropy prior sensitivity

## 2.1 Setting and notation

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 $p(\theta|X) \propto \pi(\theta) L(\theta|X)$. Suppose we are interested in the posterior of a function $\psi : \Theta \to \Psi$ of the parameter $\theta$, where $\Psi$ may

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 $\mathbb{E}_f[\cdot]$, and define the objective function $\gamma_\psi : \Psi \to \mathbb{R}$ so that $\mathbb{E}_p[\gamma_\psi(\psi)]$ captures the property of the posterior of $\psi$ 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 $\psi$. 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 $\psi$ evaluated at $\psi^*$. We denote $\gamma(\theta) \equiv \gamma_\psi(\psi(\theta))$, and study how $\mathbb{E}_p[\gamma(\theta)]$ depends on $\pi$.

To study the dependence of the posterior $p$ on the prior $\pi$, we need to describe the distorted posterior implied by an alternative prior. In particular, given an alternative prior $\tilde{\pi}$ that is absolutely continuous with respect to $\pi$, we can write:

$$\tilde{\pi}(\theta) \equiv M(\theta)\,\pi(\theta), \tag{2.1}$$

where $M$ is the Radon-Nikodym derivative of $\tilde{\pi}$ with respect to $\pi$. Since $\tilde{\pi}$ is a probability distribution, we have $M > 0$ and $\mathbb{E}_\pi[M] = 1$. Given the likelihood $L$, the prior $\tilde{\pi}$ implies the distorted posterior:

$$\tilde{p}(\theta|X) = \frac{M(\theta)}{\mathbb{E}_p[M]} p(\theta|X). \tag{2.2}$$

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.

## 2.2 Setup

To analyze the sensitivity of the posterior estimates to the prior, we search across a set of alternative priors that are close to the original prior in relative entropy, seeking the worst-case prior that yields the largest change in the posterior mean of the objective function $\gamma$. Comparing the change in the prior to the change in the posterior mean of $\gamma$ tells us of how much the posterior mean of $\gamma$ depends on the prior. Comparing the worst-case and original priors reveals parts of the prior that are important for determining the posterior mean of $\gamma$.

**Primal problem.** Formally, we consider:

$$\min_{M(\theta): M>0, \mathbb{E}_\pi[M]=1} \mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\gamma(\theta)\right] \tag{2.3}$$

$$\text{s.t. } \mathbb{E}_\pi[M(\theta)\log M(\theta)] \leq \mathcal{R}. \tag{2.4}$$

The minimization over $M$ satisfying $M > 0$ and $\mathbb{E}_\pi[M] = 1$ is equivalent to minimizing over alternative priors, as the random variable $M$ indexes the possible priors. We choose the prior that minimizes $\mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\gamma(\theta)\right]$, the distorted posterior mean of $\gamma$. Replacing the minimization operator with maximization gives the upper bound for the posterior mean. The left-hand side of (2.4) is the relative entropy or Kullback-Leibler divergence of the alternative prior relative to the original prior. The constant $\mathcal{R} \in \mathbb{R}_+$ provides an upper bound on the relative entropy, limiting us to priors that are statistically difficult to distinguish from the original prior, which we implicitly assume to contain useful information about the distribution of $\theta$. 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 $\mathcal{R}$.

There are several reasons for using relative entropy to constrain the set of priors. Firstly, relative entropy has theoretical justication. It measures the information that a Bayesian with prior $\pi$ needs to gather to change her beliefs to the alternative prior $M\pi$ and is invariant to the parameterization of $\theta$. Secondly, relative entropy does not impose parametric restrictions on the prior distortions, allowing us to analyze how distributional assumptions on the prior affect the posterior estimate. For instance, even if $\pi$ were an independent Gaussian prior, the relative entropy set of priors would include non-Gaussian and correlated priors. Thirdly, the functional form for relative entropy delivers an analytic solution to (2.3)-(2.4) that allows REPS to maintain tractability in large models. Finally, the use of relative entropy implies relatively weak conditions for the solution to (2.3)-(2.4) to be well-defined. Section 2.3 elaborates on the latter two points.

The problem (2.3)-(2.4) is related to the constraint problem of Hansen and Sargent (2001) and the prior robustness problem from Hansen and Sargent (2007), but differs because the objective function in (2.3) and the relative entropy in (2.4) 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 $\gamma$, 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.

**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:

$$\min_{M(\theta):M>0,\mathbb{E}_\pi[M]=1} \mathbb{E}_\pi[M(\theta)\log M(\theta)] \tag{2.5}$$

$$\text{s.t. } \mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\gamma(\theta)\right] = \tilde{\gamma}. \tag{2.6}$$

We now search across priors that imply that $\gamma$ has posterior mean $\tilde{\gamma}$, and picking the one that is closest to the original prior in terms of relative entropy.[2] We will justify the formulation (2.5)-(2.6), argue that it simplifies the solution, and explain how one can move seamlessly between the primal and dual problems in practice.

## 2.3 Solution

The solution to (2.5)-(2.6) has the form:

$$M(\theta) \propto \exp\left[\lambda L(\theta|X)\left(\gamma(\theta) - \tilde{\gamma}\right)\right], \tag{2.7}$$

where $\lambda \in \mathbb{R}$ is a constant to be solved for from the constraint (2.6). 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 $\theta$. This is key to making REPS feasible in large models.

The distortion $M$ depends on the parameter $\theta$ through the objective function $\gamma$ and likelihood $L$. The worst-case distortion $M$ reweights based on $\gamma$ because the statistic of interest is the posterior mean of $\gamma$. The direction and degree of reweighting depends on $\lambda$, which is the Lagrange multiplier on (2.6) scaled by $\mathbb{E}_p[M]$.[3] If $\tilde{\gamma} < \mathbb{E}_p[\gamma(\theta)]$, then in order to reduce the posterior mean of $\gamma$, we require $\lambda < 0$ so that the worst-case prior places more weight on smaller values of $\gamma$.

The difference between the solution (2.7) and the worst-case distortion in Hansen and Sargent (2001) is that the distortion in (2.7) is scaled by the likelihood $L$, which captures the role of the data in (2.5)-(2.6). The worst-case distortion depends on the likelihood because the expectations in (2.5) and (2.6) 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 $\lambda$, substitute (2.7) into (2.6):

$$\tilde{\gamma} = \mathbb{E}_p\left[\underbrace{\frac{\exp\left[\lambda L(\theta|X)\left(\gamma(\theta) - \tilde{\gamma}\right)\right]}{\mathbb{E}_p\left[\exp\left[\lambda L(\theta|X)\left(\gamma(\theta) - \tilde{\gamma}\right)\right]\right]}}_{\text{change of measure}}\gamma(\theta)\right]. \tag{2.8}$$

[2] 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.

[3] $\lambda$ is analogous to the penalty parameter in the multiplier problem of Hansen and Sargent (2001).

The right-hand side is the posterior expectation of $\gamma$ after a change of measure that depends on $\lambda$. As $\lambda$ increases, the change of measure places more weight on large values of $\gamma$, increasing the right-hand side. Since the left-hand side is a constant and the right-hand side is increasing in $\lambda$, (2.8) implies a unique solution for $\lambda$ that is straightforward to solve for numerically.[4]

**Comparison to the primal problem.** Appendix A shows that (2.3)-(2.4) also produces a solution of the form (2.7), but requires us to solve for both $\lambda$ and $\tilde{\gamma}$. The multiplier representation of (2.3)-(2.4), which is the Lagrangian problem, specifies $\lambda$ and requires us to solve for $\tilde{\gamma}$. We favor the dual representation (2.5)-(2.6) because $\tilde{\gamma}$ has a straightforward interpretation, while $\lambda$ 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 $\gamma$, then check how much the prior needs to be distorted in order to generate this change.

In practice, it will not matter whether one picks $\mathcal{R}$ and finds the worst-case posterior mean, or one picks $\tilde{\gamma}$ and finds the corresponding relative entropy. The sequential Monte Carlo algorithm in Section 5 yields a sequence of $(\mathcal{R}, \tilde{\gamma})$ pairs, allowing one to trace out the mapping between $\mathcal{R}$ and $\tilde{\gamma}$. We thus obtain $\tilde{\gamma}$ as a function of $\mathcal{R}$ despite solving the dual problem initially.

**Regularity conditions.** For the solution (2.7) to be valid, we require that the $\mathbb{E}_\pi[M] = 1$ constraint in (2.3) and (2.5) can be satisfied, and that $\mathbb{E}_p[\exp[\lambda L(\gamma - \tilde{\gamma})]]$ in (2.8) exists. A sufficient condition for this is that $L(\theta|X)(\gamma(\theta) - \tilde{\gamma})$ is bounded, so that the right-hand side of (2.7) is bounded for any given $\lambda$.

As an illustration, consider the prior $\theta \sim \mathcal{N}(0,1)$, and suppose we are interested in $\gamma(\theta) \equiv \theta^k$, where $k \geq 3$ is an odd integer. If the likelihood has Pareto tails proportional to $|\theta|^{-(\alpha+1)}$, then the above condition requires that $\alpha + 1 > k$. On the other hand, suppose the likelihood were flat. Then the solution to (2.3)-(2.4) is not well-defined.[5] Accordingly, the constraint $\mathbb{E}_\pi[M] = 1$ can no longer be satisfied with $M$ satisfying (2.7). However, the infeasibility is symptomatic of fat tails in the likelihood, and is thus an indication that the posterior mean is especially sensitive to the prior.

Relative entropy allows for relatively weak regularity conditions because it heavily penalizes increases in the mass of the tail of the original prior, thus favoring distortions closer to

[4]While the worst-case distribution is unique, one may find an alternative prior in the same set that implies a 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.

[5]To see this, recall that the $k$th moment of a t-distribution with $\nu < k$ degrees of freedom does not exist. Hence the moment of interest would not exist in an alternative prior with the tail of that t-distribution. To obtain an arbitrarily small relative entropy, we can distort the prior appropriately far out in the tail.

the mode of the original prior.[6] A statistical divergence that penalizes tail distortions less (e.g., total variation distance) would in general require more stringent regularity conditions.

## 2.4 Extensions

The REPS framework allows for flexibility in application. We now consider several prior sensitivity problems that can be incorporated into (2.5)-(2.6).

**Credible intervals.** One can adapt the constraint (2.6) to analyze the prior sensitivity of quantiles of the function of interest $\psi(\theta)$, producing bounds on the credible intervals for $\psi$ given a set of deviations in the prior. In particular, suppose we are interested in the dependence of the $q$th quantile of $\psi$ on the prior. Then we can solve (2.5) subject to:

$$\mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\mathbf{1}\left\{\psi(\theta)<\tilde{\psi}\right\}\right]=q, \tag{2.9}$$

where $\tilde{\psi}$ is the worst-case quantile of $\psi$. Since (2.9) is (2.6) with $\tilde{\gamma}=q$ and $\gamma(\theta)=\mathbf{1}\left\{\psi(\theta)<\tilde{\psi}\right\}$, the solution has the form (2.7), with the appropriate substitution for $\tilde{\gamma}$ and $\gamma$. Fixing $\tilde{\gamma}=q$, the sequential Monte Carlo algorithm yields a mapping between $\mathcal{R}$ and $\tilde{\psi}$.

With the above substitutions, we have $\gamma(\theta)-\tilde{\gamma}\in\{1-\tilde{\gamma},-\tilde{\gamma}\}$. The distortion $M$ therefore only takes on extreme values if $\theta$ has high likelihood. As a result, the conditions that $\mathbb{E}_\pi[M]=1$ and that $\mathbb{E}_p[\exp[\lambda L(\gamma-\tilde{\gamma})]]$ exists are satisfied even if the likelihood is flat, since

$$M(\theta)\propto\begin{cases}\exp[\lambda(1-\tilde{\gamma})] & \psi(\theta)<\tilde{\psi}\\ \exp[-\lambda\tilde{\gamma}] & \psi(\theta)\geq\tilde{\psi}\end{cases}. \tag{2.10}$$

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

**Subspaces.** Taking the expectations in (2.5)-(2.6) over the marginal prior and posterior of a subspace $\Theta^*$ of $\Theta$ allows us to study the dependence of the posterior on the marginal prior over $\Theta^*$ instead of the entire space $\Theta$.[7] Such an exercise can be useful if there is a natural partition for $\theta$. 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 $\Theta^*$ also simplifies the analysis of the worst-case distortions.

[6] Intuitively, express relative entropy as $\int\tilde{\pi}(\theta)\log[\tilde{\pi}(\theta)/\pi(\theta)]\,d\theta$, and notice that $\log[\tilde{\pi}/\pi]\to\infty$ as $\pi\to 0$. This also implies that we consider alternative priors with the same support as $\pi$.

[7] In the context of a partially identified model, the current approach differs from Giacomini, Kitagawa, and Uhlig (2019), who would distort the prior of $\theta^*$ conditional on the identified parameters.

More formally, consider

$$\min_{M(\theta^*):\mathbb{E}_{\pi^*}[M]=1} \mathbb{E}_{\pi^*}\left[M\left(\theta^*\right)\log M\left(\theta^*\right)\right] \tag{2.11}$$

$$\text{s.t. } \mathbb{E}_{p^*}\left[\frac{M\left(\theta^*\right)}{\mathbb{E}_{p^*}\left[M\right]}\mathbb{E}_p\left[\gamma\left(\theta\right)|\theta^*\right]\right] = \tilde{\gamma}, \tag{2.12}$$

where $\pi^*$ and $p^*$ are the marginal prior and posterior over $\Theta^*$. The constraint (2.12) arises by applying the law of iterated expectations to (2.6) and noting that $M$ now depends on $\theta^*$ only. Define $\gamma^*\left(\theta^*\right) \equiv \mathbb{E}_p\left[\gamma\left(\theta\right)|\theta^*\right]$. The solution to (2.11)-(2.12) is:

$$M\left(\theta^*\right) \propto \exp\left[\lambda L^*\left(\theta^*|X\right)\left(\gamma^*\left(\theta^*\right) - \tilde{\gamma}\right)\right], \tag{2.13}$$

where $L^*\left(\theta^*|X\right) \equiv p^*\left(\theta^*|X\right)/\pi^*\left(\theta^*\right)$ is the marginal likelihood of $\theta^*$. The solution (2.13) to the subspace problem is similar to the original solution (2.7), with the likelihood $L$ replaced with the marginal likelihood $L^*$ and the objective function $\gamma\left(\theta\right)$ replaced by its expectation conditional on $\theta^*$.

**Additional constraints.** We can further restrict the set of permissible priors by including prior or posterior moment restrictions to (2.5)-(2.6), 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 the additional equation with which to solve for the new unknown. See Appendix A for details.

One such moment restriction constrains the marginal data density. In particular, express the ratio of the marginal data density of the worst-case prior to that of the original prior as:

$$\mathbb{E}_p\left[M\left(\theta\right)\right] = \frac{\int L\left(\theta|X\right)\tilde{\pi}\left(\theta\right)d\theta}{\int L\left(\theta|X\right)\pi\left(\theta\right)d\theta} \tag{2.14}$$

and restrict $\mathbb{E}_p\left[M\right]$. The quantity $\mathbb{E}_p\left[M\right]$ is easily computed by taking an average of $M\left(\theta\right)$ across Monte Carlo draws. Berger et al. (1994) discusses why one might want to include the marginal data density as a criterion to ensure that the alternative priors considered are plausible. A small $\mathbb{E}_p\left[M\right] \ll 1$ suggests that $\tilde{\pi}$ is strongly rejected by the data, a large $\mathbb{E}_p\left[M\right] \gg 1$ could be evidence of $\tilde{\pi}$ being overfitted to the data, while $\mathbb{E}_p\left[M\right] = 1$ indicates that the data favor neither the original nor the alternative prior.

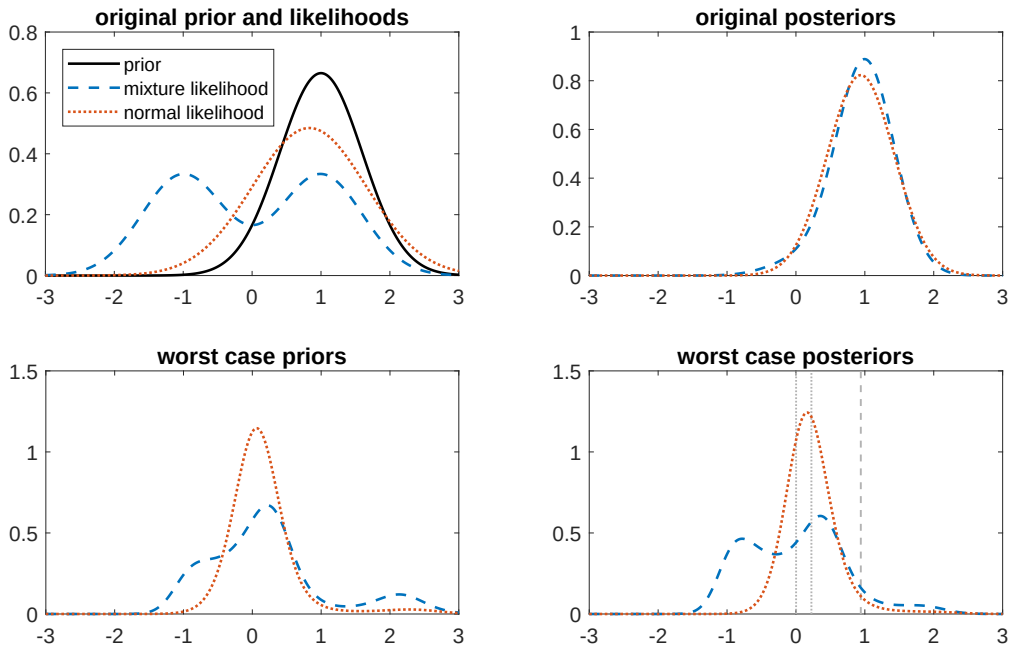


Figure 3.1: Likelihoods, priors and posteriors for mixture and normal likelihoods. Blue dashed lines correspond to mixture likelihood; red dotted lines correspond to normal likelihood. **Top left:** Original prior and likelihoods; **Top right:** Original posteriors; **Bottom left:** Worst case priors; **Bottom right:** Worst case posteriors, with original and worst case means in gray.

# 3 Two illustrative examples

We now present two stylized examples to illustrate how REPS can diagnose dependence on the prior that may be hard to detect otherwise and are prevalent in many applications. The first example shows that REPS accounts for behavior of the likelihood in the tail of the posterior, which may be hard to distinguish from visual inspection. The second shows that even if the prior and likelihood are Gaussian, REPS detects that the sensitivity to the prior depends on the object of interest $\psi$.

## 3.1 Example 1: multiple modes in the likelihood

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

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

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 posteriors have mean 0.942 and standard deviation 0.485. The top right panel of Figure 3.1 shows that the two posteriors are 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 $\theta$ has greater sensitivity to changes in the prior under the mixture likelihood. In particular, fixing $\gamma(\theta) = \theta$ and $\mathcal{R} = 1.25$, we solve (2.3)-(2.4) 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 REPS, the similarity of the two posteriors might mislead one to think that the posteriors 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 by Müller (2012), suggests that the posterior means in the two cases are 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 range of alternative priors considered by REPS. With the normal likelihood, the worst-case prior for $\theta$ is approximately normal, centered around 0. The worst-case posterior is also approximately normal, 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 is 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 that 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. Such multimodality is often 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

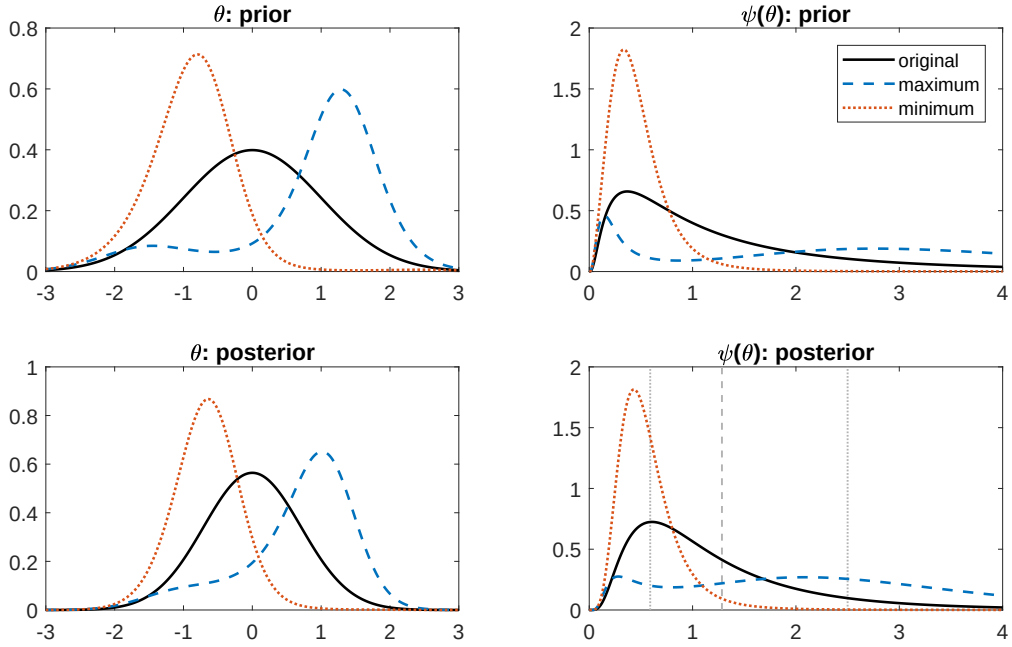


Figure 3.2: Prior and posterior of $\theta$ and $\psi(\theta)$. Black solid lines correspond to original distribution, blue dashed lines correspond to distributions that maximize mean; red dotted lines correspond to distributions that minimize mean. **Top left:** Prior of $\theta$; **Top right:** Prior of $\psi(\theta)$ **Bottom left:** Posterior of $\theta$; **Bottom right:** Posterior of $\psi(\theta)$.

provides a systematic approach to prior sensitivity analysis that accounts for potentially subtle features of the likelihood if they are important for one's posterior results.

## 3.2 Example 2: log-normal distribution

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 $\psi$ is log-normal with mean 1.28 and standard deviation 1.03. Figure 3.2 shows that the posterior of $\psi$ is skewed even though the posterior for $\theta$ is symmetric.

REPS shows that this asymmetry matters for the sensitivity of the posterior mean of $\psi$ to changes in the prior. Fixing $\gamma(\theta) = \psi(\theta)$ and taking $\mathcal{R} = 0.57$, which corresponds to a one standard deviation change in the posterior mean of $\theta$, the posterior mean of $\psi$ 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 $\psi$ is bounded below by zero but has a posterior with a fat right tail.

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 $\theta$ on the mean of $\psi$. The asymmetry arises despite the symmetry of the Gaussian prior and likelihood. Since relative entropy is invariant to one-to-one transformations of $\theta$, the set of priors does not depend on the parameterization of the problem.

These insights generalize to more complex settings where $\psi$ may be a complicated function whose sensitivity to the prior may be hard to analyze. For example, if $\psi$ 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 $\theta$ is high-dimensional. Even if visual inspection of the posterior of $\psi$ reveals that it could be sensitive to the prior, further analysis is needed to determine which parts of the prior of $\theta$ are important. REPS accounts for the function of interest through the $\gamma$ term in the solution (2.7) while checking across a wide range of alternative priors.

# 4 Quantifying the change in prior

To quantify prior sensitivity, we need to gauge how much the prior has changed to produce the specified change in posterior mean. In this section, I provide a formula that summarizes these relationships and give practitioners a rule of thumb for what is a large or small value of $\mathcal{R}$, taking the Gaussian location model as a benchmark.

## 4.1 Intuition

There are two challenges in gauging the size of $\mathcal{R}$. Firstly, the worst-case distortions are nonparametric, making it hard for a practitioner to have an intuition for whether the change in the prior is large or small. Secondly, because the distortions are concentrated in the high likelihood region, existing approaches such as the error detection probabilities (Hansen and Sargent (2008)) produce misleading conclusions.[8]

Instead, one needs to account for the concentration of the likelihood and the pdf of the prior around the high likelihood region when interpreting $\mathcal{R}$. To see this, notice that we can write the relative entropy as an integral over the prior, and recall that the worst-case distortions are concentrated in the high likelihood region. As the likelihood becomes more concentrated, the volume of the high likelihood region shrinks, reducing the effective region over which the integral is computed, thus decreasing the relative entropy. Within the high likelihood region, the integral is scaled by the prior. Intuitively, we have more prior knowledge about regions of high prior probability, making it more costly to change our beliefs about those regions.

[8] In our setting, such approaches can imply more sensitivity even when the likelihood is more concentrated.

## 4.2 Gaussian location model

The above intuition applies for the asymptotic behavior of the solution for (2.5)-(2.6) in the Gaussian location model, which we then use as a benchmark for $\mathcal{R}$. Appendix C verifies that the asymptotics provide a good approximation even for a relatively small sample size.

Consider a $d$-dimensional vector $\theta = (\theta_1, ..., \theta_d)'$ whose true value is $\theta_0$. Suppose we have prior $\theta \sim \mathcal{N}(0, \Sigma_\pi)$ and observe $T$ iid realizations of $X \sim \mathcal{N}(\theta_0, \Omega)$ with sample mean $\overline{X}_T$. Assume $\Omega$ is full rank. Then we have the 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 $\theta_i$ by $\sigma_{i,p,T}$.

**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 (2.5)-(2.6) satisfies $T^{\frac{d}{2}}\mathcal{R}_T \xrightarrow{a.s.} \overline{R}\pi(\theta_0)|\Omega|^{-\frac{1}{2}}$ for some $\overline{R}$.*

Lemma 1 states that the relative entropy $\mathcal{R}_T$ needed to shift the posterior mean by $c$ posterior standard deviations declines at rate $T^{\frac{d}{2}}$, and it depends on the variance of the data and the prior at $\theta_0$. The scaling factor $\overline{R}$ varies depending on the number of dimensions $d$ and amount of distortion $c$. See Appendix A for the proof.

As the sample size increases, 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 $\theta_0$, the asymptotic relative entropy is scaled by the prior $\pi(\theta_0)$ at $\theta_0$. The same asymptotics apply for the mean or quantile of any linear combination of $\theta$.[9]

## 4.3 Rule of thumb for $\mathcal{R}$

We calibrate $\mathcal{R}$ by taking a Gaussian approximation of the prior and posterior, then comparing the relative entropy to the corresponding Gaussian location model. In particular, for prior and posterior means $(\mu_\pi, \mu_p)$ and variances $(\Sigma_\pi, \Sigma_p)$, 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:

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

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

[9]For other models that satisfy the Bernstein-von Mises theorem, we have a similar asymptotic relative entropy, with $|\Omega|$ replaced by the Fisher information matrix.

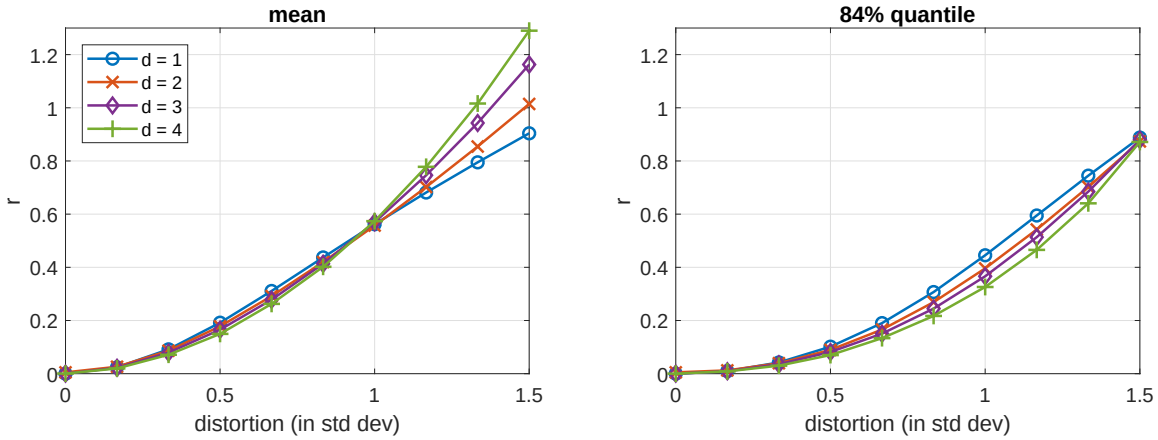


Figure 4.1: Scaled relative entropy $r$ and distortion in Gaussian location model with $\theta \sim \mathcal{N}(0, I)$, $\overline{X}_T = 0$, $T = 100$, $d \in \{1, ..., 4\}$. All distortions are scaled by posterior standard deviation. **Left:** Increase in posterior mean of $\theta_1$ by one posterior standard deviation; **Right:** Increase in 84% quantile of $\theta_1$ by one posterior standard deviation.

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 account for the concentration of the likelihood (relative to the prior) and the pdf of the prior around the high likelihood region. The $1.6^{\delta}$ scaling accounts for further differences across dimensions. The remaining free parameter $r$ controls the size of the change in the prior and posterior.

Intuitively, $r$ captures the sensitivity to the prior given the prior and posterior variances. While the prior and posterior variances are sufficient to determine prior sensitivity in a Gaussian location model, they do not account for the behavior of the likelihood in the tail of the posterior or correlations in the likelihood. In Section 6, we show that such features are important for determining the sensitivity to the prior, and that REPS accounts for them.

A practitioner can compute $r$ from (4.4), then compare the change in the posterior for the estimation of interest to the change that $r$ would imply in a Gaussian location model. Figure 4.1 shows how $r$ corresponds to changes in the mean and 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. For the mean and 84% quantile of the Gaussian model, $r = 0.05$ and $r = 0.50$ correspond approximately to 1/4 standard deviation and one standard deviation changes, respectively.[10]

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

# 5 Implementation

We now discuss the numerical implementation of the calculations in Section 2. We assume that we have Monte Carlo draws from both the prior and the posterior.

## 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 $\lambda$, we can evaluate the right-hand side of (2.8), approximating the expectations with Monte Carlo sample averages. We can then solve (2.8) using this Monte Carlo approximation. Given the solution for $\lambda$, we can now compute $M$ for any value of $\theta$. Reweighting the original Monte Carlo draws by $M$ then gives us the worst-case prior and posterior.

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

$$\mathbb{E}_\pi \left[ M(\theta) \log M(\theta) \right] = -\log \mathbb{E}_\pi \left[ \exp \left[ \lambda L(\theta|X) \left( \gamma(\theta) - \tilde{\gamma} \right) \right] \right]. \tag{5.1}$$

The right-hand side is negative log of the normalizing constant that ensures $M\pi$ integrates to one, which is straightforward to evaluate 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 $\Theta$ or the number of observations. The solution (2.7) 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 than 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

scale relative entropy according to the dispersion of the likelihood. For example, if a model is not point identified and the likelihood is flat in some identified set, such an approach may lead $|\Sigma_\ell|$ to be large even if the identified set is small.

**Algorithm 1:** Sequential Monte Carlo for 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 worst-case prior and posterior.
**Initialize:** Set $\{\theta_{\pi_0,j}\}_{j=1}^{N_{\mathrm{P}}} = \{\theta_{\pi,j}\}_{j=1}^{N_{\mathrm{P}}}$ and $\{\theta_{p_0,j}\}_{j=1}^{N_{\mathrm{P}}} = \{\theta_{p,j}\}_{j=1}^{N_{\mathrm{P}}}$.
**for** $i = 1$ **to** $N_{\mathrm{SMC}}$ **do**
  **compute weights:** Solve for $\lambda_i$ and compute $m_i \equiv \pi_i/\pi_{i-1}$ for each draw.
  **selection:** Draw from $\{\theta_{\pi_{i-1},j}\}_{j=1}^{N_{\mathrm{P}}}$ and $\{\theta_{p_{i-1},j}\}_{j=1}^{N_{\mathrm{P}}}$ 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 $N_{\mathrm{MH}}$ Metropolis-Hastings steps.
**end**
**return** $\{\theta_{\tilde{\pi},j}\}_{j=1}^{N_P} = \{\theta_{\pi_{N_{\mathrm{SMC}}},j}\}_{j=1}^{N_P}$ and $\{\theta_{\tilde{p},j}\}_{j=1}^{N_P} = \{\theta_{p_{N_{\mathrm{SMC}}},j}\}_{j=1}^{N_P}$.

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.

Our algorithm is based on Herbst and Schorfheide (2014), who begin with draws from a prior and transition to the posterior by constructing bridge distributions that are proportional to the product of the prior and the likelihood raised to an exponent. To adapt the algorithm, we construct bridge distributions for our setting and compute 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 (2.5)-(2.6) for a sequence of intermediate worst-case posterior means $\tilde{\gamma}_0 > ... > \tilde{\gamma}_{N_{\mathrm{SMC}}}$, with $\tilde{\gamma}_0 = \mathbb{E}_p[\gamma(\theta)]$ 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 in Algorithm 1. We leave the details to Appendix B.

A key by-product of the SMC algorithm is that it provides the choice of whether to fix the relative entropy and obtain worst-case means or quantile, or fix the worst-case quantities and obtain the associated relative entropy. By producing a sequence of worst-case distortions and solving for the associated relative entropies, the SMC algorithm allows the user to map the relationship between the relative entropy $\mathcal{R}$ and worst-case mean $\tilde{\gamma}$ (or quantile $\tilde{\psi}$). As emphasized in Section 2.3, the dual problem is then a computational device and does not force the user to choose the worst-case quantity instead of the relative entropy.

**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 $\pi$ 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 $(\pi, p, L^*, \gamma)$ 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

The main computational challenge in Algorithm 1 is the computation of $L$ and $\gamma$ in the mutation step. In particular, let the number of particles and Metropolis-Hastings mutation steps be $N_{\mathrm{P}}$ and $N_{\mathrm{MH}}$, respectively. Obtaining particle approximations of the worst-case prior and posterior each require us to compute $L$ and $\gamma$ for $N_{\mathrm{P}} \times N_{\mathrm{MH}} \times N_{\mathrm{SMC}}$ different parameter values. Both $L$ and $\gamma$ may be computationally expensive to compute. If we are interested in more than one statistic, we also need to repeat the SMC algorithm for each objective function we are interested in.[11] In addition, to apply Algorithm 1 to (2.11)-(2.12), we require the marginal likelihood $L^*$ and conditional expectation $\gamma^*$, both of which may be difficult to compute. To overcome these, we use an approximation to the REPS calculations, which we refer to as *approximate relative entropy prior sensitivity* (AREPS).

The main idea of AREPS is to replace $\pi$, $p$, $L^*$, and $\gamma^*$ with approximations $\hat{\pi}$, $\hat{p}$, $\widehat{L}$, and $\hat{\gamma}$. In particular, using the Monte Carlo draws from the original estimation as observations, we fit a set of basis functions to obtain approximations of $\hat{\pi}$, $\hat{p}$, and $\hat{\gamma}$ respectively for $\pi$, $p$, and $\gamma$.[12] If we are doing the REPS analysis over the entire $\Theta$, then $\widehat{L}$ and $\hat{\gamma}$ correspond to approximations for $L$ and $\gamma$, 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. From the practical perspective, many of these methods are straightforward to implement in most statistical software using built-in commands. The steps are described in Algorithm 2.

With the appropriate approximations, the approximate likelihood $\widehat{L}$ for a set of particles

[11] For example, if $\gamma$ is an impulse response function in a DSGE model, then one needs to solve the model for each draw of $\theta$ in order to compute the impulse response. To compute the likelihood $L$, one needs to run a Kalman filter using the solved model. If one were interested in the error bands for a set of impulse response functions, one would need to repeat Algorithm 1 for each impulse response at multiple horizons.

[12] 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.

can be computed in vectorized form. If one has multiple objectives (e.g., multiple horizons of an impulse response), one can parallelize across SMC algorithms. In addition, since we no longer need to compute the true $L^*$ or $\gamma^*$, output from packages such as Dynare can be directly fed into the algorithm.

For computational efficiency, the approximations $\hat{\pi}$, $\hat{p}$, and $\hat{\gamma}$ should be fast to evaluate. In our application in Section 6, we use a Gaussian mixture model and a quadratic logit to approximate $p$ and $\gamma^*$, respectively. For numerical accuracy, $L^*$ and $\gamma^*$ 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 $\pi^*$, $p^*$, or $\gamma^*$. This problem tends to be less severe when $\gamma^*$ is bounded. For example, AREPS would generally produce more accurate results when studying quantiles, since $\gamma^* \in [0, 1]$ and $M$ takes on extreme values in the high likelihood regions, which tend to be near the posterior mode.

If closed-form expressions for $L^*$ and $\gamma^*$ are available, we can reweight draws from the approximate worst-case prior and posterior to obtain draws from the true worst-case prior and posterior. See Appendix B for details.

# 6 Application: Smets and Wouters (2007)

Our main application is the workhorse DSGE model from Smets and Wouters (2007). Despite the size of the 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 even with the use of local prior sensitivity methods. The upper bound of the error bands is highly sensitive to the prior, especially to the nominal frictions parameters, of which the prior on the wage rigidity parameter is particularly important. We discuss the worst-case distortions in detail to provide support for the validity of the methodology and approximations.

## 6.1 Model and estimation

Smets and Wouters (2007) presents a medium-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.[13] The model has thirty-six parameters.

[13]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.

We use quarterly data from 1984Q1 to 2007Q4 of the Federal Reserve Economic Data (FRED) for GDP growth, consumption growth, investment growth, 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. Our original prior is from Smets and Wouters (2007). We 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.

## 6.2 Prior sensitivity

**Object of interest.** We construct 68% *robust error bands* for the impulse response of output to a one percentage point decrease in interest rates up to five years from impact. These are bounds that uniformly contain all error bands arising from the chosen set of priors. To that end, we solve the REPS problem separately for each horizon and posterior quantile determining the error band.

**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:

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

where $\hat{r}_t$ is the interest rate, $\hat{\pi}_t$ is the inflation rate, $\Delta \hat{y}_t$ is the output gap, and $\varepsilon_t^r$ 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. Each has a range $[0, 1]$, with 0 corresponding to the flexible wage and price benchmarks.

Economic theory suggests that both are important for determining the response of output to monetary policy. The Taylor rule captures the persistence of the monetary policy shock, and how the monetary authority responds to dampen future deviations in inflation and the output gap arising from the initial monetary policy shock. Economic agents 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 adjust in response to monetary policy, and thus how much output responds in equilibrium. However, because the impulse response function is determined in equilibrium, it is hard to make precise analytical statements about the effect of changing any of these parameters on the impulse response.

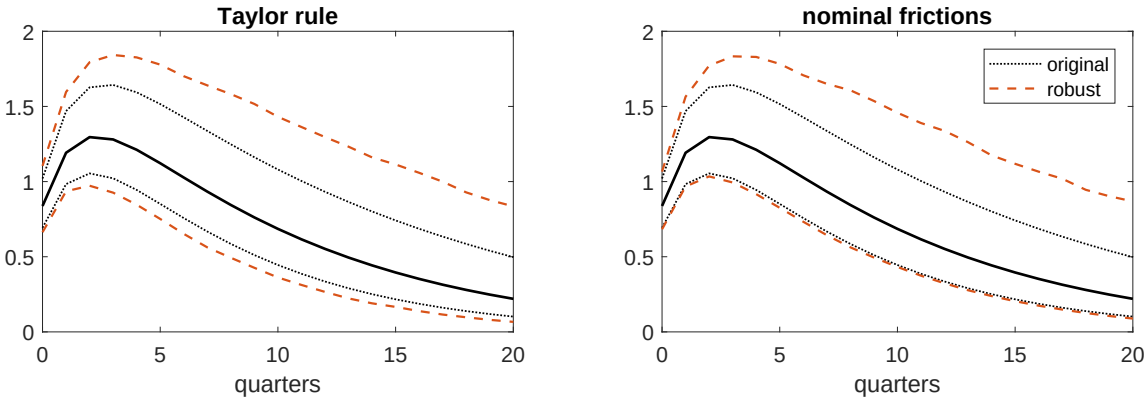


Figure 6.1: Robust 68% error bands. Black lines show original median (solid) and error bands (dotted); red dashed lines show robust error bands. **Left:** Taylor rule prior, with $r = 5 \times 10^{-3}$; **Right:** nominal frictions prior, with $r = 2.5 \times 10^{-4}$.

**Computational details.** 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 $R^2$s of between 0.97 to 0.98. Finally, the conditional probability that the impulse response is less than a given cutoff is estimated with a quadratic logit.[14] For the SMC, we use $5 \times 10^4$ particles, 250 SMC stages, and 10 Metropolis-Hastings draws at each stage. We average the results across 10 runs of the SMC. A single run of the SMC in MATLAB takes approximately three hours. We parallelize the computations by horizon across 21 cores. Appendix D provides further details, including a comparison of the exact and approximate posteriors.

## 6.3 Robust error bands

Figure 6.1 plots the original and robust 68% error bands for the impulse response. 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. As noted in Section 2.3 and 5.2, we could also have chosen a fixed relative entropy since the SMC produces the full mapping between the relative entropy and worst-case quantiles. We use the ability to move between relative entropy and worst-case quantile in order to fix the relative entropy across horizons in Figure 6.1.

Our analysis adds to the literature documenting the sensitivity of Bayesian DSGE estimates to the prior. 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,

[14] As discussed in Section 5, MATLAB has built-in commands for these approximations.

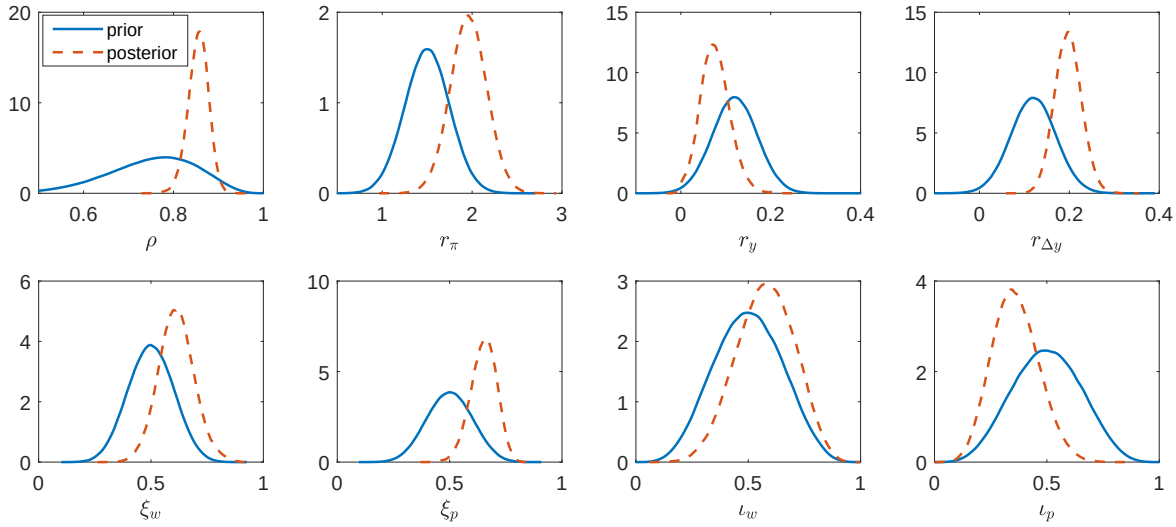


Figure 6.2: Original prior and posterior of Taylor rule parameters $\{\rho, r_\pi, r_y, r_{\Delta y}\}$ and nominal frictions parameters $\{\xi_w, \xi_p, \iota_w, \iota_p\}$.

quantile, or subset of parameters considered.

**Sensitivity.** We use $r = 5 \times 10^{-3}$ for the Taylor rule and $r = 2.5 \times 10^{-4}$ for the nominal frictions prior, where $r$ is defined in equation (4.4). These values of $r$ are one and two orders of magnitude smaller than the $r = 0.05$ benchmark respectively. In a Gaussian location model, $r = 5 \times 10^{-3}$ and $r = 2.5 \times 10^{-4}$ would respectively correspond asymptotically to 0.14 and 0.01 posterior standard deviation changes in the quantile. This is at least an order of magnitude smaller than the differences in the upper bound for most horizons, but similar to the differences in the lower bound shown in Figure 6.1.

The first reason for the sensitivity is that the likelihoods for both the Taylor rule and the nominal frictions parameters are dispersed, allowing a small change in the prior to generate a large change in the posterior of the parameters. Figure 6.2 shows that their marginal posteriors are not much more concentrated than their priors. To study the joint concentration of the likelihood, we measure the dispersion of the likelihood relative to the prior using the statistic $\sqrt{|\Sigma_\ell| / |\Sigma_\pi|}$, where $\Sigma_\ell$ is defined in (4.3). 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 $X \sim \mathcal{N}(\theta, \omega^2 I)$ with $\omega = 0.69$ and $\omega = 1.11$, respectively. With such a dispersed likelihood, the asymptotics in Lemma 1 overpredict the relative entropy needed to change the posterior estimates. In particular, in Figure C.1, the relative entropy with $T = 1$ is an order of magnitude smaller than what is predicted by the log-log trend for $T > 10$. The greater dispersion in the marginal likelihood of the nominal frictions parameters accounts

for the greater sensitivity to their prior relative to the prior on the Taylor rule parameters.

In addition, 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 quadratic regressions of the impulse response 20 quarters after impact on the Taylor rule parameters and nominal frictions parameters yield $R^2$s of 0.82 and 0.48, respectively. As a result, a given change in the posterior for either set of parameters shifts the posterior for the impulse response substantially.

**Asymmetry.** The sensitivity to the prior is uniform across neither horizons nor bounds. For both sets of parameters, the upper bound is substantially more sensitive to changes in the prior than the lower bound. In other words, given the model and data, a policymaker should be more concerned about underestimating rather than overestimating the 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. 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.

The robust and original error bands diverge 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 for short horizons.[15]

## 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 posterior standard deviation and increase the upper bound by one posterior standard deviation, respectively. For the nominal frictions prior, we consider the worst-case distortions that decrease the lower bound by 1/20 posterior standard deviation and increase the upper bound by one posterior standard deviation, respectively. These deviations correspond approximately to the robust error bands in Figure 6.1.

We summarize the distortions by the changes in the prior and posterior means (normalized

[15]The impulse response at short horizons are also less well-predicted by the parameters of interest. For example, quadratic regressions of the impulse response on impact on the Taylor rule and nominal frictions parameters respectively yield $R^2$s of 0.27 and 0.05, which are substantially smaller than those for the impulse response 20 quarters after impact.

| Parameter | | Lower bound $\frac{\mathbb{E}_{\tilde{\pi}}[\cdot]-\mathbb{E}_{\pi}[\cdot]}{\sigma_{\pi}[\cdot]}$ | Lower bound $\frac{\mathbb{E}_{\tilde{p}}[\cdot]-\mathbb{E}_{p}[\cdot]}{\sigma_{p}[\cdot]}$ | Upper bound $\frac{\mathbb{E}_{\tilde{\pi}}[\cdot]-\mathbb{E}_{\pi}[\cdot]}{\sigma_{\pi}[\cdot]}$ | Upper bound $\frac{\mathbb{E}_{\tilde{p}}[\cdot]-\mathbb{E}_{p}[\cdot]}{\sigma_{p}[\cdot]}$ |
|---|---|---|---|---|---|
| **Taylor rule** | | | | | |
| $\rho$ | persistence | 0.007 | –0.165 | –0.003 | 0.125 |
| $r_{\pi}$ | 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** | | | | | |
| $\xi_{w}$ | wage rigidity | –0.009 | –0.162 | 0.006 | 0.517 |
| $\xi_{p}$ | price rigidity | –0.040 | –0.076 | –0.000 | –0.041 |
| $\iota_{w}$ | wage indexation | –0.001 | –0.048 | –0.002 | 0.037 |
| $\iota_{p}$ | 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.

by the respective standard deviations) in Table 6.1. The prior means change less than the posterior means 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 prior in different ways.

Among the nominal rigidity parameters, we find that the posterior is especially dependent on the wage rigidity parameter, as the upper bound worst-case prior is heavily distorted in the direction that changes the corresponding posterior mean. In contrast, the price rigidity and wage indexation priors do not appear as important for the impulse response. For the Taylor rule parameters, the prior on the inflation coefficient is especially important for the lower bound of the error band, but the worst-case prior for the upper bound does not distort especially strongly in the direction of any one of the parameters.

In what follows, we show that the worst-case distortions reveal information about the

model and likelihood that is difficult to uncover using existing approaches. The methods used to understand the results are useful but heuristic instruments to analyze the results ex-post, and do not inform the researcher ex-ante about which parts of the prior are important for the posterior estimates. Much of the analysis conditions on the size of the impulse response. 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 (2.13) 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 (2.13) 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. In order to shift the lower (upper) bound of the error band, we need to shift mass to the left (right) tail of the distribution of the impulse response. We thus restrict the regressions to draws for which the impulse response in one standard deviation below (above) the mean to understand the lower (upper) bound of the impulse response. 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 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. 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 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 $r_\pi$ or $r_y$, and increases in response to an increase in $\rho$ 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 is stronger when the Taylor rule is more persistent or less responsive to changes in the output gap and inflation.

| Parameter | | Lower bound | Upper bound |
|---|---|---|---|
| **Taylor rule** | | | |
| $\rho$ | persistence | 0.088 | 0.232 |
| $r_\pi$ | 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** | | | |
| $\xi_w$ | wage rigidity | 0.049 | 0.260 |
| $\xi_p$ | price rigidity | 0.019 | –0.060 |
| $\iota_w$ | wage indexation | 0.016 | 0.053 |
| $\iota_p$ | 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.

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

We begin by comparing the original and worst-case marginal posteriors for $r_\pi$, 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 $r_\pi$. The observation corroborates results from Herbst and Schorfheide (2014), who find that the posterior mean of $r_\pi$ moves from 2.04 to 2.78 when one replaces the prior from Smets and Wouters (2007) with a more diffuse one.[16] The prior from Smets and Wouters (2007) shrinks toward smaller values of $r_\pi$, making it difficult to detect the possibility of an additional mode without reestimating the model with a new prior. REPS detects that such shrinkage is important for the posterior outcomes of interest relative to other features of the prior. On the other hand, the REPS

[16]The additional mode in the likelihood arises from fitting the data to the Taylor rule, as evidenced by the large inflation coefficient of 2.59 when we use data for the federal funds rate, inflation, and output gap to estimate the Taylor rule using linear regression (ignoring autocorrelation in the monetary policy shock). This large value arises due to the low-frequency variation in the data. We run the regression using the trend and cyclical components of HP-filtered data and find coefficients of 2.66 and 0.92, respectively. Sala (2015) estimates a similar DSGE model in the frequency domain and finds a posterior estimates of 1.81 and 1.12 for the low-frequency and high-frequency components, respectively.

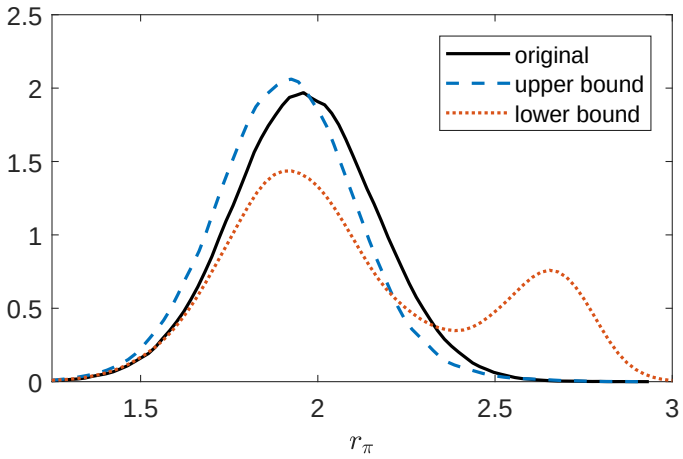


Figure 6.3: Original and worst-case posteriors of $r_\pi$. **Black solid line:** original posterior; **Blue dashed line:** worst-case posterior for upper bound; **Red dotted line:** worst-case posterior for lower bound.

analysis also eases concerns from the results of Herbst and Schorfheide (2014) by showing that a prior favoring larger values of $r_\pi$ does not substantially change the posterior of the impulse response, as seen from the relatively narrow gap between the lower bounds of the robust and original error bands in Figure 6.1.

The likelihood reveals two reasons for the small distortions in $\rho$. Firstly, Figure 6.2 suggests that $\rho$ is relatively sharply identified by the likelihood—the ratio of posterior to prior standard deviation for $\rho$ is smaller than that of $r_\pi$, $r_y$, and $r_{\Delta y}$, with a value of 0.23 as compared to 0.82, 0.66, and 0.60, respectively. As a consequence, larger changes in the prior are needed to produce the same change in the posterior or $\rho$, resulting in greater relative entropy cost. In addition, the correlations of $\rho$ with $r_\pi$, $r_y$, and $r_{\Delta y}$ are in conflict with their effect on the impulse response. In particular, the regression results suggest that $\rho$ should be distorted in the opposite direction from $r_\pi$ and $r_y$, but in the same direction as $r_{\Delta y}$. Such distortions are costly in terms of relative entropy as they run against the likelihood, as $\rho$ has a positive correlation with $r_\pi$ and $r_y$ of 0.18 and 0.33, respectively, and a negative correlation of $-0.10$ with $r_{\Delta y}$. It is thus optimal to distort $\rho$ less than other parameters. In general, one may understate the dependence of the posterior on the prior if one considers the effects of the parameters on the object of interest but not the likelihood.

The likelihood also supports the large distortions in $r_y$ and $r_{\Delta y}$ for the upper bound relative to the lower bound. Figure 6.2 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 $\xi_w$ is distorted relatively more, especially for the upper bound. Next, the posterior means for price rigidity $\xi_p$ and wage indexation $\iota_w$ move in contradictory directions when we go from the lower bound to the upper bound. Finally, the worst-case posterior mean for price indexation $\iota_p$ 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 $\xi_w$, which partly explains why the large distortion for $\xi_w$. Moreover, the regression coefficient on $\xi_p$ is negative in the regression for the upper bound, rationalizing the counterintuitive direction of the distortion of the prior on $\xi_p$. The negative coefficient arises because of an omitted variable bias—$\xi_p$ 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 keep the prior of all other parameters unchanged. REPS accounts for the fact that changing the prior for $\xi_p$ changes the posterior of the impulse response through both the marginal effect of $\xi_p$ and the effect of any other correlated parameters.

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

**Likelihood.** The worst-case posteriors for $\xi_w$, shown in Figure 6.4, provide an explanation for the especially large change in the posterior mean of $\xi_w$ for the upper bound. In particular, the worst-case posterior for the upper bound is bimodal, with a new mode around $\xi_w = 0.90$.

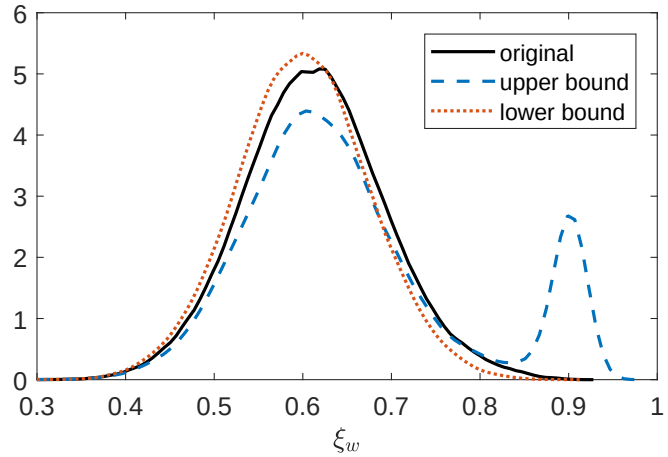


Figure 6.4: Original and worst-case posteriors for $\xi_w$. **Black solid line:** original posterior; **Blue dashed line:** worst-case posterior for upper bound; **Red dotted line:** worst-case posterior for lower bound.

**Lower bound**

| | $\xi_w$ | $\xi_p$ | $\iota_w$ | $\iota_p$ |
|---|---|---|---|---|
| $\xi_w$ | 1 | | | |
| $\xi_p$ | 0.19 | 1 | | |
| $\iota_w$ | –0.13 | –0.23 | 1 | |
| $\iota_p$ | –0.07 | –0.23 | –0.15 | 1 |

**Upper bound**

| | $\xi_w$ | $\xi_p$ | $\iota_w$ | $\iota_p$ |
|---|---|---|---|---|
| $\xi_w$ | 1 | | | |
| $\xi_p$ | –0.20 | 1 | | |
| $\iota_w$ | –0.05 | –0.31 | 1 | |
| $\iota_p$ | –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.

As with the lower bound worst-case posterior of $r_\pi$, this is in line with the diffuse prior estimates of Herbst and Schorfheide (2014). In particular, the posterior mean of $\xi_w$ 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 $\xi_p$, $\iota_w$, and $\iota_p$, 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, REPS accounts for peaks in the likelihood that are hard to detect without reestimating the model under the appropriate prior. Unlike the additional mode for $r_\pi$, this new mode in the posterior for $\xi_w$ substantially shifts the error bands for the impulse response. Indeed, the regression coefficient for $\xi_w$ in Table 6.2 is larger in magnitude than that for $r_\pi$.

The posterior correlations, reported in Table 6.3, help to account for the counterintuitive distortions in price indexation $\iota_p$. The negative correlation of $\iota_p$ with $\xi_w$, $\xi_p$, and $\iota_w$ implies that increases in these parameters correspond on average to a decrease in $\iota_p$. Hence the likelihood favors moving $\iota_p$ in the opposite direction from $\xi_w$, $\xi_p$ and $\iota_w$. In addition, Figure 6.5 shows that under the worst-case posterior for the upper bound, the new mode for $\xi_w$ corresponds to small values of $\iota_p$, decreasing the posterior mean of $\iota_p$. More generally, these

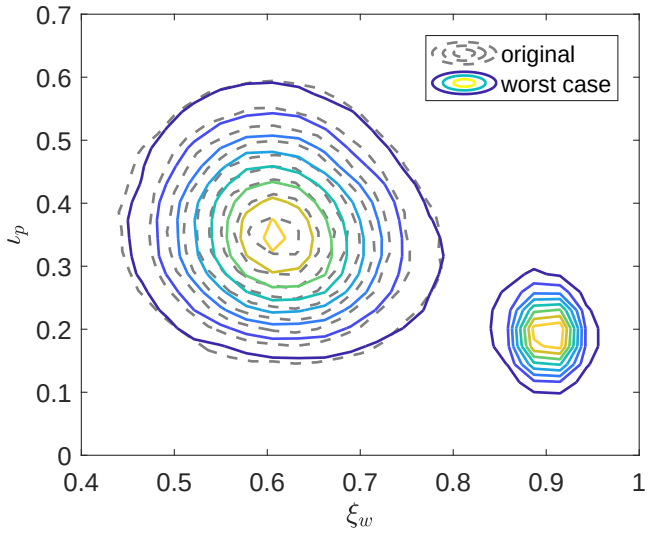


Figure 6.5: Original and lower bound worst-case posteriors for $(\xi_w, \iota_p)$. **Gray dashed lines:** original posterior; **Colored solid lines:** worst-case posterior.

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 without accounting for correlations across parameters, we may understate the sensitivity of the posterior.

The posterior correlations for $\xi_p$ further emphasize this point and provide an additional explanation for the inconsistent direction of distortions for the two worst-case posteriors of $\xi_p$. The rigidity parameters $\xi_w$ and $\xi_p$ 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 $\xi_p$ in the same direction as $\xi_w$ for the lower bound, and in the opposite direction from $\xi_w$ for the upper bound. Given the distortions for $\xi_w$, it is therefore optimal to decrease $\xi_p$ for both the upper and lower bound. In addition, $\xi_p$ is negatively correlated with $\iota_w$ for both the lower 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 $\xi_p$ and $\iota_w$ are in opposite directions.

## 6.5 Comparison to local methods

The wide robust error bands stand in contrast to Müller (2012), who finds that the impulse responses are relatively insensitive to the priors for the structural parameters. The difference arises because the REPS analysis allows for changes in the prior that are not considered by Müller (2012). The worst-case priors reveal the importance of the correlations and tail behavior of the prior, neither of which are accounted for by the perturbations considered

by Müller (2012). These features arise from assumptions made for convenience, making it crucial to understand how they matter for the posterior.

Local methods use derivatives that also fail to capture the asymmetry in the sensitivity to the prior. Figure 6.1 shows that prior sensitivity depends on the direction in which we wish to change the impulse response. Such non-linearity in the prior sensitivity should be expected more broadly, given the irregularity of the likelihood and the complicated mapping from the parameters to the function of interest $\psi$ in many applications.

# 7 Conclusion

To understand how the data inform the posterior estimates, one needs to disentangle the roles of the prior and the likelihood. Despite numerous assumptions made when writing down priors and frequent disagreement over what priors to use, such analysis is often either absent or ad hoc. Reestimating the model for a full set of alternative priors is often computationally demanding and even infeasible. Nevertheless, few prior sensitivity tools have been developed for broad economic applications.

REPS 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. REPS reduces the problem of checking across an infinite dimensional set of priors to that of solving for one unknown from one equation, allowing us to complete the computations in a similar amount of time it would have taken to estimate the model once.

The New Keynesian model of Smets and Wouters (2007) provides a laboratory to show how REPS can reveal properties of the posterior that are sensitive to changes in the prior and parts of the prior are important for these properties. 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. In parallel work, Ho (2020) uses REPS to uncover the data required to identify the effects of demographic changes on long-run interest rates. There is much future work to be done applying REPS to a wider range of applications, both as a robustness check and as a tool to understand how the data inform our estimates.

# References


Abdulkadiroglu, A., N. Agarwal, and P. Pathak (2017). The Welfare Effects of Coordinated School Assignment: Evidence from the New York City High School Match. *American Economic Review 107*(12), 3635–3689.

Avery, C. N., M. E. Glickman, C. M. Hoxby, and A. Metrick (2013). 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 $\epsilon$-Contaminated Priors. *The Annals of Statistics 14*(2), 461–486.

Berger, J. O., E. Moreno, L. R. Pericchi, M. J. Bayarri, J. M. Bernardo, J. A. Cano, J. De la Horra, J. Martín, D. Ríos-Insúa, B. Betrò, et al. (1994). An Overview of Robust Bayesian Analysis. *Test 3*(1), 5–124.

Bidder, R., R. Giacomini, and A. McKenna (2016). Stress Testing with Misspecified Models. *Federal Reserve Bank of San Francisco, Working Paper Series 2016-26*.

Cai, M., M. Del Negro, E. Herbst, E. Matlin, R. Sarfati, and F. Schorfheide (2019). Online Estimation of DSGE Models. *FRB of New York Staff Report* (893).

Canova, F. and L. Sala (2009). Back to Square One: Identification Issues in DSGE Models. *Journal of Monetary Economics 56*(4), 431–449.

Chamberlain, G. and E. E. Leamer (1976). Matrix Weighted Averages and Posterior Bounds. *Journal of the Royal Statistical Society: Series B (Methodological) 38*(1), 73–84.

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. In *Handbook of Macroeconomics* (1 ed.), Volume 2, pp. 527–724. Elsevier B.V.

Giacomini, R. and T. Kitagawa (2018). Robust Bayesian Inference for Set-Identified Models. *Cemmap Working Paper*.

Giacomini, R., T. Kitagawa, and M. Read (2019). Robust Bayesian Inference in Proxy SVARs. *Working paper*.

Giacomini, R., T. Kitagawa, and H. Uhlig (2019). Estimation under Ambiguity. *Working paper*.

Giannone, D., M. Lenza, and G. E. Primiceri (2018). Priors for the Long Run. *Journal of the American Statistical Association*.

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: Papers & Proceedings 91*(2).

Hansen, L. P. and T. J. Sargent (2007). Recursive Robust Estimation and Control Without Commitment. *Journal of Economic Theory 136*(1), 1–27.

Hansen, L. P. and T. J. Sargent (2008). *Robustness*. Princeton University Press.

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

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

Ho, P. (2020). Estimating the Effects of Demographics on Interest Rates: A Robust Bayesian Perspective. *Working paper*.

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 and Economic Statistics 31*(3), 300–314.

Leamer, E. E. (1982). Sets of Posterior Means with Bounded Variance Priors. *Econometrica: Journal of the Econometric Society*, 725–736.

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., 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. *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 (2.3)-(2.4) and (2.5)-(2.6) both have the form (2.7).*

*Proof.* Denote the marginal data density under prior $\pi$ 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 (2.5)-(2.6). Attaching the multiplier $\lambda \mathbb{E}_p[M] \zeta$ to (2.6) and the multiplier $\mu$ 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 (2.7).

Now consider the primal problem (2.3)-(2.4). Attaching the multiplier $1/(\lambda \mathbb{E}_p[M] \zeta)$ to (2.4) 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 (2.7) once we replace $\tilde{\gamma}$ with the full expression for the worst-case posterior mean of $\gamma$. □

### A.2 Subspaces

**Lemma 3.** *The solution for $M$ is problem (2.11)-(2.12) is (2.13).*

*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 (2.11)-(2.12) 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 (2.13). □

## 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 (2.5)-(2.6) 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 $\mu$ 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 (2.5)-(2.6) is:*

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

*Proof.* Define $\kappa \equiv 1/\mathbb{E}_\pi \left[ \exp \left[ \lambda L(\theta|X) \left( \gamma(\theta) - \tilde{\gamma} \right) \right] \right]$. Taking logs of the solution (2.7) yields:

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

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

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

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 (2.6) and the fact that $\mathbb{E}_p \left[ \frac{M(\theta)}{\mathbb{E}_p[M]} \gamma(\theta) \right] = \int \gamma(\theta) \, \tilde{p}(\theta|X) \, d\theta$. □

## 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 $\Omega$ diagonal, which implies $\Sigma_{\ell,T}$ is diagonal. Define $\Delta(\theta) \equiv \theta - \theta_0$. Abusing notation, we can write the likelihood as a function of $\Delta$:

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

Abusing notation again, write (2.7) as a function of $\Delta$:

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

As $T \to \infty$, since $\frac{T\sigma_{1,p,T}}{\omega_1} \to 1$, there exists $\lambda^*$, $M^*$ 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 (2.6) 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

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

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

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

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

When $\Omega$ is not diagonal, we first note that we can decompose the likelihood of $\theta$ into the marginal likelihood of $\theta_1$ 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 $\Omega$. Finally, the proof follows through with general $\overline{X}_T$ since $\overline{X}_T \xrightarrow{\text{a.s.}} \theta_0$. □

# B Sequential Monte Carlo

## B.1 Implementation details

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

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

A smaller value of $\nu$ corresponds to larger initial steps, and smaller steps toward the end of the SMC algorithm.[17] Substituting $\tilde{\gamma}_i$ into the right-hand side of (2.6) for each $i$ yields a

[17] Cai et al. (2019) propose an adaptive algorithm to select the step sizes.

sequence of distortions:

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

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 $\pi_i/\pi_{i-1}$ and $p_i/p_{i-1}$, both of which are proportional to:

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

Given $\lambda_{i-1}$ and draws from $\pi_{i-1}$ and $p_{i-1}$, the only unknown remaining is $\lambda_i$, which we solve for from (2.7), which we rewrite as:

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

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

**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 $N_{\text{P}}$, the number of Metropolis-Hastings mutation steps $N_{\text{MH}}$, and the number of SMC steps $N_{\text{SMC}}$. 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 $\lambda_i$. Similarly, $N_{\text{MH}}$ must be sufficiently large in order to solve for $\lambda_i$ accurately. If $\lambda_i$ is computed accurately, the posterior mean of $\gamma$ evaluated from the particles before and after the mutation step should be identical up to sampling error. We check if $N_{\text{SMC}}$ 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

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

instead of (2.13). 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

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

where we can solve for $\lambda^*$ using the constraint (2.12). Once we solve for $\lambda^*$, 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:

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

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

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

from which we obtain:

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

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

# C Gaussian location model finite sample performance

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 $\theta_1$ by one posterior standard deviation; or
2. increase the 84% quantile of $\theta_1$ by one posterior standard deviation,

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

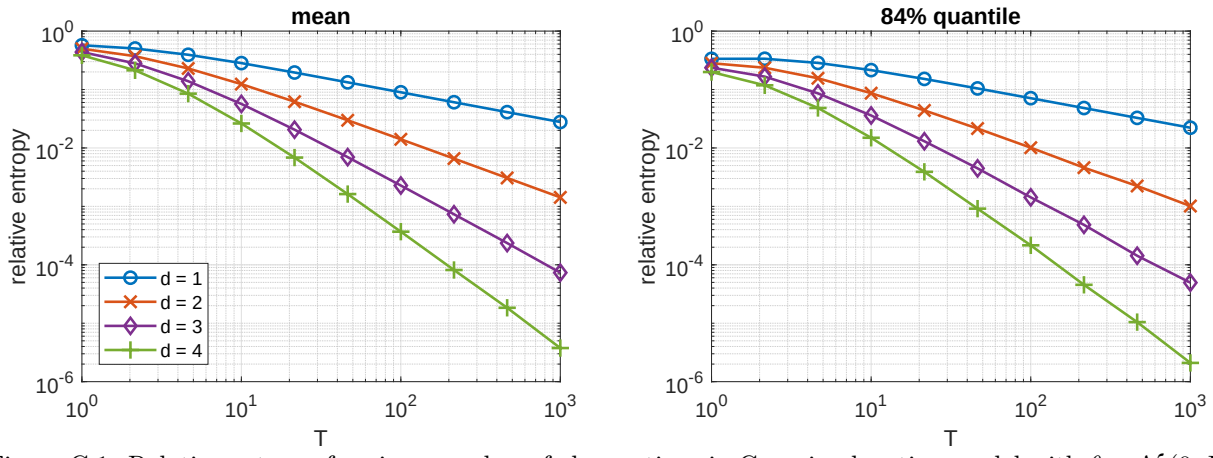


Figure C.1: Relative entropy for given number of observations in Gaussian location model with $\theta \sim \mathcal{N}(0, I)$, $\overline{X}_T = 0$, $d \in \{1, ..., 4\}$. **Left:** increase in posterior mean of $\theta_1$ by one posterior standard deviation; **Right:** increase in 84% quantile of $\theta_1$ by one posterior standard deviation.

All calculations for the Gaussian location model for dimensionality $d = 1$ and $d = 2$ are done using grids. For $d = 1$, the grid has range $[-8, 8]$ and has $10^5 + 1$ 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 $(N_{\mathrm{P}}, N_{\mathrm{MH}}, N_{\mathrm{SMC}}) = (d \times 10^5, 10, 100)$, and average across 25 runs.

**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 C.1 shows the relative entropy for different values of $T$ 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 have 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 C.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 $\overline{X}_T$.

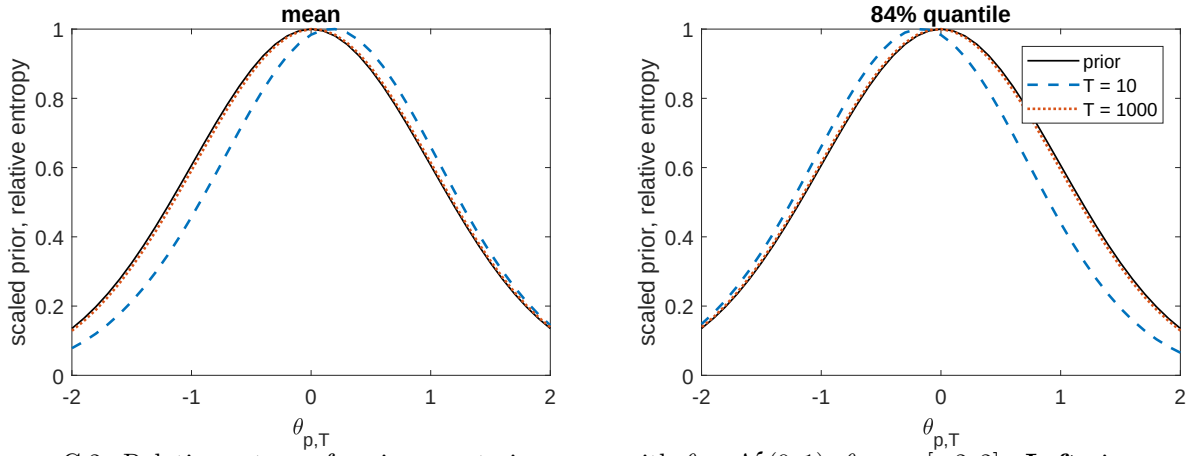


Figure C.2: Relative entropy for given posterior mean with $\theta \sim \mathcal{N}(0,1)$, $\theta_{p,T} \in [-2,2]$. **Left:** increase in posterior mean of $\theta_1$ by one posterior standard deviation; **Right:** increase in 84% quantile of $\theta_1$ by one posterior standard deviation.

# 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 $\theta$:

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

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.

## 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 $q$th quantile of $\psi_h$ under the

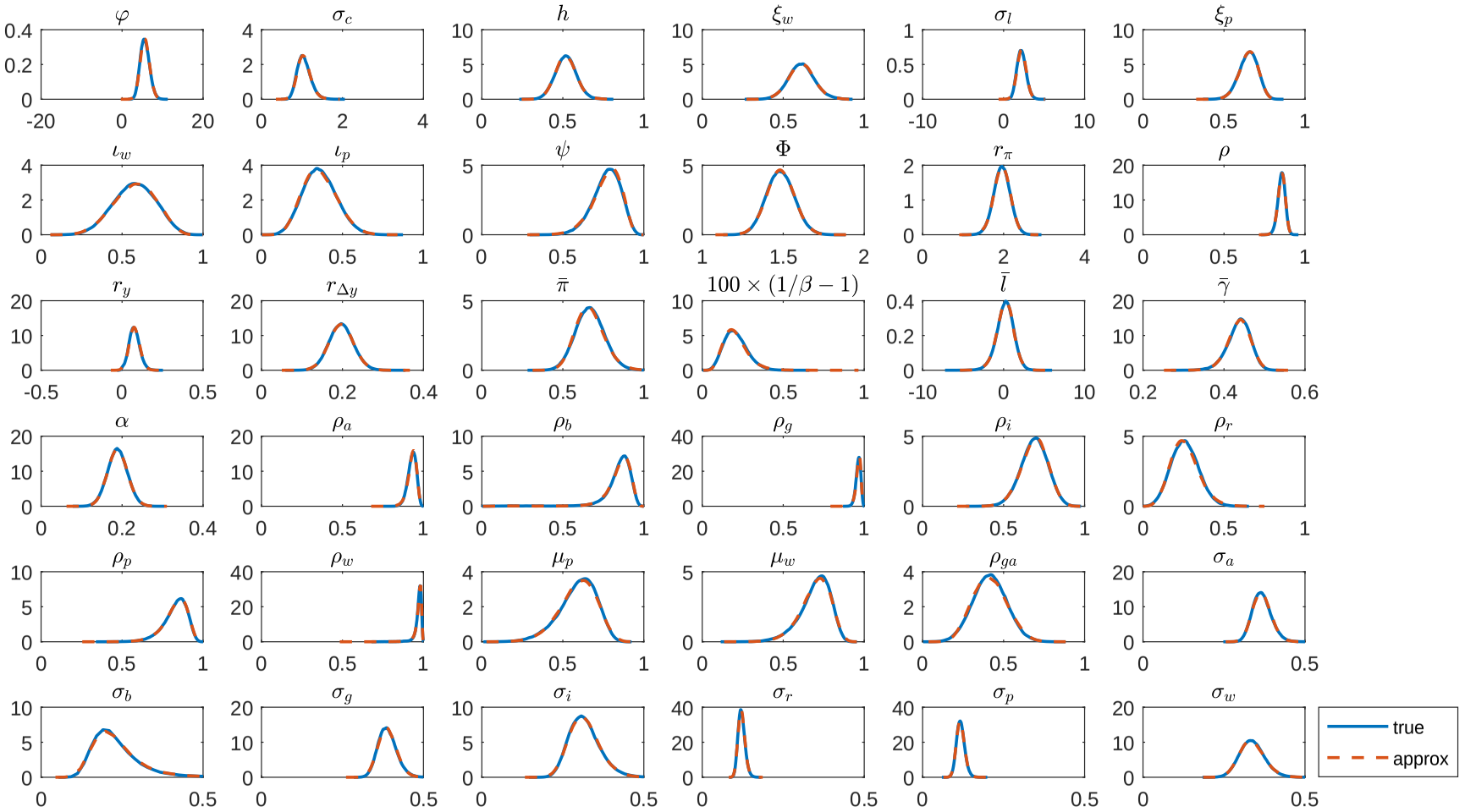


Figure D.1: Original and approximate marginal posteriors. Blue solid line corresponds to true posterior; red dashed line corresponds to Gaussian mixture approximation.

distribution $f$, and let $\sigma_f[\cdot]$ denote the standard deviation under distribution $f$. We set the worst-case $q$th quantile $\tilde{\psi}_h^q$ for horizon $h$ as follows. For the Taylor rule prior, for the lower bound we choose:

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

and for the upper bound we choose

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

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

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

and for the upper bound we choose

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

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):

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

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.

## 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 do not find substantive differences.