From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2019

Essays on Uncertainty

Ho, Paul Cher Wei

Princeton, NJ : Princeton University

http://arks.princeton.edu/ark:/88435/dsp01bg257h945

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS ON UNCERTAINTY

PAUL CHER WEI HO

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF ECONOMICS

ADVISOR: MARK WATSON

JUNE 2019

© Copyright by Paul Cher Wei Ho, 2019.

All rights reserved.

Abstract

This dissertation studies the measurement of uncertainty.

Chapter 1 develops a tool for global prior sensitivity analysis in large Bayesian models, which measures how informative data are about particular features of a model. Without imposing parametric restrictions, the methodology provides bounds for posterior statistics 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 illustrate the methodology using a canonical New Keynesian model, and show that the error bands for the impulse response of output to a monetary policy shock depend asymmetrically on the prior through features of the likelihood that are hard to account for with existing approaches.

Chapter 2 uses the tool developed in Chapter 1 to show that data on capital and savings are important for quantifying the effects of an aging population on interest rates. Using macroeconomic and demographic data from Japan, we apply Bayesian methods to estimate the effects of demographics in a parsimonious overlapping generations model. The prior sensitivity analysis reveals that without data on capital and savings over the life cycle, the discount rate, intertemporal elasticity of substitution, and capital depreciation rate are not well-identified, and the posterior for the effects of demographics is heavily dependent on the prior for these parameters. Quantitative exercises that ignore these data thus yield potentially inaccurate estimates.

Chapter 3 disentangles different sources of volatility using a model where levels and volatilties each follow factor structures, applied to a large panel of macroeconomic indicators. Principal component analysis provides an overview of the data, which informs model specification and priors for the Bayesian estimation. We identify the sources of volatility measured by existing uncertainty indices and show how different sources of volatility interact with macroeconomic levels in ways neglected by existing uncertainty indices.

Acknowledgments

I am deeply grateful to Mark Watson for his guidance and encouragement, and am indebted to Jaroslav Borovička, Nobuhiro Kiyotaki, Ulrich Müller, Mikkel Plagborg-Møller and Chris Sims for their invaluable advice. This dissertation has also benefited greatly from conversations with many others, especially SeHyoun Ahn, Timothy Christensen, Liyu Dou, Jesús Fernández-Villaverde, Federico Huneeus, Ezra Oberfield, Frank Schorfheide, Denis Tkachenko, and Kenneth West. In addition, I am indebted to conference and seminar participants too numerous to list for their comments and suggestions. Finally, I thank the National University of Singapore for hosting me for several periods during which a significant part of the research here was completed.

The first chapter of this dissertation received generous financial support from the Alfred P. Sloan Foundation, the CME Group Foundation, Fidelity Management & Research, the Macro Financial Modeling Initiative, the International Association for Applied Econometrics, and the Gregory C. Chow Econometric Research Program.

Contents

Abstract iii

Acknowledgments iv

1 Global Robust Bayesian Analysis in Large Models 1

1.1 Introduction 1

1.2 Relative entropy prior sensitivity 7

1.2.1 Setting and notation 7

1.2.2 Setup 8

1.2.3 Solution 10

1.2.4 Extensions 12

1.3 Two illustrative examples 14

1.3.1 Example 1: multiple modes in the likelihood 14

1.3.2 Example 2: log-normal distribution 17

1.4 Quantifying the change in prior 19

1.4.1 Intuition 19

1.4.2 Gaussian location model 20

1.4.3 Rule of thumb for $\mathcal{R}$ 21

1.5 Implementation 22

1.5.1 Importance sampling 22

1.5.2 Sequential Monte Carlo 23

1.5.3 Approximate REPS 25

1.6 Application: Smets and Wouters (2007) 27
1.6.1 Model and estimation 28
1.6.2 Prior sensitivity 28
1.6.3 Robust error bands 30
1.6.4 Worst-case distributions 33
1.6.5 Comparison to local methods 42
1.7 Conclusion 43
References 44
1.A Proofs 48
1.A.1 Solution to primal and dual problems 48
1.A.2 Subspaces 49
1.A.3 Additional constraints 50
1.A.4 Evaluating relative entropy 51
1.A.5 Asymptotics 51
1.B Sequential Monte Carlo 53
1.B.1 Implementation details 53
1.B.2 Evaluating relative entropy 55
1.C Gaussian location model finite sample performance 56
1.D Smets and Wouters (2007) 58
1.D.1 Gaussian mixture approximation of posterior 58
1.D.2 Sequential Monte Carlo 58
1.D.3 Robust error bands 61

2 Estimating the Effects of Demographics on Interest Rates: A Robust Bayesian Perspective **62**

2.1 Introduction 62
2.2 Econometric framework 67
2.2.1 Setup 68

2.2.2 Relation to other approaches 69
2.3 Overlapping generations model 70
2.3.1 Households 71
2.3.2 Production and market clearing 74
2.3.3 Structural break 74
2.3.4 Demographics and interest rates 75
2.4 Bayesian estimation 78
2.4.1 Estimation 78
2.4.2 Results 81
2.5 Prior sensitivity 87
2.5.1 Methodology 88
2.5.2 Results 90
2.6 Implications for estimation and calibration 96
2.6.1 Including data on capital-output ratio 97
2.7 An alternative parametric prior 105
2.7.1 Prior 105
2.7.2 Results 109
2.8 Conclusion 111
References 112
2.A Steady state of overlapping generations model 117
2.A.1 Production 117
2.A.2 Households 118
2.B Bayesian estimation 121
2.B.1 Data 121
2.B.2 Prior 122
2.B.3 Markov chain Monte Carlo 125
2.C Prior sensitivity 126

3 Factor Volatilities and Volatility Factors **127**

3.1 Introduction . . . 127

3.1.1 Related literature . . . 129

3.1.2 Outline . . . 131

3.2 Setup . . . 131

3.2.1 Model . . . 131

3.2.2 Decomposing volatility . . . 133

3.2.3 Estimation . . . 135

3.3 Principal component analysis . . . 136

3.3.1 Estimation . . . 136

3.3.2 Results . . . 138

3.4 Bayesian estimation . . . 143

3.4.1 Functional forms . . . 143

3.4.2 Prior . . . 145

3.4.3 Estimation . . . 148

3.4.4 Results . . . 148

3.5 Existing uncertainty indices . . . 155

3.5.1 Methodology . . . 155

3.5.2 Results . . . 156

3.6 Volatility and the Macroeconomy . . . 161

3.6.1 Setup . . . 161

3.6.2 Results . . . 163

3.7 Conclusion . . . 165

References . . . 165

3.A Prior . . . 169

3.B Gibbs Sampler . . . 170

3.B.1 Mixture-of-normals approximation . . . 170

3.B.2 Gibbs sampler 171

Chapter 1

Global Robust Bayesian Analysis in Large Models

1.1 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 simple low-dimensional settings.

A previous version of this paper was presented at the North American Summer Meeting of the Econometric Society and the International Association for Applied Econometrics Annual Conference.

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 particular features of the prior that are important for these results. For example, if the econometrician reports the posterior mean of an elasticity, she would search for the priors that respectively minimize and maximize this mean for a given relative entropy, thus obtaining bounds on the posterior mean. The worst-case prior will differ most from the original prior in dimensions that are most important for the posterior mean. These are parts of parameter space that are not well-identified by the likelihood but matter for the posterior mean. 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 different the original and worst-case priors need to be to generate a given change in the posterior estimate, and comparing the sensitivity to a Gaussian location model with a large number of observations.

Our main application is the impulse response of output to a monetary policy shock in the workhorse 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 compare the bounds when we distort the prior for the Taylor rule parameters to when we distort the prior for the nominal frictions parameters. The impulse response is more sensitive to the prior of the nominal frictions parameters because the likelihood for these parameters is less concentrated. For both sets of parameters, the upper bound of the error bands is

more sensitive to changes to the prior and the impulse response is more dependent on the prior at long horizons.

REPS detects dependence on the prior in the Smets and Wouters (2007) estimation that would be hard to discern with existing approaches to prior sensitivity. 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. On the other hand, 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. Local methods (e.g. Gustafson (2000); Müller (2012)) consider derivatives of specific posterior quantities with respect to the prior. These menthods not only focus on infinitesimal changes in the prior and posterior, but are also often restricted to parametric changes in the prior. As we show in both our main application and stylized examples, both of these restrictions can lead to 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 simple 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.

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 et al. (2016) construct a relative entropy ball around the prior for a partially identified parameter. Because the parameter is not identified without further restrictions, this reduces to the usual robust control problem of Hansen and Sargent (2001) conditional on the identified parameters in the model. This paper develops a more general methodology that applies to any Bayesian estimation problem, in contrast to Giacomini et al. (2016) who focus on partial identification problems.

The importance of prior sensitivity analysis is especially salient in Bayesian 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 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 (2018) applies the REPS methodology to an overlapping generations model, revealing the importance of capital and household savings 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 and Kitagawa (2018) show that priors are especially 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 1.2 and demonstrate the methodology using two stylized examples in Section 1.3. Section 1.4 derives the rule of thumb to quantify the difference between priors. I discuss the numerical implementation in Section 1.5. In Section 1.6, we turn to the DSGE model of Smets and Wouters (2007) as our main application. Section 1.7 concludes.

1.2 Relative entropy prior sensitivity

1.2.1 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:

$$p(\theta|X) \propto \pi(\theta) L(\theta|X). \tag{1.1}$$

Suppose we are interested in the posterior of a function $\psi : \Theta \to \Psi$ of the parameter θ, where Ψ 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 $\mathbb{E}_f[\cdot]$. 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 ψ that we are interested in. For instance, if we set $\gamma_\psi(\psi) = \psi$, then $\mathbb{E}_p[\gamma_\psi(\psi)]$ is the posterior mean of ψ. Altenatively, if we take $\gamma_\psi(\psi) = \mathbf{1}\{\psi \leq \psi^*\}$, then $\mathbb{E}_p[\gamma_\psi(\psi)]$ is the cumulative distribution function of ψ evaluated at ψ^*. We denote $\gamma(\theta) \equiv \gamma_\psi(\psi(\theta))$, and study how $\mathbb{E}_p[\gamma(\theta)]$ depends on π.

Alternative priors and implied posteriors. To study the dependence of the posterior p on the prior π, 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 π, we can write:

$$\tilde{\pi}(\theta) \equiv M(\theta) \pi(\theta), \tag{1.2}$$

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

prior $\tilde{\pi}$ implies the distorted posterior:

$$\tilde{p}(\theta|X) = \frac{M(\theta)}{\mathbb{E}_p[M]} p(\theta|X). \tag{1.3}$$

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.

1.2.2 Setup

I now describe the optimization problem used to analyze the sensitivity of the posterior estimates to the prior. We search across a set of alternative priors, seeking the worst-case prior that yields the largest change in the posterior mean of the objective function γ. We restrict the set of priors to be close to the original prior in relative entropy. Comparing the change in the prior to the change in the posterior mean of γ tells us of how much the posterior mean of γ depends on the prior. Comparing the worst-case prior to the original prior indicates parts of the prior that are important for determining the posterior mean of γ.

Primal problem. Formally, we consider:

$$\min_{M(\theta):\mathbb{E}_\pi[M]=1} \mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\gamma(\theta)\right] \tag{1.4}$$

$$\text{s.t. } \mathbb{E}_\pi[M(\theta)\log M(\theta)] \leq \mathcal{R} \tag{1.5}$$

The minimization over M satisfying $\mathbb{E}_\pi[M] = 1$ is equivalent to minimizing over alternative priors, as the random variable M indexes the possible priors. We choose the prior that minimizes $\mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\gamma(\theta)\right]$, which is the distorted posterior mean of γ. Replacing the minimization operator with a maximization operator gives the upper bound for the posterior mean. The left-hand side of the constraint (1.5) is the relative

entropy or Kullback-Leibler divergence of the alternative prior relative to the original prior. $\mathcal{R}$ is a positive constant that provides an upper bound on the relative entropy, limiting us to priors that are statistically difficult to distinguish from the original prior, which typically contains useful information about the distribution of θ. As $\mathcal{R} \to 0$, the worst-case and original priors converge as we are restricted to choosing $M = 1$. Section 1.4 gives benchmarks for large and small values of $\mathcal{R}$.

There are several reasons for using relative entropy in (1.5) to constrain the set of priors. Firstly, there is theoretical justication for relative entropy. It measures the information that a Bayesian with the original prior π needs to gather to change her beliefs to the alternative prior $M\pi$ and is invariant to the parameterization of θ. Secondly, the use of relative entropy ensures that the solution to (1.4)-(1.5) is well-defined. In particular, relative entropy heavily penalizes alternative priors that have large probability mass in regions in the tail of the original prior, ruling out distortions where one shifts mass arbitrarily far out into the tail. Such distortions can lead to degenerate or trivial solutions under some other measures of statistical distance. Finally, the functional form for relative entropy delivers an analytic solution to (1.4)-(1.5) that allows REPS to maintain tractability in large models.

(1.4)-(1.5) 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 (1.4) and the relative entropy in (1.5) are taken under different probability measures. This difference in probability measures arises because we are interested in how ex-ante beliefs affect ex-post estimates. Since we wish to report the *posterior* estimate of $\mathbb{E}_p[\gamma(\theta)]$, our objective function is the distorted *posterior* mean of γ, which conditions on the observed data. However, we wish to consider small changes in the *prior*, and are thus led to restrict the relative entropy of the alternative *priors* with respect to the original prior.

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

$$\min_{M(\theta):\mathbb{E}_\pi[M]=1} \mathbb{E}_\pi \left[M(\theta) \log M(\theta) \right] \tag{1.6}$$

$$\text{s.t. } \mathbb{E}_p \left[\frac{M(\theta)}{\mathbb{E}_p[M]} \gamma(\theta) \right] = \tilde{\gamma} \tag{1.7}$$

We are now searching across priors that imply that γ has posterior mean $\tilde{\gamma}$, and picking the one that is closest to the original prior in terms of relative entropy[1]. We justify the formulation (1.6)-(1.7) below and argue that it simplifies the solution.

1.2.3 Solution

The solution to (1.6)-(1.7) has the form:

$$M(\theta) \propto \exp\left[\lambda L(\theta|X) \left(\gamma(\theta) - \tilde{\gamma} \right) \right], \tag{1.8}$$

where $\lambda \in \mathbb{R}$ is a constant to be solved for from the constraint (1.7). We have therefore reduced the minimization over a nonparametric set of priors to a problem with one equation and one unknown, regardless of the dimensionality of θ. This is the key feature that makes REPS feasible in large models.

The distortion M depends on the parameter θ through the objective function γ and likelihood L. The worst-case distortion M reweights based on γ because the statistic of interest is the posterior mean of γ. The direction and degree of reweighting depends on λ, which is the Lagrange multiplier on (1.7) scaled by $\mathbb{E}_p[M]$. λ is analogous to the penalty parameter in the multiplier problem of Hansen and Sargent (2001). If

[1]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.

$\tilde{\gamma} < \mathbb{E}_p[\gamma(\theta)]$, then in order to reduce the posterior mean of γ, we require $\lambda < 0$ so that the worst-case prior places more weight on smaller values of γ.

The difference between the solution (1.8) and the worst-case distortion in Hansen and Sargent (2001) is that the distortion in (1.8) is scaled by the likelihood L, which captures the role of the data in (1.6)-(1.7). The worst-case distortion depends on the likelihood because the expectations in (1.6) and (1.7) are taken under different probability measures. Since the posterior is proportional to the product of the prior and the likelihood, concentrating distortions in the high likelihood regions generates large changes in the posterior from small distortions of the prior. If the likelihood is flat, we return to the standard exponential tilt of Hansen and Sargent (2001).

To solve for λ, substitute (1.8) into (1.7):

$$\tilde{\gamma} = \mathbb{E}_p\left[\underbrace{\frac{\exp\left[\lambda L(\theta|X)\left(\gamma(\theta)-\hat{\gamma}\right)\right]}{\mathbb{E}_p\left[\exp\left[\lambda L(\theta|X)\left(\gamma(\theta)-\hat{\gamma}\right)\right]\right]}}_{\text{change of measure}}\gamma(\theta)\right]. \tag{1.9}$$

The right-hand side is the posterior expectation of γ after a change of measure that depends on λ[2]. As λ increases, the change of measure places more weight on large values of γ, thus increasing the right-hand side. Since the left-hand side is a constant and the right-hand side is increasing in λ, (1.9) implies a unique solution for λ that is straightforward to solve for numerically[3].

Appendix 1.A shows that (1.4)-(1.5) also produces a solution of the form (1.8), but requires us to solve for both λ and $\tilde{\gamma}$. The multiplier representation of (1.4)-(1.5), which is the Lagrangian problem, specifies λ and requires us to solve for $\tilde{\gamma}$. We favor the dual representation (1.6)-(1.7) because $\tilde{\gamma}$ has a straightforward interpretation, while λ has no clear economic interpretation and is difficult to specify ex-ante. It is

[2]The posterior expectation $\mathbb{E}_p\left[\exp\left[\lambda L(\theta|X)\left(\gamma(\theta)-\hat{\gamma}\right)\right]\right]$ typically exists since the likelihood vanishes in the tail of the posterior, and the exponential function further increases the rate of decay.

[3]While the worst-case distribution is uniquely determined, 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.

therefore more convenient to specify a reasonable change in the posterior mean of γ, then check how much the prior needs to be distorted in order to generate this change. Section 1.4 suggests ways to quantify the change in the prior for a sequence of values for $\tilde{\gamma}$, in order to establish the relationship between the set of priors and worst-case posterior mean $\tilde{\gamma}$.

1.2.4 Extensions

The REPS framework allows for flexibility in application. We now consider several prior sensitivity problems that can be incorporated into (1.6)-(1.7).

Credible intervals. One can adapt the constraint (1.7) to analyze the prior sensitivity of quantiles of the function of interest $\psi(\theta)$. This allows us to produce credible intervals for ψ that are robust to deviations in the prior. In particular, suppose we are interested in the dependence of the qth quantile of ψ on the prior. Then we can solve (1.6) subject to:

$$\mathbb{E}_p\left[\frac{M(\theta)}{\mathbb{E}_p[M]}\mathbf{1}\left\{\psi(\theta)<\tilde{\psi}\right\}\right]=q, \tag{1.10}$$

where $\tilde{\psi}$ is the worst-case quantile of ψ. Since (1.10) is (1.7) with $\tilde{\gamma}=q$ and $\gamma(\theta)=\mathbf{1}\left\{\psi(\theta)<\tilde{\psi}\right\}$, the solution has the form (1.8), with the appropriate substitution for $\tilde{\gamma}$ and γ.

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

$$M(\theta)=\begin{cases}\exp[\lambda(1-\tilde{\gamma})] & \psi(\theta)<\tilde{\psi}\\ \exp[-\lambda\tilde{\gamma}] & \psi(\theta)\geq\tilde{\psi}\end{cases} \tag{1.11}$$

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

Subspaces. Taking the expectations in (1.6)-(1.7) over the marginal prior and posterior of a subspace Θ^* of Θ allows us to study the dependence of the posterior on the marginal prior over Θ^* instead of the entire space Θ. Such an exercise can be useful if there is a natural partition for θ. For example, in a New Keynesian model, one may be especially concerned about a subset of parameters whose priors are hard to calibrate from existing data. The reduced dimensionality of Θ^* can also simplify the analysis of the worst-case distortions.

More formally, consider

$$\min_{M(\theta^*):\mathbb{E}_{\pi^*}[M]=1} \mathbb{E}_{\pi^*}\left[M\left(\theta^*\right)\log M\left(\theta^*\right)\right] \tag{1.12}$$

$$\text{s.t. } \mathbb{E}_{p^*}\left[\frac{M\left(\theta^*\right)}{\mathbb{E}_{p^*}\left[M\right]}\mathbb{E}_p\left[\gamma\left(\theta\right)|\theta^*\right]\right] = \tilde{\gamma} \tag{1.13}$$

where π^* and p^* are the marginal prior and posterior over Θ^*. The constraint (1.13) arises by applying the law of iterated expectations to (1.7) and noting that M now depends on θ^* only. Define $\gamma^*\left(\theta^*\right) \equiv \mathbb{E}_p\left[\gamma\left(\theta\right)|\theta^*\right]$. The solution to (1.12)-(1.13) is:

$$M\left(\theta^*\right) \propto \exp\left[\lambda L^*\left(\theta^*|X\right)\left(\gamma^*\left(\theta^*\right)-\tilde{\gamma}\right)\right], \tag{1.14}$$

where $L^*\left(\theta^*|X\right) \equiv p^*\left(\theta^*|X\right)/\pi^*\left(\theta^*\right)$ is the marginal likelihood of θ^*. The solution (1.14) to the subspace problem is similar to the original solution (1.8), with the likelihood L replaced with the marginal likelihood L^* and the objective function $\gamma\left(\theta\right)$ replaced by its expectation conditional on θ^*. Section 1.5 proposes approximations for both of these objects when implementing these calculations in practice.

(1.12)-(1.13) nests the problem in Giacomini et al. (2016) under a quadratic loss function. Since the posterior expectation minimizes expected quadratic loss, equation

(1.13) is equivalent to:

$$\min_{\gamma'} \left(\gamma' - \mathbb{E}_p \left[\frac{M(\theta)}{\mathbb{E}_p[M]} \gamma(\theta) \right] \right)^2 = \tilde{\gamma} \tag{1.15}$$

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

Additional constraints. We can further restrict the set of permissible priors by including prior or posterior moment restrictions to (1.6)-(1.7), 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 1.A for details.

1.3 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. Such features are prevalent in many applications.

1.3.1 Example 1: multiple modes in the likelihood

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.

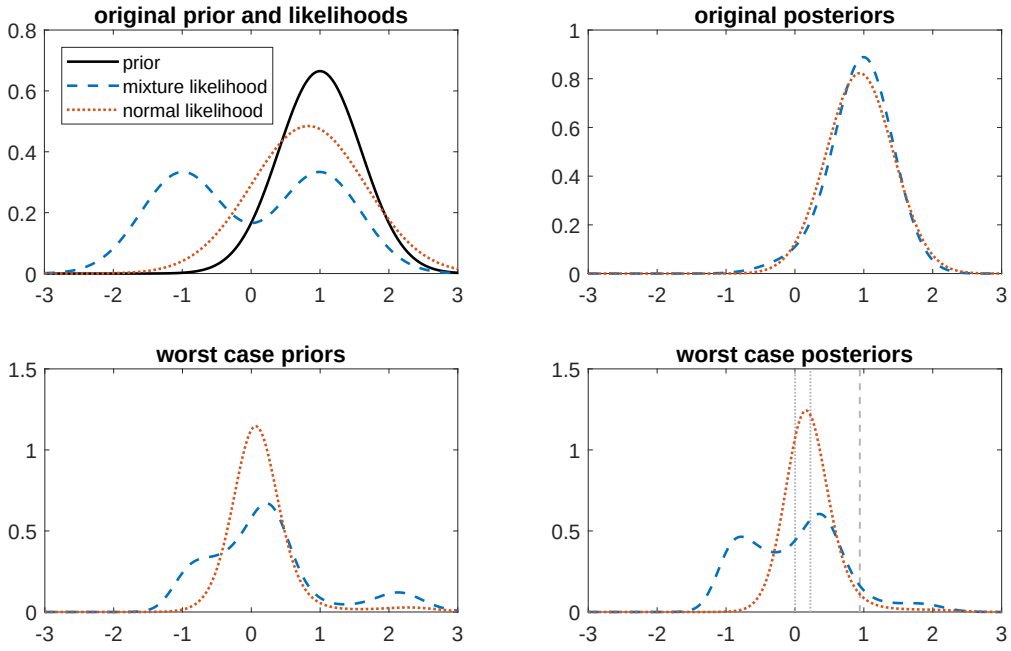


Figure 1.1: Likelihoods, priors and posteriors for mixture and normal likelihoods. Blue dashed lines correspond to mixture likelihood, red dotted lines correspond to normal likelihood. **Top left:** Original prior and likelihoods; **Top right:** Original posteriors; **Bottom left:** Worst case priors; **Bottom right:** Worst case posteriors, with original and worst case means in gray.

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:

$$X \sim \begin{cases} \mathcal{N}(-\theta, 0.6^2) & \text{w.p. } 0.5 \\ \mathcal{N}(\theta, 0.6^2) & \text{w.p. } 0.5 \end{cases} \tag{1.16}$$

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

$$X \sim \mathcal{N}\left(\theta, 0.678^2\right) \tag{1.17}$$

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 1.1 shows that the two posteriors are hard to distinguish visually. On the

other hand, the top left panel of Figure 1.1 shows that the mixture likelihood has modes at -1 and 1, while the normal likelihood has only one mode at 0.831.

REPS shows that the posterior mean of θ has greater sensitivity to changes in the prior under the mixture likelihood. In particular, fixing $\gamma(\theta) = \theta$ and $\mathcal{R} = 1.25$, we solve (1.4)-(1.5) for both models. Since the original prior and $\mathcal{R}$ are identical across the models, we are choosing from the same set of priors in both cases. However, the mixture model's worst-case posterior mean of 0.002 is substantially lower than the normal model's worst-case posterior mean of 0.224. Without prior sensitivity analysis, the similarity of the 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 in 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 1.1, indicate the importance of the range of alternative priors considered by REPS. With the normal likelihood, the worst-case prior for θ is approximately normal, centered around 0. The worst-case posterior is also approximately normal, also centered around the new mean of 0.2. In contrast, with the mixture likelihood, the worst-case prior is flatter, and places a relatively large mass around $\theta = -1$. The worst-case posterior now bimodal, with a second mode around $\theta = -1$ corresponding to the second mode in the likelihood, which was not visible from the original posterior. The worst-case distortions are informative about parts of the prior the one should be concerned about even if one does not regard the exact shape of the worst-case prior as being plausible.

This example illustrates how the robustness of a posterior estimate to changes in the prior can depend on peaks in the likelihood that are dampened by the original prior. In such cases, visual inspection of the prior and posterior can mislead one to believe a result is more robust than it actually is. Herbst and Schorfheide (2014)

show that under more diffuse priors, the DSGE models of Smets and Wouters (2007) and Schmitt-Grohé and Uribe (2012) produce multimodal posteriors that can alter inference relative to a tighter prior. We find that these features matter for posterior inference in our main application to the DSGE model of Smets and Wouters (2007) in Section 1.6. In general, such multimodality is hard to detect without reestimating the model for different priors. Flatter marginal priors may not reveal these modes, since the parameterization and independence assumptions matter for how flattening the marginals impacts the posterior of the object of the interest. REPS provides a systematic approach to prior sensitivity analysis that accounts for features of the likelihood in the tail of the posterior, such as multiple modes, if they are important for one's posterior results.

1.3.2 Example 2: log-normal distribution

We now show that even if the prior and likelihood are Gaussian, REPS detects that the sensitivity to the prior depends on the object of interest ψ.

Suppose $\theta \in \mathbb{R}$ and we have prior $\theta \sim \mathcal{N}(0,1)$, $X \sim \mathcal{N}(\theta,1)$, and observe data $X = 0$. The posterior is $\theta \sim \mathcal{N}\left(0, \frac{1}{2}\right)$. Suppose we wish to do REPS analysis on $\psi(\theta) = \exp(\theta)$, so that ψ is log-normal with mean 1.28 and standard deviation 1.03. Figure 1.2 shows that the posterior of ψ is skewed even though the posterior for θ is symmetric.

REPS shows that this asymmetry matters for the sensitivity of the posterior mean of ψ to changes in the prior. In particular, fix $\gamma(\theta) = \psi(\theta)$, and take $\mathcal{R} = 0.57$, which corresponds to a one standard deviation change in the posterior mean of θ. For this value of $\mathcal{R}$, the posterior mean of ψ has a maximum value of 2.50 (an increase of 1.18 standard deviations) and minimum value of 0.59 (a decrease of 0.67 standard deviations). The asymmetry in sensitivity arises because the ψ is bounded below by zero, but has a posterior with a fat right tail.

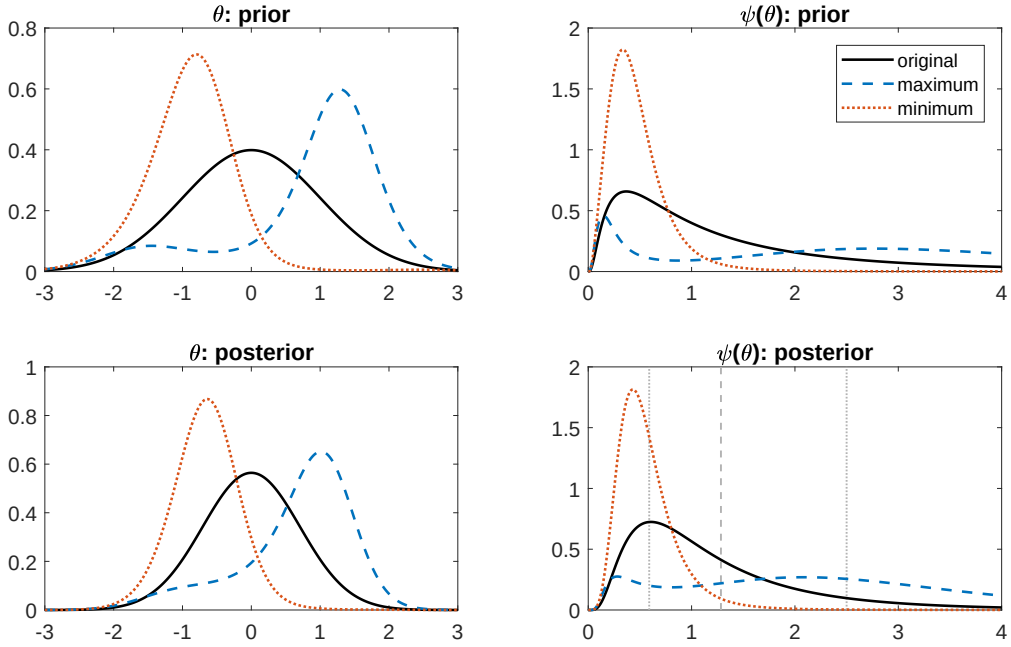


Figure 1.2: Prior and posterior of θ and $\psi(\theta)$. Black solid lines correspond to original distribution, blue dashed lines correspond to distributions that maximize mean, red dotted lines correspond to distributions that minimize mean. **Top left:** Prior of θ; **Top right:** Prior of $\psi(\theta)$ **Bottom left:** Posterior of θ; **Bottom right:** Posterior of $\psi(\theta)$.

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

These insights generalize to more complex settings, where ψ may be a complicated function whose sensitivity to the prior may be hard to analyze. For example, if ψ is an impulse response in a DSGE model, one would need to solve the model and then evaluate the impulse response. The challenge is compounded when θ is high-dimensional. Even if visual inspection of the posterior of ψ reveals that it could

be sensitive to the prior, it is not clear without further analysis which parts of the prior of θ are important. REPS accounts for the function of interest through the γ term in the solution (1.8) while checking across a wide range of alternative priors.

1.4 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 a large or small value of $\mathcal{R}$, taking the Gaussian location model as a benchmark.

1.4.1 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.

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. The relative entropy increases with the prior in the high likelihood region because it is harder to change our beliefs in a region that we have knowledge about.

1.4.2 Gaussian location model

We first show how the above intuition applies for the asymptotic behavior of the solution for (1.6)-(1.7) in the Gaussian location model, which we then use as a benchmark for $\mathcal{R}$. Appendix 1.C verifies that the asymptotics provide a good approximation even for a relatively small sample size T.

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

$$\theta_{p,T} = T\Sigma_{p,T}\Omega^{-1}\overline{X}_T \tag{1.18}$$

$$\Sigma_{p,T} = \left(\Sigma_\pi^{-1} + T\Omega^{-1}\right)^{-1} \tag{1.19}$$

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

Lemma 1.1. *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 (1.6)-(1.7) satisfies $T^{\frac{d}{2}}\mathcal{R}_T \to \overline{R}\pi(\theta_0)|\Omega|^{-\frac{1}{2}}$ for some $\overline{R}$.*

Lemma 1.1 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 depends on the variance of the data and the prior at θ_0. The scaling factor $\overline{R}$ varies depending on the number of dimensions d and amount of distortion c. See Appendix 1.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 θ_0, the asymptotic relative entropy is scaled by by the prior $\pi(\theta_0)$ at θ_0. For other models that satisfy the Bernstein-von Mises theorem, we have the same asymptotic result, with $|\Omega|$ replaced by the Fisher information matrix. In addition, the same asymptotics apply for the mean or quantile of any linear combination of θ.

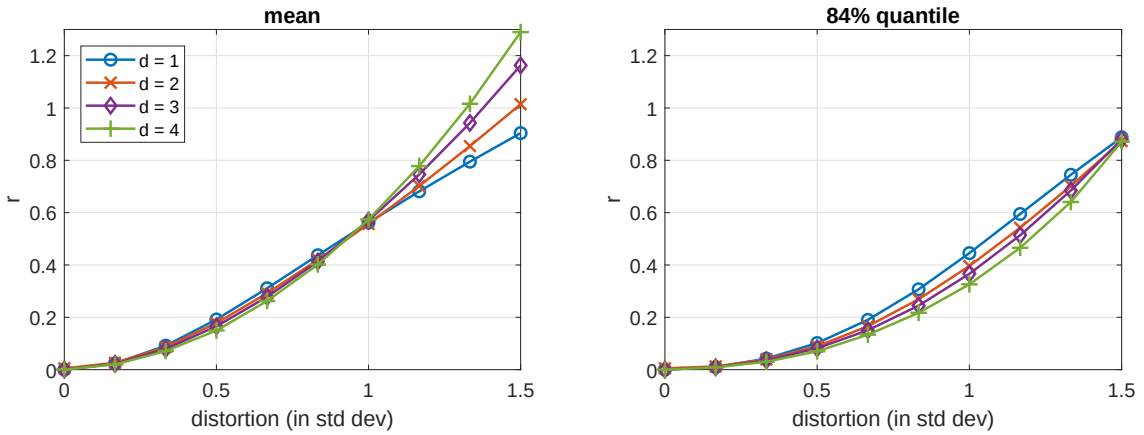


Figure 1.3: Scaled relative entropy r and distortion in Gaussian location model with $\theta \sim \mathcal{N}(0, I)$, $\overline{X}_T = 0$, $T = 100$, $d \in \{1, ..., 4\}$. All distortions are scaled by posterior standard deviation. **Left:** increase in posterior mean of θ_1 by one posterior standard deviation; **Right:** increase in 84% quantile of θ_1 by one posterior standard deviation.

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

More generally, we calibrate $\mathcal{R}$ by taking a Gaussian approximation of the prior and posterior, then picking the relative entropy to match the corresponding Gaussian location model. In particular, for prior and posterior means (μ_π, μ_p) and variances (Σ_π, Σ_p), consider the approximation $\mathcal{N}(\mu_\pi, \Sigma_\pi)$ and $\mathcal{N}(\mu_p, \Sigma_p)$ for the prior and posterior respectively. Define the dispersion of the likelihood:

$$\Sigma_\ell \equiv \left(\Sigma_p^{-1} - \Sigma_\pi^{-1}\right)^{-1}. \tag{1.20}$$

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

We parameterize the relative entropy as:

$$\mathcal{R} = 1.6^\delta r \frac{\pi(\mu_p)}{\pi(\mu_\pi)} \sqrt{\frac{|\Sigma_\ell|}{|\Sigma_\pi|}} \quad \text{where} \quad \delta = \begin{cases} 0 & d = 1 \\ d & d > 1 \end{cases} \tag{1.21}$$

The last two terms arise from asymptotics in Lemma 1.1, 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^{δ} scaling accounts for further differences across dimensions. The remaining free parameter r controls the size of the change in the prior and posterior.

A practitioner can compute r from (1.21), then compare the change in the posterior for the particular estimation she is considering to the change that r would imply in a Gaussian location model. Figure 1.3 shows the values of r for different changes in the mean of 84% quantile in the Gaussian location model with $\Sigma_\pi = \Omega = I$, and $T = 100$. As a rule of thumb, $r < 0.05$ is small and $r > 0.50$ is large. For the mean and 84% quantile of the Gaussian model, $r = 0.05$ and $r = 0.50$ correspond approximately to a 1/4 standard deviation change and a one standard deviation change respectively.

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

1.5 Implementation

We now discuss the numerical implementation of the calculations in Section 1.2. We assume that we have Monte Carlo draws from both the prior and the posterior.

1.5.1 Importance sampling

If the distribution of the worst-case distortion M does not have fat tails, then we can solve for M and evaluate the worst-case prior and posterior using importance sampling. In particular, for any λ, we can evaluate the right-hand side of (1.9), approximating the expectations with Monte Carlo sample averages. We can then

solve (1.9) using this Monte Carlo approximation. Given the solution for λ, we can now compute M for any value of θ. Reweighting the original Monte Carlo draws by M then gives us the worst-case prior and posterior.

In addition, Lemma 1.4 in Appendix 1.A shows that we can evaluate relative entropy using the expression:

$$\mathbb{E}_\pi \left[M(\theta) \log M(\theta) \right] = -\log \mathbb{E}_\pi \left[\exp \left[\lambda L(\theta|X) \left(\gamma(\theta) - \tilde{\gamma} \right) \right] \right] \tag{1.22}$$

The right hand side is negative log of the normalizing constant that ensures $M\pi$ integrates to one. It is therefore straightforward to evaluate the right-hand side of (1.22) using draws from the prior. This produces more precise estimates than averaging $\log M$ across the draws from the worst-case prior because such a calculation would require the normalizing constant as well.

Importance sampling performs poorly when the distribution of M has fat tails, which occurs when the likelihood is sharply peaked. This problem becomes more severe as we increase the dimensionality of Θ or the number of observations. The solution (1.8) 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.

1.5.2 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

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_{SMC} **do**
  compute weights: Solve for λ_i and compute $m_i \equiv \pi_i/\pi_{i-1}$ for each draw.
  selection: Draw from $\{\theta_{\pi_{i-1},j}\}_{j=1}^{N_{\mathrm{P}}}$ 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_{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 $\{\theta_{\tilde{p},j}\}_{j=1}^{N_P} = \{\theta_{p_{N_{\mathrm{SMC}}},j}\}_{j=1}^{N_P}$.

the original and worst-case distributions. Beginning with draws from the original prior and posterior, referred to as *particles*, we iteratively construct particle approximations of the bridge distributions, before arriving at a particle approximation of the worst-case prior and posterior.

Algorithm. 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 (1.6)-(1.7) 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 below in Algorithm 1. We leave the details to Appendix 1.B.

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

Algorithm 1 poses two computational challenges. Firstly, if we are interested in more than one statistic, we need to repeat the SMC algorithm for each objective function we are interested in. For instance, if one were interested in the error bands for a set of impulse response functions in a DSGE model, one would need to repeat Algorithm 1 for each impulse response function at multiple horizons. For large-scale DSGE models, this can be computationally demanding. Secondly, to apply Algorithm 1 to problem (1.12)-(1.13), we require the marginal likelihood L^* and conditional expectation γ^*, both of which may be difficult to compute.

1.5.3 Approximate REPS

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

Algorithm 2: Approximate REPS

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

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

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

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

Algorithm. The main idea of AREPS is to replace π, p, L^* and γ^* with approximations $\hat{\pi}$, $\hat{p}$, $\widehat{L}$ and $\hat{\gamma}$. In particular, 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 π, p and γ.[4] If we are doing the REPS analysis over the entire Θ, then $\widehat{L}$ and $\hat{\gamma}$ correspond to approximations for L and γ respectively. Analogously to semiparametric methods, we use basis functions to allow for flexibility while mitigating the curse of dimensionality that arises in fully nonparametric estimation.

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

[4]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.

Computational details. For computational efficiency, the approximations $\hat{\pi}$, $\hat{p}$ and $\hat{\gamma}$ should be fast to evaluate. For example, in our application in Section 1.6, we use a Gaussian mixture model and a quadratic logit to approximate p and γ^* respectively.

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

If closed form expressions for L^* and γ^* are available, we can reweight draws from the approximate worst-case prior and posterior to obtain draws from the true worst-case prior and posterior. See Appendix 1.B for details.

1.6 Application: Smets and Wouters (2007)

For our main application, we consider the workhorse DSGE model from Smets and Wouters (2007). Even in such a large model, REPS is not only feasible, but also accounts for features of the likelihood that are especially hard to diagnose in such high-dimensional settings, even with the use of local prior sensitivity methods. Among the various results, we detect features of the likelihood that have been corroborated by existing estimations of the model with alternative priors, providing support for the validity of both the overall methodology and the numerical approximations.

1.6.1 Model and estimation

Model. 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[5]. This gives a total of thirty-six parameters.

Original estimation. We use quarterly data from 1984Q1 to 2007Q4 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.

1.6.2 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 the set of all error bands arising from the chosen set of priors. In particular, 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

[5]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.

$\{\rho, r_\pi, r_y, r_{\Delta y}\}$ from the monetary policy rule:

$$\hat{r}_t = \rho \hat{r}_{t-1} + (1-\rho)\left[r_\pi \hat{\pi}_t + r_y\left(\Delta \hat{y}_t\right) + r_{\Delta y}\left[\left(\Delta \hat{y}_t\right) - \left(\Delta \hat{y}_{t-1}\right)\right]\right] + \varepsilon_t^r, \tag{1.23}$$

where $\hat{r}_t$ is the interest rate, $\hat{\pi}_t$ is the inflation rate, $\Delta \hat{y}_t$ is the output gap, and ε_t^r is an exogenous AR(1) shock process. The second set of parameters is $\{\xi_w, \xi_p, \iota_w, \iota_p\}$, which determine the level of wage rigidity, price rigidity, wage indexation, and price indexation 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. Households have rational expectations about this future path of interest rates, and make decisions that determine the response of output in equilibrium. Similarly, the nominal frictions determine how much prices 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 these parameters on the impulse response.

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^2s of between 0.97 to 0.98. Finally, the conditional probability that the impulse response is less than a given cutoff is estimated with a quadratic logit. For the SMC, we use 5×10^4 particles, 250 SMC stages, and 10 Metropolis-Hastings draws at each stage. We average the results across 10 runs of the SMC. A single run of the SMC in MATLAB takes approximately

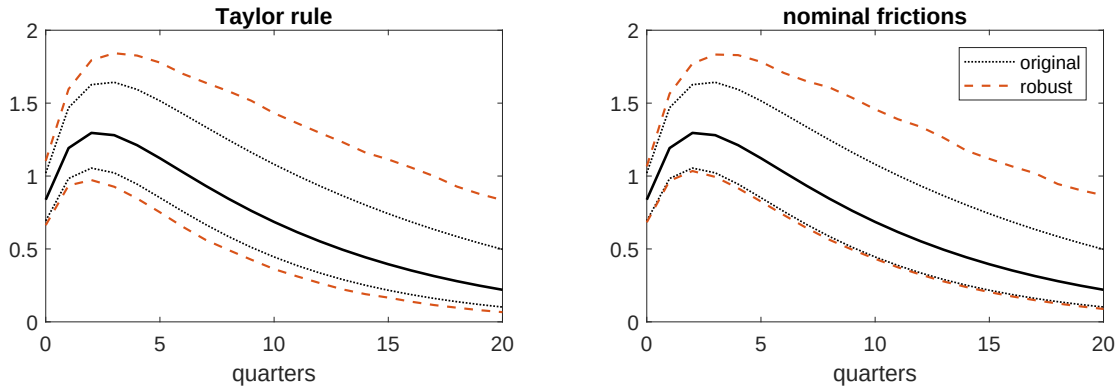


Figure 1.4: Robust 68% error bands. Black lines show original median (solid) and error bands (dotted); red dashed lines show robust error bands. **Left:** Taylor rule prior, with $r = 5 \times 10^{-3}$; **Right:** nominal frictions prior, with $r = 2.5 \times 10^{-4}$.

three hours. We parallelize the computations by horizon across 21 cores. Appendix 1.D provides further details, including checks for the accuracy of the approximations.

1.6.3 Robust error bands

Figure 1.4 plots the original 68% error bands for the impulse response, together with the robust error bands.

Sensitivity of error bands. For each set of parameters, we fix the relative entropy so that the maximum distance between the original and robust error bands is approximately one posterior standard deviation. Using the rule of thumb in equation (1.21), this requires $r = 5 \times 10^{-3}$ for the Taylor rule and $r = 2.5 \times 10^{-4}$ for the nominal frictions prior, indicating a high degree of sensitivity to the prior. 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, which is at least an order of magnitude smaller than the largest differences shown in Figure 1.4.

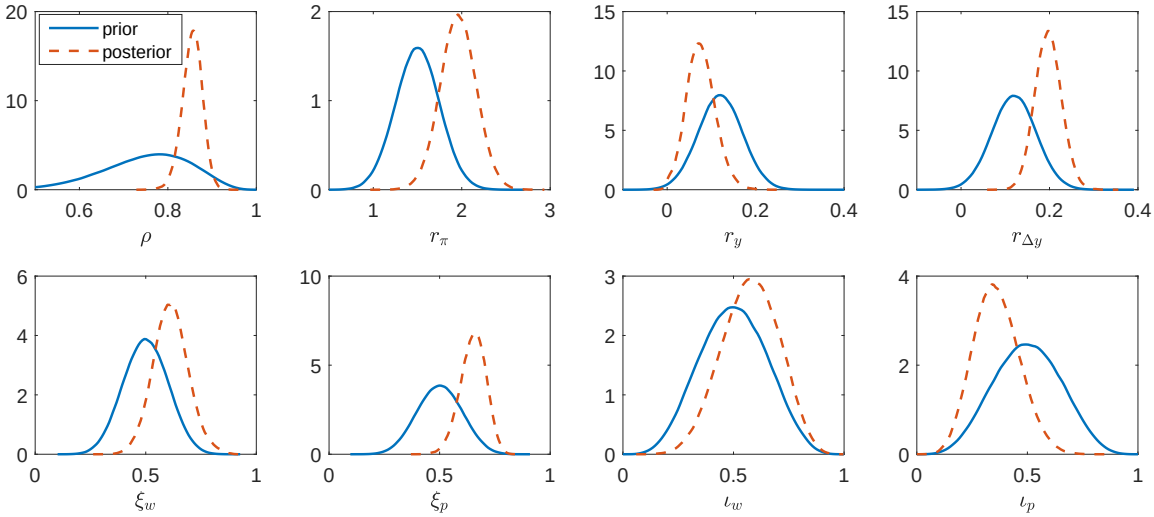


Figure 1.5: Original prior and posterior of Taylor rule parameters $\{\rho, r_\pi, r_y, r_{\Delta y}\}$ and nominal frictions parameters $\{\xi_w, \xi_p, \iota_w, \iota_p\}$.

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 1.5 shows that their marginal posteriors are not much more concentrated than their priors. To study the joint concentration of the likelihood, we consider the statistic $\sqrt{|\Sigma_\ell| \,/\, |\Sigma_\pi|}$, where Σ_ℓ is defined in (1.20), measures the dispersion of the likelihood relative to the prior. This statistic is 0.23 for the Taylor rule and 1.53 for the nominal frictions parameters. In the Gaussian location model, these values correspond to having a standard normal prior and observing just one observation of $X \sim \mathcal{N}(\theta, \omega^2 I)$ with $\omega = 0.69$ and $\omega = 1.11$ respectively. With such a dispersed likelihood, the asymptotics in Lemma 1.1 overpredicts the relative entropy needed to change the posterior estimates. In particular, in Figure 1.9, the relative entropy with $T = 1$ is an order of magnitude smaller than what is predicted by the log-log trend for $T > 10$.

In addition, both the Taylor rule and nominal frictions parameters are good predictors of the impulse response function. For example, using the Monte Carlo

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

Asymmetry. The sensitivity to the prior is uniform across neither horizons nor bounds. The robust error bands diverge from the original error bands as we increase the horizon, indicating that the impulse response depends more on the prior at longer horizons. Intuitively, since low frequency fluctuations are estimated less precisely than high frequency fluctuations, the marginal likelihood for the impulse response is more concentrated for short horizons. Moreover, the impulse response at short horizons are 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^2s of 0.27 and 0.05. These R^2s are substantially smaller than those for the impulse response 20 quarters after impact.

In addition, the upper bound is more sensitive to changes in the prior than the lower bound for both sets of parameters. Hence, given the model and data, a policymaker should be more concerned about underestimating than overestimating the effects of a surprise decrease in interest rates. Like the log-normal example shown in Figure 1.2, the original posterior of the impulse response is skewed to the right, which partially explains the greater sensitivity of the upper bound of the error bands. However, the shape of the posterior for the impulse response does not fully capture the asymmetry in sensitivity. In particular, the asymmetry is more pronounced for

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					
ρ	persistence	0.007	–0.165	–0.003	0.125
r_{π}	inflation coef.	0.000	0.743	–0.004	–0.234
r_{y}	output gap coef.	0.005	0.192	–0.001	–0.282
$r_{\Delta y}$	output gap growth coef.	–0.002	–0.072	–0.005	0.204
Nominal frictions					
ξ_{w}	wage rigidity	–0.009	–0.162	0.006	0.517
ξ_{p}	price rigidity	–0.040	–0.076	–0.000	–0.041
ι_{w}	wage indexation	–0.001	–0.048	–0.002	0.037
ι_{p}	price indexation	0.004	0.108	0.008	–0.212

Table 1.1: 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.

the nominal frictions parameters, reflecting differences in the mapping from the two sets of parameters to the impulse response.

The wide robust error bands add 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, quantile, or subset of parameters considered.

1.6.4 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 1.4.

We summarize the distortions by the changes in the prior and posterior means in Table 1.1. 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.

Next, we show that the worst-case distortions reveal information about the model and likelihood that is difficult to uncover by 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. 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, and therefore focus on draws for which the impulse response in one standard deviation below (above) the mean. Comparing the behavior of the left and right tails of the impulse response reveals the reasons for the asymmetry in distortions.

To understand the worst-case distortions, we use the fact that the distortion (1.14) depends on the parameters through the objective function $\gamma^*(\theta^*)$ and the marginal

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

Table 1.2: 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 $L^*(\theta^*|X)$. 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 (1.14) is a conditional expectation, this regression captures both the direct effect of the parameters on the impulse response, as well as the indirect effect from the conditional distribution of the remaining parameters under the posterior. The impulse response and parameters are normalized to mean zero and standard deviation one, so the coefficients can be interpreted as the number of standard deviations the impulse response increases by in response to a one standard deviation increase in the corresponding parameter on average. The results are reported in Table 1.2. We then analyze the original and worst-case distributions in order to understand the shape of the likelihood.

Taylor rule distortions

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

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

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

We begin by comparing the original and worst-case marginal posteriors for r_π, shown in Figure 1.6. The worst-case posterior for the lower bound is bimodal, with an additional peak around $r_\pi = 2.75$, revealing a high likelihood in the right tail of r_π. The observation corroborates the results of Herbst and Schorfheide (2014), who find that the posterior mean of r_π moves from 2.04 to 2.78 when one replaces the prior from Smets and Wouters (2007) with a more diffuse one. This mode in the likelihood arises from fitting the data to the Taylor rule. In particular, using data for the

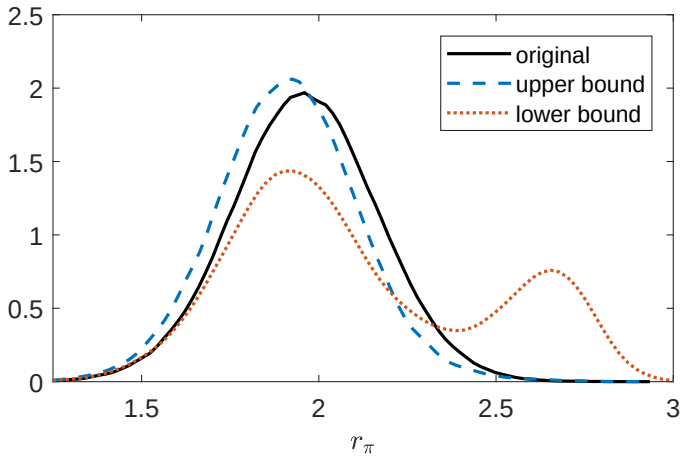


Figure 1.6: Original and worst-case posteriors of r_π. **Black solid line:** original posterior; **Blue dashed line:** worst-case posterior for upper bound; **Red dotted line:** worst-case posterior for lower bound.

federal funds rate, inflation, and output gap from FRED and estimating the Taylor rule using a linear regression, we find a large inflation coefficient of 2.59[6]. This large value arises due to the low frequency variation in the data. As a heuristic exercise, we run the regression of the trend and cyclical components of HP filtered data, and find coefficients of 2.66 and 0.92 respectively. Sala (2015) formally estimates a similar DSGE model in the frequency domain, and finds a posterior estimate of 1.81 for the low frequency component, which is significantly larger than the estimate of 1.12 for the high frequency component[7]. The prior from Smets and Wouters (2007) shrinks more towards smaller values of r_π, making it difficult to detect the possibility of an additional mode without reestimating the model with a new prior. However, one may not know ex-ante which priors might result in these changes in the posterior, or whether the changes are consequential to the impulse response. REPS flexibly accounts for such features in the likelihood if they are important for the posterior outcomes of interest.

[6]The regression ignores the autocorrelation of 0.26 in the monetary policy shock.

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

The likelihood reveals two reasons for the small distortions in ρ. Firstly, Figure 1.5 suggests that ρ is relatively sharply identified by the likelihood. In particular, the ratio of posterior to prior standard deviation for ρ is smaller than that of r_π, r_y and $r_{\Delta y}$, with a value of 0.23 as compared to 0.82, 0.66 and 0.60 respectively. As a result, larger changes in the prior are needed to produce the same change in the posterior or ρ, which would result in a greater relative entropy cost. In addition, the correlations of ρ with r_π, r_y and $r_{\Delta y}$ are in conflict with their effect on the impulse response. In particular, the regression results suggest that ρ should be distorted in the opposite direction from r_π and r_y, but in the same direction as $r_{\Delta y}$. Such distortions are costly in terms of relative entropy as they run against the likelihood, as ρ has a positive correlation with r_π 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 ρ 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 1.5 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.

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

Dependence of impulse response on parameters. The largest coefficient from the regression reported in Table 1.2 is the one on wage rigidity ξ_w, which partly explains why the large distortion for ξ_w. Moreover, the regression coefficient on ξ_p is negative in the regression for the upper bound, rationalizing the counterintuitive direction of the distortion of the prior on ξ_p. The negative coefficient arises because of an omitted variable bias — ξ_p is correlated with other parameters that also affect the impulse response, biasing the regression coefficient relative to what we would have found if we controlled for all the parameters in the model. The coefficient obtained without controlling for the remaining parameters is the relevant one here because we are leaving the prior of all other parameters unchanged. REPS accounts for the fact that changing the prior for ξ_p changes the posterior of the impulse response through both the marginal effect of ξ_p and the effect of any other correlated parameters.

On the other hand, the coefficient of wage indexation ι_w for the upper bound regression contradicts the shift of the posterior distribution towards smaller values.

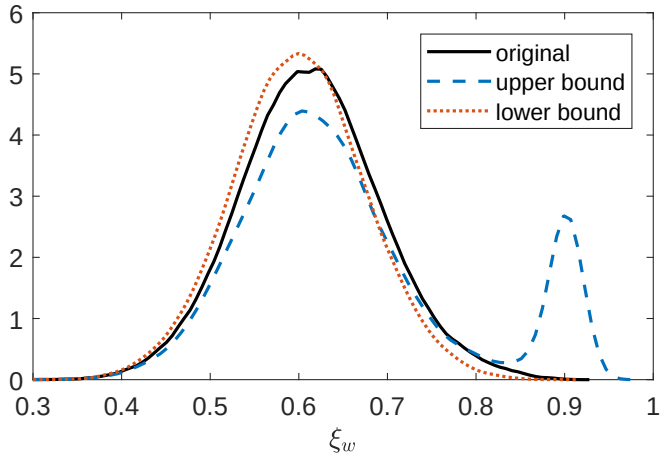


Figure 1.7: Original and worst-case posteriors for ξ_w. **Black solid line:** original posterior; **Blue dashed line:** worst-case posterior for upper bound; **Red dotted line:** worst-case posterior for lower bound.

Moreover, the small regression coefficients on price indexation ι_p are inconsistent with both the direction and magnitude of the worst-case distortions.

Likelihood. The worst-case posteriors for ξ_w, shown in Figure 1.7, provide an explanation for the especially large change in the posterior mean of ξ_w for the upper bound. In particular, the worst-case posterior for the upper bound is bimodal, with a new mode around $\xi_w = 0.90$. As with the lower bound worst-case posterior of r_π, this is in line with the diffuse prior estimates of Herbst and Schorfheide (2014). In particular, the posterior mean of ξ_w shifts from 0.70 under the Smets and Wouters (2007) prior to 0.93 under the diffuse prior. This is a larger change than that of ξ_p, ι_w, and ι_p, whose posterior means move from 0.66, 0.59, and 0.22 to 0.72, 0.73 and 0.11 respectively under the diffuse prior. Again, REPS accounts for peaks in the likelihood that are hard to detect without reestimating the model under the appropriate prior. The asymmetry between the upper and lower robust error bands also suggests that this new mode in the posterior for ξ_w is more consequential for the impulse response than the new mode in the lower bound worst-case posterior of r_π.

	Lower bound			
	ξ_w	ξ_p	ι_w	ι_p
ξ_w	1			
ξ_p	0.19	1		
ι_w	–0.13	–0.23	1	
ι_p	–0.07	–0.23	–0.15	1

	Upper bound			
	ξ_w	ξ_p	ι_w	ι_p
ξ_w	1			
ξ_p	–0.20	1		
ι_w	–0.05	–0.31	1	
ι_p	–0.08	–0.15	–0.14	1

Table 1.3: 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.

Indeed, the regression coefficient for ξ_w in Table 1.2 is larger in magnitude than that for r_π.

The posterior correlations, reported in Table 1.3, help to account for the counter-intuitive distortions in price indexation ι_p. The negative correlation of ι_p with ξ_w, ξ_p and ι_w imply that increases in these parameters correspond on average to a decrease in ι_p. Hence the likelihood favors moving ι_p in the opposite direction from ξ_w, ξ_p and ι_w. In addition, Figure 1.8 shows that under the worst-case posterior for the upper bound, the new mode for ξ_w corresponds to small values of ι_p, decreasing the posterior mean of ι_p. More generally, these distortions indicate that if we check the robustness of the posterior by changing the prior of a set of parameters in the direction suggested by economic intuition without accounting for correlations across parameters, we may understate the sensitivity of the posterior.

The posterior correlations for ξ_p further emphasize this point and provide an additional explanation for the inconsistent direction of distortions for the two worst-case posteriors of ξ_p. The rigidity parameters ξ_w and ξ_p have a positive correlation of 0.19 conditional on the impulse response being small, and a negative correlation of -0.20 conditional on the impulse response being large. This drives ξ_p in the same direction as ξ_w for the lower bound, and in the opposite direction from ξ_w for the upper bound. Given the distortions for ξ_w, it is therefore optimal to decrease ξ_p for both the upper and lower bound. In addition, ξ_p is negatively correlated with ι_w for both the lower and upper bounds. This implies that distorting their priors

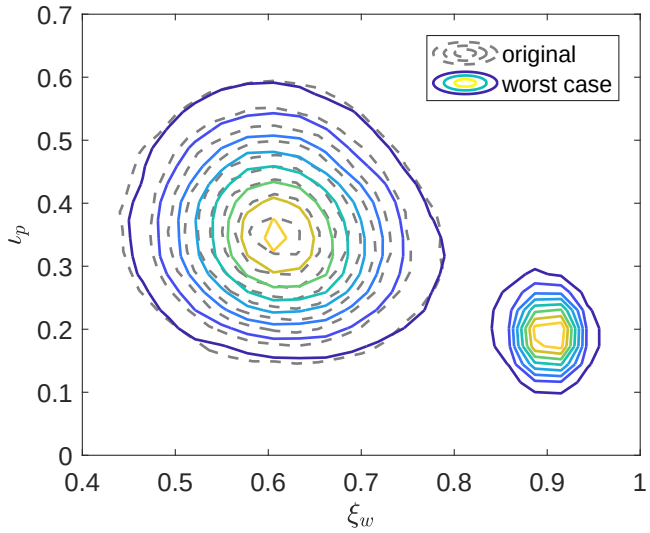


Figure 1.8: Original and lower bound worst-case posteriors for (ξ_w, ι_p). **Gray dashed lines:** original posterior; **Colored solid lines:** worst-case posterior.

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

1.6.5 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. Müller (2012) also concludes that the posterior for the structural parameters depends heavily on the prior, but finds little sensitivity of the impulse responses to the priors of 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). Since the correlations and tail behavior of the prior used in Smets and Wouters (2007) come from assumptions made for convenience, it is especially important that they are allowed to change when we do prior sensitivity analysis.

Local methods use derivatives that also fail to capture the asymmetry in the sensitivity to the prior. Figure 1.4 shows that prior sensitivity depends on the direction in which we wish to change the impulse response. Such non-linearity in the prior sensitivity might also be expected more broadly, given the irregularity of the likelihood and the complicated mapping from the parameters to the function of interest ψ in many applications (e.g. Ho (2018)), whether we focus on means or quantiles.

1.7 Conclusion

To understand how the data inform the posterior estimates, one needs to disentangle the roles of the prior and the likelihood. Despite the numerous assumptions made when writing down priors and the disagreement over what priors to use for various parameters, such analysis is often either absent or ad hoc. Reestimating the model for many alternative priors can be computationally demanding, and it is infeasible to reestimate the model for the full set of possible priors. On the other hand, few prior sensitivity tools have been developed for economic applications.

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. The REPS methodology reduces the complex 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 that shows how REPS can reveal properties of the posterior are sensitive to changes in the

prior, and which parts of the prior are most important for these properties. Moreover, the worst-case distortions uncover features of the likelihood that are important for posterior inference yet hard to detect. These are useful diagnostics for any Bayesian estimation. In related work, Ho (2018) uses REPS to obtain constructive conclusions about 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 is informing 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 ϵ- Contaminated Priors. *The Annals of Statistics 14*(2), 461–486.

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

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

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

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

Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016). Solution and Estimation Methods for DSGE Models. 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 H. Uhlig (2016). 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. (2018). 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.

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

1.A Proofs

1.A.1 Solution to primal and dual problems

Lemma 1.2. *The solutions for M in problems (1.4)-(1.5) and (1.6)-(1.7) both have the form (1.8).*

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

$$p(\theta|X) = \frac{\pi(\theta) L(\theta)}{\zeta} \tag{1.24}$$

Consider the dual problem (1.6)-(1.7). Attaching the multiplier $\lambda \mathbb{E}_p[M] \zeta$ to (1.7) and the multiplier μ to the constraint $\mathbb{E}_\pi[M] = 1$, we have the first-order condition:

$$0 = \mu + 1 + \log M(\theta) - \lambda L(\theta|X) (\gamma(\theta) - \tilde{\gamma}), \tag{1.25}$$

which we can rearrange to obtain (1.8).

Now consider the primal problem (1.4)-(1.5). Attaching the multiplier $1/(\lambda \mathbb{E}_p[M] \zeta)$ to (1.5) and the multiplier $\mu/(\lambda \mathbb{E}_p[M] \zeta)$ to the constraint $\mathbb{E}_\pi[M] = 1$, we have the

first-order condition:

$$\begin{aligned} 0 &= \mu + 1 + \log M(\theta) - \lambda \mathbb{E}_p[M] L(\theta|X) \left(\frac{\gamma(\theta)}{\mathbb{E}_p[M]} - \frac{\mathbb{E}_p[M(\theta)\gamma(\theta)]}{\mathbb{E}_p[M]^2} \right) \\ &= \mu + 1 + \log M(\theta) - \lambda L(\theta|X) \left(\gamma(\theta) - \mathbb{E}_p \left[\frac{M(\theta)}{\mathbb{E}_p[M]} \gamma(\theta) \right] \right). \end{aligned} \tag{1.26}$$

Rearranging (1.26), we have:

$$M(\theta) \propto \exp \left[\lambda L(\theta|X) \left(\gamma(\theta) - \mathbb{E}_p \left[\frac{M(\theta)}{\mathbb{E}_p[M]} \gamma(\theta) \right] \right) \right], \tag{1.27}$$

which has the same form as (1.8) once we replace $\tilde{\gamma}$ with the full expression for the worst-case posterior mean of γ. □

1.A.2 Subspaces

Lemma 1.3. *The solution for M is problem (1.12)-(1.13) is (1.14).*

Proof. Notice that

$$\mathbb{E}_{p^*}[M(\theta^*)\gamma(\theta)] = \mathbb{E}_{p^*}[M(\theta^*)\mathbb{E}_p[\gamma(\theta)|\theta^*]] \tag{1.28}$$

It immediately follows that the first-order condition of (1.12)-(1.13) is:

$$0 = \mu + 1 + \log M(\theta) - \lambda L(\theta|X) (\mathbb{E}_p[\gamma(\theta)|\theta^*] - \tilde{\gamma}), \tag{1.29}$$

which simplifies to (1.14). □

1.A.3 Additional constraints

Consider the constrained optimization problem:

$$\min_{M(\theta):\mathbb{E}_\pi[M]=1} \mathbb{E}_\pi \left[M(\theta) \log M(\theta) \right] \tag{1.30}$$

$$\text{s.t. } \mathbb{E}_p \left[\frac{M(\theta)}{\mathbb{E}_p[M]} \gamma(\theta) \right] = \tilde{\gamma} \tag{1.31}$$

$$\mathbb{E}_p \left[\frac{M(\theta)}{\mathbb{E}_p[M]} g_{p,k}(\theta) \right] = \tilde{g}_{p,k} \text{ for } k = 1, ..., K \tag{1.32}$$

$$\mathbb{E}_\pi \left[M(\theta) g_{\pi,l}(\theta) \right] = \tilde{g}_{\pi,l} \text{ for } l = 1, ..., L \tag{1.33}$$

This is the problem (1.6)-(1.7) augmented by the additional moment conditions (1.32)-(1.33).

Attaching multipliers $\lambda \mathbb{E}_p[M] \zeta$, $\mu_{p,k} \mathbb{E}_p[M] \zeta$ and $\mu_{\pi,l}$ to constraints (1.31), (1.32) and (1.33) respectively, we obtain the first-order condition:

$$\begin{aligned} 0 =& \mu + 1 + \log M(\theta) - \lambda L(\theta|X) (\gamma(\theta) - \tilde{\gamma}) \\ & - \sum_{k=1}^{K} \mu_{p,k} L(\theta|X) (g_{p,k}(\theta) - \tilde{g}_{p,k}) - \sum_{l=1}^{L} \mu_{\pi,l} (g_{\pi,l}(\theta) - \tilde{g}_{\pi,l}), \end{aligned} \tag{1.34}$$

where μ is the multiplier on the constraint $\mathbb{E}_\pi[M] = 1$. We rearrange (1.34) to obtain the solution:

$$\begin{aligned} M(\theta) \propto & \exp\left[\lambda L(\theta|X) (\gamma(\theta) - \tilde{\gamma}) \right] \\ & \times \exp\left[L(\theta|X) \sum_{k=1}^{K} \mu_{p,k} (g_{p,k}(\theta) - \tilde{g}_{p,k}) + \sum_{l=1}^{L} \mu_{\pi,l} (g_{\pi,l}(\theta) - \tilde{g}_{\pi,l}) \right], \end{aligned} \tag{1.35}$$

where the second term introduces $K+L$ additional unknowns arising from the moment conditions (1.32)-(1.33).

1.A.4 Evaluating relative entropy

Lemma 1.4. *The solution for the minimum relative entropy in (1.6)-(1.7) is:*

$$\mathbb{E}_\pi \left[M(\theta) \log M(\theta) \right] = -\log \mathbb{E}_\pi \left[\exp \left[\lambda L(\theta|X) \left(\gamma(\theta) - \tilde{\gamma} \right) \right] \right] \tag{1.36}$$

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 (1.8) yields:

$$\log M(\theta) = \log \kappa + \lambda L(\theta|X) \left(\gamma(\theta) - \tilde{\gamma} \right). \tag{1.37}$$

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:

$$\begin{aligned}
\mathbb{E}_\pi \left[M(\theta) \log M(\theta) \right] &= \int M(\theta) \log M(\theta) \, \pi(\theta) \, d\theta \\
&= \int \left[\log \kappa + \lambda L(\theta|X) \left(\gamma(\theta) - \tilde{\gamma} \right) \right] \tilde{\pi}(\theta) \, d\theta \\
&= \log \kappa + \int \lambda \zeta \left(\gamma(\theta) - \tilde{\gamma} \right) \frac{\tilde{\pi}(\theta) L(\theta|X)}{\zeta} d\theta \\
&= \log \kappa + \lambda \zeta \left(\int \gamma(\theta) \, \tilde{p}(\theta|X) \, d\theta - \tilde{\gamma} \right) = \log \kappa
\end{aligned} \tag{1.38}$$

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

1.A.5 Asymptotics

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

Proof. (Lemma 1.1) Assume $\overline{X}_T = \theta_0$, and first consider the case with Ω 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 Δ:

$$L\left(\Delta;\Sigma_{\ell}\right)=\left|2\pi\Sigma_{\ell,T}\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\Delta'\Sigma_{\ell,T}^{-1}\Delta\right]=\left|\Sigma_{\ell}^{T}\right|^{-\frac{1}{2}}L\left(\Sigma_{\ell,T}^{-\frac{1}{2}}\Delta;I\right) \tag{1.39}$$

Abusing notation again, write (1.8) as a function of Δ:

$$M\left(\Delta;\Sigma_{\ell}\right)\propto\exp\left[\lambda\left(\Sigma_{\ell}^{T}\right)L\left(\Delta;\Sigma_{\ell,T}\right)\left(\Delta_{1}+\left(1-\frac{T\sigma_{p,T}^{1}}{\omega_{1}}\right)\theta_{0}+c\sigma_{p,T}^{1}\right)\right] \tag{1.40}$$

As $T\to\infty$, since $\frac{T\sigma_{p,T}^{1}}{\omega_{1}}\to1$, there exists λ^{*}, M^{*} such that:

$$\left|\Sigma_{\ell,T}\right|^{-\frac{1}{2}}\lambda\left(\Sigma_{\ell,T}\right)\to\lambda^{*} \tag{1.41}$$

$$M\left(\Sigma_{\ell,T}^{\frac{1}{2}}\Delta;\Sigma_{\ell,T}\right)\to M^{*}\left(\Delta\right) \tag{1.42}$$

for (1.7) and $\mathbb{E}_{\pi}\left[M\left(\Delta\right)\right]=1$ to be satisfied. In particular, we have

$$M^{*}\left(\Delta\right)\propto\exp\left[\lambda^{*}L\left(\Delta;I\right)\left(\Delta_{1}+c\omega_{1}\right)\right] \tag{1.43}$$

Denoting $\hat{\pi}\left(\Delta\right)\equiv\pi\left(\Delta+\theta_{0}\right)$, the relative entropy is

$$T^{\frac{d}{2}}\mathcal{R}_{T}=T^{\frac{d}{2}}\int M\left(\Delta;\Sigma_{\ell,T}\right)\log\left[M\left(\Delta;\Sigma_{\ell,T}\right)\right]\hat{\pi}\left(\Delta\right)d\Delta \tag{1.44}$$

$$\approx T^{\frac{d}{2}}\int M^{*}\left(\Sigma_{\ell,T}^{-\frac{1}{2}}\Delta\right)\log\left[M^{*}\left(\Sigma_{\ell,T}^{-\frac{1}{2}}\Delta\right)\right]\hat{\pi}\left(\Delta\right)d\Delta \tag{1.45}$$

$$\approx\overline{R}\pi\left(\theta_{0}\right)T^{\frac{d}{2}}\left|\Sigma_{\ell,T}\right|^{-\frac{1}{2}}=\overline{R}\pi\left(\theta_{0}\right)\left|\Omega\right|^{-\frac{1}{2}} \tag{1.46}$$

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

When Ω is not diagonal, we first note that we can decompose the likelihood of θ into the marginal likelihood of θ_{1} and the conditional likelihood of $\theta_{2:d}|\theta_{1}$, both of which remain Gaussian. An eigendecomposition of $\theta_{2:d}|\theta_{1}$ reparameterizes the

likelihood in terms of orthogonal components, after which we can apply the proof for diagonal Ω. Finally, the proof follows through with general $\overline{X}_T$ since $\overline{X}_T \to \theta_0$. $\square$

1.B Sequential Monte Carlo

1.B.1 Implementation details

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

$$\gamma_i = \mathbb{E}_p\left[\gamma\left(\theta\right)\right] + \left(\tilde{\gamma} - \mathbb{E}_p\left[\gamma\left(\theta\right)\right]\right)\left(\frac{i}{N_{\text{SMC}}}\right)^{\nu}. \tag{1.47}$$

A smaller value of ν corresponds to larger initial steps, and smaller steps toward the end of the SMC algorithm. Substituting $\tilde{\gamma}_i$ into the right-hand side of (1.7) for each i yields a sequence of distortions:

$$M_i\left(\theta\right) \propto \exp\left[\lambda_i L\left(\theta|X\right)\left(\gamma\left(\theta\right) - \tilde{\gamma}_i\right)\right], \tag{1.48}$$

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

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

$$m_i\left(\theta\right) \equiv \frac{M_i\left(\theta\right)}{M_{i-1}\left(\theta\right)} \propto \exp\left[L\left(\theta|X\right)\left[\lambda_i\left(\gamma\left(\theta\right) - \tilde{\gamma}_i\right) - \lambda_{i-1}\left(\gamma\left(\theta\right) - \tilde{\gamma}_{i-1}\right)\right]\right]. \tag{1.49}$$

Given λ_{i-1} and draws from π_{i-1} and p_{i-1}, the only unknown remaining is λ_i, which we solve for from (1.8), which we rewrite as:

$$\tilde{\gamma} = \mathbb{E}_{p_{i-1}}\left[\frac{\exp\left[L\left(\theta|X\right)\left[\lambda_i\left(\gamma\left(\theta\right)-\tilde{\gamma}_i\right)-\lambda_{i-1}\left(\gamma\left(\theta\right)-\tilde{\gamma}_{i-1}\right)\right]\right]}{\mathbb{E}_{p_{i-1}}\left[\exp\left[L\left(\theta|X\right)\left[\lambda_i\left(\gamma\left(\theta\right)-\tilde{\gamma}_i\right)-\lambda_{i-1}\left(\gamma\left(\theta\right)-\tilde{\gamma}_{i-1}\right)\right]\right]\right]}\gamma\left(\theta\right)\right]. \quad (1.50)$$

With a sufficiently large N_{SMC}, importance sampling of p_i from p_{i-1} is feasible. We can then solve for λ_i in (1.50) 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_{P}, the number of Metropolis-Hastings mutation steps N_{MH}, and the number of SMC steps N_{SMC}. Relative to Herbst and Schorfheide (2014), it is more important here to have a large number of particles, so that the expectations in equation (1.50) are evaluated accurately when solving for λ_i. Similarly, N_{MH} must be sufficiently large in order to solve for λ_i accurately. If λ_i is computed accurately, the posterior mean of γ evaluated from the particles before and after the mutation step should be identical up to sampling error. We can check if N_{SMC} is sufficiently large by ensuring that at each stage, the distribution of m_i is well-behaved in the tails.

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

$$\widetilde{M}\left(\theta^*\right) \propto \exp\left[\tilde{\lambda}\widehat{L}\left(\theta^*|X\right)\left(\hat{\gamma}\left(\theta^*\right)-\tilde{\gamma}\right)\right] \quad (1.51)$$

instead of (1.14). 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

$$\frac{M(\theta^*)}{\widetilde{M}(\theta^*)} \propto \frac{\exp\left[\lambda^* L^*(\theta^*|X)\left(\gamma^*(\theta^*) - \tilde{\gamma}\right)\right]}{\exp\left[\tilde{\lambda}\widehat{L}(\theta^*|X)\left(\hat{\gamma}(\theta^*) - \tilde{\gamma}\right)\right]} \tag{1.52}$$

where we can solve for λ^* using the constraint (1.13). Once we solve for λ^*, we can begin with the approximate worst-case draws, then use the selection and mutation steps from Algorithm 1 to obtain draws from the true worst-case distributions.

1.B.2 Evaluating relative entropy

We now use the output from the SMC in Algorithm 1 together with Lemma 1.4 to evaluate the relative entropy of the worst-case prior relative to the original prior.

To use Lemma 1.4, recall that the intermediate weights in Algorithm 1 have the form:

$$m_i(\theta) \equiv \frac{M_i(\theta)}{M_{i-1}(\theta)} = \frac{\kappa_i}{\kappa_{i-1}} \exp\left[L(\theta|X)\left[\lambda_i\left(\gamma(\theta) - \tilde{\gamma}_i\right) - \lambda_{i-1}\left(\gamma(\theta) - \tilde{\gamma}_{i-1}\right)\right]\right]. \tag{1.53}$$

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

$$\frac{\kappa_i}{\kappa_{i-1}} = \mathbb{E}_{\pi_{i-1}}\left[\exp\left[L(\theta|X)\left[\lambda_i\left(\gamma(\theta) - \tilde{\gamma}_i\right) - \lambda_{i-1}\left(\gamma(\theta) - \tilde{\gamma}_{i-1}\right)\right]\right]\right], \tag{1.54}$$

from which we obtain:

$$\mathbb{E}_\pi\left[M_i(\theta)\log M_i(\theta)\right] = \sum_{\iota=1}^{i} \log\frac{\kappa_\iota}{\kappa_{\iota-1}} \tag{1.55}$$

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

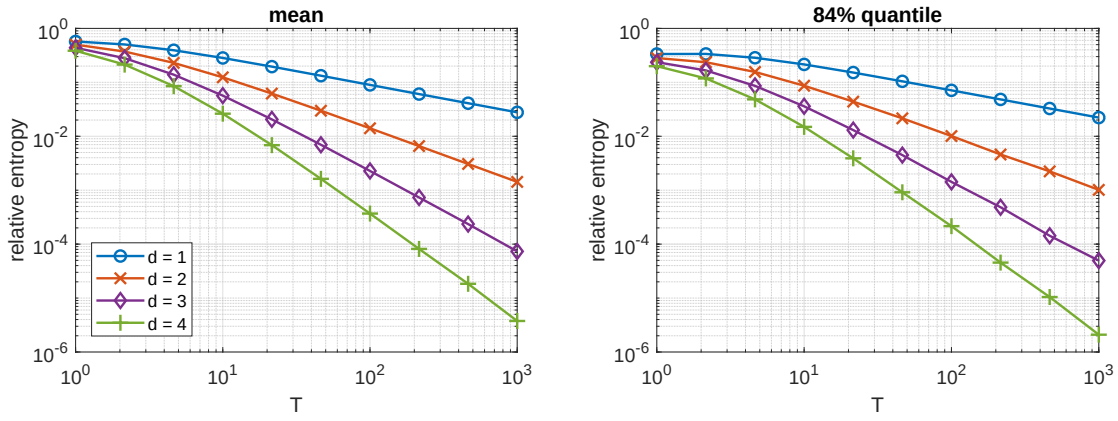


Figure 1.9: Relative entropy for given number of observations in Gaussian location model with $\theta \sim \mathcal{N}(0, I)$, $\overline{X}_T = 0$, $d \in \{1, ..., 4\}$. **Left:** increase in posterior mean of θ_1 by one posterior standard deviation; **Right:** increase in 84% quantile of θ_1 by one posterior standard deviation.

1.C Gaussian location model finite sample performance

We now show that Lemma 1.1 provides a good approximation for the relative entropy in the Gaussian location model even for relatively small values of T. In each case, we set $\Sigma_\pi = \Omega = I$ and show that the relative entropy needed to:

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

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

approximately scales with $T^{-\frac{d}{2}}$ and $\pi(\theta_0)$, as predicted by Lemma 1.1.

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 $10^5 + 1$ 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 1.9 shows the relative

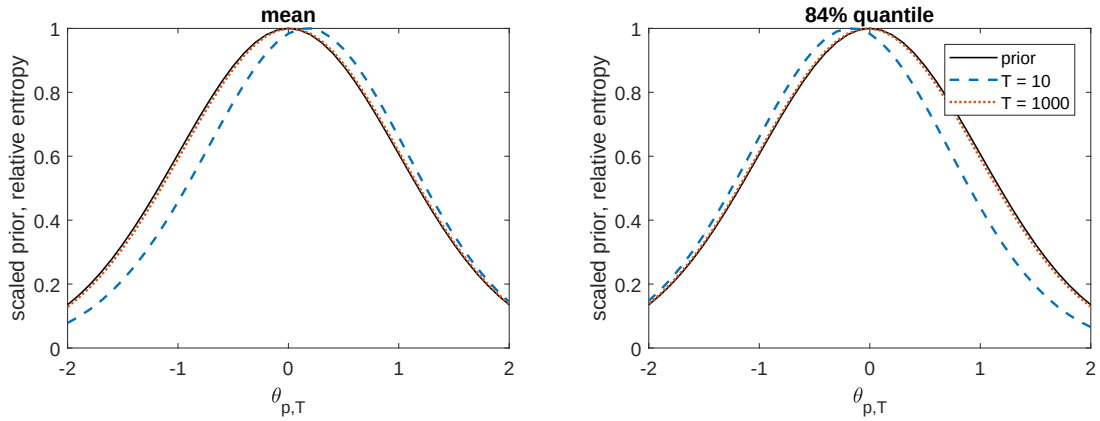


Figure 1.10: Relative entropy for given posterior mean with $\theta \sim \mathcal{N}(0,1)$, $\theta_{p,T} \in [-2,2]$. **Left:** increase in posterior mean of θ_1 by one posterior standard deviation; **Right:** increase in 84% quantile of θ_1 by one posterior standard deviation.

entropy for different values of T and d. Firstly, notice that Lemma 1.1 gives an accurate approximation for the behavior of relative entropy as T increases for $T \geq 10$, with a gradient of $-\frac{d}{2}$ when we plot $\log \mathcal{R}$ against $\log T$. When T is small, the data has not yet swamped the prior. The resulting greater sensitivity to the prior is reflected in the relative entropy for $T < 10$ being small compared to when $T \geq 10$, relative to what is predicted by Lemma 1.1.

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 1.10 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 $\overline{X}_T$.

1.D Smets and Wouters (2007)

1.D.1 Gaussian mixture approximation of posterior

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

$$\hat{\theta}_i = \begin{cases} \theta_i & \theta_i \in (-\infty, \infty) \\ \log(\theta_i) & \theta_i \in (0, \infty) \\ \log\left(\frac{1}{\theta_i} - 1\right) & \theta_i \in (0, 1) \end{cases} \tag{1.56}$$

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 1.11 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.

1.D.2 Sequential Monte Carlo

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

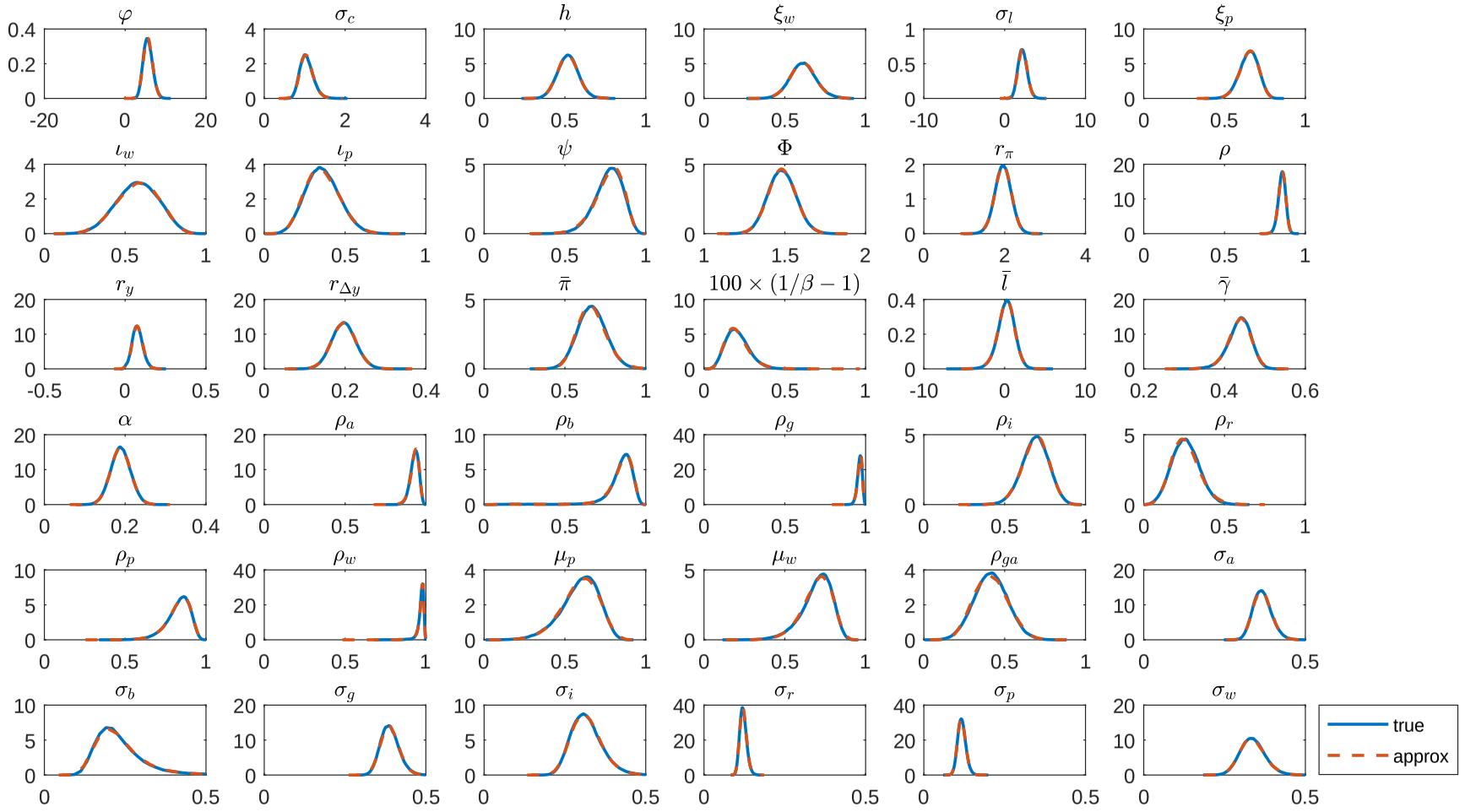


Figure 1.11: Original and approximate marginal posteriors. Blue solid line corresponds to true posterior; red dashed line corresponds to Gaussian mixture approximation.

the Taylor rule prior, for the lower bound we choose:

$$\tilde{\psi}_h^{0.16} = \begin{cases} Q_p\left[\psi_h; 0.16\right] - \frac{1}{8}\sigma_p\left[\psi_h\right] & h \leq 2 \text{ or } h \geq 16 \\ Q_p\left[\psi_h; 0.16\right] - \frac{1}{5}\sigma_p\left[\psi_h\right] & 13 \leq h \leq 15 \\ Q_p\left[\psi_h; 0.16\right] - \frac{1}{4}\sigma_p\left[\psi_h\right] & h = 3 \text{ or } 10 \leq h \leq 12 \\ Q_p\left[\psi_h; 0.16\right] - \frac{1}{3}\sigma_p\left[\psi_h\right] & \text{otherwise} \end{cases} \tag{1.57}$$

and for the upper bound we choose

$$\tilde{\psi}_h^{0.84} = \begin{cases} Q_p\left[\psi_h; 0.84\right] + \frac{1}{2}\sigma_p\left[\psi_h\right] & h \leq 4 \\ Q_p\left[\psi_h; 0.84\right] + \frac{3}{4}\sigma_p\left[\psi_h\right] & 5 \leq h \leq 8 \\ Q_p\left[\psi_h; 0.84\right] + \sigma_p\left[\psi_h\right] & \text{otherwise} \end{cases} \tag{1.58}$$

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

$$\tilde{\psi}_h^{0.16} = \begin{cases} Q_p\left[\psi_h; 0.16\right] - \frac{1}{240}\sigma_p\left[\psi_h\right] & h \leq 2 \\ Q_p\left[\psi_h; 0.16\right] - \frac{1}{80}\sigma_p\left[\psi_h\right] & h = 3 \\ Q_p\left[\psi_h; 0.16\right] - \frac{1}{40}\sigma_p\left[\psi_h\right] & 10 \leq h \leq 21 \\ Q_p\left[\psi_h; 0.16\right] - \frac{1}{20}\sigma_p\left[\psi_h\right] & \text{otherwise} \end{cases} \tag{1.59}$$

and for the upper bound we choose

$$\tilde{\psi}_h^{0.84} = \begin{cases} Q_p\left[\psi_h; 0.84\right] + \frac{1}{2}\sigma_p\left[\psi_h\right] & h \leq 4 \\ Q_p\left[\psi_h; 0.84\right] + \frac{3}{4}\sigma_p\left[\psi_h\right] & 5 \leq h \leq 8 \\ Q_p\left[\psi_h; 0.84\right] + \sigma_p\left[\psi_h\right] & \text{otherwise} \end{cases} \tag{1.60}$$

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 (1.47):

$$\tilde{\psi}^q_{h,i} = Q_p\left[\psi_h; q\right] + \left(\tilde{\psi}^q_h - Q_p\left[\psi_h; q\right]\right)\left(\frac{i}{N_{\text{SMC}}}\right)^{\nu} \tag{1.61}$$

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.

1.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.

Chapter 2

Estimating the Effects of Demographics on Interest Rates: A Robust Bayesian Perspective

2.1 Introduction

With the secular changes in both demographics and interest rates in developed countries over the past thirty years, quantifying the effects of an aging population on the interest rate has become crucial for forecasting and policy analysis. While economic theory proposes several channels through which demographics can impact the interest rate, the size of these effects is determined by underlying structural parameters whose values need to be disciplined by data.

This paper shows that data on capital and household savings over the life cycle are important for accurately quantifying the effects of demographics. Without these data, the estimated effects of demographics on interest rates in an overlapping generations (OLG) model vary substantially with the prior for the discount rate, intertemporal elasticity of substitution, and capital depreciation rate, even though other parameters

are well-identified. We establish these insights using robust Bayesian analysis. First, we use Bayesian methods to estimate the structural parameters of a parsimonious OLG model, and show that the discount rate, intertemporal elasticity of substitution, and capital depreciation rate are not well-identified by the data. We then use nonparametric prior sensitivity analysis techniques from Ho (2019) to show how the prior for these parameters influences the estimated effects of demographics on interest rates. These results show that calibrating or estimating OLG models without data on capital and savings can lead to misleading conclusions about the quantitative effect of demographics on both interest rates and other macroeconomic variables.

Our econometric framework is a vector autoregression (VAR) with a structural break that captures the secular changes in demographics, interest rates, and other macroeconomic aggregates. The long-run averages of the VAR are determined by the steady state of an OLG model that captures the economic effects of demographics on the natural interest rate. We use the OLG model from Gertler (1999), which captures the main effects of demographics on macroeconomic variables but remains tractable and transparent. Given the structural parameters, the model implies an interest rate that we refer to as the natural interest rate, as in Laubach and Williams (2003). To measure the historical effect of demographics, we take the difference between the estimated natural interest rate and the counterfactual natural interest rate that would have arisen if specific demographic parameters had not experienced a structural break. In particular, we study the effects of population growth, life expectancy, and the relative productivity of retirees. No further restrictions are placed on the dynamics of the data, allowing flexibility in the high frequency variation.

We use Bayesian methods to fit the model to eight macroeconomic and demographic time series from 1980-2013 for Japan, where secular macroeconomic and demographic changes have been especially pronounced. While the starkness of these changes improves identification, there remains substantial posterior uncertainty about

both the effects of demographics and the underlying parameters. Our estimates of the realized and counterfactual natural interest rates have relatively wide 95% credible intervals of up to 1.5 percentage points. While the time-varying parameters and structural break date are tightly identified by the data, the likelihood for the time-invariant parameters is dispersed. In particular, the posterior standard deviations of these time-invariant parameters are at most 35% smaller than the prior standard deviations. Since the effects of demographics in the model are determined by these underlying parameters, the results suggest that the estimated effects are strongly influenced by the prior, and that more data are required to pin down the effects.

To formalize this result, we use the relative entropy prior sensitivity (REPS) methodology from Ho (2019) to determine how much the estimated effects of demographics on interest rates depend on the prior for the three fixed parameters — the discount rate, intertemporal elasticity of substitution, and capital depreciation rate. REPS considers a nonparametric set of priors that are close to the original prior in relative entropy, and seeks a *worst-case* prior that changes the posterior results the most. The worst-case priors yield bounds that contain the posterior credible intervals for any prior in that set, and identify parts of the prior that are important for the posterior estimates. The methodology thus measures how much the effects of demographics depend on the fixed parameters, and how informative the likelihood is in the relevant directions.

The REPS analysis shows that a small change in the prior can lead to substantial changes in the estimated effects of demographics on interest rates. Many features of the worst-case prior are plausible ex ante, emphasizing the concern that the estimated effects are not well-identified by the data. To show that the results are not dependent on the specific class of priors considered by REPS, we reestimate the model using a different parameteric prior that is motivated by the REPS analysis, but chosen to be consistent with existing calibrations of OLG models. The new prior changes

the posterior mean for the effects of demographics by over one posterior standard deviation, or up to 0.6 percentage points.

The large role of the prior for the posterior in our baseline estimation highlights the need to include additional data to discipline the parameters. Intuitively, the effects of demographics on interest rates are determined by the capital demand and savings supply functions. Capital and savings data provide measures of quantities, while the interest rate provides a measure of prices. For example, by measuring savings and investment responses to changes in macroeconomic conditions, aggregate data on capital improves estimates of the intertemporal elasticity of substitution and depreciation rate. In addition, savings rates over the lifecycle reveal how households trade off current and future consumption, which is informative for the discount rate and intertemporal elasticity of substitution.

As an illustration, we incorporate the capital-output ratio into the original set of time series data. The additional data reduces the standard deviation of the intertemporal elasticity of substitution and depreciation rate by two and five times respectively, but does not change the precision of the discount rate estimate. We obtain a more precise estimate for the effect of the relative productivity of retirees on interest rates, corroborating the result from REPS analysis that the depreciation rate is especially important for determining this effect. In addition, the new data shifts the posterior means of the effects of demographics by up to 1.5 percentage points. We thus propose that future work uses data on both capital and savings over the lifecycle to discipline the discount rate, intertemporal elasticity of substitution, and depreciation rate, so that the implied effects of demographics are more precisely and accurately estimated.

Related literature. This paper shows that when quantifying the effects of demographics on interest rates using OLG models, it is essential to pin down the discount

rate, intertemporal elasticity of substitution, and capital depreciation rate using the appropriate data on capital and savings. Numerous papers calibrate OLG models (e.g. Aksoy et al. (2015); Carvalho et al. (2016); Gagnon et al. (2016); Eggertsson et al. (2017)). In particular, Ikeda and Saito (2014) and Sudo and Takizuka (2018) calibrate OLG models for Japan. The analysis here suggests that the numerical results of these papers may be sensitive to their calibration strategies, as there is a wide range of possible parameter values consistent with the time series in our baseline estimation. Future calibrations should therefore discipline the parameters using additional restrictions from capital and household savings data.

More broadly, there have been numerous papers seeking to understand the secular decline in interest rates in developed countries through a range of theories. For example, Summers (2015) connects the decline to secular stagnation, while Caballero et al. (2008) present a model of global imbalances. Krugman (1998) discusses the case of Japan. The empirical analysis here allows for both demographic and non-demographic parameter changes. The relative contribution of each component is determined jointly by the data and the OLG model structure. Including additional determinants of the interest rate would not change the qualitative conclusions of this paper. Regardless of the sophistication of the model, one expects the discount rate, intertemporal elasticity of substitution and depreciation rate to be important for determining the effects of demographics. The estimates would only become sharper if the model introduces new restrictions on these parameters.

Our results are also relevant for the large literature that seeks to quantify the effect of demographics on other aspects of the economy. For example, Fujiwara and Teranishi (2008) and Wong (2018) consider the transmission of monetary policy shocks, while Attanasio et al. (2007) and Kitao (2017) study the social security reform. Many of these papers use OLG models similar to the one here, and one may expect

their quantitative results to also rely on the discount rate, intertemporal elasticity of substitution and depreciation rate.

Finally, the econometric framework here contributes to a large literature on estimating models using specific frequencies of the data. We discuss the connections in detail in Section 2.2.

Outline. The organization of the paper is as follows. I introduce the econometric model in Section 2.2 and outline the OLG model in Section 2.3. Section 2.4 describes the Bayesian estimation and Section 1.2 describes the prior sensitivity analysis. In Section **??**, I discuss how additional data can improve the estimates for the effects of demographics. Section 2.7 shows that the results are robust to using a parametric prior motivated by the REPS analysis. Section 1.7 concludes.

2.2 Econometric framework

We now present the econometric framework that we use to disentangle the low frequency secular changes in the data from the high frequency variation. The secular change is modeled as a change in the steady state of an OLG model that occurs in response to a structural break in the macroeconomic and demographic parameters. The data is annual and the high frequency variation is modeled as a mean zero VAR(1) process appended to a constant term whose values are determined by the OLG model's steady states. The empirical analysis will examine how the data informs estimates of both the time-varying and fixed parameters, and how the effects of demographics on interest rates in the OLG model depend on these parameters.

2.2.1 Setup

We observe annual data y_t that follows the model:

$$y_t = \mu(s_t) + v_t \tag{2.1}$$

$$v_t = \Phi(s_t) v_{t-1} + u_t \tag{2.2}$$

$$u_t \sim \mathcal{N}(0, \Sigma(s_t)) \tag{2.3}$$

where $s_t = \mathbf{1}\{t \geq t^*\}$ is an indicator for the structural break. The data y_t has a mean μ and dynamics v_t, which follow a mean zero VAR(1) process. The mean and VAR process change after the structural break in period t^*. The structural change in μ captures the secular changes in levels observed in the data. The structural break in the VAR process in period t^* allows for differences in the dynamics, including changes in the volatility, persistence, or comovement of the data after the structural break. Including a second break does not materially affect results.

We are interested in modeling the secular changes in the data, represented by μ. To that end, we assume that μ is determined by the steady state of the OLG model with parameters that depend on s_t. We impose these cross-equation restrictions on μ in order to account for economic forces when estimating the secular movements in y_t. The OLG model also allows us to compute counterfactuals for μ. On the other hand, we allow for flexibility in the dynamics by modeling v_t as a reduced form VAR. Since the OLG model plays no role in the high frequency variation of the data, we can avoid modeling the frictions necessary to match business cycle dynamics, which keeps the OLG model tractable both for estimation and interpretation. Instead, we isolate features of the data that the simple OLG model is best suited to explain. Nevertheless, the results here extend to more sophisticated OLG models.

Estimating the econometric framework (2.1)-(2.3) is analogous to fitting the steady state of the OLG model to data from the start and end of the sample. However,

instead of using data from a specific year, we estimate these steady states based on the full time series of data. By estimating the break date t^*, we allow the data to determine which periods correspond to the steady state before and after the structural break. The VAR flexible accounts for correlation across time and variables.

2.2.2 Relation to other approaches

The model (2.1)-(2.3) decomposes the data into a long-run mean component $\mu(s_t)$ and a high frequency component v_t. The long-run mean captures the low frequency changes in the data. We now discuss related approaches to extracting the low frequency variation from the data.

By using the OLG model to discipline only the long-run averages instead the full variation of the data, the approach described above is similar to the literature on limited information Bayes estimation of DSGE models. For example, Christiano et al. (2010) estimate a DSGE model using only impulse responses to a subset of identified structural shocks, thus focusing on the effect of these shocks in the model while ignoring predictions of the model that are of less interest to the researcher. Since we are interested in secular changes in the economy here, we only use the steady state of the model to discipline the means μ. Sala (2015) estimates a DSGE model in the frequency domain, restricting the estimation to components of the likelihood corresponding to particular frequencies. In (2.1)-(2.3), the VAR captures the high frequency variation, and the structural estimates are informed by the low frequency component consisting of two long-run means, with a structural break date that is estimated.

In related work, Del Negro and Schorfheide (2004) use information from a DSGE model to inform the priors of a VAR. In our econometric framework, instead of using the OLG model to construct the prior, we directly model the mean of the VAR as coming from the steady state of the OLG model. While the object of interest for Del

Negro and Schorfheide (2004) is the VAR estimates, in this paper we are interested in the estimates of the OLG model's structural parameters and implied counterfactuals.

If we had fitted both low and high frequency variation to the OLG model, we would have required a rich set of frictions in the OLG model to produce realistic dynamics, then incorporate these dynamics in the estimation. Smets and Wouters (2007) do this in the context of a representative agent DSGE model. Log-linearizing such a model would yield the system (2.1)-(2.3) without the regimes s_t, but would impose restrictions on Φ and Σ arising from the equilibrium conditions in the structural model. For a given set of structural parameters, our approach yields the same μ if the frictions affect the dynamics but not the steady state. However, we remain agnostic about the modeling of frictions. Moreover, the approach here ensures that the structural parameter estimates are primarily driven by the low-frequency variation in the data, which is the object of interest. Sala (2015) shows that the posterior estimates depend on the frequencies used in the estimation, suggesting that the distinction between μ and v_t is important for our purposes.

One could also filter out the low frequency fluctuations in the data using weighted averages without reference to any model (e.g. Hodrick and Prescott (1997), Müller and Watson (2018)), then use these low frequency fluctuations as input to estimate the parameters of the OLG model. However, we find that the high frequency observations are informative for estimating the time-varying parameters as well as the structural break date.

2.3 Overlapping generations model

The OLG model is taken from Gertler (1999). It is a neoclassical growth model with endogenous labor supply, stochastic retirement and death, as well as recursive preferences. This model is a parsimonious way to capture the key economic forces

linking demographics to interest rates. In this section, we describe the setup of the model and focus on equilibrium conditions that are essential for understanding the relationship between demographics and interest rates. The full set of steady state equilibrium conditions is listed in Appendix 2.A, and Gertler (1999) provides details on derivations.

2.3.1 Households

Life cycle and population growth

Each individual is born as a worker. At the end of each period, a worker has probability $1-\omega$ of retiring and a retiree has probability $1-\gamma$ of dying. Denoting the stock of workers by N_t, we assume $(1-\omega+n)\,N_t$ new workers are born each period, so that the workforce grows at a constant rate n. The ratio of retirees to workers ψ_t satisfies:

$$\psi_{t+1}N_{t+1} = \gamma\psi_t N_t + (1-\omega)\,N_t. \tag{2.4}$$

Retirees

Retirees have preferences:

$$V_t^{r,i,j} = \left[\left[\left(C_t^{r,i,j}\right)^{\nu}\left(1-L_t^{r,i,j}\right)^{1-\nu}\right]^{\rho} + \beta\gamma\left(V_{t+1}^{r,i,j}\right)^{\rho}\right]^{\frac{1}{\rho}}, \tag{2.5}$$

where i and j indicate the birth and retirement cohort respectively. $C_t^{r,i,j}$ and $L_t^{r,i,j}$ are consumption and labor of retirees respectively. These preferences imply risk neutrality with respect to wealth, which allows for aggregation across cohorts. In addition, the intertemporal elasticity of substitution, $\sigma = \frac{1}{1-\rho}$, controls the desire to smooth consumption over time, which is a key force for determining households' propensity to save and hence the interest rate. The survival rate γ augments the discount factor β, so that retirees have an effective discount rate of $\beta\gamma$.

We assume perfect annuity markets (Yaari (1965); Blanchard (1985)) that insure against the risk arising from the uncertain time of death. Each retiree places her wealth in a mutual fund that invests its proceeds. The surviving fraction of retirees γ receive all the returns, while those who die receive nothing. Retirees are therefore subject to the budget constraint:

$$A_{t+1}^{r,i,j} = \frac{R_t}{\gamma} A_t^{r,i,j} + W_t \xi L_t^{r,i,j} - C_t^{r,i,j}, \tag{2.6}$$

where $A_t^{r,i,j}$ is the level of assets and W_t is the wage per effective unit of labor. $\xi \in [0,1]$ is the productivity of a retiree relative to a worker and R_t/γ is the return to a surviving retiree.

Workers

Workers have preferences:

$$V_t^{w,i} = \left[\left[\left(C_t^{w,i} \right)^{\nu} \left(1 - L_t^{w,i} \right)^{1-\nu} \right]^{\rho} + \beta \left[\omega V_{t+1}^{w,i} + (1-\omega) V_{t+1}^{r,i,t+1} \right]^{\rho} \right]^{\frac{1}{\rho}} \tag{2.7}$$

and are subject to the budget constraint:

$$A_{t+1}^{w,i} = R_t A_t^{w,i} + W_t L_t^{w,i} - C_t^{w,i}. \tag{2.8}$$

The worker has probability $1-\omega$ of becoming a retiree in the next period. The presence of the retiree's value function $V_{t+1}^{r,i,t+1}$ in (2.7) implies that workers take retirement into account when making savings decisions. Therefore, the value of γ and ξ implicitly enter into the worker's decisions.

Aggregation

Retirees. Retiree consumption is linear in the sum of assets and discounted expected value of labor income, with marginal propensity to consume $\epsilon_t \pi_t$ that does not depend on cohort. This allows us to write aggregate consumption as:

$$C_t^r = \epsilon_t \pi_t \left(\frac{R_t}{\gamma} A_t^r + H_t^r \right), \tag{2.9}$$

where human wealth:

$$H_t^r = W_t \xi L_t^r + \frac{\psi_t}{\psi_{t+1}(1+n)} \frac{H_{t+1}^r}{R_{t+1}} \tag{2.10}$$

is defined as the presented discounted value of expected labor income for the entire population of retirees. The first term is the labor income earned in period t, while the $\frac{\psi_t}{\psi_{t+1}(1+n)}$ term accounts for the population growth of retirees.

Workers. Similarly, workers have common a marginal propensity to consume π_t. Aggregate worker consumption can be written as:

$$C_t^w = \pi_t \left(R_t A_t^w + H_t^w \right), \tag{2.11}$$

where human wealth defined as:

$$H_t^w = W_t L_t^w + \omega \frac{1}{1+n} \frac{H_{t+1}^w}{R_{t+1} \Omega_{t+1}} + (1-\omega) \xi^{\nu-1} \epsilon_{t+1}^{\frac{1}{1-\sigma}} \frac{1}{\psi_{t+1}(1+n)} \frac{H_{t+1}^r}{R_{t+1} \Omega_{t+1}}. \tag{2.12}$$

The first term is the labor income of workers. The next two terms are the expected present discounted value of human wealth in period $t+1$. Ω_{t+1} augments R_{t+1} to account for the possibility of retiring at the end of the period and is defined as:

$$\Omega_{t+1} = \omega + (1-\omega) \xi^{\nu-1} \epsilon_{t+1}^{\frac{1}{1-\sigma}}. \tag{2.13}$$

Workers adjust their valuation of future labor income in response to two changes that happen when they retire — they become less productive by an exogenous factor of ξ, and their marginal propensity to consume increases by an endogenously determined factor of ϵ_{t+1}.

State variables. The linearity of the consumption decisions implies that we do not need to keep track of individual cohorts when solving for the aggregate steady state or dynamics. Instead, we can aggregate across workers and retirees respectively, which reduces the number of states and makes Bayesian estimation possible.

2.3.2 Production and market clearing

Aggregate output follows the Cobb-Douglas production function:

$$Y_t = (X_t L_t)^{\alpha} K_t^{1-\alpha}, \tag{2.14}$$

where $L_t \equiv L_t^w + \xi L_t^r$ and α determines the labor share. The labor augmenting productivity grows at a constant rate x and capital depreciates at rate δ. Market clearing for the capital and goods markets implies:

$$A_t = K_t \tag{2.15}$$

$$Y_t = K_{t+1} - (1-\delta) K_t + C_t^w + C_t^r \tag{2.16}$$

2.3.3 Structural break

We assume that the long-run mean $\mu(s_t)$ in equation (2.1) is determined by the steady state of the OLG model above. When the structural break occurs in period t^*, a subset of the structural parameters are redrawn from the same distribution that generated the parameters before t^*. This new set of parameters yields a new steady

state and long-run mean. The demographic parameters affected by the structural break are the working population growth n, the survival rate of retirees γ, and the relative productivity of retirees ξ. The non-demographic parameters affect are the productivity growth x, the labor share α, and the parameter ν controlling the disutility of labor.

2.3.4 Demographics and interest rates

An aging demographic affects the steady state interest rate in the OLG model through several channels. The strength of each of these channels depends nonlinearly on the structural parameters. In what follows, we drop the t subscript to denote steady state variables.

Firstly, as the share of retirees increases, the supply of savings decreases, thus raising the interest rate. This *savings composition channel* arises because retirees have a higher marginal propensity to consume, as can be seen from the steady state expressions:

$$\epsilon\pi = 1 - \left(\frac{R}{(1+x)^{1-\nu}}\right)^{\sigma-1} \beta^{\sigma}\gamma \tag{2.17}$$

$$\pi = 1 - \left(\frac{R\Omega}{(1+x)^{1-\nu}}\right)^{\sigma-1} \beta^{\sigma} \tag{2.18}$$

for retiree and worker marginal propensities to consume $\epsilon\pi$ and π respectively. Retirees have effective discount rate $\beta\gamma < \beta$, which induces them to consume more from their wealth. Moreover, workers save to smooth consumption into retirement, since retirement leads to a lower productivity and hence lower expected wealth. This force is captured by the Ω adjustment defined in (2.13) to the interest rate R in equation (2.18).

Secondly, increases in the survival rate γ or decreases in the relative productivity ξ cause decreases in the marginal propensity to consume within groups, which decreases

interest rates through an increase in the supply of savings. We refer to this as the *within-group savings channel.* When retirees have a higher probability of survival, their effective discount rate $\beta\gamma$ increases, which increases the incentive to save in order to smooth consumption. Workers anticipate the reduced propensity to consume as a retiree, reflected by the $V_t^{r,i,t+1}$ term in their continuation value (2.7), and respond by decreasing their marginal propensity to consume as well. A lower relative productivity of retirees ξ implies a greater drop in human wealth upon retirement, leading to increased saving by workers to smooth consumption.

Changes in demographics also affect interest rates through a *capital demand channel*, which is captured by the equilibrium condition for capital in steady state:

$$R = (1 - \alpha) k^{-1} + (1 - \delta), \tag{2.19}$$

where k is the steady state capital-output ratio K_t/Y_t. In particular, when the share of retirees increases or the relative productivity of retirees decreases, the average household in the economy is less productive. In the absence of adjustments to labor supply, this lowers the marginal product of capital $(1 - \alpha) k^{-1}$, which leads to a fall in the demand for capital that pushes interest rates downwards. On the other hand, when workers expect to have a longer or less productive retirement, they increase their labor supply in order to accumulate wealth for retirement. This can dampen or even reverse the capital demand effect.

Finally, there is a *general equilibrium channel*, as defined in Carvalho et al. (2016). In response to a decrease in interest rates from the savings composition, within-group savings, or capital demand channels, households decrease savings and firms increase their use of capital. These forces amplify the direct effect of the savings composition, within-group savings, and capital demand channels.

In section 2.5, we analyze the sensitivity of our results to the prior for the discount factor β, intertemporal elasticity of substitution σ, and capital depreciation rate δ, because these three parameters influence the effects described above. A higher discount rate reduces the incentive to smooth over time, since households place a lower weight on future utility. This greater discounting weakens the savings composition and within-group savings channels. Since the two channels produce opposite effects, the net effect of changing β depends on the parameters of the model. A higher intertemporal elasticity of substitution σ strengthens the effect of the discount factor β, but decreases the sensitivity of the worker's marginal propensity to consume π to the interest rate adjustment Ω. In particular, when $\sigma = 1$, Ω no longer shows up in (2.18), and the incentive to smooth consumption into retirement vanishes, reducing the savings composition and within-group savings effects. In addition, a larger σ decreases households' response to changes in the interest rate, thus reducing the general equilibrium effect. A higher capital depreciation rate δ increases the amount of investment necessary to maintain the steady state level of capital, hence amplifying the capital demand effect.

Any demographic change influences interest rate through a combination of the channels described above. The relative importance of these channels determines the equilibrium response of interest rates. To quantify the effects of demographics, one needs to use data to discipline the parameters in the model that control the strength of each channel. Without more formal analysis, it is hard to establish which combinations of parameter values are supported by the data and how these change the quantitative effects of demographics. These challenges remain or are exacerbated in larger OLG models.

In what follows, we estimate the model to quantify the posterior uncertainty of (β, σ, δ) as well as the rest of the model parameters. The prior sensitivity analysis

will then show that the data is not informative about (β, σ, δ), which implies a wide range of possible estimates for the effect of demographics on interest rates.

2.4 Bayesian estimation

We now use Bayesian methods to estimate the model (2.1)-(2.3), with μ determined by the steady state of the OLG model in Section 2.3. Given the prior, the posterior distribution concentrates around parameter values supported by the data. The posterior credible intervals indicate the uncertainty in the estimates given the prior and data. Section 2.5 uses prior sensitivity analysis to distinguish the contribution of the data from that of the prior.

2.4.1 Estimation

Data

We use annual data from 1980-2013 for Japan. We focus on data for Japan because the macroeconomic and demographic changes there have been especially pronounced.

The time series we observe are GDP growth, working population growth, the share of workers as a share of the population, the share of workers among the employed, the employment-population ratio, real wage growth, real interest rate, and labor share. In the data, we take the working population to be individuals between 15-64 years old, and take the retirement population to be individuals aged 65 and above. To match our choice of 15-64 year olds as workers, ω is calibrated to 0.98 so that workers in the model have an average working life of 50 years.

The data are chosen to be similar to what is used in the literature to calibrate models quantifying the effect of demographics on interest rates. For example, numerous papers calibrate the discount rate to match a given interest rate (e.g. Ikeda and Saito (2014); Carvalho et al. (2016); Eggertsson et al. (2017)). Our results will

		type	mean	std dev
$100\left(\beta^{-1}-1\right)$	discount rate	Gamma	3.50	0.50
σ	intertemporal elasticity of substitution	Normal	0.35	0.15
δ	depreciation rate	Beta	0.08	0.02

Table 2.1: Prior for fixed structural parameters (β, σ, δ).

show that without additional data on capital and savings, these calibrations can produce misleading results, because the effects of demographics are not well-identified. Moreover, we show in Section 2.6 that adding the capital-output ratio into the estimation can sharpen the estimates for the effects of demographics.

Several papers (e.g. Attanasio et al. (2007); Kitao (2017); Sudo and Takizuka (2018)) include the capital-output ratio in their calibration, but do not directly target the interest rate. The results in these papers suggest that omitting interest rates may also lead to misleading results. The calibrations in Attanasio et al. (2007) and Kitao (2017) produce interest rates that are roughly 5 percentage points higher the real return on government bonds[1], while Sudo and Takizuka (2018) find a real interest rate that is up to 2.5 percentage points higher than the natural interest rate estimated using the methodology of Laubach and Williams (2003). Since we wish to estimate the effects of demographics on interest rates, it is natural to include interest rates in the baseline estimation. The results in Section 2.6 show that even if we use data on both interest rates and the capital-output ratio, we still require additional data for accurate estimates.

We also include data that directly inform us of the path of the time-varying parameters. Table 2.6 in Appendix 2.B describes the mapping between observables and model variables.

Prior

We focus on the prior for (β, σ, δ), which we report in Table 2.1, and describe the rest of the prior in Appendix 2.B. The prior for (β, σ, δ) are of particular interest because our results will show that the data are especially uninformative about these three parameters, and the prior sensitivity analysis in Section 2.5 will show that the prior for these parameters is important for the estimated effects of demographics.

Our baseline prior is based on values used in existing calibrations of the Gertler (1999) OLG model. The prior for β implies a mean discount rate of 3.5%, which is close to existing calibrations of the same model. For σ, we choose a prior with mean 0.35 and standard deviation 0.15 to match the calibration of similar models used in papers based on the same model (e.g. Gertler (1999); Fujiwara and Teranishi (2008) set $\sigma = 0.25$, while Carvalho et al. (2016); Ferrero et al. (2017) set $\sigma = 0.50$). Finally, the prior for δ has a mean of 0.08 and standard deviation of 0.02, which allows for the range of calibrations in the literature (e.g. Gertler (1999); Fujiwara and Teranishi (2008); Carvalho et al. (2016) set $\delta = 0.10$, while Kara and von Thadden (2016) set $\delta = 0.05$). We keep the prior independent across parameters, as is often done in the estimation of structural models. Figure 2.10 in Appendix 2.B compares the prior to these calibrations from the literature.

While we have chosen this prior to match the literature that is closest to the OLG model here, one could have picked a different prior that was equally plausible. For example, the empirical literature has found a wide range of estimates for the intertemporal elasticity of substitution σ ranging from 0 to 2. Similarly, the depreciation rate one obtains can depend on subjective choices about the measurement process, such as how much to aggregate across different types of capital (see e.g. Fraumeni (1997); Feenstra et al. (2015)). In Section 2.5, we show that the particular prior we have

[1] Attanasio et al. (2007) argue that such an interest rate is reasonable when compared to the return on equity.

chosen does impact posterior inference, which motivates the call for incorporating additional data into the estimation.

Markov chain Monte Carlo

To sample from the posterior, we use a Metropolis-within-Gibbs algorithm described in Appendix 2.B. We take 2×10^5 burn-in draws, which we use to calibrate the proposal density. We then take 2.5×10^6 draws, keeping every 25th draw to save memory. To check for convergence, we partition the draws into four blocks and ensure that the posterior moments and marginals are similar across blocks.

2.4.2 Results

Structural break and long-run means

Figure 2.1 plots the estimated long-run means $\mu(s_t)$, with 68% error bands. The structural break date t^* is estimated to be 1991. The long-run means before and after the break are distinct even after accounting for the error bands. Intuitively, μ is identified by 11 periods of data before t^* and 33 periods of data after t^*. In addition, the OLG model places cross-equation restrictions on the comovement of μ. In particular, variables such as the interest rate and employment-population ratio are endogenous objects in the OLG model. The long-run means μ are therefore jointly identified by the data and the equilibrium conditions of the OLG model. The fact that the long-run means track the data indicates that there exist parameter combinations that allow the model to fit the data well.

Natural interest rate and counterfactuals

We now focus on our objects of interest — the natural interest rate and the effect of demographics.

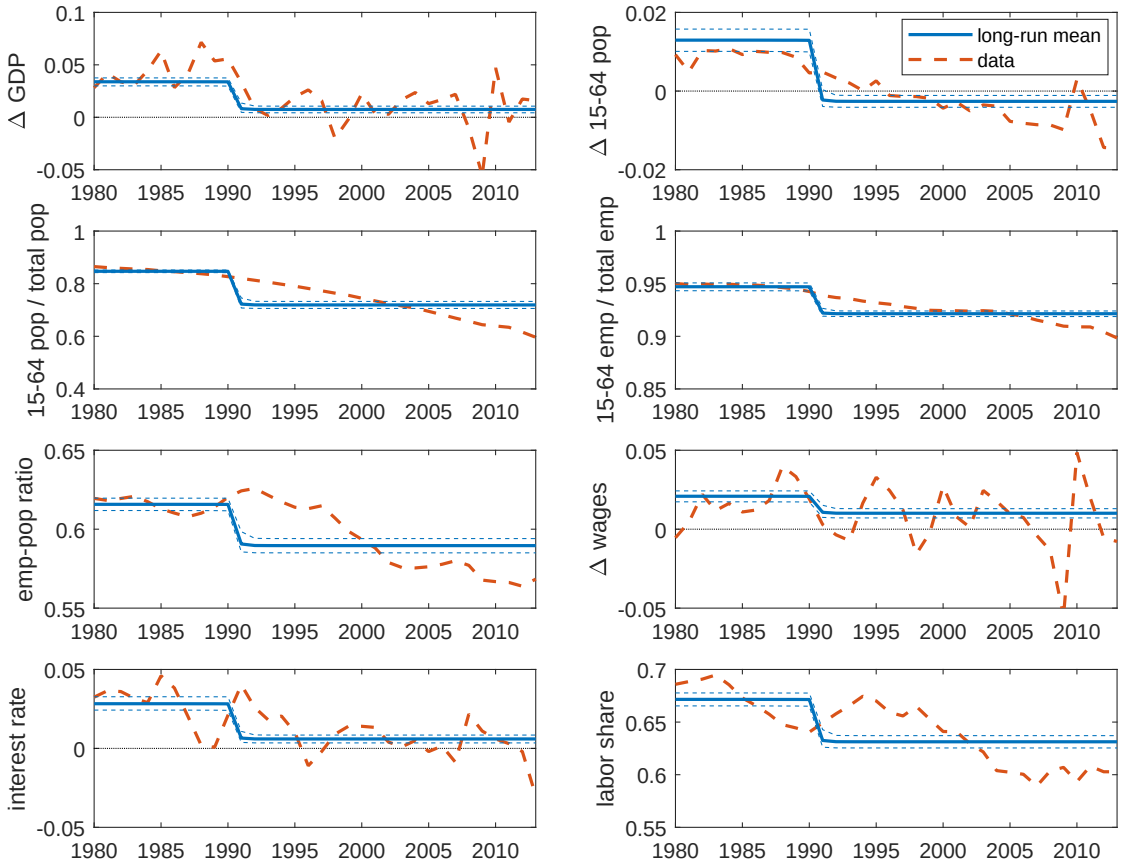


Figure 2.1: Estimated long-run means. **Blue lines:** median (solid) and 68% error bands (dashed) of long-run means; **Red dashed lines:** data.

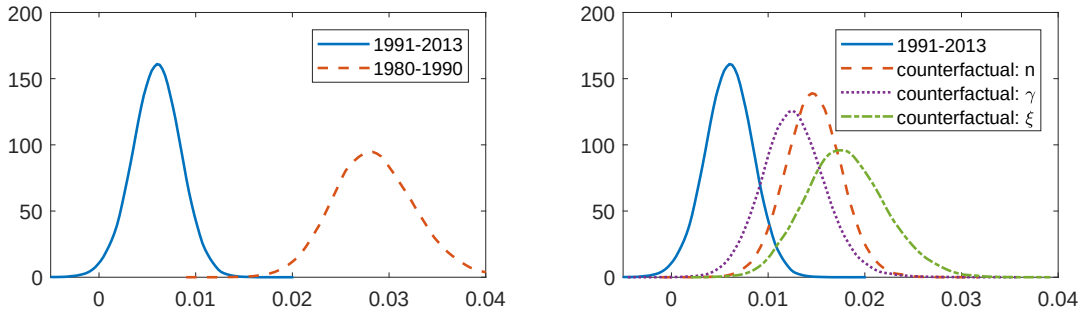


Figure 2.2: Posterior of natural interest rate and counterfactuals. **Left:** natural interest rate before and after structural break; **Right:** natural interest rate after structural break and counterfactuals.

Natural interest rate. We define the natural interest rate to be the interest rate implied by the OLG model for given parameter values. In particular, define $\theta^* \equiv (\beta, \sigma, \delta)'$ and $\zeta \equiv (x, \nu, \alpha)'$. For any set of parameters $(\theta^*, \zeta, n, \gamma, \xi)$, we can compute the steady state interest rate implied by the structural model, which we denote by $R(\theta^*, \zeta, n, \gamma, \xi)$. For each period t , define the natural interest rate:

$$R_t \equiv R(\theta^*, \zeta(s_t), n(s_t), \gamma(s_t), \xi(s_t)).$$

This is similar to the existing literature extracting the natural interest rate (e.g. Laubach and Williams (2003); Holston et al. (2016); Del Negro et al. (2017)) using equilibrium conditions of a DSGE model. Here, we focus on two long-run average interest rates since we are concerned with the long-run trend in interest rates. The focus on the long run allows the OLG model to exclude frictions that are normally included in DSGE models used for extracting the natural interest rate at business cycle frequencies.

The left panel of Figure 2.2 shows that the posterior mean of the natural interest rate decreased from 2.85% to 0.60% after the structural break. The posteriors of the two interest rates are distinct, providing statistical evidence that the real interest rate has declined since the 1980s. The widths of the 68% credible intervals are 0.84 percentage points before the break and 0.49 percentage points after the break, which is comparable to the error bands that Del Negro et al. (2017) find for the low frequency component of the natural interest rate in the United States. These credible intervals are wide enough to imply substantial uncertainty about the effects of a given path of interest rates.

Counterfactuals. We use the counterfactual natural interest rate to quantify the contribution of the population growth rate n, survival rate γ, and relative productivity

of retirees ξ. In particular, we consider the counterfactual interest rates:

$$\widehat{R}_t^n \equiv R\left(\theta^*, \zeta\left(s_t\right), \hat{n}, \gamma\left(s_t\right), \xi\left(s_t\right)\right)$$
$$\widehat{R}_t^\gamma \equiv R\left(\theta^*, \zeta\left(s_t\right), n\left(s_t\right), \hat{\gamma}, \xi\left(s_t\right)\right)$$
$$\widehat{R}_t^\xi \equiv R(\theta^*, \zeta\left(s_t\right), n\left(s_t\right), \gamma\left(s_t\right), \hat{\xi})$$

where we pick the counterfactual parameter values $\hat{\cdot}$ to be the median estimate for the parameter in 1980. These counterfactuals change one of the demographic parameters, while keeping all other parameters identical. Given the Monte Carlo draws for each of the parameters, we can construct the posterior distributions for $\widehat{R}_t^n$, $\widehat{R}_t^\gamma$ and $\widehat{R}_t^\xi$.

The right panel of Figure 2.2 shows the posterior distribution of $\widehat{R}_T^n$, $\widehat{R}_T^\gamma$ and $\widehat{R}_T^\xi$, i.e., the counterfactual natural interest rate at the end of the sample had one of the three demographic parameters remained at its median value before the structural break. They have means of 1.47%, 1.27% and 1.80% respectively, thus explaining between one third to one half of the decline in interest rates. Note that the effects are not additive — the effect from changing more than one parameter is not the sum of the effects of changing each of those parameters individually. The widths of the 68% credible intervals range are 0.58, 0.64 and 0.83 percentage points respectively.

Structural parameters

The above estimates reveal substantial posterior uncertainty in the estimates for both the natural interest rate and the counterfactuals. To better understand the sources of uncertainty, we now turn to the estimates of the underlying structural parameters.

Fixed parameters. Figure 2.3 shows that the marginal posteriors for the fixed parameters (β, σ, δ) are relatively close to their priors, indicating a dispersed likelihood. This is especially true for the discount rate β and capital depreciation rate δ. Intuitively, (β, σ, δ) are identified from two steady state observations. The

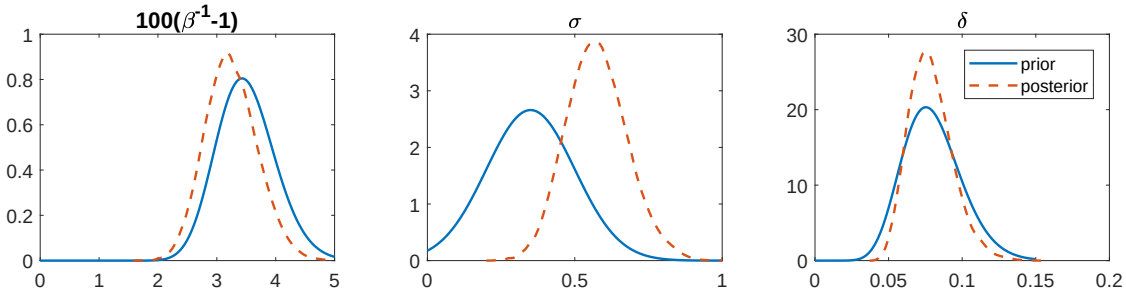


Figure 2.3: Marginal priors and posteriors of fixed parameters. **Solid blue line:** prior; **Dashed red line:** posterior.

dispersed posteriors arise from the OLG model not placing substantial restrictions on the possible values of (β, σ, δ) individually given the estimated long-run means. Nevertheless, the likelihood does place restrictions on the joint distribution of (β, σ, δ). For example, the correlation between σ and δ is -0.66.

One implication of these results is that one of the three fixed parameters could be well-identified given the other two parameters. However, fitting these parameters jointly to the data could yield a much wider range of possible values. Section 2.5 shows that the parameter values supported by the data can imply varied effects of demographics in the OLG model.

The dispersed marginal posteriors are not unique to the model and data here. For example, **?** state that they calibrate δ because it is difficult to estimate with the data they use. In addition, they obtain relatively diffuse estimates for β and σ relative to their prior even though they use a longer time series of quarterly data and estimate the model using all frequencies in the data. While we lose information from business cycle fluctuations, we also have a more parsimonious model that has fewer parameters to be estimated.

Time-varying parameters. In contrast to the fixed parameters, the time-varying parameters have marginal posteriors that are substantially tighter than their priors, as shown in Figure 2.4. Moreover, the posteriors for each of the parameters before

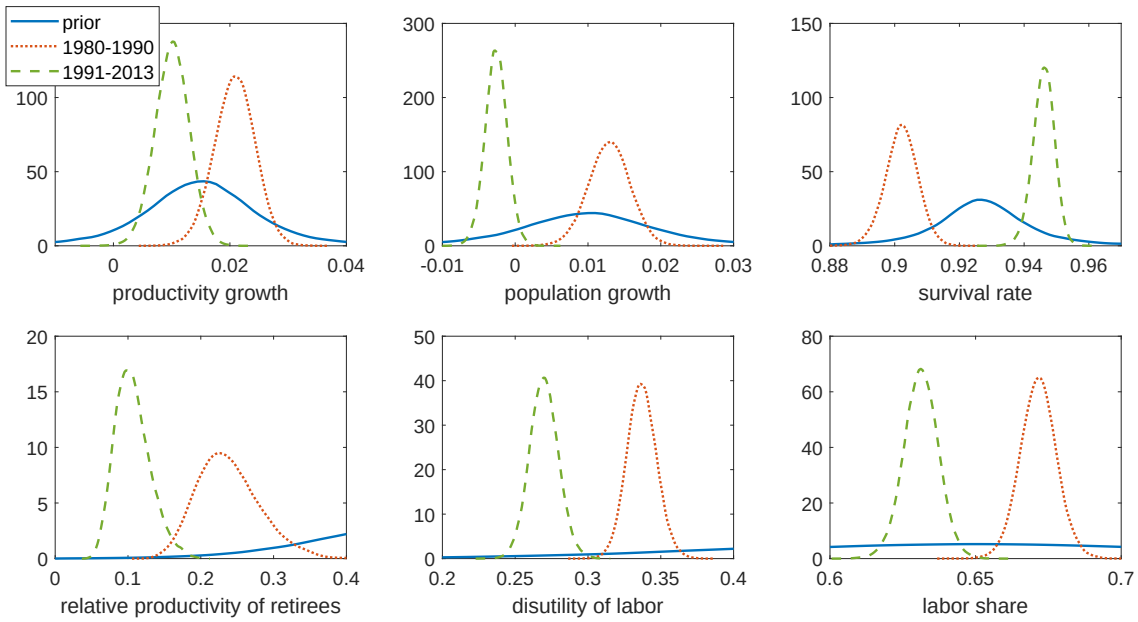


Figure 2.4: Marginal priors and posteriors of time-varying parameters. **Solid blue line:** prior; **Dotted red line:** posterior for parameter before structural break; **Dashed green line:** posterior for parameter after structural break.

and after the structural break are distinct from each other. The tightness of the posterior reflects the fact that the long-run averages of the data are tightly estimated relative to the prior, and each of the time-varying parameters is closely connected to one of the time series.

Working population growth is directly observed, while the survival rate can be inferred from the fraction of workers in the population given population growth. In the OLG model, productivity growth is equal to real wage growth as well as per capita GDP growth. The relative productivity of retirees is closely related to the fraction of workers among the employed, and the disutility of labor is similarly connected to the employment-population ratio.

In general, the estimates match the historical narrative of Japan's economy in 1980-2013. The decline in productivity growth was a symptom of the lost decade. The decrease in population growth corresponds to the declining birth rates since the early 1970s, while the increase in survival rate matches the growth in life expectancy.

The relative productivity of retirees is estimated to be lower after the break date due to the decrease in the fraction of the workforce below the age of 65. Even though the aggregate employment rate decreased, the employment rate by age group has increased since the 1980s, indicating a decrease in the disutility of labor. Finally the decline in the labor share has been well documented by Karabarbounis and Neiman (2014) and others.

2.5 Prior sensitivity

We now analyze the sensitivity of the estimated counterfactual interest rates to changes in the prior of (β, σ, δ) using the relative entropy prior sensitivity (REPS) methodology introduced by Ho (2019). REPS explores a nonparametric set of priors that are close to the original prior in terms of relative entropy, and finds bounds for the posterior results. In particular, we compute 68% *robust credible intervals*, defined as bounds that contain the equal-tailed 68% posterior credible interval for any prior in a given set of priors. In particular, we find the upper (lower) bound for the posterior 84% (16%) quantile given a set of priors that is close to the original prior.

Section 2.3 showed theoretically that the three parameters (β, σ, δ) have a role in determining the effect of demographics on interest rates, while Section 2.4 showed that these parameters are not well-identified by the data. Because the interest rate is determined jointly by all the equilibrium conditions, it is difficult to predict how these estimated effects change if we jointly change the prior of (β, σ, δ). Changing the prior, as opposed to exogenously specifying new values of (β, σ, δ), ensures that the new posterior continues to be disciplined by the data, respecting the joint likelihood of both (β, σ, δ) and the time-varying parameters. While it is impossible to reestimate the model for all possible priors on (β, σ, δ), REPS provides bounds on our results for a nonparametric set of priors close to the original prior.

In what follows, we present the REPS methodology. We then show how REPS reveals high sensitivity of the estimated effects to the prior for (β, σ, δ), and identifies features of the prior that are important for the estimated effects.

2.5.1 Methodology

Denote the full vector of parameters by θ, the prior by π and the posterior by p. As before, define $\theta^* \equiv (\beta, \sigma, \delta)'$ and let the marginal prior and posterior of θ^* be π^* and p^* respectively. We are interested in how much the qth quantile of a function $\psi(\theta)$ can change as we change the marginal prior π^*.

In our setting, ψ is the difference between the counterfactual and realized natural interest rate:

$$\psi(\theta) \in \left\{ \widehat{R}_T^n - R_T, \widehat{R}_T^\gamma - R_T, \widehat{R}_T^\xi - R_T \right\} \tag{2.20}$$

These three definitions of ψ measure the effects of population growth n, the survival rate γ, and the relative productivity of retirees ξ on the interest rate. Each choice of ψ depends on the prior in a distinct way because n, γ and ξ influence the interest rate in different ways. The conclusions we yield using these choices of ψ can be extended to other functions of of the structural parameters that may arise in settings such as forecasting or policy analysis.

We solve for an alternative prior $\tilde{\pi}^*$ for θ^* that minimizes the qth quantile:

$$\min_{\tilde{\pi}^*} \tilde{\psi} \text{ s.t. } \mathcal{R} \geq \mathbb{E}_{\tilde{\pi}^*} \left[\frac{\tilde{\pi}^*(\theta^*)}{\pi^*(\theta^*)} \log \left(\frac{\tilde{\pi}^*(\theta^*)}{\pi^*(\theta^*)} \right) \right] \tag{2.21}$$

$$q = \mathbb{E}_{\tilde{p}} \left[\mathbf{1} \left\{ \psi(\theta) \leq \tilde{\psi} \right\} \right] \tag{2.22}$$

where $\tilde{p}$ is the posterior arising from the worst case marginal prior $\tilde{\pi}^*$, keeping the conditional prior of the remaining parameters and the likelihood the same. Solving (2.21)-(2.22) involves searching for a *worst-case* prior that minimizes the qth quantile

of ψ. Replacing the minimization with maximization instead yields the upper bound for the quantile. (2.22) states that $\tilde{\psi}$ is the qth posterior quantile of ψ. The constraint in (2.21) limits the relative entropy or Kullback–Leibler divergence of $\tilde{\pi}^*$ relative to π^* to be less than some constant $\mathcal{R} \geq 0$, restricting us to choose among priors that are statistically difficult to distinguish from the original prior π^*. Problem (2.21)-(2.22) does not place parametric restrictions on the alternative prior. In particular, the feasible set of priors includes nonparametric distributions that could introduce correlations across parameters even though we started with a parametric and independent prior on θ^*. In our application here, we seek bounds for the 68% equal-tailed credible interval for effects of n, γ and ξ. We therefore take $q \in \{0.16, 0.84\}$ in (2.22), with minimization replaced with maximization for $q = 0.84$.

The optimization problem (2.21)-(2.22) yields the solution:

$$\frac{\tilde{\pi}^*(\theta^*)}{\pi^*(\theta^*)} \propto \exp\left[\lambda L^*(\theta^*|X)\left(\mathbb{E}_{\tilde{p}}\left[\mathbf{1}\left\{\psi(\theta) \leq \tilde{\psi}\right\}|\theta^*\right] - q\right)\right], \tag{2.23}$$

where $L^*(\theta^*|X)$ is the marginal likelihood of θ^*, and $\lambda \geq 0$ is a scalar that can be solved for. REPS therefore reduces the infinite dimensional problem of searching for a worst-case prior to a one-dimensional problem. The $\mathbb{E}_{\tilde{p}}\left[\mathbf{1}\left\{\psi(\theta) \leq \tilde{\psi}\right\}|\theta^*\right]$ term in (2.23) indicates that the worst-case prior places more weight on values of θ^* for which the conditional probability of $\psi \leq \tilde{\psi}$ is greater. The L^* term arises because it is optimal to concentrate the distortions in the high likelihood region of θ^*. 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.

To gauge the size of $\mathcal{R}$, Ho (2019) recommends the rule:

$$\mathcal{R} = 1.6^d r \frac{\pi(\mu_p)}{\pi(\mu_\pi)} \sqrt{\frac{|\Sigma_\ell|}{|\Sigma_\pi|}} \tag{2.24}$$

$$\Sigma_\ell \equiv \left(\Sigma_p^{-1} - \Sigma_\pi^{-1}\right)^{-1} \tag{2.25}$$

where μ_p and μ_π are the posterior and prior means for θ^* respectively, Σ_p and Σ_π are the posterior and prior variances for θ^* respectively, and d is the dimensionality of θ^*. Σ_ℓ is a measure of how dispersed the likelihood is. The choice of r determines how large the relative entropy is, with $r < 0.05$ corresponding to small levels of relative entropy[2]. We shall pick r to correspond to a one to two standard deviation change in the quantiles on average.

We implement the above computations using the sequential Monte Carlo and numerical approximations described in Ho (2019). See Appendix 2.C for details.

2.5.2 Results

Robust credible intervals

Figure 2.5 shows that the robust credible intervals are wide, indicating that a small change in the prior for (β, σ, δ) can produce a large change in the posterior estimates for the effects of demographics. In particular, with $r = 0.005$, the robust credible intervals are between 3.7 to 5.5 posterior standard deviations wider than the corresponding original credible intervals. Relative to the $r = 0.05$ benchmark, $r = 0.005$ corresponds to a very small set of priors, which shows that the estimated effects of demographics are very sensitive to the prior.

[2]Suppose ψ is a linear combination of θ^*, and we estimate $\theta^* \equiv (\beta, \sigma, \delta)'$ from a large number of observations $\theta^* + \varepsilon_t$, with $\varepsilon_t \sim \mathcal{N}(0, \Omega)$ and Ω is known. Then $r = 0.05$ and $r = 0.005$ would correspond to changes in the quantile of approximately 0.4 and 0.1 posterior standard deviations respectively. Ho (2019) shows that if the variance of ε_t is equal to the prior variance of θ^*, then ten observations is sufficient for this to be a good approximation.

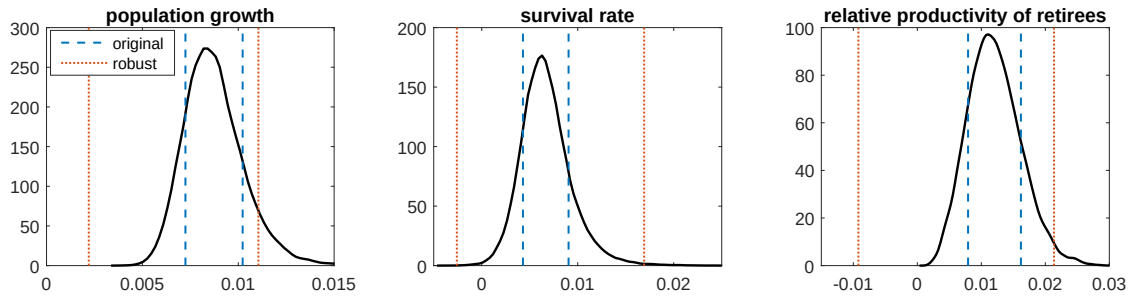


Figure 2.5: Robust credible intervals for difference between counterfactual and realized natural interest rate, with $r = 0.005$. **Dashed blue line:** original 68% credible interval; **Dotted red line:** robust 68% credible interval; **Solid black line:** posterior density.

There are two reasons for the high degree of prior sensitivity. Firstly, as suggested in Figure 2.3, the likelihood for (β, σ, δ) is dispersed. Defining Σ_π and Σ_ℓ as in (2.25), we find $\sqrt{|\Sigma_\ell| \,/\, |\Sigma_\pi|} = 2.2$, which indicates that the likelihood is more dispersed than the prior. Secondly, (β, σ, δ) are strong predictors of the demographic effects $\widehat{R}_T^n - R_T$, $\widehat{R}_T^\gamma - R_T$, and $\widehat{R}_T^\xi - R_T$. Quadratic regressions of $\widehat{R}_T^n - R_T$, $\widehat{R}_T^\gamma - R_T$, and $\widehat{R}_T^\xi - R_T$ on $(\beta^{-1}, \sigma, \delta)$ yield R-squares of 0.67, 0.74, and 0.93 respectively. Changing the prior for (β, σ, δ) thus results in a large change in the posterior for both (β, σ, δ) and the effects of demographics.

These results therefore show that the data are uninformative about structural parameters that are important for determining the effects of demographics on interest rates in the OLG model. Moreover, because the dispersion of the likelihood is a feature of the model and data only, the prior sensitivity applies to any other function of the model parameters that is well-predicted by (β, σ, δ).

Worst-case posteriors

To understand how the estimates depend on each part of the prior, we compare the original and worst-case posteriors for the 16% and 84% quantiles, shown in Figures 2.6 and 2.7 respectively. The worst-case priors for the 16% quantiles place greater weight on parameter values that produce smaller effects of demographics, while the

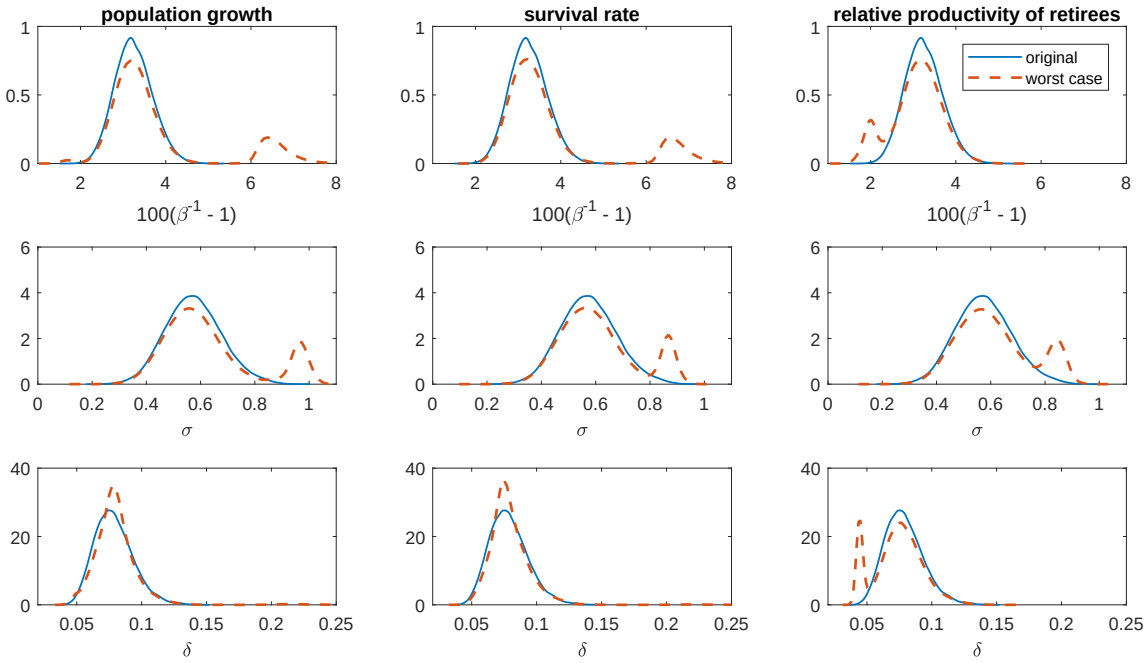


Figure 2.6: Original and worst-case posteriors generating a one posterior standard deviation decrease in the 16% quantile. **Left:** $\widehat{R}_T^n - R_T$; **Center:** $\widehat{R}_T^\gamma - R_T$; **Right:** $\widehat{R}_T^\xi - R_T$.

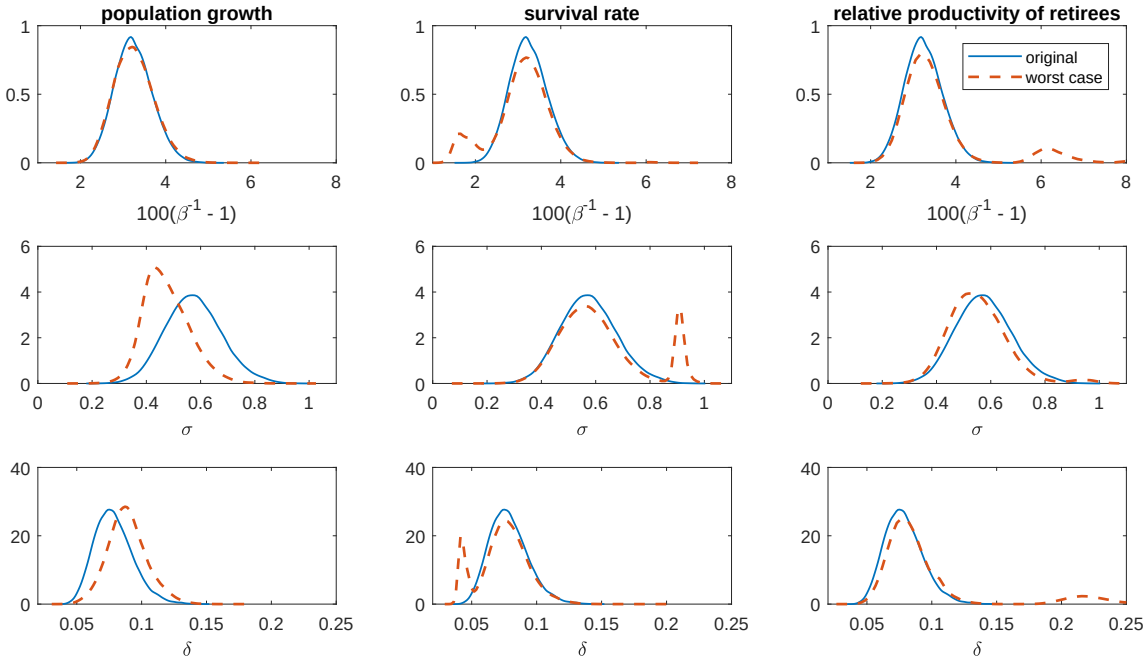


Figure 2.7: Original and worst-case posteriors generating a one posterior standard deviation increase in the 84% quantile. **Left:** $\widehat{R}_T^n - R_T$; **Center:** $\widehat{R}_T^\gamma - R_T$; **Right:** $\widehat{R}_T^\xi - R_T$.

worst-case priors for the 84% quantiles place greater weight on parameter values that imply larger effects. While these are not necessarily the only distortions that shift the effects in the desired direction, each worst-case prior provides one example of an alternative prior that has a large impact on the estimated effects of demographics. The worst-case distortions are also indicative of the relative importance of the savings composition, within-group savings, capital demand, and general equilibrium effects, as well as their sensitivity to changes in (β, σ, δ).

Population growth. The worst-case posteriors corresponding to the effect of population growth involve especially large distortions to the marginal of the intertemporal elasticity of substitution σ. In particular, the worst-case distortions place more weight on large (small) values of σ to decrease (increase) the estimated effect of population. A change in the working population growth n affects the interest through the savings composition, capital demand, and general equilibrium channels. Increasing σ reduces the worker's incentive to save for retirement, which decreases the savings composition effect, thus strengthening the net effect of population growth. However, a large σ also dampens the effect of population growth by decreasing the general equilibrium effect. That the worst-case posterior concentrates around large values of σ to decrease the 16% quantile suggests that the general equilibrium effect is especially sensitive to σ.

The distortions for β^{-1} are asymmetric. To decrease the 16% quantile, the worst-case distortion places more weight on large values of β^{-1}. In contrast, the marginal for β^{-1} is relatively unchanged for the 84% quantile. The asymmetry highlights the nonlinearity in the mapping from (β, σ, δ) to the effect of population growth. For example, the effect of β on the marginal propensities to consume (2.17) and (2.18) is amplified by a larger value of σ. As a result, changing β has a smaller effect when accompanied by the decrease in σ for the 84% quantile worst-case distortion.

For both the 16% and 84% quantiles, the distortion of δ is relatively small, suggesting that the capital demand channel is less important than the effects originating from household savings decisions. The small increase in δ for the 84% quantile arises due to the negative correlation of -0.66 between σ and δ.

Survival rate. The worst-case posterior for the 16% quantile for the effect of the survival rate is similar to that of population growth, while the distortions for the 84% quantile differ sharply. To increase the 84% quantile for the effect of the survival rate, the worst-case distortions put greater weight on small values of β^{-1}, large values of σ, and small values of δ.

Decreasing β^{-1} increases the within-group saving effect. As we decrease β^{-1}, we increase the change in the effective discount rate $\beta\gamma$ arising from a change in the survival rate γ. Retirees thus increase savings by a greater amount in response to the same increase in γ. Moreover, decreasing β^{-1} increases the worker's incentive to smooth consumption into retirement. As a result, workers' marginal propensity to consume declines more in response to an increase in γ.

On its own, increasing σ serves to dampen the effect of the survival rate on interest rates. However, increasing σ also amplifies the effects of changing β. That the worst-case distortions for the 84% quantile involves an increase in σ suggests that the interaction with β is more important than the direct effect of σ. The nonlinearity emphasizes the importance of studying the joint distribution and effect of (β, σ, δ) rather than analyzing each parameter independently.

The worst-case distortions for the 84% quantile result in a new posterior mode for δ around 0.05. The decrease in δ dampens the capital demand channel. This is consistent with an increase in the effect of the survival rate γ because the increase in worker labor supply more than offsets the increased fraction of retirees in the economy. As a result, the average labor productivity rises, which increases the

marginal product of capital. Decreasing δ reduces the resulting increase in capital demand, thus strengthening the effect of γ on interest rates.

Relative productivity of retirees. The worst-case posterior for the 16% quantile for the effect of the relative productivity ξ is similar to that of the 84% quantile for the effect of the survival rate γ. These distortions therefore increase the effect of γ but decrease the effect of ξ. The contrasting dependence on the prior shows that the estimates of both the absolute and relative magnitudes of the different effects of demographics are sensitive the prior.

The worst-case posterior indicates the importance of the capital demand channel for determining the effect of the relative productivity of retirees. In addition, it shows that the capital demand effect serves to decrease interest rates in response to a decrease in ξ even though it increases interest rates in response to a decrease in γ.

Decreasing β^{-1} increases the incentive for workers to increase their labor supply in response to a decrease in ξ, hence decreasing the downward pressure on interest rates from the capital demand channel. This effect dominates the increase in the within-group savings effect arising from the reduction of β^{-1}. The decrease in δ also dampens the capital demand effect. On the other hand, the increase in σ decreases the incentive for workers to increase their labor supply, but also decreases the within-group savings effect. The effect of σ on the latter channel dominates. The worst-case posterior for the 84% quantile also suggests an important role for the capital demand channel, as more weight is put on large values of β^{-1} and δ.

The fact that the marginal distortions push the various effects of ξ in different directions emphasizes the nonlinear dependence on (β, σ, δ). The analysis here accounts for this nonlinearity while remaining disciplined by the likelihood. In particular, the endogenous labor supply response to changes in ξ and γ are disciplined by the data on the fraction of workers among the employed and the employment-population ratio.

2.6 Implications for estimation and calibration

We have shown that the time series data used here are not informative about (β, σ, δ), and that these parameters are important for estimating the effects of demographics on interest rates. More data is therefore needed to inform the analysis about these parameters, in order for estimates of the effects to be determined primarily by the data rather than the prior. In contrast, the time-varying parameters are well-identified by the data. To quantify the effects of demographics on interest rates, one should therefore focus on including additional data to discipline the values of (β, σ, δ).

The dependence of our estimates on the prior for (β, σ, δ) is not unique to the model considered here. The savings composition, within-group savings, capital demand, and general equilibrium channels are present in many OLG models that have been used to quantify the effects of demographic changes on the interest rate. Additional channels such as financial frictions (Ikeda and Saito (2014); Wong (2018)), social security schemes (Muto et al. (2016); Sudo and Takizuka (2018)), or international capital flows (Brooks (2003); Attanasio et al. (2006)) also depend on (β, σ, δ). These extensions will only tighten the likelihood of (β, σ, δ) if they place additional cross-equation restrictions that discipline the range of plausible values for (β, σ, δ). In such cases, it would be important to understand the sensitivity of these restrictions to the details of the extensions and to show empirical support for the relevant parts of the model.

The importance of (β, σ, δ) informs us of the data needed to tighten our estimates of the effects of demographics. To better identify β and σ, one can incorporate data on saving rates over the life cycle. In particular, such cross-sectional data can help us to identify how households respond to their lifecycle tranjectory of wages, which helps to quantify the savings composition effect. The evolution of these savings patterns over time provides information on the exact response of households' marginal propensities to consume to changes in life expectancy or future wages, which in turn helps to

identify the strength of the within-group savings effect. Including aggregate data on capital and investment, such as the capital-output ratio, can help identify σ and δ by providing the response of savings and investment to interest rate changes.

An alternative approach would be to include a longer time series or use a panel of countries. Increasing the length of the time series would increase the effective number of steady state observations if we observe structural breaks prior to 1980. However, constructing a sufficiently long time series to narrow estimates on (β, σ, δ) may be challenging given the relative scarcity of data before 1950. Including a larger panel of countries would similarly allow us to observe more regimes. For such an exercise to narrow our estimates of (β, σ, δ), we require that (β, σ, δ) are identical or relatively similar across countries.

The results here suggest that when calibrating OLG models, one should consider a range overidentifying restrictions to pin down (β, σ, δ). For example, for a given σ and δ, even though it is possible to pin down β given a particular interest rate, our analysis has shown that the data support a range of possible interest rates one can use. Once one acknowledges the uncertainty in σ and δ, one has an even wider range of possible combinations. Our results show that these parameter combinations can produce different conclusions. Additional data is required to determine the appropriate parameter values with more precision.

2.6.1 Including data on capital-output ratio

To illustrate the effect of including the appropriate data, we now study how including the capital-output ratio changes the estimation and prior sensitivity results. First, we reestimate the model using the capital-output ratio along with the original eight time series, with the original prior from Section 2.4. Next, we repeat the REPS analysis. Both exercises show that the capital-output ratio is highly informative about the effects of demographics.

Bayesian estimation

Table 2.2 shows that the inclusion of the capital-output ratio tightens the posterior estimate of the structural parameters. The inclusion of data on the capital-output ratio decreases the posterior standard deviation of δ by nearly five times, while the standard deviation of the intertemporal elasticity of substitution σ is halved. The standard deviation of the discount rate β^{-1} is roughly unchanged. The posterior means of δ and σ change by over two posterior standard deviations, while the mean of the discount rate is relatively unaffected by the inclusion of the capital-output ratio.

The changes in the posterior indicate that the capital-output ratio adds substantial information about the values of σ and δ, but not β. Equation (2.19) shows that given the interest rate $R-1$ and the labor share α, the capital-output ratio pins down the capital depreciation rate δ. Similarly, the responsiveness of savings to interest rates is determined by σ, as shown by marginal propensities to consume (2.17)-(2.18). Data on interest rates measures the price of savings and capital, while the capital-output ratio measures the response of the quantities. As a result, fitting the model to both of these variables gives us more precise estimates of σ and δ.

With the new data, the posterior standard deviation of the effect of the relative productivity of retirees ξ is halved, but the posterior standard deviation of the effects of population growth n and the survival rate γ are approximately tripled. The estimated effect of ξ is more precise because of the tighter estimate of δ. In particular, Figures 2.6 and 2.7 show that the worst-case posteriors for the effect of ξ involve large distortions in the marginal of δ, suggesting that the estimate of δ is important for the estimated effect of ξ. The marginal of δ was distorted less in the worst-case posteriors corresponding to the effects of n and γ. The posterior for the effects of n and γ become more dispersed partly because of the new estimates for the structural parameters. The decrease in σ and increase in δ make savings

		original			**with K/Y ratio**		
		mean	sd	68% C.I.	mean	sd	68% C.I.
Structural parameters							
$100\,(\beta^{-1}-1)$	discount rate	3.245	0.442	(2.804, 3.690)	3.443	0.427	(3.013, 3.873)
σ	IES	0.574	0.103	(0.471, 0.677)	0.282	0.053	(0.230, 0.335)
100δ	depreciation rate	7.861	1.485	(6.393, 9.317)	11.394	0.300	(11.098, 11.692)
Effects of demographics							
$100\left(\widehat{R}_T^n - R_T\right)$	population growth	0.875	0.159	(0.725, 1.023)	1.480	0.421	(1.122, 1.817)
$100\left(\widehat{R}_T^\gamma - R_T\right)$	survival rate	0.671	0.266	(0.429, 0.905)	2.140	0.808	(1.401, 2.858)
$100\left(\widehat{R}_T^\xi - R_T\right)$	relative productivity	1.206	0.412	(0.793, 1.618)	1.423	0.202	(1.240, 1.604)

Table 2.2: Posterior statistics for fixed structural parameters and effects of demographics, including capital-output ratio in data.

and investment more responsive to changes in the interest rate, which increases the general equilibrium channel. The increased general equilibrium channel amplifies the effects of demographics on interest rates, which increases the difference between the net effect of small and large partial equilibrium effects. This amplification outweighs the effect of the increased precision in the estimates.

The posterior means of the effects also change substantially with the introduction of the capital-output ratio into the data. The new estimates emphasize that the appropriate data can alter the conclusions of quantitative exercises on demographics and interest rates. In particular, the differences are large enough to impact policy analysis or forecasting.

Cross-sectional data on different age groups in the population can further tighten these estimates. With observations of how households discount over their life cycle, or how savings for different age groups respond to changes in macroeconomic variables, we have a more direct measure of the relative value households attach to current and future consumption. This information helps to pin down β and σ. However, with the inclusion of cross-sectional data, care should be taken to weigh the different series appropriately to avoid overfitting. In addition, the dynamics in the current econometric framework would have to be modified to account for the potentially large number of time series. This is left to future work.

Prior sensitivity

Figure 2.8 shows that the inclusion of the capital-output ratio results in robust credible intervals that are much closer to the original credible intervals, which indicates that the data are more informative in this estimation. For comparison, we consider the robust credible intervals with $r = 0.005$ with and without the capital-output ratio.

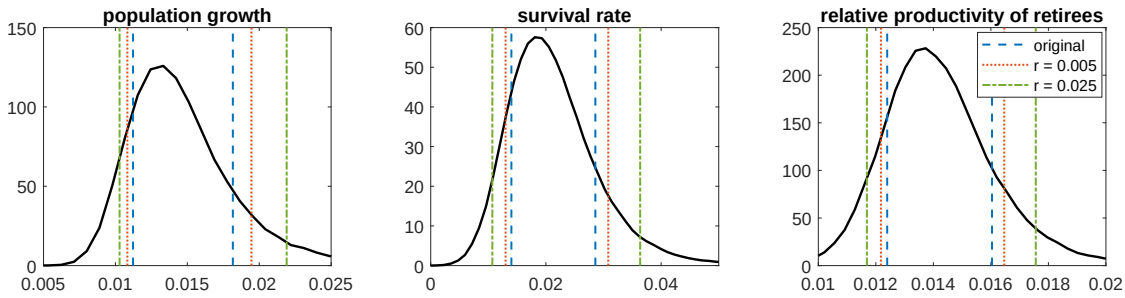


Figure 2.8: Robust credible intervals for difference between counterfactual and realized natural interest rate with $r \in \{0.005, 0.025\}$ in estimation with capital-output ratio. **Dashed blue line:** original 68% credible interval; **Dotted red line:** robust 68% credible interval with $r = 0.005$; **Dash-dot green line:** robust 68% credible interval with $r = 0.025$; **Solid black line:** posterior density.

Table 2.3 reports the statistic:

$$\rho(\psi) \equiv \frac{\chi_r(\psi)}{\chi_p(\psi)} - 1,$$

where χ_p and χ_r are the widths of the original and robust 68% credible intervals respectively for each definition of ψ in (2.20). ρ measures how much wider the robust credible interval is compared to the original credible interval, thus providing a measure of the sensitivity of the credible interval to the prior. Including the capital-output ratio decreases ρ from roughly 2.5 to 0.2. The REPS analysis therefore shows that the prior now plays a smaller role in determining the estimated effects of population growth n and the survival rate γ, even though the credible intervals for these estimates are wider than in the estimation without the capital-output ratio. These results support the earlier claim that these wider posterior credible intervals arise due to the new estimates for σ and δ.

We also find that the amount by which the robust credible intervals tighten is consistent with the worst-case distortions. In particular, we define ρ_k and ρ_o to be the values of ρ for the estimation with and without the capital-output ratio respectively, and take ρ_o/ρ_k as a measure of how much additional information the

		68% C.I. (χ_p)		**rob. C.I. (χ_r)**		**relative change (ρ)**		
		orig.	K/Y	orig.	K/Y	orig.	K/Y	ratio ($\frac{\rho_o}{\rho_k}$)
Effects of demographics								
$100\left(\widehat{R}_T^n - R_T\right)$	population growth	0.298	0.694	0.885	0.863	1.965	0.243	8.081
$100\left(\widehat{R}_T^\gamma - R_T\right)$	survival rate	0.475	1.458	1.951	1.782	3.105	0.223	13.954
$100\left(\widehat{R}_T^\xi - R_T\right)$	relative productivity	0.826	0.365	3.056	0.429	2.702	0.177	15.307

Table 2.3: Widths of 68% credible intervals and robust credible intervals with $r = 0.005$, for estimations with and without capital-output ratio. **Left panel:** width of 68% credible interval; **Center panel:** width of robust 68% credible interval; **Right panel:** difference between widths of credible interval and robust credible interval, normalized by width of 68% credible interval.

capital-output ratio provides for a given ψ. This ratio is largest for the effect of the relative productivity of retirees ξ. The worst-case posteriors for the effect of ξ involve largest distortions in the distribution of δ, suggesting that the importance of information about δ for the estimated effects of demographics is greatest for the effect of ξ. The change in ρ is greatest for the effect of ξ since the capital-output ratio tightens the estimate for δ more than β or σ. On the other hand, the worst-case posteriors for the effect of population growth n involve minimal distortions to the distribution of δ, consistent with ρ_o/ρ_k being the smallest for this effect.

Nevertheless, the robust credible intervals remain wide relative to the rule of thumb in Ho (2019). Figure 2.8 shows that with $r = 0.025$, which is half the $r = 0.05$ benchmark, the robust credible intervals are between 1.1 to 1.4 posterior standard deviations wider than the corresponding original credible intervals. The change is about 1.5 times as large as what one would expect from the $r = 0.05$ benchmark (see footnote 2 for details). These results show that even though the capital-output ratio greatly improves estimates, the augmented data remain relatively uninformative about the effects of demographics.

In particular, Table 2.4 shows that the worst-case distortions are concentrated in the prior of (β, σ) in this estimation with the capital-output ratio[3]. In contrast, Figures 2.6 and 2.7 show that without the capital-output ratio, the worst-case distortions in δ were relatively more important, especially for the estimated effect of the relative productivity ξ. Since δ is more precisely estimated with the inclusion of the capital-output ratio, the worst-case distortions indicate that it is now more difficult to change the estimated effects by changing the prior of δ. For future work estimating the effects of demographics, it is therefore important to incorporate additional data, such as household savings rates, that can further discipline β and σ.

[3]The only exception is the worst-case distortions corresponding to the 16% quantile for the effect of population growth n. However, since Figure 2.8 shows that this quantile is especially robust to changes in the prior, we focus our discussion on the other five worst-case priors.

		pop. growth (n)		**surv.rate (γ)**		**rel. prod. (ξ)**	
		16% qtl	84% qtl	16% qtl	84% qtl	16% qtl	84% qtl
Structural parameters							
$100\,(\beta^{-1}-1)$	discount rate	0.082	–0.039	0.701	–0.073	1.162	–0.254
σ	IES	1.364	–0.165	0.830	–0.182	0.676	–0.306
100δ	depreciation rate	2.029	–0.059	0.109	–0.049	–0.203	0.017

Table 2.4: Difference between worst-case and original posterior mean of $(\beta^{-1}, \sigma, \delta)$, normalized by original posterior standard deviation. Worst-case posteriors correspond to half posterior standard deviation change in quantile.

2.7 An alternative parametric prior

To demonstrate that the sensitivity of the estimated effects to the prior does not depend on the nonparametric nature of REPS, we now show that the posterior results also change substantially when we estimate the model using a different parametric prior. The new prior incorporates the worst-case distortions for the 16% quantile for the effect of the relative productivity of retirees ξ, and the 84% quantile for the effect of the survival rate γ. The parametric assumptions ensure that the prior retains potentially desirable smoothness properties that can be violated by the worst-case prior from REPS. In addition, we show that the distortions put greater weight on ex ante reasonable parameter values by comparing the prior to calibrations of OLG models in the literature. Note that the exercise does not imply a preference for either prior. Instead, the goal is to show that the dependence on the prior in the above analysis applies even when we impose further restrictions on the set of permissible priors.

2.7.1 Prior

We now take the prior of (β, σ) to be:

$$\begin{bmatrix} 100\,(\beta^{-1}-1) \\ \sigma \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 2.00 \\ 0.50 \end{bmatrix}, \begin{bmatrix} 2.00^2 & -0.315 \\ -0.315 & 0.45^2 \end{bmatrix} \right).$$

The new prior for δ is an independent Beta distribution that has mean 0.06 instead of 0.08, and has the same standard deviation as original prior of 0.02.

The new prior places greater weight on regions that the REPS analysis suggest would be particularly important for the estimated effects of the survival rate γ and the relative productivity of retirees ξ. This prior places more prior weight on regions of the parameter space with small values of β^{-1} and δ, as well as large values of σ.

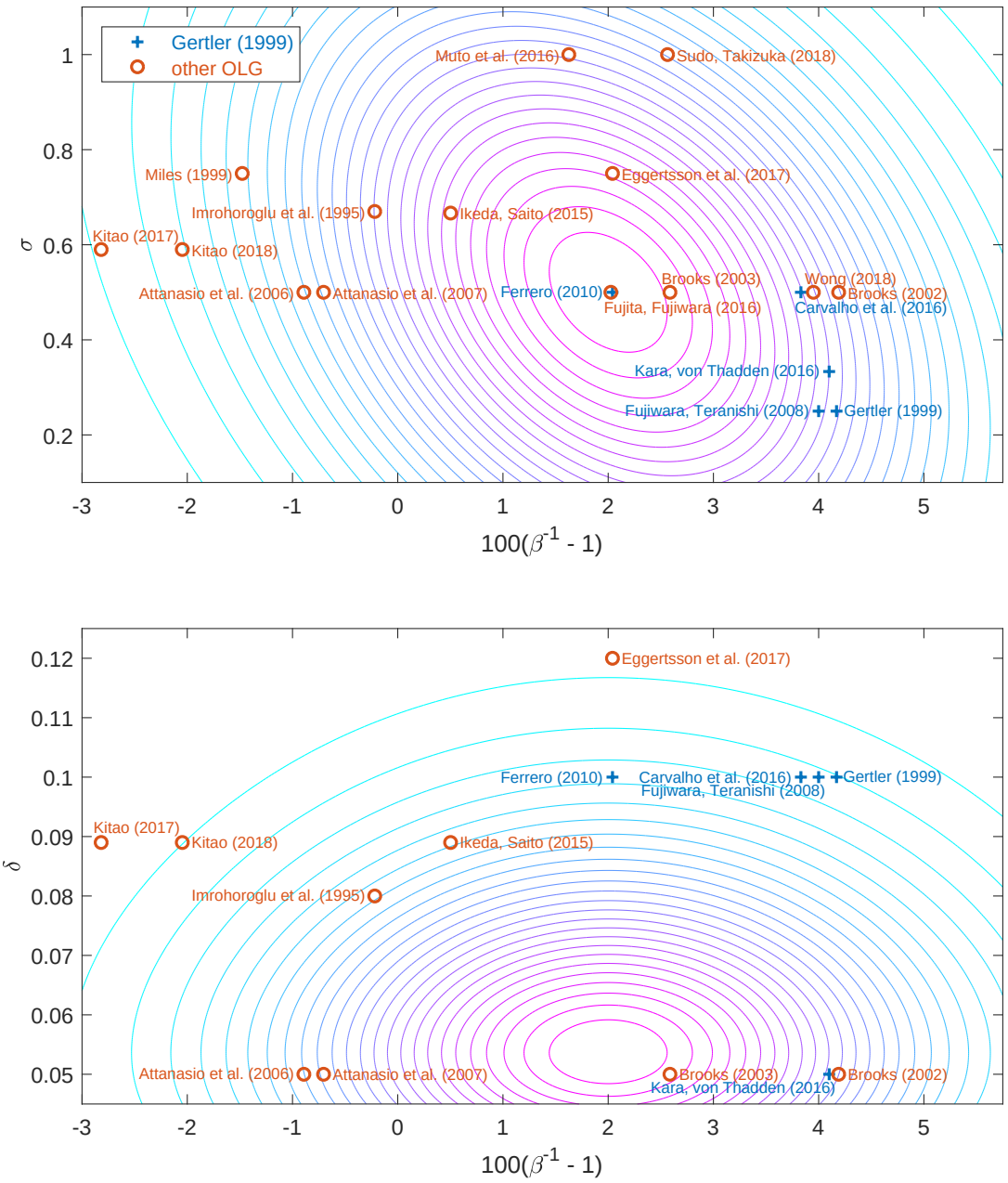


Figure 2.9: Existing calibrations of (β, σ, δ). Blue crosses indicate calibrations of models based on Gertler (1999), red circles indicate calibrations of other OLG models, lines are level curves for the joint distribution under the alternative prior. **Top:** β and σ; **Bottom:** β and δ.

These regions coincide the worst-case distortions for the 16% quantile for the effect of ξ, and the 84% quantile for the effect of γ. Moreover, the REPS analysis also showed that increasing the prior mass on large values of σ reduces the estimated effect of population growth n.

While the worst-case distortions indicate features of the prior that are important for our estimates, some of thes distortions appear implausible and inconsistent with the existing literature. For example, a discount rate $\beta < 0.95$ or depreciation rate $\delta > 0.20$ are unlikely. On the other hand, there has been a wide range of empirical estimates for σ ranging from 0 to 2. We now show that the new prior for (β, σ, δ) is reasonable based on the range of values used in the OLG literature, shown in Figure 2.9.

We decrease the prior mean and increase the prior variance of β^{-1}. In addition, the Gaussian prior on β^{-1} no longer imposes the $\beta < 1$ restriction of the original Gamma distribution prior. The restriction that $\beta < 1$ for representative agent models no longer applies in OLG models since the stochastic death introduces discounting on top of the time preference β. Imrohoroglu et al. (1995) argue that a value of $\beta > 1$ matches the empirical evidence and allows their model to fit the US wealth-income ratio. Nevertheless, we keep a mean discount rate of 2% to reflect the positive discount rates used by a majority of the literature.

We increase the prior mean and variance of σ to reflect the wide range of values in the literature beyond versions of the Gertler (1999) model. In particular, Figure 2.9 shows sixteen papers with $\sigma \geq 0.5$. Not only does REPS analysis indicate that placing more weight on large values of σ can affect posterior inference, but the existing OLG literature is consistent with such a prior.

Under the new prior, β and σ have a correlation of -0.35, which matches the correlation for the sample of calibrations shown in Figure 2.9. Most of these papers pick σ independently, then calibrate β to match a steady state statistic such as the

interest rate in a given year or the capital-output ratio. The correlation reflects the fact that β and σ jointly determine these moments. For example, in a representative agent neoclassical growth model with constant relative risk aversion, the household's Euler equation implies a negative relationship between β and σ to jointly match consumption growth and interest rates. In the model here, the dependence is summarized by the expressions (2.17) and (2.18) for the marginal propensities to consume. Since the REPS analysis has shown that the original assumption of independence between β and σ is not innocuous, we include this correlation in our new prior.

Finally, to illustrate the importance of δ for the estimated effects of the survival rate γ and relative productivity ξ, we consider a new prior for δ that has a smaller mean. The prior mean of 0.06 is closer to the World Penn Tables estimate of 0.05 for the depreciation rate in Japan. Figure 2.9 shows that this is a plausible prior mean, with values of δ ranging between 0.05 to 0.12 in the literature. We keep δ independent from β and σ for two reasons. Firstly, the correlation between δ and (β, σ) among the papers considered in Figure 2.9 is smaller than the correlation between β and σ. Secondly, while β and σ are both preference parameters that directly influence households' savings decisions, δ governs the production side of the economy.

While this prior has been chosen to emphasize the results from the REPS analysis, it is one that would be plausible without observing the data. The prior for (β, σ) was chosen based on existing calibrations, with no reference to the data. The prior for δ is also plausible given existing empirical work that does not use the data we consider. We do not take a stand on which prior should be preferred. Instead, we emphasize that a complete analysis of the data requires understanding the range of results that can arise from different possible priors, and take the alternative prior as an illustrative example. In particular, although there are numerous other plausible priors, we pick this particular one because the REPS results show that these specific changes in the prior can lead to large changes in the posterior estimates.

2.7.2 Results

Table 2.5 shows that changing the prior substantially changes the estimates for the effects of demographics on the interest rate. In all three cases, the mean changes by more than one posterior standard deviation, and the new mean is not contained in the original 68% credible interval. The change in the posterior mean of up to 0.6 percentage points is also economically significant. In comparison, Smets and Wouters (2007) estimate the standard deviation of the monetary policy shock to be 0.2 percentage points, and that a one standard deviation monetary policy shock leads to a 3% decline in output.

The changes in the estimated effects are driven by large changes in the posterior estimates for (β, σ, δ). The mean for the discount rate falls from 3% to 0%, the mean for the intertemporal elasticity of substitution rises from 0.57 to 0.79, and the mean depreciation rate falls from 8% to 3%[4]. Relative to the posterior standard deviations, these are statistically large changes. Moreover, the resulting changes in the estimated effects of demographics indicate that these are also economically significant. In contrast to the posterior for (β, σ, δ), the estimates for the long-run mean μ and the time-varying parameters are relatively unaffected by the change in the prior.

The directions in which the estimates change are consistent with the worst-case posteriors in Figures 2.6 and 2.7. The estimate for the effect of population growth decreases because it depends negatively on σ. The increased weight on small values of β^{-1}, large values of σ, and small values of δ is in line with the distortions that increase the 84% quantile for the effect of the survival rate, and thus increase the overall estimate of that effect. Similarly, the estimate for the effect of the relative

[4]The fact that the new posterior estimate for the depreciation rate is smaller than typical values in the literature should not be viewed as a criticism of a prior that was deemed plausible before observing the data. Rather, the result is further evidence that the current data are insufficient to discipline the estimate of δ.

		original			with alternative prior		
		mean	sd	68% C.I.	mean	sd	68% C.I.
Structural parameters							
$100\left(\beta^{-1}-1\right)$	discount rate	3.245	0.442	(2.804, 3.690)	0.022	1.353	(-1.365, 1.425)
σ	IES	0.574	0.103	(0.471, 0.677)	0.790	0.149	(0.624, 0.959)
100δ	depreciation rate	7.861	1.485	(6.393, 9.317)	3.328	0.890	(2.433, 4.282)
Effects of demographics							
$100\left(\widehat{R}_T^n - R_T\right)$	population growth	0.875	0.159	(0.725, 1.023)	0.696	0.136	(0.571, 0.829)
$100\left(\widehat{R}_T^\gamma - R_T\right)$	survival rate	0.671	0.266	(0.429, 0.905)	1.110	0.225	(0.891, 1.332)
$100\left(\widehat{R}_T^\xi - R_T\right)$	relative productivity	1.206	0.412	(0.793, 1.618)	0.635	0.465	(0.184, 1.118)

Table 2.5: Posterior statistics for effects of demographics, using alternative parametric prior.

productivity of retirees decreases as the alternative prior places greater weight on similar regions to the worst-case distortions for the 16% quantile for that effect.

We have attained these changes in the posterior using an economically plausible alternative prior that lies in a commonly used distributional family. The prior sensitivity identified by REPS therefore remains an issue even when we restrict ourselves to parametric alternative priors. In addition, many of the calibrations considered in Figure 2.9 fall around the high posterior density region for (β, σ, δ), suggesting that the results in the literature are also sensitive to the exact calibration strategy. For example, while it is common to calibrate β to match the interest rate in a given year (e.g. Ikeda and Saito (2014); Carvalho et al. (2016)), we find a large dispersion in the likelihood for β here that translates into imprecise estimates for the effect of demographics. Once we account for uncertainty in the natural interest rate as well as the joint dependence of all the observables on the full parameter vector, we find that the likelihood supports a wide range of parameter combinations that are consistent with the data. Our analysis has shown that the estimated effects of demographics varies substantially with parameter values in this range.

2.8 Conclusion

The aging populations and falling interest rates in developed countries have led to a large literature seeking to quantify the effects of demographics on interest rates through the lens of overlapping generations models. The analysis here shows that these results may be fragile. In particular, we have shown that the estimated effects of demographics depend crucially on the discount rate, intertemporal elasticity of substitution, and capital depreciation rate. Moreover, without the appropriate data, these parameters are not well-identified. As a result, a large set of parameter values can be justified, leading to a wide range of possible estimated effects. We have

argued that including data on savings and capital to discipline these parameters can help produce accurate and precise estimates in future work.

While this paper focuses on the effect of demographics on interest rates, the results also apply to the broader literature on demographics. For example, the consequences of social security policy (e.g. Imrohoroglu et al. (1995); Gertler (1999); Kitao (2017)) depends on the savings decisions of households and how these decisions respond to policy. Quantifying these consequences again relies on estimates for the discount rate and intertemporal elasticity of substitution. Unless the inclusion of social security into the model places restrictions that help narrow these estimates, the sensitivity of the estimates for these parameters to the prior continues to hold. How this translates to sensitivity of the statistics of interest is left to future work.

We have also provided an example of how REPS allows a researcher to gain a better understanding of the model and the data. Firstly, prior sensitivity analysis can reveal how much one's estimation results depend on one's prior. Secondly, the analysis can provide avenues for future research sharpening the estimates. Moreover, the results can help inform calibration exercises as well, by suggesting appropriate moments to use. The methodology is not specific to the model here, and can provide insights into other models and empirical questions as well.

References

Aksoy, Y., H. S. Basso, R. Smith, and T. Grasl (2015). Demographic Structure and Macroeconomic Trends. *Banco de Espana Working Papers 1528*.

Attanasio, O., S. Kitao, and G. L. Violante (2007). Global Demographic Trends and Social Security Reform. *Journal of Monetary Economics 54*(1), 144–198.

Attanasio, O. P., S. Kitao, and G. L. Violante (2006). Quantifying the Effects of the Demographic Transition in Developing Economies. *Advances in Macroeco-*

nomics 6(1).

Blanchard, O. J. (1985). Debt, Deficits, and Finite Horizons. *Journal of Political Economy 93*(2), 223–247.

Brooks, R. (2003). Population Aging and Global Capital Flows in a Parallel Universe. *IMF Staff Papers 50*(2), 200–221.

Caballero, R. J., E. Farhi, and P.-O. Gourinchas (2008). An Equilibrium Model of "Global Imbalances" and Low Interest Rates. *American Economic Review 98:1*, 358–393.

Carvalho, C., A. Ferrero, and F. Nechio (2016). Demographics and Real Interest Rates : Inspecting the Mechanism. *European Economic Review 88*, 208–226.

Christiano, L. J., M. Trabandt, and K. Walentin (2010). DSGE Models for Monetary Policy Analysis. In *Handbook of Monetary Economics*, Volume 3, pp. 285–367. Elsevier Ltd.

Del Negro, M., D. Giannone, M. P. Giannoni, and A. Tambalotti (2017). Safety, Liquidity, and the Natural Rate of Interest. *Brookings Papers on Economic Activity* (Spring), 235–316.

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

Eggertsson, G. B., N. R. Mehrotra, and J. A. Robbins (2017). A Model of Secular Stagnation: Theory and Quantitative Evaluation. *NBER Working Paper 23093*.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The Next Generation of the Penn World Table. *American Economic Review 105*(10), 3150–3182.

Ferrero, G., M. Gross, and S. Neri (2017). On Secular Stagnation and Low Interest Rates: Demography Matters. *Banca d'Italia Working Papers 1137*(September 2017).

Fraumeni, B. M. (1997). The Measurement of Depreciation in the U.S. National Income and Product Accounts. *Survey of Current Business 77*, 7–23.

Fujiwara, I. and Y. Teranishi (2008). A Dynamic New Keynesian Life-cycle Model: Societal Aging, Demographics, and Monetary Policy. *Journal of Economic Dynamics and Control 32*(8), 2507–2511.

Gagnon, E., B. K. Johannsen, and D. Lopez-Salido (2016). Understanding the New Normal: The Role of Demographics. *Finance and Economics Discussion Series 2016-080*.

Gertler, M. (1999). Government Debt and Social Security in a Life-cycle Economy. *Carnegie-Rochester Conference Series on Public Policy 50*, 61–110.

Ho, P. (2019). Global Robust Bayesian Analysis in Large Models. *Working paper*.

Hodrick, R. J. and E. C. Prescott (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Monetary Economics 29*(1), 1–16.

Holston, K., T. Laubach, and J. C. Williams (2016). Measuring the Natural Rate of Interest : International Trends and Determinants. *Federal Reserve Bank of San Francisco Working Paper 2016*(11), 38.

Ikeda, D. and M. Saito (2014). The Effects of Demographic Changes on the Real Interest Rate in Japan. *Japan and the World Economy 32*, 37–48.

Imrohoroglu, A., S. Imrohoroglu, and D. H. Joines (1995). A Life Cycle Analysis of Social Security. *Economic Theory 6*, 83–114.

Kara, E. and L. von Thadden (2016). Interest Rate Effects of Demographic Changes in a New-Keynesian Framework. *Macroeconomic Dynamics 20*(1), 120–164.

Karabarbounis, L. and B. Neiman (2014). The Global Decline of the Labor Share. *Quarterly Journal of Economics 129*(1), 61–103.

Kitao, S. (2017). When Do We Start? Pension Reform in Aging Japan. *Japanese Economic Review 68*(1), 26–47.

Krugman, P. R. (1998). It's Baaack: Japan's Slump and the Return of the Liquidity Trap. *Brookings Papers on Economic Activity 1998*(2), 137.

Laubach, T. and J. C. Williams (2003). Measuring the Natural Rate of Interest. *Review of Economics and Statistics 85*(4), 1063–1070.

Müller, U. K. and M. W. Watson (2018). Long-Run Covariability. *Econometrica 86*(3), 775–804.

Muto, I., T. Oda, and N. Sudo (2016). Macroeconomic Impact of Population Aging in Japan: A Perspective from an Overlapping Generations Model. *IMF Economic Review 64*(3), 408–442.

Sala, L. (2015). DSGE Models in the Frequency Domain. *Journal of Applied Econometrics 30*, 219–240.

Smets, F. and R. Wouters (2007). Shocks and Frictions in US Business Cycles : A Bayesian DSGE Approach. *American Economic Review 97*(3), 586–606.

Sudo, N. and Y. Takizuka (2018). Population Aging and the Real Interest Rate in the Last and Next 50 Years — A Tale Told by an Overlapping Generations Model. *Bank of Japan Working Paper Series 18-E-1.*

Summers, L. H. (2015). Demand Side Secular Stagnation. *American Economic Review: Papers & Proceedings 105*(5), 60–65.

Wong, A. (2018). Transmission of Monetary Policy to Consumption and Population Aging. *Working paper*.

Yaari, M. E. (1965). Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *The Review of Economic Studies 32*(2), 137–150.

Appendix

2.A Steady state of overlapping generations model

Denote $\ell_t = L_t/N_t$ and $\ell_t^w \equiv L_t^w/L_t$. In addition, we use lower case letters to denote variables normalized by y and drop the time subscripts for steady states. We need to solve for:

$$\{\psi, c^r, c^w, h^r, h^w, \ell, \ell^w, \epsilon, \pi, \Omega, \lambda, k, R\}$$

2.A.1 Production

Firm capital decision The firm capital decision is:

$$\begin{aligned} R_t &= (1-\alpha)\frac{Y_t}{K_t} + (1-\delta) \\ &= (1-\alpha)\, k_t^{-1} + (1-\delta) \end{aligned}$$

which yields the steady state condition:

$$k = \frac{1-\alpha}{R-1+\delta} \tag{2.26}$$

Resource constraint We have the aggregate resource constraint:

$$Y_t = K_{t+1} - (1-\delta)\, K_t + C_t^w + C_t^r.$$

Normalizing by Y_t, we have, in steady state:

$$1 = (x + n + \delta)\, k + c^w + c^r \tag{2.27}$$

2.A.2 Households

Population The population of retirees is:

$$\psi_{t+1} N_{t+1} = \psi_{t+1}\,(1+n)\, N_t = \gamma \psi_t N_t + (1-\omega)\, N_t$$

which yields the steady state condition:

$$\psi = \frac{1-\omega}{1+n-\gamma} \tag{2.28}$$

Wealth share of retirees We have:

$$\begin{aligned}
\lambda_{t+1} A_{t+1} &= \lambda_t R_t A_t + W_t \xi L_t^r - C_t^r + (1-\omega)\left[(1-\lambda_t)\, R_t A_t + W_t L_t^w - C_t^w\right] \\
&= (1 - \omega + \omega\lambda_t)\, R_t A_t + \alpha Y_t \frac{\xi L_t^r}{L_t} - C_t^r + (1-\omega)\left(\alpha Y_t \frac{L_t^w}{L_t} - C_t^w\right)
\end{aligned}$$

Normalizing by Y_t, we have, in steady state:

$$(1 + x + n)\, \lambda k = (1 - \omega + \omega\lambda)\, Rk + \alpha - \alpha\omega \ell^w - c^r - (1-\omega)\, c^w \tag{2.29}$$

Consumption of retirees We have:

$$C_t^r = \epsilon_t \pi_t \left(\lambda_t R_t A_t + H_t^r\right)$$

Normalizing by Y_t, we have, in steady state:

$$c^r = \epsilon \pi \left(\lambda R k + h^r\right) \tag{2.30}$$

Human wealth of retirees If retirees choose to supply labor, we have the law-of-motion:

$$\begin{aligned} H_t^r &= \xi L_t^r W_t + \frac{\psi_t N_t}{\psi_{t+1} N_{t+1}} \frac{H_{t+1}^r}{R_{t+1}/\gamma} \\ &= \alpha Y_t \frac{\xi L_t^r}{L_t} + \frac{\gamma}{1+n} \frac{\psi_t}{\psi_{t+1}} \frac{H_{t+1}^r}{R_{t+1}} \end{aligned}$$

Normalizing by Y_t, we have, in steady state:

$$h^r \left(1 - \frac{(1+x)\,\gamma}{R}\right) = \alpha \left(1 - l^w\right)$$

Consumption of workers We have:

$$C_t^w = \pi_t \left[(1-\lambda_t)\, R_t A_t + H_t^w\right]$$

Normalizing by Y_t, we have, in steady state:

$$c^w = \pi \left[(1-\lambda)\, Rk + h^w\right] \tag{2.31}$$

Human wealth of workers We have the law-of-motion:

$$\begin{aligned} H_t^w &= L_t^w W_t + \omega \frac{N_t}{N_{t+1}} \frac{H_{t+1}^w}{R_{t+1}\Omega_{t+1}} + (1-\omega)\, \xi^{\nu-1} \epsilon_{t+1}^{\frac{1}{1-\sigma}} \frac{N_t}{\psi_{t+1} N_{t+1}} \frac{H_{t+1}^r}{R_{t+1}\Omega_{t+1}} \\ &= \alpha Y_t \frac{L_t^w}{L_t} + \frac{\omega}{1+n} \frac{H_{t+1}^w}{R_{t+1}\Omega_{t+1}} + \frac{1-\omega}{1+n} \frac{\xi^{\nu-1} \epsilon_{t+1}^{\frac{1}{1-\sigma}}}{\psi_{t+1}} \frac{H_{t+1}^r}{R_{t+1}\Omega_{t+1}} \end{aligned}$$

Normalizing by Y_t, we have, in steady state:

$$h^w \left(1 - \frac{(1+x)\,\omega}{R\Omega}\right) = \alpha l^w + \frac{(1+x)\,(1-\omega)\,\xi^{\nu-1} \epsilon^{\frac{1}{1-\sigma}} h^r}{\psi R\Omega} \tag{2.32}$$

Labor supply Define $\phi \equiv \frac{1-\nu}{\nu}$. The first-order condition for an individual worker's labor is (abusing notation):

$$L_t^w = 1 - \phi \frac{C_t^w}{W_t}$$

Aggregating, we have:

$$L_t^w = N_t - \frac{\phi}{\alpha} \frac{L_t}{Y_t} C_t^w.$$

Dividing by L_t, we have, in steady state:

$$\ell^{-1} = \ell^w + \frac{\phi}{\alpha} c^w \tag{2.33}$$

The first-order condition for an individual retiree's labor is (abusing notation):

$$L_t^r = 1 - \phi \frac{C_t^r}{\xi W_t}$$

Aggregating, we have:

$$L_t^r = \psi_t N_t - \frac{\phi}{\xi \alpha} \frac{L_t}{Y_t} C_t^r$$

Dividing by L_t/ξ, we have:

$$\xi \psi_t \frac{N_t}{L_t} = 1 - \ell_t^w + \frac{\phi}{\alpha} c_t^r$$

Therefore, in steady state:

$$(1 + \xi\psi)\, \ell^w = 1 + \frac{\phi}{\alpha} \left(c^r - \xi\psi c^w\right) \tag{2.34}$$

Retiree propensity to consume The retiree propensity to consume follows:

$$\epsilon_t \pi_t = 1 - \left[\left(\frac{W_t}{W_{t+1}} \right)^{1-\nu} R_{t+1} \right]^{\sigma - 1} \beta^\sigma \gamma \frac{\epsilon_t \pi_t}{\epsilon_{t+1} \pi_{t+1}}$$

The firm labor decision $W_t L_t = \alpha Y_t$ implies

$$\frac{W_t}{W_{t+1}} = \frac{Y_t}{Y_{t+1}} \frac{L_{t+1}}{L_t} = \frac{1}{1+x+n} \frac{\ell_{t+1}}{\ell_t}.$$

Hence, in steady state,

$$\epsilon\pi = 1 - \left(\frac{R}{(1+x)^{1-\nu}} \right)^{\sigma-1} \beta^\sigma \gamma \tag{2.35}$$

Worker propensity to consume The worker propensity to consume follows:

$$\pi_t = 1 - \left[\left(\frac{W_t}{W_{t+1}} \right)^{1-\nu} R_{t+1} \Omega_{t+1} \right]^{\sigma-1} \beta^\sigma \frac{\pi_t}{\pi_{t+1}}$$

where

$$\Omega_{t+1} = \omega + (1-\omega)\, \xi^{\nu-1} \epsilon_{t+1}^{\frac{1}{1-\sigma}}.$$

This yields steady state conditions:

$$\pi = 1 - \left(\frac{R\Omega}{(1+x)^{1-\nu}} \right)^{\sigma-1} \beta^\sigma \tag{2.36}$$

$$\Omega = \omega + (1-\omega)\, \xi^{\nu-1} \epsilon^{\frac{1}{1-\sigma}} \tag{2.37}$$

2.B Bayesian estimation

2.B.1 Data

Table 2.6 lists the data series that we use as observables, as well as their mapping to the variables or parameters in the OLG model. For labor quantity variables, we focus on employment at the extensive margin, so that labor supply in the model is taken to be the fraction of workers and retirees respectively who are employed. For wages,

data series	model counterpart	source
GDP growth	$x+n$	FRED
age 15-64 pop. growth	n	OECD
age 15-64 pop. / total 15+ pop.	$1/(1+\psi)$	OECD
age 15-64 emp. / total 15+ emp.	$\xi\ell^w/[1-(1-\xi)\ell^w]$	OECD
employment-population ratio	$[1-(1-\xi)\ell^w]/\xi$	OECD
real wage growth	x	FRED
real interest rate	$R-1$	Bank of Japan, IMF DSBB
labor share	α	Penn World Table 9.0
capital-output ratio	k	Penn World Table 9.0

Table 2.6: Observables, model counterparts, and data sources

we use hourly earnings for manufacturing, which is close to the series for monthly private sector earnings.

2.B.2 Prior

Structural parameters. The structural parameters and their priors are summarized in Table 2.7. For each regime, we estimate the time-varying parameters $(x, n, \gamma, \xi, \nu, \alpha)$. These are drawn iid for each regime from a distribution whose mean and variance are estimated. The priors for these means and variances are also reported in Table 2.7. Figure 2.10 compares the prior for (β, σ, δ) to existing calibrations of the Gertler (1999) model.

VAR parameters. For the VAR parameters $\{\Phi(s_t), \Sigma(s_t)\}$, we use a Normal-inverse Wishart prior that shrinks towards white noise, to ensure that v_t captures primarily high frequency variation. The Normal-inverse Wishart prior makes it straightforward to integrate out the VAR parameters. The estimation results are not significantly affected by the particular Normal-inverse Wishart prior used. The coefficients are drawn iid across regime.

Conditional on $\mu(s_t)$, we have a prior that $v_{t^*} \sim \mathcal{N}(\mu(0) - \mu(1), \Omega)$ and $v_T \sim \mathcal{N}(0, \Omega)$, where Ω is a diagonal matrix of variances that we calibrate to be equal to

		type	mean	std dev
Structural Parameters				
fixed				
ω	one minus retirement rate	calibrated	0.98	–
$100\left(\beta^{-1}-1\right)$	discount rate	Gamma	3.50	0.50
σ	intertemporal elasticity of substitution	Normal	0.35	0.15
δ	depreciation rate	Beta	0.08	0.02
time-varying				
$100x$	productivity growth	Normal	μ_x	σ_x
$100n$	population growth	Normal	μ_n	σ_n
$100\left(\gamma^{-1}-1\right)$	probability of death	Gamma	μ_γ	σ_γ
ξ	relative productivity of retirees	Beta	μ_ξ	σ_ξ
ν	one minus disutility of labor	Beta	μ_ν	σ_ν
α	labor share	Normal	μ_α	σ_α
Hyperparameters				
means				
μ_x	productivity growth	Normal	1.50	0.50
μ_n	population growth	Normal	1.00	0.50
μ_γ	probability of death	Gamma	8.00	0.50
μ_ξ	relative productivity of retirees	Beta	0.50	0.10
μ_ν	one minus disutility of labor	Beta	0.50	0.10
μ_α	labor share	Normal	0.65	0.05
standard deviations				
σ_x	productivity growth	Inv. Gamma	1.00	1.00
σ_n	population growth	Inv. Gamma	1.00	1.00
σ_γ	probability of death	Inv. Gamma	2.00	2.00
σ_ξ	relative productivity of retirees	Inv. Gamma	0.10	0.10
σ_ν	one minus disutility of labor	Inv. Gamma	0.10	0.10
σ_α	labor share	Inv. Gamma	0.10	0.10

Table 2.7: Prior for structural parameters.

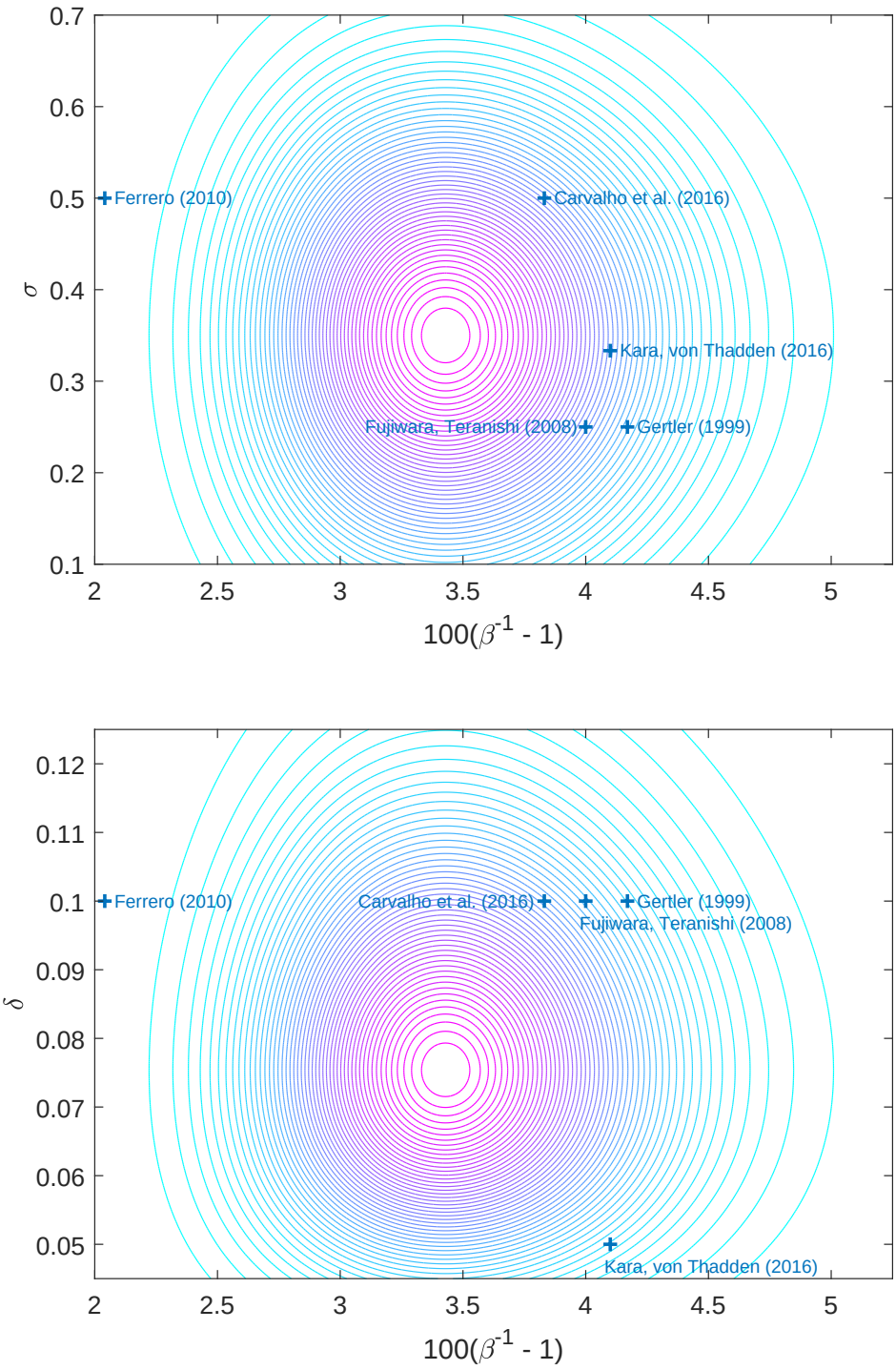


Figure 2.10: Original prior and existing calibrations of (β, σ, δ) for the Gertler (1999) model. Blue crosses indicate calibrations, lines are level curves for the joint distribution under the prior. **Top:** β and σ; **Bottom:** β and δ.

the variances of each series in $\{y_t\}$. This prior on the initial and terminal condition implies that we expect the measurement error to be large immediately after the regime change, but to shrink towards the end of the sample. Intuitively, y_t starts out of steady state in t^*, but converges towards its steady state. The vector autoregression allows this convergence to be modeled flexibly. The results do not change substantially if we ignore the prior for v_{t^*} and v_T.

Structural break. The prior for t^* is flat for $t^* \in [1985, 2009]$ and zero otherwise. The restriction ensures that t^* does not lie in the initial or final periods of the sample. In the estimation, t^* is tightly estimated to lie around 1991.

2.B.3 Markov chain Monte Carlo

The MCMC algorithm to estimate the model (2.1)-(2.3) has two main blocks. First we draw t^* given the structural parameters. Next, we draw the structural parameters given t^*.

Each step involves one or more Metropolis-Hastings draws. Given the structural parameters, we can compute $\overline{y}$ by solving numerically solving for the steady state of the OLG model given those parameters. Given $\overline{y}$, we can obtain v_t. Since v_t follows a VAR(1) with a Normal-inverse Wishart prior, it is straightforward to integrate out the VAR parameters when evaluating the posterior. This improves reduces the size of the parameter space, improves convergence, and is inconsequential for our results since the VAR parameters are not our main focus.

For the burn-in draws, we break the second step into several blocks. We make draws for the fixed and time-varying parameters in separate blocks, and break the time-varying parameters into two blocks corresponding to their values before and after the structural break. The blockwise draws improves convergence when our proposal density is not yet optimized. For the main draws, we draw all the structural

parameters in a single block, and scale the covariance of the burn-in draws for the proposal density.

2.C Prior sensitivity

We use the SMC algorithm from **?**, which is an extension of the SMC algorithm for Bayesian estimation in **?**. We use 10^5 particles, 500 SMC steps and 5 Metropolis-Hastings steps, and repeat this 20 times in parallel. This takes approximately 15 hours to complete. However, we are able to obtain qualitatively similar results with even half the SMC steps, which would require less than half the time. In comparison, the main Bayesian estimation takes approximately five hours to complete.

To produce Figures 2.5 and 2.8, we extrapolate the output from the SMC algorithm. First, we take the median relative entropy of the 20 runs of the SMC algorithm for each SMC step. The ith SMC step corresponds to the same worst-case quantile across runs. Next, we run a polynomial regression of the worst-case quantile on the median relative entropy. Finally, we use the regression to predict the worst-case quantiles for the $r = 0.005$ or $r = 0.025$.

Chapter 3

Factor Volatilities and Volatility Factors

3.1 Introduction

Volatility in the macroeconomy arises from several sources[1], and the effect of volatility on the economy depends on its source. Cross-sectional variance can impact decisions differently from the variance of aggregate shocks. Similarly, the variance of productivity shocks can have different effects from the variance of price shocks. Exisiting measures of uncertainty mix the contributions of various sources, making their interpretation problematic. This paper provides a framework to disentangle the sources of volatility.

The framework is a factor model with stochastic volatility in both factor and idiosyncratic shocks describing a large panel of macroeconomic indicators. The factors consist of a small number of aggregate shocks that drive most of the variation in the economy, while idiosyncratic shocks capture residual series-specific variation. An

[1]Much of the related literature uses the broader term "uncertainty". I use the term "volatility" to emphasize the distinction from Knightian uncertainty, since all uncertainty in the model here arises from the variance of shocks.

increase in aggregate volatility can arise from two sources. If any factor becomes more volatile, then the economy as a whole will become more volatile since the factor drives a substantial amount of movement in numerous variables. Alternatively, comovement in the variance of idiosyncratic shocks can increase the volatility of the economy by increasing the idiosyncratic variation across a large number of variables. In order to capture the comovement of the volatility of different shocks, I assume the volatilities follow a factor structure, similar to Carriero et al. (2016b) and Gorodnichenko and Ng (2017). The *volatility factors* capture the comovement of the volatilities.

I estimate the model using a two step approach that combines principal component anlaysis (PCA) and full-information Bayesian methods. While the literature has used PCA and Bayesian methods independently, I use them as complements. PCA requires minimal structural restrictions and does not require a prior, avoiding two challenges faced by Bayesian methods and making it useful for initial data analysis and model specification. However, PCA relies on asymptotics and produces noisy estimates. The additional structure from a full Bayesian analysis improves on the PCA estimates.

The analysis reveals one volatility factor capturing comovement in the stochastic volatility of idiosyncratic shocks and one volatility factor capturing common stochastic volatility in the factor shocks, and substantial stochastic volatility not captured by the volatility factors. The decomposition provides a more nuanced historical narrative of volatility in the U.S. macroeconomy than univariate measures. Moreover, I show that existing indices fail to capture some sources of volatility. Augmenting a standard vector autoregression (VAR) with a reduced-rank regression of the various volatilities, I find evidence that more than one dimension of volatility is important for explaining the movement of macroeconomic variables.

3.1.1 Related literature

The closest papers to this one are Carriero et al. (2016b) and Gorodnichenko and Ng (2017), who also assume that stochastic volatility is driven by a small number of factors. Carriero et al. (2016b) extract a macroeconomic volatility factor and a financial volatility factor by using Bayesian methods to estimate a factor-augmented vector autoregression model with stochastic volatility. Gorodnichenko and Ng (2017) obtain a volatility factor by combining PCA methods with the structure of the second-order approximation of dynamic stochastic general equilibrium (DSGE) models. Unlike the estimation here, neither approach makes a distinction between a joint increase in shock variances and an increase in volatility of a specific aggregate shock.

Jurado et al. (2015) also estimate a factor model with a large dataset, and measure aggregate uncertainty by averaging the estimated volatility of the unforecastable component of future values across series. However, they do not disentangle the components of the forecast errors in their estimation. Ludvigson et al. (2015) distinguish between measures of macro, financial, and real volatility by using alternative datasets or averaging across forecast errors in different variables. However, their measures do not account for the comovement in the volatilties, which the framework here is designed to do.

The broader literature measuring uncertainty typically fails to disentangle different sources of uncertainty. The ARCH literature growing out of Engle (1982) focused on low-dimensional systems where the distinction between sources of uncertainty was harder to make. Bloom (2009)'s scalar measure of uncertainty using VIX combines volatility from a wide range of aggregate and idiosyncratic shocks. Papers that seek to isolate sources of uncertainty potentially face a problem of interpretation if they do not account for other sources of uncertainty in their estimation. For example, Baker et al. (2016) and Fernández-Villaverde et al. (2015) focus on measuring policy uncertainty, but potentially captures overall economic uncertainty if the policy uncertainty arises

during unusual economic periods. The analysis here allows us to check if these measures are subject to such criticisms. Rossi et al. (2017) decompose uncertainty into Knightian uncertainty and risk, but do not further decompose this risk.

When incorporating stochastic volatility into macroeconomic models (see Fernández-Villaverde and Rubio-Ramírez (2013)), one must choose which shocks to possess time-varying volatility and then model the correlation structure of the volatilities. The framework here is a reduced form approach to inform some of these decisions. In related work, Justiniano and Primiceri (2008) and Diebold et al. (2017) study the relative importance of stochastic volatility in different shocks in DSGE models. Primiceri (2005) and Sims and Zha (2006) study changes in monetary policy using VARs. However, by focusing on a small number of time series, these papers do not distinguish between the volatility in aggregate shocks as opposed to idiosyncratic shocks. Boivin and Giannoni (2006) connect reduced form factor models to DSGE models.

In terms of methodology, Carriero et al. (2016a) and Carriero et al. (2016b) use equation-by-equation estimation to make Bayesian methods feasible for a large-scale VAR with time-varying volatility. Gorodnichenko and Ng (2017) take advantage of the second-order approximation structure to apply PCA methods. I combine the two approaches to leverage on the strengths of each. My PCA procedure differs from Gorodnichenko and Ng (2017) by directly using fitted residuals implied by the model for levels. The assumption of a factor model for the levels greatly reduces the number of parameters to be estimated relative to Carriero et al. (2016a) and Carriero et al. (2016b), which overcomes the computational challenges discussed in Carriero et al. (2016a).

3.1.2 Outline

Section 3.2 sets up the model. In Section 3.3, I use PCA to get an overview of the data. In Section 3.4, I present a Bayesian estimation of the model. Section 3.5 connects the sources of volatility to other uncertainty indices. Section 3.6 studies the effect of the sources of volatility on the macreconomy. Lastly, Section 3.7 concludes.

3.2 Setup

3.2.1 Model

We observe a N-dimensional vector $y_t = (y_{1,t}, ..., y_{N,t})'$ for $t = 1, ..., T$. We assume that y_t follows a factor structure with stochastic volatility:

$$y_t = H^y f_t + u_t \tag{3.1}$$

$$f_t = A^f(L) f_t + \left(B^f\right)^{-1} \Sigma_t^f \varepsilon_t^f \tag{3.2}$$

$$u_{i,t} = A_i^u(L) u_{i,t-1} + \sigma_i^u \varepsilon_{i,t}^u \tag{3.3}$$

where $\varepsilon_t^f, \varepsilon_{i,t}^u \overset{iid}{\sim} \mathcal{N}(0, I)$. Each element $y_{i,t}$ consists of a small number $M^f < N$ of common factors f_t, and an orthogonal idiosyncratic component $u_{i,t}$. The factors capture the common comovement of the data. The idiosyncratic components capture the dynamics of the data that are orthogonal to the factors. These models possess desireable forecasting properties (Eickmeier and Ziegler (2008), Stock and Watson (2012b)), thus providing a reasonable setting to measure volatility. Σ_t^f is a diagonal matrix whose diagonal elements contain the standard deviations $\sigma_{j,t}^f$ of the innovations to the factors f_t. $\sigma_{i,t}^u$ is the standard deviation of the innovation to the ith idiosyncratic component $u_{i,t}$.

When N is large, modelling the evolution of the variances in a way that reduces the dimension of the problem is useful for the estimation and interpretation of results.

We assume the variances follow a factor structure:

$$\log \sigma_{j,t}^f = k_j^f \left(g_t, v_{j,t}^f \right) \tag{3.4}$$

$$\log \sigma_{i,t}^u = k_i^u \left(g_t, v_{i,t}^u \right) \tag{3.5}$$

$$g_t = a^g \left(g_{t-1}, w_t^g \right) \tag{3.6}$$

$$v_{j,t}^f = a_j^{v,f} \left(v_{j,t-1}^f, e_{j,t}^{v,f} \right) \tag{3.7}$$

$$v_{i,t}^u = a_i^{v,u} \left(v_{i,t-1}^u, e_{i,t}^{v,u} \right) \tag{3.8}$$

The standard deviations are driven by a small number M^g of common factors g_t, as well as an orthogonal idiosyncratic component. g_t captures the comovement of the volatilities. Henceforth we shall refer to $\log \sigma_{j,t}^f$ as *factor volatilities* and $\log \sigma_{i,t}^u$ as *idiosyncratic volatilities*. We shall refer to g_t as the *volatility factor*. Volatility in y_t can be driven by an increase in factor volatility or a joint increase in the idiosyncratic volatilities. The joint increase is captured by an increase (decrease) in the volatility factors on which the idiosyncratic volatilities have positive (negative) factor loadings. One can interpret factor volatilies as the volatility of aggregate shocks, and the idiosyncratic volatilities as measuring, for example, cross-sectional dispersion.

To establish a causal relationship between the macroeconomy and the volatility of a specific level shock, it is important to focus on the orthogonal components $v_{j,t}^f$ and $v_{i,t}^u$. Regressing a dependent variable on a direct measure of the log standard deviations $\log \sigma_{j,t}^f$ and $\log \sigma_{i,t}^u$ results in ommited variable bias - the estimated coefficient will be contaminated due to the comovement of the volatilities with each other. If g_t fully captures the comovement in volatilities, then $v_{j,t}^f$ and $v_{i,t}^u$ extract the orthogonal variation of the log standard deviations and can be used as instruments for the volatilities.

The factor structure for both the levels and variances drastically reduces the number of parameters to be estimated while still accounting for comovement in the volatilities.The framework thus allows estimation using a data set with a large number of time series in a tractable setting where the likelihood is informative. In particular, the factor model for the levels allows us to distinguish changes in the volatility of factor shocks from idiosyncratic shocks. PCA in Section 3.3 supports the assumption of a factor structure in the data used here. Lastly, the framework is reduced form and makes minimal assumptions.

3.2.2 Decomposing volatility

To illustrate the contribution of different sources of volatility to aggregate volatility, consider a simple version of (3.1)-(3.8). In particular, suppose f_t and g_t are scalars, and consider the model:

$$y_{i,t} = h_i^y f_t + \sigma_i^u \varepsilon_{i,t} \tag{3.9}$$

$$f_t = a^f f_{t-1} + \sigma^f \varepsilon_t^f \tag{3.10}$$

$$\log \sigma_t^f = k^f g_t + v_t^f \tag{3.11}$$

$$\log \sigma_{i,t}^u = k_i^u g_t + v_{i,t}^u \tag{3.12}$$

where $\varepsilon_{i,t}, \varepsilon_t^f, v_t^f, v_{i,t}^u \overset{iid}{\sim} \mathcal{N}(0,1)$.

Denoting the history of f_t in period t by f^t and assuming all parameters are known, we can decompose the forecast error in $y_{i,t}$ conditional on f^{t-1} into the contributions from the factor innovation and the idiosyncratic innovation:

$$y_{i,t} - \mathbb{E}\left[y_{i,t} | f^{t-1}\right] = h_i^y \sigma_t^f \varepsilon_t^f + \sigma_i^u \varepsilon_{i,t} \tag{3.13}$$

where h_i^y is the ith element of H^y. Hence, the average volatility in the economy can be decomposed as

$$\frac{1}{N}\sum_{i=1}^{N} V\left[y_{i,t} - \mathbb{E}\left[y_{i,t}|f^{t-1}, u^{t-1}\right]\right] \tag{3.14}$$

$$= \underbrace{\frac{1}{N}\sum_{i=1}^{N}\left(h_i^y \sigma_t^f\right)^2}_{\text{factor vol.}} + \underbrace{\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_{i,t}^u\right)^2}_{\text{idiosyncratic vol.}} \tag{3.15}$$

$$= \underbrace{\frac{\sum_{i=1}^{N}\left(h_i^y\right)^2}{N}\exp\left(2v_t^f\right)}_{\text{idiosyncratic factor vol.}}\underbrace{\exp\left(2k^f g_t\right)}_{\text{vol. factor}} + \frac{\sum_{i=1}^{N}\exp\left(2v_{i,t}^u\right)}{N}\underbrace{\exp\left(2k_i^u g_t\right)}_{\text{vol. factor}} \tag{3.16}$$

Equation (3.15) breaks up the average variance into factor and idiosyncratic volatility. Equation (3.16) shows that an increase in average volatility can arise from an increase in v_t^f or an increase in g_t. If the change in $\exp\left(2v_{i,t}^u\right)$ has mean zero, then the contribution of $v_{i,t}^u$ to average volatility is small when N is large because $v_{i,t}$ is independent across i.

The effect of the idiosyncratic component of factor volatility v_t^f does not vanish when N is large since a large shock to f_t results in a large change for all observed variables $y_{i,t}$ with non-zero factor loadings $h_i^y \neq 0$. In particular, notice from equation (3.16) that the effect of a change in v_t^f is scaled by the mean squared factor loading on f_t. In response to variation in f_t, all variables move in parallel by a magnitude scaled by their factor loadings. If f_t were a vector, the volatility of the different components of f_t would not necessarily move in parallel or have the same effects on the economy. The model then helps us to extract the volatility of each component of f_t.

On the other hand, if g_t increases and $k^f, k_i^u > 0$, then both factor and idiosyncratic innovations have higher variance. The increased variance of the idiosyncratic innovations implies that less of the variation in $y_{i,t}$ is explained by f_t. For example, if f_t captures aggregate productivity and the idiosyncratic components correspond to industry-specific shocks, then an increase in g_t would not only increase the volatility

	eigenvalues					
y_t	38.76	11.40	9.74	5.97	4.92	4.48
$\log \sigma_t^{f,\text{MA}}$	2.03	0.82	0.60	0.55	-	-
$\log \sigma_t^{u,\text{MA}}$	26.75	14.21	8.10	7.33	7.19	6.21

Table 3.1: Largest eigenvalues of the standardized y_t, $\log \sigma_t^{f,\text{MA}}$ and $\log \sigma_t^{u,\text{MA}}$.

of aggregate productivity, but also the variance of cross-sectional shocks. The latter leads to an increase in cross-sectional dispersion. By construction, g_t is orthogonal to v_t^f, and the two need not have the same impact on the economy.

3.2.3 Estimation

Data I use a panel of 145 quarterly time series of macroeconomic and financial indicators from Stock and Watson (2012a). The data runs from 1960Q1 to 2014Q4. Following Stock and Watson (2012a), each series is transformed to become stationary by taking logs and differencing. A biweight kernel eliminates long run trends that are arguably predictable. All variables are then standardized to have mean zero and standard deviation two. Full details of the transformations can be found in the appendix of Stock and Watson (2012a). I form a balanced panel by removing variables with missing observations, which does not significantly change the composition of variables.

Identification We require further assumptions and normalizations in (3.1)-(3.8) for estimation and identification. I assume that y_t is driven by $M^f = 4$ factors, which is a reasonable number based on scree plots (see Table 3.1). I assume the factors f_t follow a VAR(4), and the idiosyncratic components each follows an independent AR(1). H^y and f_t cannot be seperately identified, since one can post-multiply H^y by some

invertible matrix R and premultiply f_t by R^{-1} and leave equation (3.1) unchanged. I therefore normalize the top 4×4 block of H^y to be the identity matrix, so the factors are associated with industrial production, consumer price index, consumer credit outstanding, and the fed funds rate respectively. These series have relatively large factor loadings on some combination of the first four principal components, represent easily interpretable parts of the economy, and are not highly correlated. Lastly, I assume that B^f is lower triangular with ones on the diagonal[2]. The lower triangularity identifies the structural shocks in equation (3.1). The ones on the diagonal identify the scale of Σ_t^f.

3.3 Principal component analysis

3.3.1 Estimation

Before proceeding to a full Bayesian estimation of the model, it is informative to use PCA to get an overview of the data. The results from PCA also provides guidance for both the model and the prior in the Bayesian estimation.

PCA is computationally convenient and requires minimal assumptions. Asymptotic theory allows one to determine the number of factors without a full model. Moreover, PCA sidesteps the problem of specifying a prior. Given the size of the model, the likelihood may not fully swamp the priors. However, we have little guidance on a reasonable prior for parameters determining the comovement of the volatilities. PCA allows us to estimate the volatility factors, then use information from these factors to inform our choice of prior.

[2]Carriero et al. (2016a) argue that given the MCMC simulations, one can also use alternative identification schemes for structural analysis using methods from Baumeister and Hamilton (2015) and Rubio-Ramirez et al. (2010). Here I will rely on the lower triangularity of B^f to interpret my results.

I first use PCA on y_t to obtain point estimates of f_t, which we shall denote by f_t^{PCA}[3]. Regressing y_t on f_t^{PCA} equation-by-equation yields fitted residuals u_t^{PCA}. Next, I use least squares to estimate

$$f_t^{\text{PCA}} = A^f(L) f_t^{\text{PCA}} + B^f \tilde{\varepsilon}_t^f \tag{3.17}$$

$$u_{i,t}^{\text{PCA}} = a_i^u u_{i,t-1}^{\text{PCA}} + \tilde{\varepsilon}_{i,t}^u \tag{3.18}$$

I take log of the squared residuals

$$\log\left(\tilde{\varepsilon}_t^{\text{PCA}} \cdot \tilde{\varepsilon}_t^{\text{PCA}}\right) = 2\log\vec{\sigma}_t^{\text{PCA}} + \log\left(\varepsilon_t^{\text{PCA}} \cdot \varepsilon_t^{\text{PCA}}\right) \tag{3.19}$$

as a measure of volatility, contaminated by log chi-square errors $\log\left(\varepsilon_t^{\text{PCA}} \cdot \varepsilon_t^{\text{PCA}}\right)$. A 9-quarter moving average of $\log\left(\tilde{\varepsilon}_t^{\text{PCA}} \cdot \tilde{\varepsilon}_t^{\text{PCA}}\right)$ smooths out the transitory variation from the log chi-square errors since we assume these errors are iid. If $\log\vec{\sigma}_t^{\text{PCA}}$ is persistent, this yields an estimate for $\log\vec{\sigma}_t^{\text{PCA}}$[4]. Denote the moving average by $\log\vec{\sigma}_t^{\text{MA}}$. I obtain estimates for $g_t^{f,\text{PCA}}$ and $g_t^{u,\text{PCA}}$ by applying PCA to the components of $\log\vec{\sigma}_t^{\text{MA}}$ corresponding to factor and idiosyncratic level shocks respectively. Although aysymptotics derived by Bai (2003) suggest that we obtain consistent estimates as $N \to \infty$, measurement error from the estimation of f_t^{PCA} and the presence of $\log\left(\varepsilon_t^{\text{PCA}} \cdot \varepsilon_t^{\text{PCA}}\right)$ add noise to the estimates.

The above procedure does not impose any additional structure on equations (3.4)-(3.8). Instead, PCA relies on large N asymptotics to estimate f_t^{PCA}, and only on functional form assumptions in equations (3.1)-(3.3) to estimate u_t^{PCA} and $\tilde{\varepsilon}_t^{\text{PCA}}$. One can question the accuracy of the results given the inherent measurement error

[3]Bai (2003) derives large N asymptotics and provides conditions under which these estimates are consistent, including under certain forms of heteroskedasticity.

[4]With the data here, doing PCA without taking moving averages gives estimates that are noisier but look similar at a low frequency.

from the two-stage approach[5]. Furthermore, since the time-variation in the variances is not accounted for when estimating (3.17) and (3.18), the estimates for $A^f(L)$ and a_i^u are inefficient. Moreover, taking a moving average removes the high-frequency variation in the volatilities. The small number of factor volatilities implies that PCA on these volatilities is subject to even greater measurement error than with the idiosyncratic volatilities. The measures of volatility we obtain from PCA are thus noisy estimates of the low frequency component of $\log \vec{\sigma}_t$. Nevertheless, PCA remains useful as a preliminary exercise for understanding the data and informing the Bayesian estimation. Bayesian estimation overcomes the shortfalls by accounting for the joint distribution of parameters and latent variables through the likelihood.

3.3.2 Results

Figure 3.1 shows the PCA moving average estimates of factor volatilities $\log \sigma_t^{f,\text{MA}}$. Visual inspection suggests a common factor with peaks in the mid 1970s, early 1980s, and late 2000s, as is found in Jurado et al. (2015). However, the substantial and persistent differences across the factor volatilities suggest non-trivial idiosyncratic variation in the volatility of the level factor shocks.

The eigenvalues of standardized $\log \sigma_t^{f,\text{MA}}$ and $\log \sigma_t^{u,\text{MA}}$ suggest a one-factor structure for $\log \sigma_t^{f,\text{MA}}$ and either a one or two-factor structure for $\log \sigma_t^{u,\text{MA}}$ (see Table 3.1). The eigenvalues measure how much of the total variation in the data is accounted for by the corresponding principal component. If the data has a factor structure with p factors, then the p largest eigenvalues will dominate asymptotically.

These results lead us to analyze the first principal component of $\log \sigma_t^{f,\text{MA}}$ and the first two principal components of $\log \sigma_t^{u,\text{MA}}$. The principal components are plotted in Figure 3.2. The principal component $g_t^{f,\text{PCA}}$ of $\log \sigma_t^{f,\text{MA}}$ confirms peaks in volatility

[5]Gorodnichenko and Ng (2017) provide an alternative approach to extract volatilities and volatility factors using PCA, which relies on the structure of the second-order approximation of a DSGE. While their approach avoids some of the measurement error that contaminates my estimates, it does not arise naturally from the reduced form framework presented here.

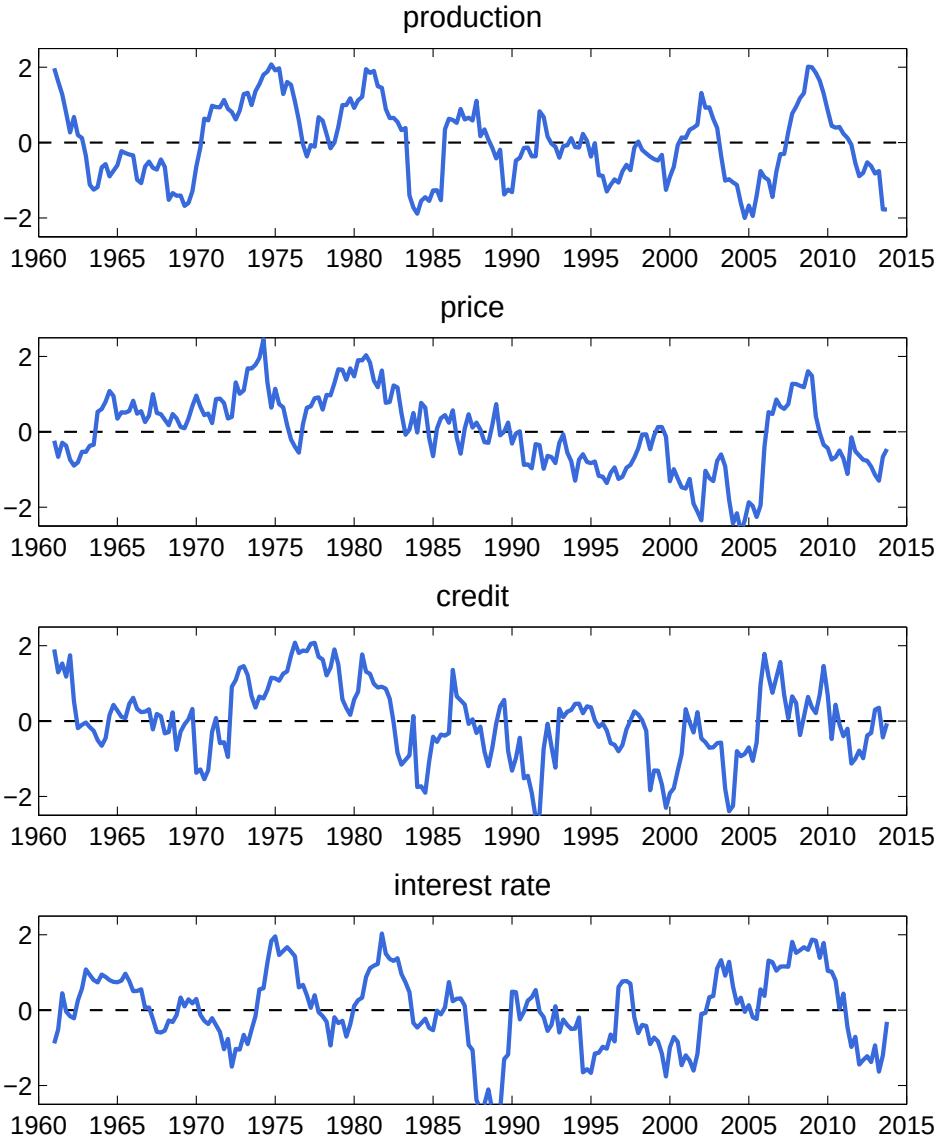


Figure 3.1: Factor volatilities estimated by PCA (normalized to have mean zero and variance one).

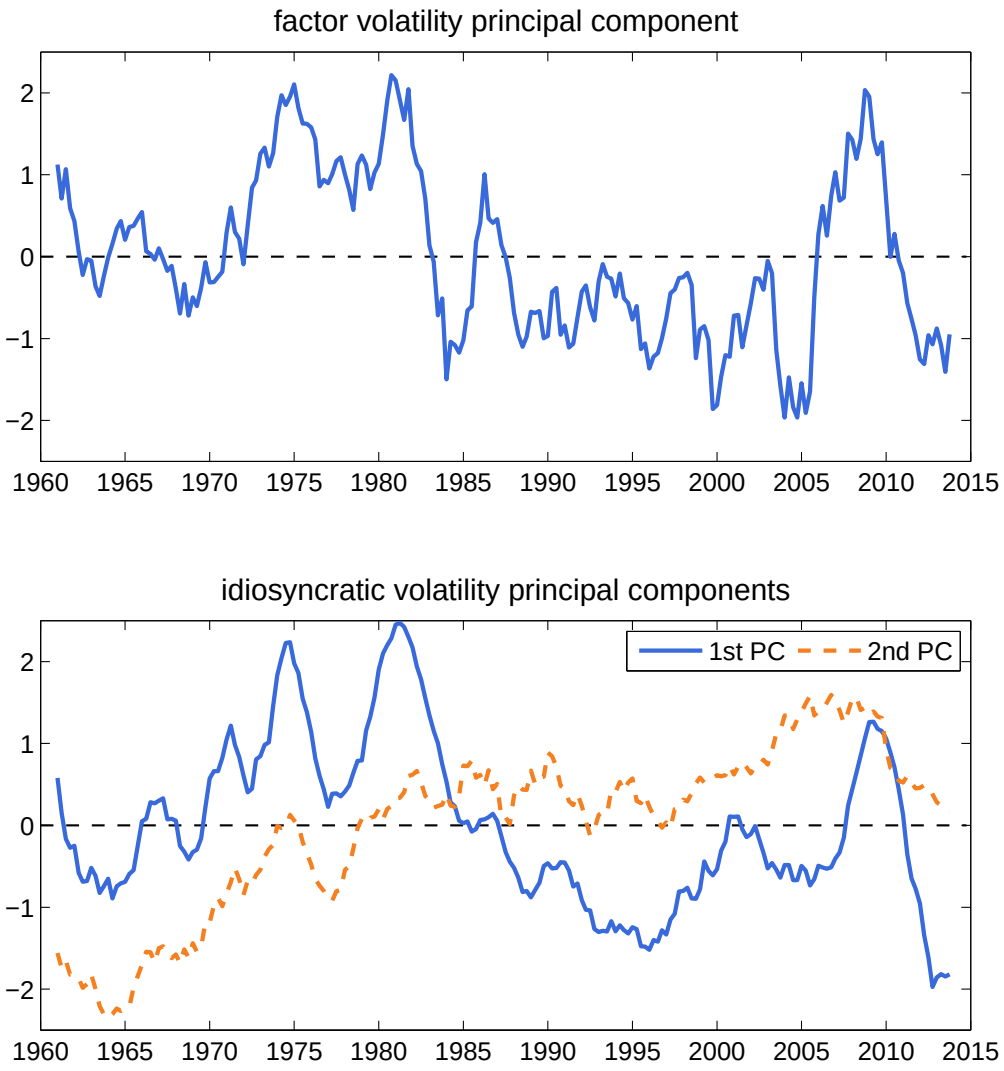


Figure 3.2: Principal components (normalized to have mean zero and variance one). **Top:** first principal component of factor volatilities. **Bottom:** first (blue solid line) and second (orange dashed line) principal component of idiosyncratic volatilities.

in the mid 1970s, early 1980s, and late 2000s. In addition, it suggests a high factor volatility regime around 1972-1982, followed by a low volatility regime ending in the mid 2000s, followed by a sharp increase in factor volatility leading into the Great Recession. The factor volatility factor $g_t^{f,\text{PCA}}$ has loadings 0.539, 0.527, 0.530, and 0.390 on the production, price, credit, and interest rate factors respectively. Despite the small number of series and possibly large measurement errors, $g_t^{f,\text{PCA}}$ has large loadings and appears reasonable.

The first principal component $g_{1,t}^{u,\text{PCA}}$ of $\log \sigma_t^{u,\text{MA}}$ evolves much like the factor volatility principal component $g_t^{f,\text{PCA}}$, with a correlation of 0.75 and the same three major peaks. The smoothness of $g_{1,t}^{u,\text{PCA}}$ relative to $g_t^{f,\text{PCA}}$ arises partly because we have a larger number of idiosyncratic volatilities than factor volatilities to average over. The top panel of Figure 3.3 shows the factor loadings for the idiosyncratic volatilities of different series. The factor loadings are mostly positive, suggesting that the three major peaks in Jurado et al. (2015) correspond to periods where not only the factor shocks became more volatile, but the idiosyncratic shocks jointly became more volatile on average. However, there is wide dispersion in the factor loadings. For example, the idiosyncratic volatility with the smallest factor loading in magnitude is the production of consumer energy products, with a factor loading of 0.0036. During periods when $g_{1,t}^{u,\text{PCA}}$ is high, the level factors f_t explain neither more nor less of such series compared to periods when $g_{1,t}^{u,\text{PCA}}$ is low. There is an overall decrease in $g_{1,t}^{u,\text{PCA}}$ after the early 1980s, which is best understood with reference to the second principal component.

The second principal component $g_{2,t}^{u,\text{PCA}}$ of $\log \sigma_t^{u,\text{MA}}$ has an increasing trend. The bottom panel of Figure 3.3 shows that the factor loadings on the real variables' idiosyncratic volatities are negative on average, the factor loadings on price idiosyncratic volatilities are a mix of positive and negative, while the remaining factor loadings are largely positive. The factor loadings suggest that $g_{2,t}^{u,\text{PCA}}$ conflates several forces. The

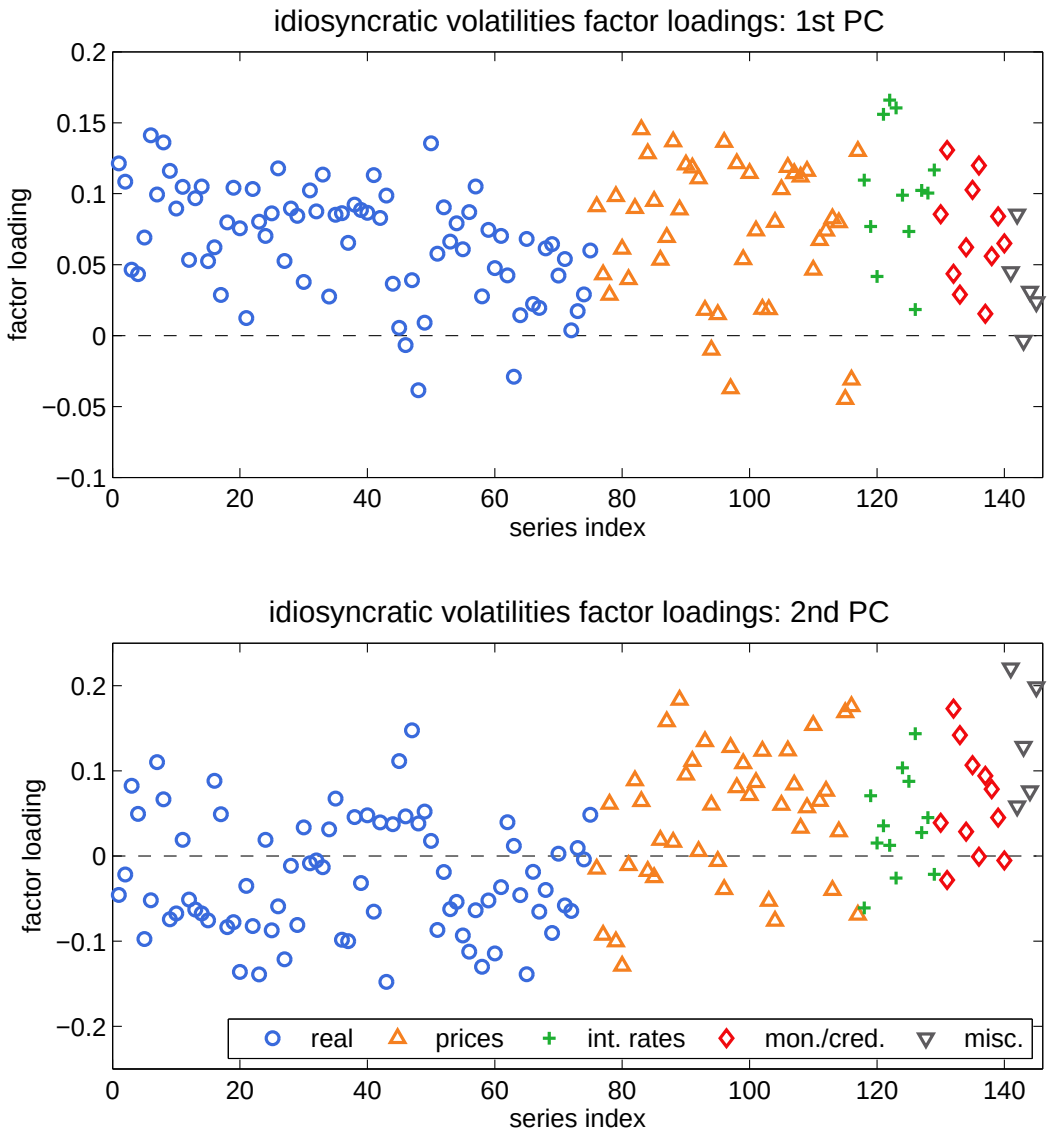


Figure 3.3: Factor loadings for idiosyncratic volatilities from PCA. Each point shows factor loading for idiosyncratic volatility of specific series. Series arranged along horizontal axis by category. **Top:** first principal component. **Bottom:** second principal component.

idiosyncratic volatilities of the variables with negative factor loadings have fallen on average, even more than suggested by the downward trend in $g_{1,t}^{f,\text{PCA}}$. During the Great Moderation, not only did factor shocks become less volatile, but the factors were also able to better explain these variables. The positive loadings on the remaining volatilities have two interpretations. If an idiosyncratic volatility was not affected by the Great Moderation, then a positive loading from the second volatility prinicpal component undoes some of the low-frequency decrease in volatility from the first volatility principal component. This is probably the case for many of the money and credit variables. A number of the the positive factor loadings on the price idiosyncratic volatilities arise because their constructions result in smaller measurement errors in the earlier periods due to smoothing of the data (Stock and Watson (2016)).

The plausibility of these findings supports the use of PCA for initial data analysis. However, given the fragility of the asymptotic assumptions that PCA rests on, a thorough analysis requires a complete estimation of a fully-specified model.

3.4 Bayesian estimation

3.4.1 Functional forms

For the Bayesian estimation, we specify the functions in (3.4)-(3.8) to form the likelihood. In particular, I assume the volatilities follow:

$$\log \sigma_{j,t}^f = k_j^f g_t^f + v_{j,t}^f \tag{3.20}$$

$$\log \sigma_{i,t}^u = k_i^u g_t^u + v_{i,t}^u \tag{3.21}$$

$$\begin{bmatrix} g_t^f & g_t^u \end{bmatrix}' = \begin{bmatrix} g_{t-1}^f & g_{t-1}^u \end{bmatrix}' + w_t^g \tag{3.22}$$

$$v_{j,t}^f = v_{j,t-1}^f + \omega_j^f e_{j,t}^f \tag{3.23}$$

$$v_{i,t}^u = v_{i,t-1}^u + \omega_i^u e_{i,t}^u \tag{3.24}$$

where $e^f_{j,t}, e^u_{i,t} \overset{iid}{\sim} \mathcal{N}(0,1)$ and $w^g_t \sim \mathcal{N}(0, \Omega^g)$.

The factor g^f_t is a scalar that captures the comovement of the factor volatilities and the factor g^u_t is a scalar that captures the comovement of the idiosyncratic volatilities. The assumption that g^f_t is a scalar is in line with the evidence presented in Section 3.3. The evidence from PCA also showed that g^u_t is either scalar or bivariate. However, only the first principle component $g^{u,\text{PCA}}_{1,t}$ showed meaningful business cycle variation. The challenge in interpreting the second idiosyncratic volatility principal component and the strength of the first principal component (from the eigenvalues) motivate the choice of modelling g^u_t as a scalar process. Our prior will shrink towards positive factor loadings for g^u_t to ensure that the estimation identifies the first component.

The volatility factors and the idiosyncratic components of volatility follow random walks. While in theory this assumption implies that g^f_t and g^u_t are nonstationary, here we only consider their evolution over a finite period, which ensures that they remain finite. In practice, the likelihood disciplines the evolution of the volatilities so that they are not explosive. The random walk allows the volatilities to be tracked flexibly, which is key to measuring volatility and its factor structure. Since the random walk assumption implies that there is no stationary distribution for g^f_t, g^u_t, $v^f_{j,t}$ and $v^u_{i,t}$, I identify the levels of g^f_t, g^u_t, $v^f_{j,t}$ and $v^u_{i,t}$ by setting $g^f_0 = g^u_0 = 0$.

I do not impose restrictions on the covariance Ω^g of w^g_t, allowing the innovations to the volatility factors to be correlated. The high correlation between $g^{f,\text{PCA}}_t$ and $g^{u,\text{PCA}}_{1,t}$ emphasize the importance of allowing for comovement in g^f_t and g^u_t. When Ω^g is diagonal, g^f_t is identified from only the comovement of $\log \sigma^f_{j,t}$, and g^u_t is identified from only the comovement of $\log \sigma^u_{i,t}$. However, when Ω^g implies a positive correlation between g^f_t and g^u_t, a joint increase in factor volatilities $\log \sigma^f_{j,t}$ leads to an increase in g^u_t on average. Hence, the volatility factors g^f_t and g^u_t are determined jointly by both $\log \sigma^f_{j,t}$ and $\log \sigma^u_{i,t}$. The restriction on the volatility factor loadings makes the interpretation of g^f_t and g^u_t straightforward. The appropriate rotation on $\left(g^f_t, g^u_t\right)'$

would generate factors that are orthogonal but that load on both the factor and idiosyncratic volatilities.

The functional form assumptions embodied in equations (3.1)-(3.8) and (3.20)-(3.24) make the contribution of the various sources of volatility transparent. The structure also makes it straightforward to draw from the posterior using a standard Gibbs sampler (e.g. Del Negro and Otrok (2008); Primiceri (2005); Stock and Watson (2016)). As is common in the literature, the model presented here does not directly measure the effect of volatilities on the levels or vice versa. Instead, we shall use a reduced-rank regression to analyze the connection between the volatilities and levels in the macroeconomy. A joint estimation of stochastic volatility and its effects on levels, as in Carriero et al. (2016b), introduces nonlinearities that complicate the estimation.

3.4.2 Prior

The existing literature offers little guidance on priors for the parameters in equations (3.20)-(3.24) governing the law-of-motion for the priors. I thus use the PCA results from Section 3.3 to inform my prior, which has a loose empirical Bayes interpretation. An empirical Bayes approach parameterizes priors by hyperparameters chosen to maximize the marginal likelihood. Instead, I match the estimates from PCA. As long as PCA gives estimates near the peak of the likelihood, the priors I choose will not be far from the values that would be implied by an empirical Bayes estimation. PCA has the advantage of being computationally trivial relative to empirical Bayes.

We parameterize the priors as follows:

$$\frac{k_j^f}{\lambda^f} \sim \mathcal{N}\left(\mu^f, \left(s^f \mu^f\right)^2\right) \qquad \left(\frac{\omega_j^f}{\lambda^f}\right)^2 \sim \Gamma^{-1}\left(3, \frac{3}{16}\right) \tag{3.25}$$

$$\frac{k_i^u}{\lambda^u} \sim \mathcal{N}\left(\mu^u, (s^u \mu^u)^2\right) \qquad \left(\frac{\omega_i^u}{\lambda^u}\right)^2 \sim \Gamma^{-1}\left(3, \frac{3}{16}\right) \tag{3.26}$$

$$\Omega^g \sim \mathcal{W}^{-1}\left(3, \frac{3}{16} I\right) \tag{3.27}$$

We need to calibrate the hyperparameters $\left\{\lambda^f, \lambda^u, \mu^f, \mu^u, s^f, s^u\right\}$ and normalize the scale of the volatility factors.

λ^f and λ^u scale equations (3.20) and (3.21) respectively to determine the variation in the log standard deviations relative to the volatility factors. μ^f and μ^u correspond to the average size of the volatility factor loadings relative to the variance of the idiosyncratic components. s^f and s^u parameterize the dispersion of volatility factor loadings, scaled by their means.

The scale matrix for the prior on Ω^g represents a belief that g_t^f and g_t^u are independent and that the annual change in each has a variance of $\frac{1}{4}$. The rate parameter for the prior on $\frac{\omega_j^f}{\lambda^f}$ and $\frac{\omega_i^u}{\lambda^u}$ implies a belief that the volatility idiosyncratic components scaled by $\frac{1}{\lambda^f}$ and $\frac{1}{\lambda^u}$ respectively also have an annual change with variance $\frac{1}{4}$. We can thus interprete the volatility factor loadings as the amount of variation in the log standard deviations captured by the volatility factor relative to the volatility idiosyncratic component.

We now use the PCA results to calibrate the hyperparameters. Denote the element of $\log \vec{\sigma}_t^{\mathrm{MA}}$ corresponding to the jth level factor shock by $\log \sigma_{j,t}^{f,\mathrm{MA}}$. Use least squares to estimate

$$\log \sigma_{j,t}^{f,\mathrm{MA}} = k_j^f g_t^{f,\mathrm{PCA}} + v_{j,t}^{f,\mathrm{PCA}} \tag{3.28}$$

Define the relative contribution of the volatility factor to the overall variation of the log standard deviations as

$$\mu_j^f \equiv \sqrt{\frac{V\left[\hat{k}_j^f g_t^{f,\text{PCA}}\right]}{V\left[\hat{v}_{j,t}^{f,\text{PCA}}\right]}} \tag{3.29}$$

I calibrate μ^f to match the median contribution of the volatility factor:

$$\mu^f = \text{median}\left[\mu_j^f\right] \tag{3.30}$$

I take the median in order to avoid overfitting for outliers[6]. I pick s^f so that $s^f \mu^f$ matches the standard deviation of relative contribution of the factor volatility factor:

$$s^f = \frac{\sqrt{V\left[\mu_j^f\right]}}{\mu^f} \tag{3.31}$$

I set λ^f to scale equations (3.20) and (3.21):

$$\lambda^f = \sqrt{\frac{1}{\left(\mu^f\right)^2 + 1}} \tag{3.32}$$

This scaling implies that the log variances have a annual change of variance one. Lastly, to identify the scale of the volatility factors, I normalize $k_j^f/\lambda^f = \mu^f$ for the j corresponding to the median observation. I use a similar approach to calibrate μ^u, s^u and λ^u. See Appendix 3.A for the calibrated values.

The prior on parameters governing the comovement of the levels (3.1)-(3.3) are standard. The prior for the factor loadings is $h_{i,j}^y \sim \mathcal{N}(0,1)$ for $i > M^f$ (the first M^f rows of H^y are normalized to identify the factors). The prior on $A^f(L)$ and B^f shrinks towards independent persistence, similar to a Minnesota prior. However, since the series have been differenced and had long-run trends removed, I center the

[6]With the data here, the mean and median are similar.

prior at a level of persistence that implies stationarity. For simplicity, we fix, for all i and j, the initial conditions $f_{j,0} = 0$. These initial conditions are reasonable given the normalization of y_t to be mean zero. Allowing a non-degenerate prior for f_0 only affects the smoothed factors in the initial periods. Further details are reported in Appendix 3.A.

3.4.3 Estimation

We draw from the posterior using a Gibbs sampler described in Appendix 3.B.

Carriero et al. (2016a) outline the computational challenges of estimating a large VAR with stochastic volatility. The factor structure for levels sidesteps the problem. A^f and B^f are straightforward to draw since the VAR in the transition equation (3.2) involves only a small number M^f variables. Moreover, the assumption that f_t fully captures the comovement of y_t implies that we can estimate H^y from the measurement equation (3.1) equation-by-equation.

I run 20 parallel chains with different initial draws. For each chain, I take 7,000 draws, of which I drop the first 2,000 and keep every subsequent fifth draw, giving 1,000 draws per chain. This takes approximately 24 hours on a dual-core Intel i7-4600U 2.10GHz processor. I run the multiple parallel chains to guard against multiple modes, and check trace plots to ensure that the 2,000 burn-in draws are sufficient.

3.4.4 Results

Factor volatilities

Figure 3.4 plots the median estimate of the idiosyncratic component $v_{j,t}^f$ of the factor volatilities alongside related macroeconomic time series. The variation in the idiosyncratic component $v_{j,t}^f$ of the factor volatilities should be interpreted as the variation in the volatility of the factor $f_{j,t}$ over and above the variation in volatility

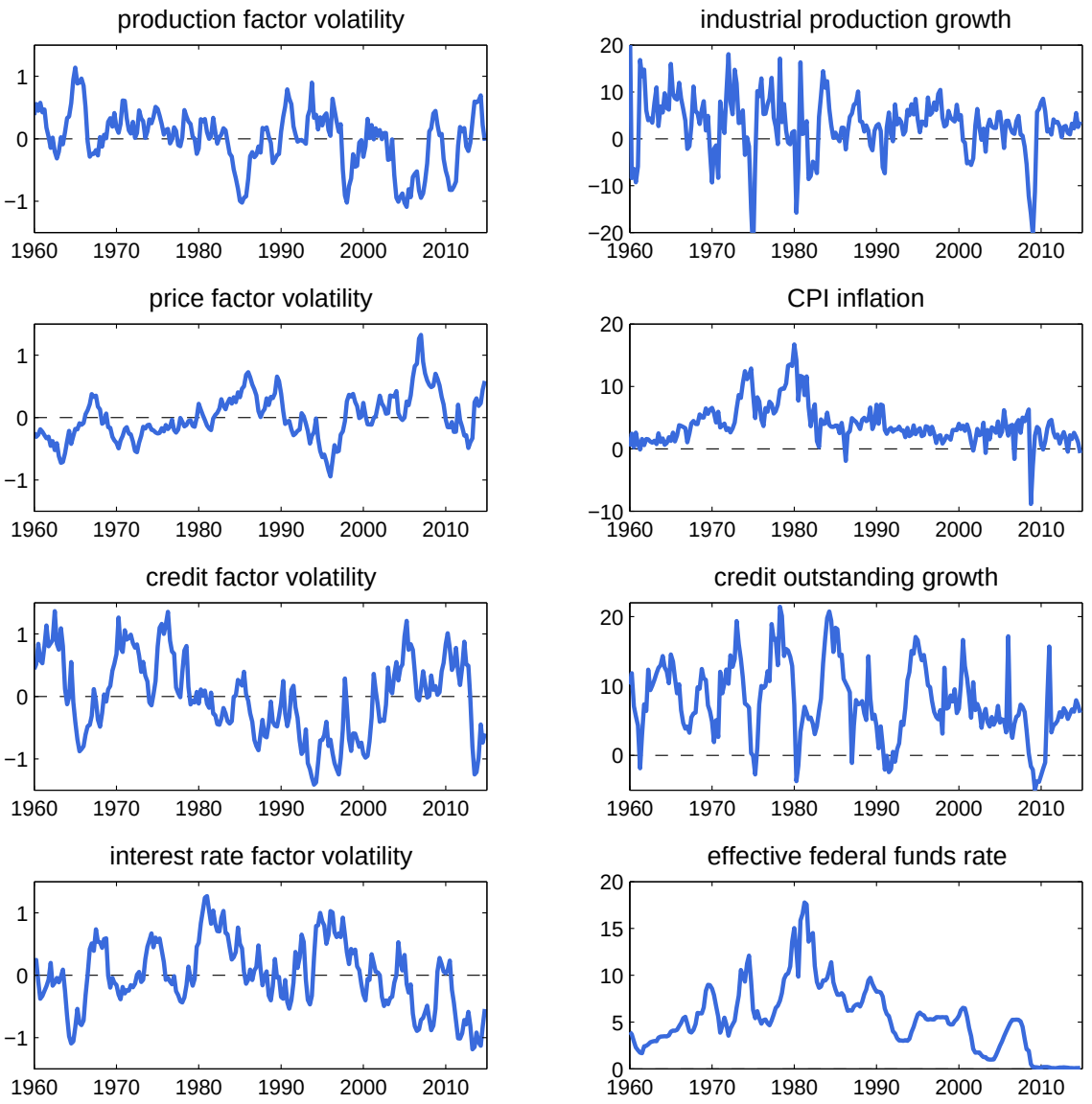


Figure 3.4: Median factor volatilities estimated by Bayesian estimation (demeaned) and associated macroeconomic time series (annualized, in percent). **Left:** Factor volatilities. **Right:** Macroeconomic time series.

accounted for by the volatility factor g_t^f. The largest correlation between the median estimates of $v_{1:T}^f$ and g_t^f is 0.35, and PCA on the idiosyncratic components $v_{1:T}^f$ yields eigenvalues of 1.47, 1.14, 1.06 and 0.32, suggesting that g_t^f has indeed accounted for much of the comovement in the factor volatilities.

The idiosyncratic component of the production factor volatility, which we denote by $v_{\text{prod},t}^f$, first peaks during the high growth period of the mid 1960s. It then remains relatively constant before dipping drastically in the mid 1980s, coinciding with a period where real variables were less volatile relative to prices and interest rates. This decrease is also present in the PCA estimates. The idiosyncratic component of the production factor volatility then increases and remains relatively constant through most of the 1990s. $v_{\text{prod},t}^f$ remains around its historic average before falling sharply in the early to mid 2000s and remaining low until the Great Recession hits in the late 2000s, a pattern observed in the PCA estimates as well.

Spikes in the idiosyncratic component of the price factor volatility, $v_{\text{price},t}^f$, tend to coincide with periods of volatile commodity prices. For example, the largest rise in $v_{\text{price},t}^f$ starts from the mid 2000s, when there were large swings in energy prices but relatively low volatility in the rest of the economy. These results suggest that energy prices display variation in volatility that is orthogonal to the volatility in other parts of the economy. $v_{\text{price},t}^f$ does not display a downward trend, which runs against the usual narrative of inflation. Moreover, there is no substantial increase in $v_{\text{price},t}^f$ in the 1970s and early 1980s. Accounting for the common comovement in volatilities thus reveals that the reduction of price volatility after the Great Moderation was part of a broader trend rather than one restricted to inflation.

The idiosyncratic component of the credit factor volatility, $v_{\text{cred},t}^f$, begins at a high level, decreases and remains low during the 1980s and 1990s, then rises toward the end of the sample. During the initial period, there are spikes in the early 1960s, early 1970s and mid 1970s, coinciding with sharp declines in the Dow Jones Industrial

Average and the S&P index. The rise at the end of the sample begins during the Dot-com bubble, and persists beyond the Great Recession.

The evolution of the idiosyncratic component of the interest rate factor volatility $v^f_{\text{int},t}$ matches the monetary history of the United States, similar to Sims and Zha (2006). For instance, the tenure of Paul Volcker was characterized by larger swings in interest rates than any Federal Reserve chair in the sample period, and is marked by a persistent increase in $v^f_{\text{int},t}$ during the first half of the 1980s. The subsequent downward trend in $v^f_{\text{int},t}$ during Alan Greenspan's tenure is interrupted in the early to mid 1990s, as interest rates were decreased in response to the 1990-1991 recession, then raised from 3% to 6% in late 1994. $v^f_{\text{int},t}$ continues to trend downward during Ben Bernake's tenure, except when the Federal Reserve cut interest rates sharply the late 2000s. This is followed by lowest level of $v^f_{\text{int},t}$ of the sample period as the zero lower bound binds.

Volatility factors

We now turn to the volatility factors g^f_t and g^u_t, plotted in Figure 3.5. As in the PCA estimation and Jurado et al. (2015), we see peaks in the mid 1970s, early 1980s, and late 2000s. The largest peaks of g^f_t and g^u_t come in the mid 1970s. g^f_t and g^u_t are positively correlated, with the estimate of Ω^g implying a median correlation in w^g_t of 0.16. In addition, the median estimates reveal substantial high frequency variation in g^f_t and g^u_t, which is smoothed out when we take the moving average for the PCA estimates.

The results in Figure 3.5 shed light on the finding in Giannone et al. (2008) and Carriero et al. (2016b) that there is weaker evidence of the Great Moderation in models with larger datasets than models with smaller datasets. In particular, there is greater evidence of the Great Moderation in g^f_t than g^u_t. From 1960 to 1980, g^f_t remains mostly above average besides temporary drops around 1963. After 1985,

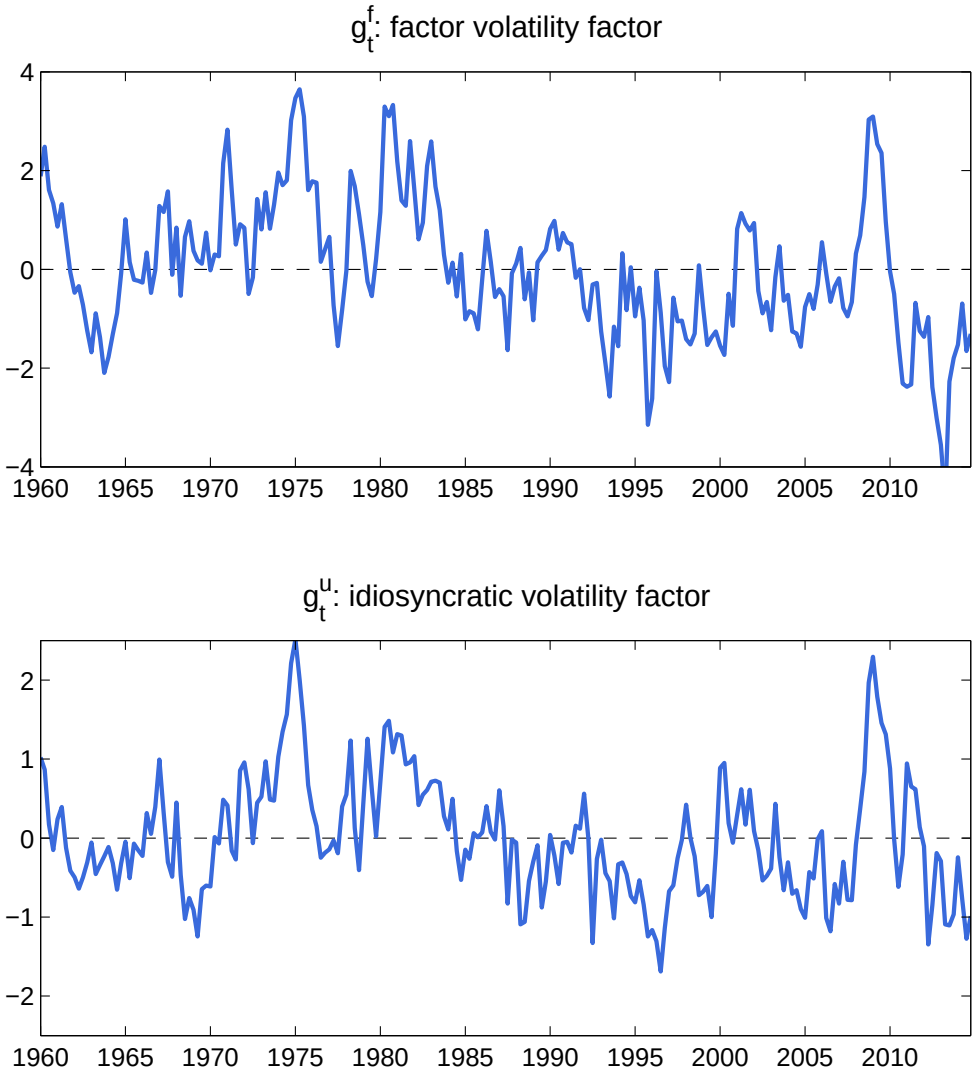


Figure 3.5: Median volatility factors from Bayesian estimation (demeaned). **Top:** factor volatility factor. **Bottom:** idiosyncratic volatility factor.

	prod	price	credit	int rate
factor loading (k_j^f)	0.903	0.753	0.603	0.709
	(0.148)	(0.000)	(0.136)	(0.170)
std dev of idiosyn innovation (ω_j^f)	0.829	0.710	1.064	0.944
	(0.340)	(0.284)	(0.441)	(0.376)
contribution of g_t^f	0.755	0.742	0.467	0.588
	(0.152)	(0.149)	(0.210)	(0.210)

Table 3.2: Factor loadings (k_j^f), standard deviation of idiosyncratic component (ω_j^f), and contribution of factor for factor volatilities based on variance decomposition from Bayesian estimation. Median estimates with standard deviations in parantheses.

g_t^f remains largely below average except during the 2008-2009 recession. On the other hand, g_t^u fluctuates around its mean during the entire sample period. The results therefore suggest that the evidence from smaller datasets reflect the Great Moderation in factor volatility. The effect is diminished in large datasets due to the relative stability of the idiosyncratic volatility factor g_t^u.

Volatility factor loadings

The estimated factor loadings and variance decompositions show that there is both comovement among the volatilities and substantial idiosyncratic movements in the volatilities. The dispersion in the posterior estimates for the factor loadings suggests that the likelihood is informative despite the size of the parameter space.

Table 3.2 reports the median and standard deviation of k_j^f, ω_j^f, as well as the contribution of g_t^f based on a variance decomposition. The factor volatility factor g_t^f loads positively on each of the volatility factors. The fact that these estimates are all significantly above zero is consistent with the estimates from the PCA analysis. The volatility factor g_t^f accounts for over half of the variation in the factor volatilities overall.

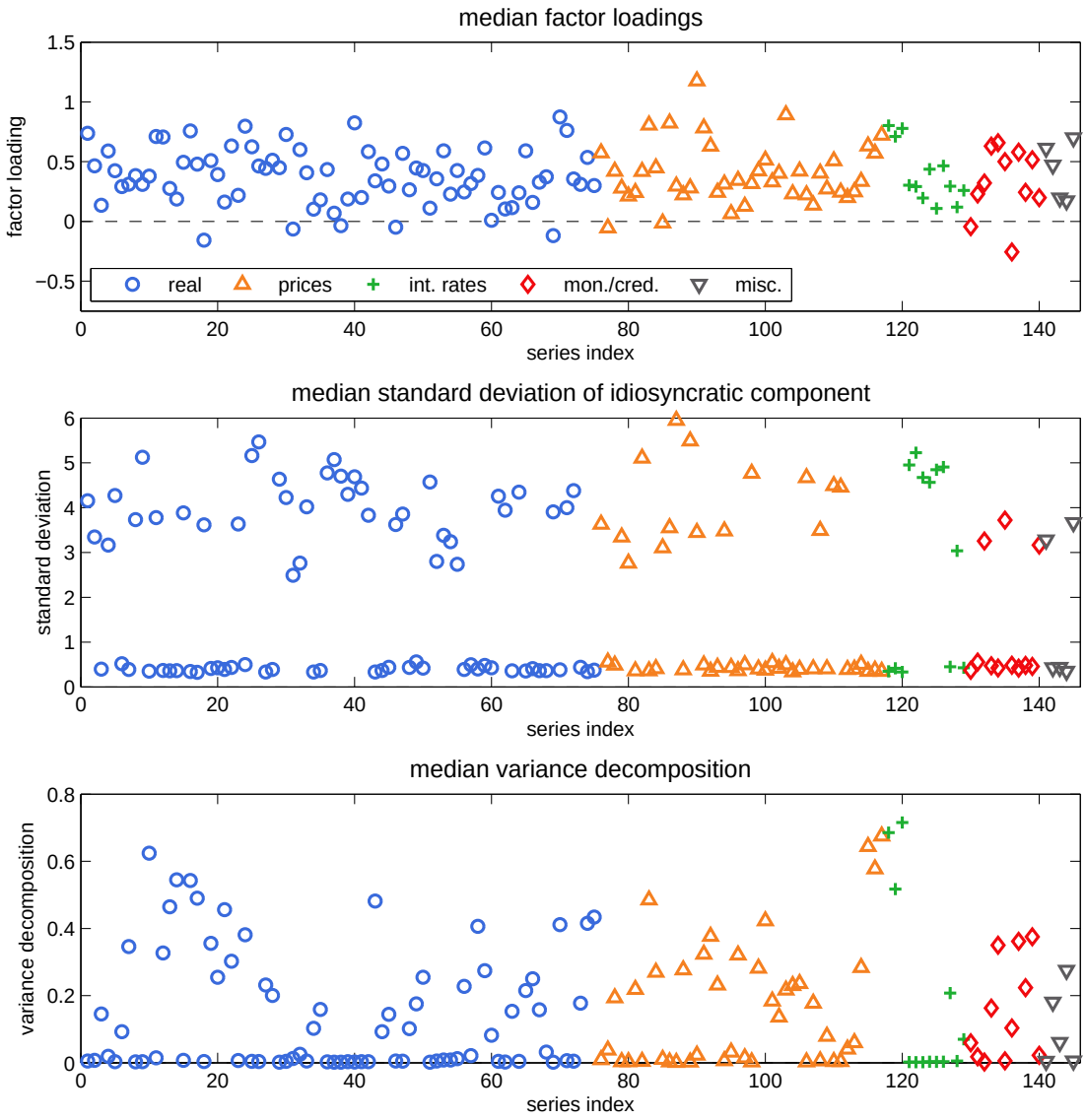


Figure 3.6: Median factor loadings (k_i^u), standard deviation of idiosyncratic component (ω_i^u), and contribution of factor for idiosyncratic volatilities from Bayesian estimation. Each point shows value for specific series. Series arranged along horizontal axis by category (real variables, prices, interest rates, money and credit, miscellaneous); marker type indicates category. **Top:** factor loadings. **Middle:** standard deviation of idiosyncratic component of idiosyncratic volatilities. **Bottom:** contribution of factor from variance decomposition.

Figure 3.6 shows that the median idiosyncratic volatility factor loadings are mostly positive but relatively dispersed, as in the PCA estimation. Similarly, the median estimates of ω_i^u vary widely across i. Variance decompositions of the idiosyncratic volatilities reveal a range of estimates for the fraction of the variation in the idiosyncratic volatilities explained by g_t^u. For a number of the series, the volatility factor explains close to none of the variation of the idiosyncratic volatilities. For example, the variation of idiosyncratic volatility of most interest rate variables is largely explained by their idiosyncratic component. Most of these variables are associated with bond yields, suggesting that a data set with more financial variables may give rise to a financial volatility factor, as proposed in Ludvigson et al. (2015). On the other hand, there are also series whose idiosyncratic volatility is largely explained by g_t^u. For instance, g_t^u explains around three quarters of the idiosyncratic volatility of the effective federal funds rate.

3.5 Existing uncertainty indices

3.5.1 Methodology

Existing mesaures of uncertainty typically aggregate different sources of uncertainty into one-dimensional indices. The decomposition here allows us to study what sources of volatility enter these uncertainty indices, which is important for interpreting the indices and connecting them to economic theory.

To that end, I regress different uncertainty indices on the volatility factors g_t and the idiosyncratic components v_t^f of the factor volatilities:

$$x_{\iota,t} = \beta_{\iota,0} + \beta'_{\iota,g} g_t + \beta'_{\iota,v} v_t^f + \epsilon_{\iota,t} \tag{3.33}$$

where $x_{\iota,t}$ is an uncertainty index, with ι indexing the different measures. The coefficients $\beta_{\iota,g}$ and $\beta_{\iota,v}$ indicate whether a particular source of volatility is captured by the uncertainty index $x_{\iota,t}$. Since we use estimates of realized volatility in a data set that focuses on macroeconomic variables, $\epsilon_{\iota,t}$ captures external sources of uncertainty (e.g. model uncertainty), volatility of variables omitted from the data set, and measurement error.

Conversely, one can ask how much of the variation in g_t and v_t^f can be accounted for by existing uncertainty indices. I approach this question by regressing the elements of g_t and v_t^f on the vector of uncertainty indices x_t.

I consider three broad groups of measures of uncertainty, and within each group consider uncertainty indices that focus on different parts of the economy:

1. Volatility: Bloom (2009) uncertainty index based on VIX; Jurado et al. (2015) macroeconomic uncertainty; Ludvigson et al. (2015) financial uncertainty; Fernández-Villaverde et al. (2015) fiscal volatility.

2. Media: Baker et al. (2016) economic policy uncertainty; Baker et al. (2016) monetary policy uncertainty.

3. Disagreement: SPF CPI forecast interquartile range; SPF real GDP forecast interquartile range.

3.5.2 Results

Table 3.3 reports the regression estimates.

The macroeconomic uncertainty index of Jurado et al. (2015) is also obtained by estimating a factor model using a large macroeconomic data set. Given the closeness of the exercise in Jurado et al. (2015) to the one here, it is unsurprising that the regression yields a large adjusted R^2 of 0.578. However, the regression only yields significant coefficients on the idiosyncratic volatility factor g_t^u, which on

	Volatility				Media		Disagreement	
	VIX	macro	finance	fiscal	EPU	MPU	CPI	GDP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VF								
factor	0.030	-0.021	0.061	-0.235	0.071	-0.259	-0.110	0.058
	(0.086)	(0.036)	(0.053)	(0.099)	(0.221)	(0.172)	(0.124)	(0.154)
id	0.151	0.083	0.061	0.184	0.092	0.266	0.232	0.138
	(0.058)	(0.017)	(0.027)	(0.053)	(0.108)	(0.071)	(0.063)	(0.096)
FV								
prod	-0.073	0.108	-0.137	0.783	0.048	0.958	0.563	0.453
	(0.357)	(0.107)	(0.159)	(0.296)	(0.622)	(0.508)	(0.335)	(0.428)
price	0.076	0.208	-0.1179	0.901	-0.119	0.621	0.651	0.194
	(0.370)	(0.111)	(0.166)	(0.302)	(0.723)	(0.559)	(0.389)	(0.477)
credit	-0.087	0.063	-0.073	0.451	0.020	0.523	0.436	0.271
	(0.193)	(0.057)	(0.089)	(0.157)	(0.353)	(0.302)	(0.179)	(0.229)
int	0.010	0.101	-0.082	0.395	0.026	0.503	0.360	0.237
	(0.222)	(0.067)	(0.107)	(0.182)	(0.457)	(0.367)	(0.254)	(0.283)
R^2	0.203	0.578	0.272	0.246	0.031	0.252	0.347	0.485
Obs	210	218	218	177	120	120	134	185

Table 3.3: Estimates from regression of existing uncertainty indices on volatility factors (VF) g_t and idiosyncratic component of factor volatilities (FV) v_t^f. Newey-West standard errors in parantheses.

its own explains 44% of the variation in the Jurado et al. (2015) macroeconomic uncertainty index. The VIX index of Bloom (2009) and the financial uncertainty index in Ludvigson et al. (2015) also yield a significant coefficients only for the idiosyncratic volatility factor g_t^u, but have smaller adjusted R^2 values of 0.203 and 0.272 respectively. These estimates emphasize the importance of including idiosyncratic or cross-sectional volatility shocks in any model that seeks to rationalize the behavior of these indices.

Fernández-Villaverde et al. (2015) estimate fiscal volatility using a linear model of taxes or government spending with stochastic volatility. I use the measure of capital tax volatility, which they use as a benchmark. The regression produces positive significant coefficients on all factor volatilities and a non-trivial adjusted R^2 of 0.246. Macroeconomic volatility may thus not have been fully controlled for in the capital tax volatility index, potentially leading to omitted variable bias.

Baker et al. (2016) use text analysis of news media reports to obtain measures of economic policy uncertainty (EPU) as well as uncertainty in other areas of the economy. The regression of the EPU index on g_t and v_t^f yields a small adjusted R^2 of 0.031 and no significant coefficients. There are several explanations for the result. Firstly, the EPU index focuses on policy variables such as taxes and government spending that are relatively underrepresented in the data set here. In addition, the EPU index seeks to measure ex-ante uncertainty reflected in media reports, while the measures of volatility here are ex-post measures. Furthermore, media reports reflect not just volatility but also Knightian uncertainty. The monetary policy uncertainty (MPU) index of Baker et al. (2016) retains the two latter distinctions, but focuses on policy that is more well-represented in our data through interest rate variables. The MPU index yields a significant coefficient on g_t^u, which is consistent with the fact that almost 80% of the volatility of the idiosyncratic component of the federal funds rate

is explained by g_t^u. In addition, the increased adjusted R^2 of 0.252 reflects the fact that monetary policy variables are more well-represented in the data here.

The final set of uncertainty indices are CPI and real GDP forecast dispersions in the SPF, motivated by Bachmann et al. (2013). They both yield relatively high adjusted R^2s of 0.347 and 0.485 for CPI and GDP respectively, suggesting that forecast dispersion is in fact predicted by a linear combination of g_t and v_t^f. The fact that forecast dispersion is larger during periods when factor shocks are jointly more volatile should be accounted for in modeling forecaster disagreement.

We now turn to the question of whether the uncertainty indices considered here fully capture the variation in g_t and v_t^f. The results are summarized in Table 3.4. Over half the variation in the volatility factors g_t are captured by the uncertainty indices, with the adjusted R^2 for g_t^f and g_t^u at 0.485 and 0.611 respectively. On the other hand, $v_{\text{prod},t}^f$ only yields an R^2 of 0.112, while $v_{\text{int},t}^f$ yields an R^2 of 0.167. These estimates suggest that while the volatility factors are well accounted for, the uncertainty indices do not fully span the space of factor volatilities.

To conclude, the information existing about sources of volatility contained in uncertainty indices is not always obvious at the outset. The analysis above reveals the volatility that enters these uncertainty indices, aiding in our understanding and interpretation of the indices. Conversely, the uncertainty indices considered above fail to capture several sources of volatility. Whether this is consequential for understanding the effects of volatility on the macroeconomy depends on how the different sources of volatility interact with the macroeconomy.

	Vol factors		Factor volatilities			
	factor	idiosyn	prod	price	credit	int rate
	(1)	(2)	(3)	(4)	(5)	(6)
Volatility						
VIX	0.260	-0.076	-0.100	-0.276	0.150	0.548
	(0.475)	(0.210)	(0.300)	(0.216)	(0.289)	(0.314)
macro	7.183	3.926	-0.800	2.457	3.149	-1.837
	(1.321)	(0.725)	(1.117)	(0.833)	(1.375)	(1.173)
finance	-0.498	1.016	0.434	-0.687	-0.890	0.512
	(0.918)	(0.434)	(0.624)	(0.533)	(0.606)	(0.653)
fiscal	0.261	-0.218	0.052	0.124	0.311	-0.223
	(0.364)	(0.199)	(0.154)	(0.110)	(0.192)	(0.209)
Media						
EPU	0.238	-0.522	-0.353	0.175	-0.084	0.364
	(0.438)	(0.148)	(0.170)	(0.109)	(0.194)	(0.245)
MPU	-0.424	0.840	0.648	-0.568	0.219	-0.367
	(0.702)	(0.194)	(0.248)	(0.181)	(0.325)	(0.333)
Disagreement						
CPI	-0.208	0.052	-0.199	0.109	0.249	-0.223
	(0.303)	(0.157)	(0.198)	(0.125)	(0.180)	(0.180)
GDP	1.113	0.279	0.181	0.222	-0.352	0.211
	(0.267)	(0.113)	(0.145)	(0.098)	(0.197)	(0.182)
Adj. R^2	0.485	0.611	0.112	0.383	0.283	0.167
Obs	120	120	120	120	120	120

Table 3.4: Estimates from regression of volatility factors g_t and idiosyncratic component of factor volatilities v_t^f on existing uncertainty indices. Newey-West standard errors in parantheses.

3.6 Volatility and the Macroeconomy

3.6.1 Setup

The results thus far raise the question of which sources of volatility are most important for explaining macroeconomic variables. I use a reduced-rank regression to estimate the most important linear combinations of volatilities for explaining the unexplained variation in macroeconomic levels, and the contribution of these linear combinations[7].

In particular, consider the reduced-rank regression model

$$y_t^{\text{VAR}} = \Phi(L)\, y_t^{\text{VAR}} + \Theta\Xi z_t + \nu_t^{\text{VAR}} \tag{3.34}$$

where $\Phi(L)$ is a lag polynomial with four lags and y_t^{VAR} is a 7-dimensional vector of industrial production, unemployment, hours, inflation, real wages, the federal funds rate, and the S&P index. The variables in y_t^{VAR} are similar to those in the VAR exercise of Bloom (2009). Define $z_t \equiv \left(\hat{g}_t, \hat{v}_t^f\right)'$, where $\hat{g}_t$ and $\hat{v}_t^f$ denote the median estimates for g_t and v_t^f. Ξ is a $r \times \left(M^g + M^f\right)$ matrix of factor loadings, which yields factors Ξz_t. Θ is a $7 \times r$ matrix that represents the dependence of y_t^{VAR} on Ξz_t. Following Anderson (1951), I estimate the model using canonical correlations.

$\Theta\Xi$ gives the variation of y_t^{VAR} associated with a change in z_t after controlling for the lags of y_t^{VAR}. With $r < M^g + M^f$, the model restricts y_t^{VAR} to depend only on an r-dimensional subspace of the volatilities z_t. The canonical correlations and variates will not depend on the rotation of z_t and by extension do not depend on the identification of the structural factor shocks in equation (3.2).

	(1)	(2)	(3)	(4)	(5)	(6)
Volatility factors						
factor	0.022	0.109	0.573	0.597	-0.550	0.014
idiosyn	-0.149	0.426	0.832	-0.069	0.212	0.234
Factor volatilities						
production	0.033	-0.713	0.172	0.427	0.115	0.516
price	0.414	0.743	-0.135	0.177	-0.336	0.338
credit	-0.461	0.446	-0.027	0.330	0.130	-0.680
interest rate	-0.166	-0.191	0.283	-0.502	-0.687	-0.364
Dependent variables						
industrial production	0.122	-0.143	0.070	0.024	-0.037	-0.054
unemployment	-0.234	-0.013	0.006	-0.008	0.048	-0.003
hours	0.329	-0.175	0.040	-0.001	-0.025	-0.010
inflation	-0.004	-0.109	0.087	0.004	0.081	0.013
real wages	0.027	0.305	0.065	0.065	0.015	-0.033
federal funds rate	0.017	-0.012	0.154	-0.126	-0.005	-0.002
S&P index	0.127	0.034	-0.139	-0.079	0.021	-0.055
canonical correlation	0.394	0.374	0.236	0.188	0.115	0.095
$\Pr > F$	2e-5	0.003	0.242	0.435	0.592	0.393

Table 3.5: Canonical correlations of y_t^{VAR} with z_t after controlling for lagged y_t^{VAR}. **Top:** correlation of canonical variates with residuals of volatility factors and factor volatilities. **Middle:** correlation of canonical variates with residuals of y_t^{VAR}. **Bottom:** canonical correlations and F-test.

3.6.2 Results

Table 3.5 reports the canonical correlations, as well as the correlations between the canonical variates and the residuals from regressions of y_t^{VAR} and z_t on lagged y_t^{VAR}, which we shall henceforth denote by $\tilde{\cdot}$.

The canonical correlations are non-trivial, suggesting that the aggregate volatility does indeed account for some of the variation of the macroeconomic variables not accounted for by their lags. An F-test rejects the null that the first two canonical variate pairs are uncorrelated but cannot reject the null that the subsequent canonical variate pairs are uncorrelated. We thus focus on the first two canonical variate pairs, plotted in Figure 3.7.

The first canonical variate has a strong positive correlation with $\tilde{v}^f_{\text{price},t}$ of 0.414, a negative correlation with $\tilde{v}^f_{\text{cred},t}$ of -0.461. It has positive correlations with the residuals of industrial production and hours, and a negative correlation with the residual of unemployment, but looks substantially different from existing uncertainty indices. It is closest to the Jurado et al. (2015) macroeconomic uncertainty index, with a correlation of -0.246.

The second canonical variate is primarily associated with a rise in real wages (with a correlation of 0.305). It is most strongly associated with low production factor volatility (with a correlation of -0.713) and high price volatility (with a correlation of 0.743). We note that production factor volatility was not well spanned by the existing uncertainty indices considered here, suggesting that the indices considered in Section 3.5 may fail to capture this dimension of volatility despite its interactions with macroeconomic levels.

[7] A full analysis of the effect of volatility on the macroeconomy would incorporate the level effect of volatility shocks into the model, as in Carriero et al. (2016b). The analysis here instead follows the spirit of Bloom (2009) and subsequent work by incorporating point estimates of measures of uncertainty into VARs.

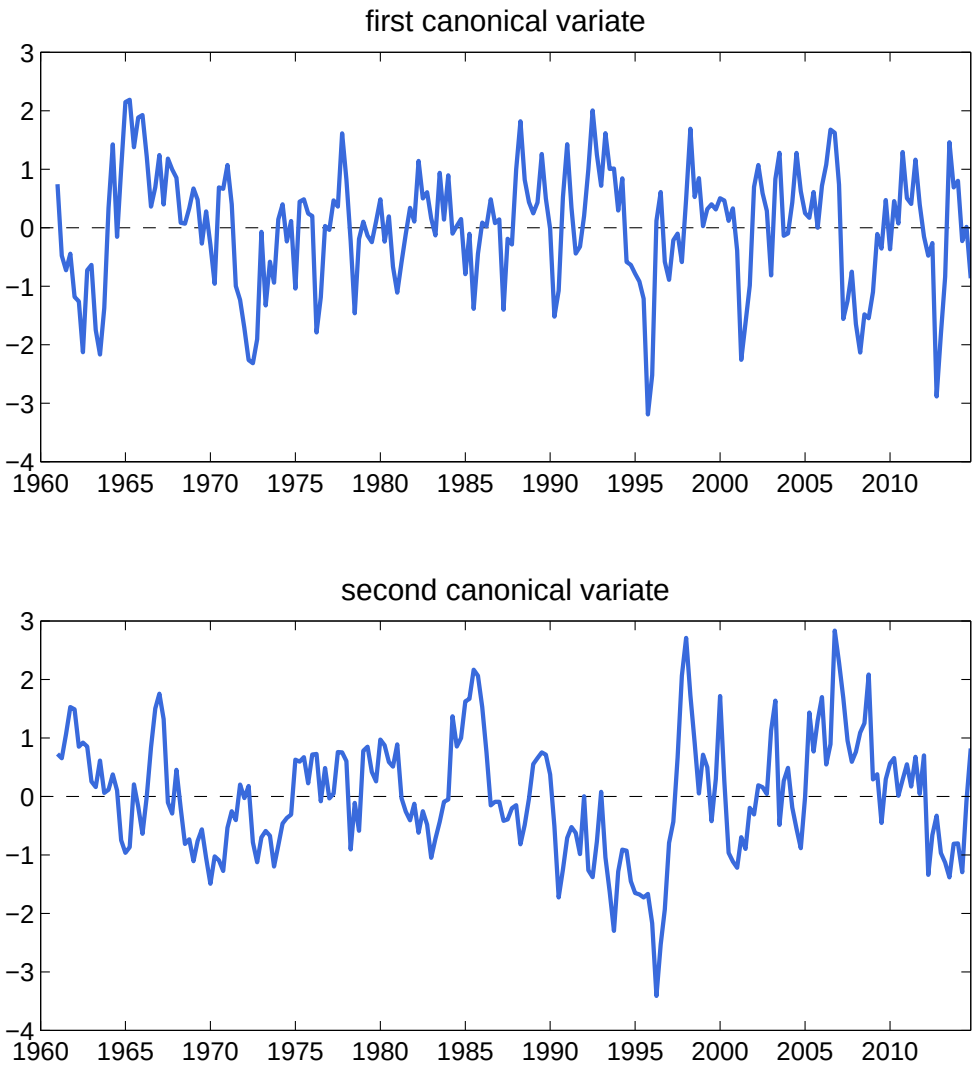


Figure 3.7: First and second canonical variates (normalized to have mean zero and standard deviation one).

While the motivation for much recent work on uncertainty is the implications for macroeconomic levels, the evidence above highlights that existing indices may fail to capture some of the components of volatility that have first moment effects. Moreover, by focusing on one-dimensional indices of uncertainty, the existing literature conflates disparate effects of volatility on the macroeconomy, making interpretation of these existing results problematic.

3.7 Conclusion

Disentangling the sources of volatility is necessary to understand the evolution and effects of volatility. The framework here can be used with large datasets to determine whether changes in volatility arose from variation in the volatility of specific factors or from the comovement of the volatility of various shocks in the economy. I utilized PCA estimates to inform the Bayesian estimation, which revealed economically substantive differences in the evolution of different volatities and shed light on which volatilities existing uncertainty indices capture. The decompositions are important when interpreting these indices and modeling their reduced form effects on the macroeconomy. Moreover, by focusing on one-dimensional indices, the existing literature potentially misses out on interesting interactions between volatilities and levels.

References

Anderson, T. W. (1951). Estimating Linear Restricctions on Regression Coefficients for Multivariate Normal Distributions. *Annals of Mathematical Statistics*, 22(3):327–351.

Bachmann, R., Elstner, S., and Sims, E. R. (2013). Uncertainty and Economic Activity: Evidence from Business Survey Data. *American Economic Journal:*

Macroeconomics, 5(2):217–249.

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71(1):135–171.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics*, 131(4):1593–1636.

Baumeister, C. and Hamilton, J. D. (2015). Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information. *Econometrica*, 83(5):1963–1999.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica*, 77(3):623–685.

Boivin, J. and Giannoni, M. P. (2006). DSGE Models in a Data-rich Environment. *NBER Working Paper*.

Carriero, A., Clark, T. E., and Marcellino, M. G. (2016a). Large Vector Autoregressions with Stochastic Volatility and Flexible Priors. *Federal Reserve Bank of Cleveland Working Paper*.

Carriero, A., Clark, T. E., and Marcellino, M. G. (2016b). Measuring Uncertainty and Its Impact on the Economy. *Federal Reserve Bank of Cleveland Working Paper*.

Del Negro, M. and Otrok, C. (2008). Dynamic Factor Models with Time-varying Parameters: Measuring Changes in International Business Cycles. *FRB of New York Staff Report*.

Diebold, F. X., Schorfheide, F., and Shin, M. (2017). Real-time Forecast Evaluation of DSGE Models with Stochastic Volatility. *Journal of Econometrics*, 201(2):322–332.

Eickmeier, S. and Ziegler, C. (2008). How Successful are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach. *Journal of Forecasting*, 27(3):237–265.

Engle, R. F. (1982). Autoregressive Conditional Heteroscadasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007.

Fernández-Villaverde, J., Guerrón-Quintana, P., Kuester, K., and Rubio-Ramírez, J. (2015). Fiscal Volatility Shocks and Economic Activity. *American Economic Review*, 105(11):3352–84.

Fernández-Villaverde, J. and Rubio-Ramírez, J. (2013). Macroeconomics and Volatility: Data, Models, and Estimation. In *Advances in Economics and Econometrics: Tenth World Congress*, pages 137–183. Cambridge University Press.

Giannone, D., Lenza, M., and Reichlin, L. (2008). Explaining the Great Moderation: It is Not the Shocks. *Journal of the European Economic Association*, 6(May 2008):621–633.

Gorodnichenko, Y. and Ng, S. (2017). Level and Volatility Factors in Macroeconomic Data. *Journal of Monetary Economics*, 91:52–68.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring Uncertainty. *American Economic Review*, 105(3):1177–1216.

Justiniano, A. and Primiceri, G. E. (2008). The Time-Varying Volatility of Macroeconomic Fluctuations. *American Economic Review*, 98(3):604–641.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65(3):361–393.

Ludvigson, S. C., Ma, S., and Ng, S. (2015). Uncertainty and Business Cycles: Ecogenous Impulse or Endogenous Response? *NBER Working Paper*.

Primiceri, G. E. (2005). Time-Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies*, 72(3):821–852.

Rossi, B., Sekhposyan, T., and Soupre, M. (2017). Understanding the Sources of Macroeconomic Uncertainty. *Working paper*.

Rubio-Ramirez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference. *Review of Economic Studies*, 77(2):665–696.

Sims, C. A. and Zha, T. (2006). Were There Regime Switches in US Monetary Policy? *American Economic Review*, 96(1):54–81.

Stock, J. H. and Watson, M. W. (2012a). Disentangling the Channels of the 2007-2009 Recession. *Brookings Papers on Economic Activity*, pages 81–135.

Stock, J. H. and Watson, M. W. (2012b). Generalized Shrinkage Methods for Forecasting Using Many Predictors. *Journal of Business and Economic Statistics*, 30(4):481–493.

Stock, J. H. and Watson, M. W. (2016). Core Inflation and Trend Inflation. *Review of Economics and Statistics*, 98(4):770–784.

Appendix

3.A Prior

Denote

$$A^f(L) = \begin{bmatrix} a^1_{1,1} & \cdots & a^1_{1,M} \\ \vdots & \ddots & \vdots \\ a^1_{M,1} & \cdots & a^1_{M,M} \end{bmatrix} L + ... + \begin{bmatrix} a^4_{1,1} & \cdots & a^4_{1,M} \\ \vdots & \ddots & \vdots \\ a^4_{M,1} & \cdots & a^4_{M,M} \end{bmatrix} L^4 \tag{3.35}$$

$$B^f = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ b_{2,1} & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ b_{M,1} & \cdots & b_{M,M-1} & 1 \end{bmatrix} \tag{3.36}$$

The parameters in equations (3.1)-(3.3) have the priors:

$$h^y_{i,1:M} \sim \mathcal{N}(0, I) \text{ for } i > M \tag{3.37}$$

$$a^l_{i,j} \sim \begin{cases} \mathcal{N}(0.75, 0.5^2) & l = 1, i = j \\ \mathcal{N}(0, 0.5^2) & \text{otherwise} \end{cases} \tag{3.38}$$

$$b_{i,j} \sim \mathcal{N}\left(0, 0.5^2\right) \tag{3.39}$$

$$a^u_i \sim \mathcal{N}\left(0.5, 0.5^2\right) \tag{3.40}$$

In addition, I impose the initial conditions:

$$y_{-3:0} = 0 \tag{3.41}$$

$$f_{-3:0} = 0 \tag{3.42}$$

The priors for the parameters in equations (3.20)-(3.24) are described in (3.25)-(3.27), where our calibration leads us to the hyperparameters:

$$\left(\lambda^f, \mu^f, s^f\right) = (0.6584, 1.1432, 0.2171) \tag{3.43}$$

$$(\lambda^u, \mu^u, s^u) = (0.9202, 0.4254, 0.7016) \tag{3.44}$$

In addition, I impose $g_0^f = g_0^u = 0$ and have the initial condition:

$$v_j^f, v_i^u \sim \mathcal{N}\left(0, 2^2\right) \tag{3.45}$$

3.B Gibbs Sampler

3.B.1 Mixture-of-normals approximation

Before describing the Gibbs sampler, it is necessary to briefly describe an approximation from Kim et al. (1998) that we shall use. The fact that $\log\left(\varepsilon_{i,t}^u\right)^2$ and $\log\left(\varepsilon_{j,t}^f\right)^2$ are iid $\log\chi_1^2$ random variables poses an inconvenience for drawing $\log\vec{\sigma}_t^y$ and $\log\vec{\sigma}_t^f$. Following Kim et al. (1998), we approximate the $\log\chi_1^2$ by a mixture of normals, i.e.

$$\log\left(\varepsilon_{j,t}^f\right)^2 \overset{approx}{\sim} \mathcal{N}\left(\mu_\chi\left(\xi_{j,t}^f\right), \sigma_\chi^2\left(\xi_{j,t}^f\right)\right) \tag{3.46}$$

$$\log\left(\varepsilon_{i,t}^u\right)^2 \overset{approx}{\sim} \mathcal{N}\left(\mu_\chi\left(\xi_{i,t}^u\right), \sigma_\chi^2\left(\xi_{i,t}^u\right)\right) \tag{3.47}$$

where $\xi_{i,t}^u$ and $\xi_{j,t}^f$ are iid random variables with support $\left\{1, ..., \overline{\xi}\right\}$, with corresponding probabilities $\left\{p\left(1\right), ..., p\left(\overline{\xi}\right)\right\}$. We follow the 10-component approximation of **?**.

Details on how the approximation allows for convenient sampling of $\log(\tilde{\sigma}_t^2)$ are provided below.

3.B.2 Gibbs sampler

Including the parameters $\xi_{i,t}^u$ and $\xi_{j,t}^f$, we can now consider the following partition of the model parameters:

$$\theta_1 = \left\{H^y, A^f, B^f, f_{1:T}, u_{1:T}\right\} \tag{3.48}$$

$$\theta_2 = \left\{\xi_{1:M,1:T}^f, \xi_{1:N,1:T}^u\right\} \tag{3.49}$$

$$\theta_3 = \left\{K^f, K^u, \Omega^g, \omega_{1:M}^f, \omega_{1:N}^u, g_{1:T}^f, g_{1:T}^u, v_{1:T}^f, v_{1:T}^u\right\} \tag{3.50}$$

θ_1 is the collection of parameters for the factor model for the levels, θ_2 is the sequence of mixture of normals indicators, θ_3 is the sequence of volatilities, and θ_4 is the collection of parameters from the factor model for the volatilities. Given the partition, we can make draws from the posterior as follows:

1. Draw $\theta_1, \theta_2|\theta_3$.
   (a) Draw $\theta_1|\theta_3$.
      i. Draw $f_{1:T}|H^y, A^f, B^f, \theta_3$.
      ii. Draw $H^y, A^f, B^f|f_{1:T}, \theta_3$.
   (b) Draw $\theta_2|\theta_1, \theta_3$.
2. Draw $\theta_3|\theta_1, \theta_2$.
   (a) Draw $g_{1:T}^f, g_{1:T}^u, v_{1:T}^f, v_{1:T}^u|K^f, K^u, \Omega^g, \omega_{1:M}^f, \omega_{1:N}^u, \theta_1, \theta_2$.
   (b) Draw $K^f, K^u, \Omega^g, \omega_{1:M}^f, \omega_{1:N}^u|g_{1:T}^f, g_{1:T}^u, v_{1:T}^f, v_{1:T}^u, \theta_1, \theta_2$.

Drawing factors from a linear Gaussian state space model

We now detail 1(a)(i) and 2(a). Both involve drawing factors from a factor model with known parameters and are special cases of the following model:

$$x_t = Hs_t + \Sigma_t^x \varepsilon_t^x \tag{3.51}$$

$$s_t = As_{t-1} + B^{-1}\Sigma_t^s \varepsilon_t^s \tag{3.52}$$

where x_t is observed data and s_t is a vector of common factors and Σ_t^x and Σ_t^s are diagonal.

Let μ_t and Ξ_t be the mean and variance of s_t conditional on information $\mathcal{I}_t = \{x_\tau : \tau \leq t\}$ up to time t, i.e.

$$s_t|\mathcal{I}_t \sim \mathcal{N}(\mu_t, \Xi_t) \tag{3.53}$$

We can solve for μ_t and Ξ_t from the Kalman filter.

The joint distribution

$$\begin{bmatrix} s_{t+1} \\ s_t \end{bmatrix} |\mathcal{I}_t \sim \mathcal{N}\left(\begin{bmatrix} A\mu_t \\ \mu_t \end{bmatrix}, \begin{bmatrix} A\Xi_t A' + \Sigma_{t+1}^s & A\Xi_t \\ \Xi_t A' & \Xi_t \end{bmatrix} \right) \tag{3.54}$$

implies that we can write

$$s_t = \mu_t + \Xi_t A' \left(A\Xi_t A' + \Sigma_{t+1}^s\right)^{-1} (s_{t+1} - A\mu_t) + \zeta_t \tag{3.55}$$

$$\zeta_t \sim \mathcal{N}\left(0, \Xi_t - \Xi_t A' \left(A\Xi_t A' + \Sigma_{t+1}^s\right)^{-1} A\Xi_t\right) \tag{3.56}$$

with ζ_t uncorrelated with past observation of x_t and with s_{t+1}. Therefore, we can draw the terminal value $s_T|\mathcal{I}_T$ from the output of the Kalman filter, and subsequently draw $s_t|s_{t+1}$ for $t < T$.

For step 1(a)(i), we cast equations (3.1)-(3.3) into the above form of equations (3.51)-(3.52) by writing the ith row of equation (3.1) as:

$$(I - a_i^u L)\, y_{i,t} = (I - a_i^u L)\, h_i^y f_t + \sigma_{i,t}^u \varepsilon_{i,t}^u \tag{3.57}$$

and writing the transition equation (3.2) in companion form. For step 2(a), we note that our mixture of normals approximation gives us a linear Gaussian state space model for (3.20)-(3.24) conditional on a given draw for the mixture weights θ_2, where the unobserved state is the vector $(g_t', v_t')'$ (see Subsection 3.B.2).

Drawing factor loadings in the presence of stochastic volatility

We now consider the first part of step 1(a)(ii) and the first part of step 2(b).

Given $s_{1:T}$ and $\Sigma_{1:T}^x$, each row of equation (3.51) is in the form of a generalized least squares equation, from which we can draw H. In particular, we can normalize $(x_{i,t}, s_t)$ for each i and t by $\sigma_{i,t}^x$, and rewrite the ith row of equation (3.51) as

$$\tilde{x}_{i,t} = h_i \tilde{s}_t + \varepsilon_{i,t}^x \tag{3.58}$$

where $\tilde{x}_{i,t}$ and $\tilde{s}_t$ are appropriately scaled. Given our normal prior for $h_i \overset{\text{prior}}{\sim} \mathcal{N}\left(h_i^\pi, (\Lambda_i^\pi)^{-1}\right)$, we now draw from the posterior of each vector $h_i \sim \mathcal{N}\left(\hat{h}_i, \widehat{\Lambda}_i^{-1}\right)$, with

$$\hat{h}_i = \widehat{\Lambda}_i^{-1} \left(\Lambda_i^\pi h_i^\pi + \tilde{s}_{1:T}' \tilde{x}_{i,1:T}\right) \tag{3.59}$$

$$\widehat{\Lambda}_i = \left(\tilde{s}_{1:T}' \tilde{s}_{1:T} + \Lambda_i^\pi\right) \tag{3.60}$$

where $\tilde{x}_{i,1:T} = (\tilde{x}_{i,1}, ..., \tilde{x}_{i,T})'$ and $\tilde{s}_{1:T} = (\tilde{s}_1, ..., \tilde{s}_T)'$. This allows use to draw H^y, K^f, K^u.

Drawing coefficients of VAR with stochastic volatility

For the remaining parameters to be drawn in step 1(a)(ii), we note that 3.20 is a VAR with stochastic volatility, which is a special case of the model in Primiceri (2005).

Given Σ_t^s in equation (3.52), we can draw A and B. In particular, write A as

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,k} \\ \vdots & \ddots & \vdots \\ a_{k,1} & \cdots & a_{k,k} \end{bmatrix} \tag{3.61}$$

We can write equation (3.52) as

$$s_t = (I_k \otimes s_t') \begin{bmatrix} a_{1,1:k}' \\ \vdots \\ a_{k,1:k}' \end{bmatrix} + B^{-1}\Sigma_t^s \varepsilon_t^s \tag{3.62}$$

where $a_{i,1:k} = (a_{i,1}, ..., a_{i,k})$.

Similarly, write B as

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ b_{2,1} & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ b_{k,1} & \cdots & b_{k,k-1} & 1 \end{bmatrix} \tag{3.63}$$

We can write equation (3.52) as:

$$B\left(s_t - A\left(L\right)s_t\right) = \Sigma_t^s \varepsilon_t^s \tag{3.64}$$

Define $\hat{\varepsilon}_t^s \equiv f_t - A(L) f_t$. Then we can write

$$\hat{\varepsilon}_t^s = - \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ \left(\hat{\varepsilon}_{1,t}^s\right)' & 0 & \cdots & 0 \\ 0 & \left(\hat{\varepsilon}_{1:2,t}^s\right)' & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \left(\hat{\varepsilon}_{1:(k-1),t}^s\right)' \end{bmatrix} \begin{bmatrix} b_{2,1}' \\ b_{3,1:2}' \\ \vdots \\ b_{k,1:(k-1)}' \end{bmatrix} + \Sigma_t^s \varepsilon_t^s \tag{3.65}$$

Both (3.62) and (3.65) are linear models with time-varying variances that can be drawn using weighted least squares.

Mixture of normals approximation

I now describe the implementation of the mixture of normals approximation from Kim et al. (1998).

With the mixture of normals approximation in (3.47), we can make posterior draws for $\log \sigma_{j,t}^f$ and $\log \sigma_{i,t}^u$ from a linear Gaussian state space model conditional on $\xi_{j,t}^f$ and $\xi_{i,t}^u$ respectively. I will describe the draws of $\log \sigma_{j,t}^f$. The process for drawing $\log \sigma_{i,t}^u$ is analogous.

In particular, we observe $\log\left(\sigma_{j,t}^f \varepsilon_{j,t}^f\right)^2$ and would like to make draws from the conditional posterior of $\log \sigma_{j,t}^f$. In order to make the approximation robust against small values of $\log\left(\varepsilon_{j,t}^f\right)^2$, I offset $\left(\sigma_{j,t}^f \varepsilon_{j,t}^f\right)^2$ by $c = 0.001$, and offset the means of the mixture of normals by 1.2704, as suggested by Kim et al. (1998). Omitting c leads to an additional peak in the likelihood at $\sigma_t^f = 0$, which the MCMC algorithm can potentially get stuck at. Let $\mathbf{1}_n$ and $\mathbf{1}_k$ denote vectors of n and k ones respectively. For a generic vector $X = (x_1, ..., x_{n^*})'$, use the notation $\log X = (\log x_1, ..., \log x_{n^*})'$. Lastly, denote the offset chi-squares by $\log\left(\tilde{\varepsilon}_{j,t}^f\right)^2$. Accounting for c, we have the

system

$$\log\left(\left(\sigma^f_{j,t}\varepsilon^f_{j,t}\right)^2 + 0.001 \times \mathbf{1}_n\right) = 2k^f_j g^f_t + 2v^f_{j,t} + \log\left(\tilde{\varepsilon}^f_{j,t}\right)^2 \tag{3.66}$$

Step 2(a) is now straightforward. Given $\xi^f_{j,1:T}$, this is a linear Gaussian state space model with known coefficients and unobserved states g^f_t and $v^f_{j,t}$, which we can draw as described in Subsection 3.B.2.

Drawing Ω^g, ω^f_j, ω^u_i

The remaining parameters Ω^g, ω^f_j, ω^u_i are straightforward to draw. The conjugate inverse-Wishart prior for Ω^g implies that the posterior for Ω^g is also an inverse-Wishart conditional on $g_{1:T}$. Similarly, $\left(\omega^f_j\right)^2$ and $(\omega^u_i)^2$ have inverse-Gamma posteriors conditional on $v^f_{j,1:T}$ and $v^u_{i,1:T}$ respectively.