Robust Bayesian Estimation and Inference for Dynamic Stochastic General Equilibrium Models*

Yizhou Kuang

This Version: February 28th, 2024
Click here for the latest version

Abstract

This paper introduces a new algorithm to conduct robust Bayesian estimation and inference in dynamic stochastic general equilibrium models. The algorithm combines standard Bayesian methods with an equivalence characterization of model solutions. This algorithm allows researchers to perform the following analysis: First, find the complete range of posterior means of both the structural parameters and any parameters of interest robust to the choice of priors in a sense I make precise. Second, derive the robust Bayesian credible region for these parameters. I prove the validity of this algorithm and apply this method to the models in Cochrane (2011), An and Schorfheide (2007) and Smets and Wouters (2007) to achieve robust estimations for structural parameters and impulse responses. In addition, I conduct a sensitivity analysis of optimal monetary policy rules with respect to the choice of priors and provide bounds to the optimal Taylor rule parameters.

KEYWORDS: DSGE models, Bayesian inference, identified set, informative priors, policy analysis, parameter uncertainty

*Department of Economics, University of Manchester. Email: yizhou.kuang@manchester.ac.uk. I would like to thank my advisors, Francesca Molinari (chair), Yongmiao Hong, José Luis Montiel Olea, and Kristoffer Nimark, for their great support and advice. This paper has also benefitted from discussions with Majid Al-Sadoon, Enghin Atalay, Ryan Chahrour, Timothy Christensen, Liyu Dou, Axelle Ferriere, Joachim Freyberger, Kyle Herkenhoff, Toru Kitagawa, Marcin Kolasa, Paul Levine, Konrad Menzel, Simon Mongey, Willi Mutschler, Johannes Pfeifer, Mikkel Plagborg-Møller, Chen Qiu, Zhongjun Qu, Jonathan Roth, Raffaele Rossi, Azeem Shaikh, Alessio Volpicella, Christian Wolf, Molin Zhong, and numerous seminar and conference participants for useful feedback and suggestions. I am especially grateful to Marcin Kolasa for kindly sharing their *SINGULAR* coding insights. All errors are my own.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have been the workhorse of modern macroeconomics. They are taught in almost all doctoral programs in economics and are used by many central banks and financial institutions, such as the US Federal Reserve, IMF and Sveriges Riksbank for monetary policy analysis.[1] Estimation and inference are typically conducted using Bayesian methods (Smets and Wouters, 2007; Christiano et al., 2010; Justiniano and Preston, 2010).

Bayesian methods are attractive to macroeconomists in estimating DSGE models for multiple reasons. One is that the advent of Markov chain Monte Carlo (MCMC) methods allows researchers to estimate and evaluate complicated models. In addition, under the Bayesian framework, researchers' prior knowledge of parameters can be incorporated into the estimation of the model. With a given prior, researchers can draw from a posterior distribution through MCMC without having to worry about identification issues; in fact, Lindley (1972) (p. 46) concludes "that unidentifiability causes no real difficulty in the Bayesian approach."

However, standard Bayesian inference can be problematic in models that are not point-identified (Poirier, 1998; Gustafson, 2009; Moon and Schorfheide, 2012; Morris, 2014). For example, Moon and Schorfheide (2012) find that any choice of priors would result in Bayesian highest-posterior-density sets that are asymptotically strictly smaller than the true identified set. Morris (2014) shows in simulation that the posterior mode will not exhibit a "jump" pattern in a bimodal likelihood case, while the prior shifts its weight continuously from one mode to another. In addition, it is often hard to even tell whether the parameters in a DSGE model are identified because of its complicated structure and the large number of parameters. These factors make the estimation result from the standard Bayesian approach potentially inconsistent: the posterior mean can fail to converge to the true value, and its probability limit is sensitive to the choice of prior even asymptotically.

In this paper, I propose a robust Bayesian algorithm that allows finding the range of posterior means over a class of priors and a robust Bayesian credible region that has credibility of at least $1-\alpha$ over the same class. Instead of committing to a prior with great confidence, researchers can start with any "reasonable" prior with positive density on the parameter values allowed by the

[1] A non-exhaustive list of institutions that have developed DSGE models includes the Bank of Canada, the Bank of England, the Bank of Israel, the Central Bank of Chile, the Central Reserve Bank of Peru, the Czech Central Bank, the European Central Bank, the International Monetary Fund, the Norges Bank, the Sveriges Riksbank, and the U.S. Federal Reserve (Adolfson et al., 2011; Del Negro et al., 2013; Lindé, 2018; Christiano et al., 2018).

model constraints. Then, for each given parameter drawn from the posterior, researchers can find the observationally equivalent set of this parameter and use it to solve an optimization problem for the upper and lower bounds of the parameters of interest. The final step is to find the mean or the quantile of the bounds attained across posterior draws, or both. This algorithm is convenient to use because it can be applied complementarily with standard Bayesian estimation methods.

On the theory side, I show that the new robust Bayesian algorithm is valid. Specifically, the expected set of means achieved from this algorithm is the same as the collection of all posterior expectations generated by an arbitrary choice of prior within a distribution class suggested by Giacomini and Kitagawa (2021). Therefore, researchers can use the algorithm to conduct robust Bayesian inference on the parameters of interest without trying out all possible priors. I also show that under some regularity conditions, the estimated range of posterior means will asymptotically converge to the true identified set.

In the second part of the paper, I apply this algorithm to multiple models to show it is useful for understanding results from the literature that are based on historical data, for conducting inference, and for analyzing policies. I start with the model in Cochrane (2011), which is simple enough to be analytically tractable. Using simulated data, I compare the theoretically identified set and the estimated range of posterior means. I show that the algorithm performs well in estimating the identified set. Then I work with a more economically meaningful class of small-scale DSGE models, sometimes referred to as the three-equation New Keynesian model. In particular, I work with models similar to those of Galí and Gertler (1999) and An and Schorfheide (2007) with i.i.d. shocks, auto-correlated shocks, and a variant with a cost-push shock. These three examples have the property that with temporary shocks, parameters are not identified, but the impulse responses are identified. With serially correlated shocks, local identification fails but does not affect policy-making. Moreover, in the case with a cost-push shock, parameter uncertainty causes uncertainty in the optimal policy.

This paper is most closely related to the literature on identification in DSGE models. In a pioneering paper, Canova and Sala (2009) summarize different identification issues that DSGE models can have and propose diagnostics to detect identification deficiencies. Iskrev (2010) provides the sufficient conditions, whereas Komunjer and Ng (2011) and Qu and Tkachenko (2012) provide the necessary and sufficient conditions for local identification taking different paths. Komunjer and Ng (2011) perform analysis from the state-space characterization, and Qu and Tkachenko (2012) use a frequency domain approach. Koop et al. (2013) also propose two Bayesian identification indicators to check local identification. Qu and Tkachenko (2017) offer a more general framework to check global identification by assessing the Kullback-Leibler distance between two parametrizations of DSGE models.

Kociecki and Kolasa (2018) offer an alternative theoretical analysis of global identification based on results from Komunjer and Ng (2011). They build a polynomial equation system to characterize the observationally equivalent state-space parameters. Kocięcki and Kolasa (2023) extend this characterization of solutions and solve them analytically by finding all the roots of a system of polynomial equations. Qu and Tkachenko (2012) also attempt to evaluate the "non-identification curve", but their method is computationally demanding and can trace only local identification failures. All these studies focus on checking identification at a given (estimated) parameter value. In this paper, I propose an easy-to-implement, robust Bayesian algorithm for finding the complete identified set of parameters consistent with the data. Although there is always a choice to modify the model (e.g., add more shocks, or fix some parameter values) whenever identification fails, the algorithm proposed in this paper allows researchers to understand better the identification power and informativeness of model assumptions and data. More importantly, when the model is point-identified, the estimation results will be the same as those of standard Bayesian methods. That is, the algorithm proposed in this paper does not have any cost beyond the computation burden. It is also a valuable tool for separating information in the data from any prior input that is not revised by the data.

Additionally, this paper also speaks to both the partial identification and the robust Bayesian literature. The literature on frequentist analysis of set-identified models is large. It dates back to Hurwicz (1950), followed by the seminal work of Manski (1995), and more recent papers, such as Horowitz and Manski (2000), Manski (2003), Imbens and Manski (2004), Chernozhukov et al. (2007), Beresteanu and Molinari (2008), Stoye (2009), Andrews and Soares (2010), Romano and Shaikh (2010), Beresteanu et al. (2011) and Kaido et al. (2019). See Molinari (2020) for a review. The robust Bayesian analysis framework, initially explored in statistics (e.g., DeRoberts and Hartigan (1981), Wasserman (1989), and Berger (1990)), has recently been applied in economics for inference in set-identified models. In the robust Bayesian analysis literature, robustness is often discussed in terms of model or loss function uncertainty (Berger, 1990), but this paper focuses specifically on the sensitivity of Bayesian answers to variations in priors. There is a growing body of literature on Bayesian inference for partially identified models. This literature includes Moon and Schorfheide (2012), Gustafson (2015), Kline and Tamer (2016), Chen et al. (2018), Liao and Simoni (2019), Florens and Simoni (2021), Giacomini and Kitagawa (2021), Ke et al. (2022), Giacomini et al. (2022), and Bacchiocchi and Kitagawa (2022).

Some of the technical details of this paper are related to the engineering and math literature. The characterization of observationally equivalent state-space models is based on theories of linear systems (Glover, 1973; Antsaklis and Michel, 1997). The solution of polynomial systems and Gröbner basis are from algebraic geometry (Cox et al., 2013).

A key contribution of this paper is that it allows estimating the true identified set in DSGE models. In the paper, I use the algorithm from Kocięcki and Kolasa (2023), which allows finding observationally equivalent semi-structural parameters, and combine it with standard Bayesian methods to perform robust Bayesian estimation under the Giacomini and Kitagawa (2021) multiple-prior structure. However, unlike in Giacomini and Kitagawa (2021), it is challenging to form an explicit prior on the point-identified parameters[2] (which are usually called reduced-form parameters), update them with data, and map them back to find structural parameters. The reason is the complexity of DSGE models makes it generally impossible to find a closed-form representation of the likelihood function in terms of structural parameters. Instead, I can draw from posteriors of structural parameters and directly obtain observationally equivalent sets from those draws. After averaging the sets, I then show that this set achieved from one given prior is the range of all posterior means for a prior class, and the estimated set will converge to the true identified set consistent with the data.

The rest of this paper is organized as follows: Section 2 presents the motivations for this study, showing that the posterior parameter estimates, impulse response functions, and optimal policies can all be sensitive to the choice of priors regardless of sample size and number of posterior draws in set-identified DSGE models. In section 3, I first illustrate the structure of a typical DSGE model and a crucial identification condition to the proposed algorithm. Then I set up the robust Bayesian framework, propose an algorithm to conduct robust Bayesian inference for DSGE models, and show key theoretical results supporting this algorithm. Section 4 discusses examples from section 2 under the robust Bayesian setting. Section 5 concludes the paper.

[2] One can consider the spectral densities of observed variables or the set of all observationally equivalent structural parameters as the point-identified parameters in DSGE models, which can be high-dimensional or even infinite-dimensional when there does not exist a clear parametrization by the parameters of a DSGE model.

2 Motivation

Estimation in DSGE models can be difficult, because they are rich in parameters and often have a complicated model structure. Standard Bayesian methods are commonly employed in estimating this class of models. However, little is known about the robustness of the estimation results when the model is not identified. Moreover, it is unknown how inference based on the estimation results, and the policy analysis, could be related to this identification problem.

The following example illustrates the identification failure, and the formal definition is given in section 3.

Example 1 (A White-noise Process) *Consider the following stochastic process,*

$$Y_t = D\varepsilon_t, \quad \varepsilon_t \sim N(0, I_{n_\varepsilon}),$$

where Y_t is a vector of observed variables at time t, and D is the coefficient matrix. What can be recovered from the data (the reduced-form parameter in Giacomini and Kitagawa (2021)) in this model is $\mathbb{E}\left[Y_t Y_t'\right] = D_0 D_0'$, where D_0 is the true coefficient. Without further assumptions, one can have any $\bar{D} = D_0 Q$, with Q an orthonormal matrix, and still have $\bar{D}\bar{D}' = D_0 D_0'$. Therefore, D is not identified.

It will become clearer later that with an arbitrary informative prior distribution on D, the posterior mode of D, just because it has a higher prior weight but has the same likelihood as D_0, can be far away from the true D_0.

In this section, I examine a few examples to show that in set-identified DSGE models, estimation, inference, and policies made based on standard Bayesian results can be sensitive to the choice of priors, regardless of sample size. To resolve this issue, I propose an algorithm to perform robust Bayesian estimation and inference.

Although the work of Kocięcki and Kolasa (2023) allows, for a given parameter value, computation of the collection of all parameters that induce the same distribution of the data, it is not clear how that method could help researchers find the set of parameters supported by both the model and the data. One possibility is to apply their procedure with the maximum likelihood estimator. I suggest an alternative algorithm that has a preferable finite-sample interpretation (see Theorem 3).

2.1 Parameter Estimation

One direct interpretation of parameter estimates is that they indicate what values with underlying economic meanings are supported in historical data. Consider the simple monetary policy model introduced in King (2000), and thoroughly discussed in Cochrane (2011); solving the model results in the autoregressive equation of order 1 (AR(1))

$$\pi_t = \rho\pi_{t-1} - \frac{1}{\phi_\pi - \rho}\varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_e^2),$$

where π_t is the inflation rate, ε_t is the shock term of an AR(1) monetary policy disturbance, ρ is the correlation coefficient, ϕ_π is the Taylor rule parameter, and σ_e is the standard error of the monetary policy shock. Only π_t is assumed to be observed. Here the structural parameters to be estimated are stacked in a vector $\theta = (\sigma_e, \phi_\pi, \rho)$; the autocorrelation function identifies $(\rho, \frac{\sigma_e}{\phi_\pi - \rho})$.

In the rest of this section and in section 6, the exercises of application to different models are done in the following way. First, start with a set of "true" values and model specifications, simulate a sample of 200 periods,[3] and save the generated observed variables. Then, from the artificial data, estimate the parameters, perform standard Bayesian analysis, and compare the results with the true values.

I run a standard Bayesian estimation of the parameters σ_e, ϕ_π, and ρ in *Dynare*, a software platform that has been used by macroeconomists for handling a wide class of economic models, including DSGE and overlapping generations (OLG) models (Adjemian et al., 2011). The reported local identification results from embedded methods based on Iskrev (2010), Komunjer and Ng (2011), and Qu and Tkachenko (2012) confirm that local identification fails because σ_e and ϕ_π are pairwise collinear. I use the "uninformative prior" such that the priors of the parameters are uniform and independent. I use 500,000 posterior draws and a 400,000 burn-in period. Ideally, the reported statistics of the posterior distribution should reflect this non-identification issue of σ_e and ϕ_π.

However, in the estimation results, the posterior mean of (σ_e, ϕ_π) does not converge to the true parameter values (see Table 1). The posterior modes (maximum a posteriori estimator, or MAP) also fail to be close to the true values. In addition, the Bayesian credible regions of (σ_e, ϕ_π) do not cover the true values.

[3] This is approximately 16 years of observations if one has monthly data, which is a reasonable amount in DSGE estimations. I also perform the same exercise with 500 periods and 1000 periods and the results and conclusions are similar to those obtained using 200 periods.

Figure 1 shows that the marginal posterior distributions for ϕ_π and σ_e are different from the prior even when neither is identified. A careful inspection of the likelihood function in terms of σ_e and ϕ_π (fixing $\rho = 0.8$) shows that the likelihood achieves its maximum on the line $\sigma_e = \phi_\pi - 0.8$. It is flatter near the maximum region when both values are high (see Figure 2). However, when the estimated parameters become high-dimensional, it is impossible to check the overall model identification by inspecting the likelihood function. That invalidates the "eyeballing" method to find the identified set of structural parameters.

The informative posteriors of the unidentified parameters also confirm that data-based learning about the identified parameters can "spill over" onto the unidentified parameters (see Koop and Poirier 1997; Koop et al. 2013). All the results are robust to the number of posterior draws and replications. I also perform an exercise to explore the behavior of the posterior mode, minimizing the negative log likelihood, penalized by the prior, with 1000 replications and a sample size ranging from 100 to 1000 observations.[4] As is shown in Figure 6 in the appendix, the pattern illustrated using MCMC from the sample above is not uncommon.

Table 1: Prior and posterior distribution of structural parameters, from a single run of the MCMC procedure on one sample

	True value	Prior distribution			Posterior distribution			
		Distr.	Mean	St. Dev.	Mode	Mean	5 percent	95 percent
σ_e	1	Uniform	4	2.02	1.85	4.43	1.95	6.77
ϕ_π	1.8	Uniform	4	1.73	2.51	4.91	2.78	7.00
ρ	0.8	Uniform	0.75	0.09	0.81	0.81	0.74	0.87

2.2 The Impulse Responses

Impulse response analysis is one of the most used tools in macroeconomics. The impulse response functions (IR, or IRF) can be used to predict the implications of an unexpected shock or a policy

[4]To do the exercise, I use the particle swarm optimizer embedded in MATLAB to find the global minimum. The initial value of the parameter vector is set at the prior mean of the uniform prior. Because the objective function always has infinite minimizers, with a uniform randomization of the starting points of particles around the prior mean, the minimizer reported is also random. Moreover, the reported numerical minimizer "picked" by the algorithm should reflect the shape of the objective function. The way the particle swarm algorithm works is similar to how MCMC with adaptive variance works. The non-identifiability of ϕ_π and σ_e can also be reflected in the trace of the MCMC draws in Figure 5, because the range of draws is wide for these two parameters.

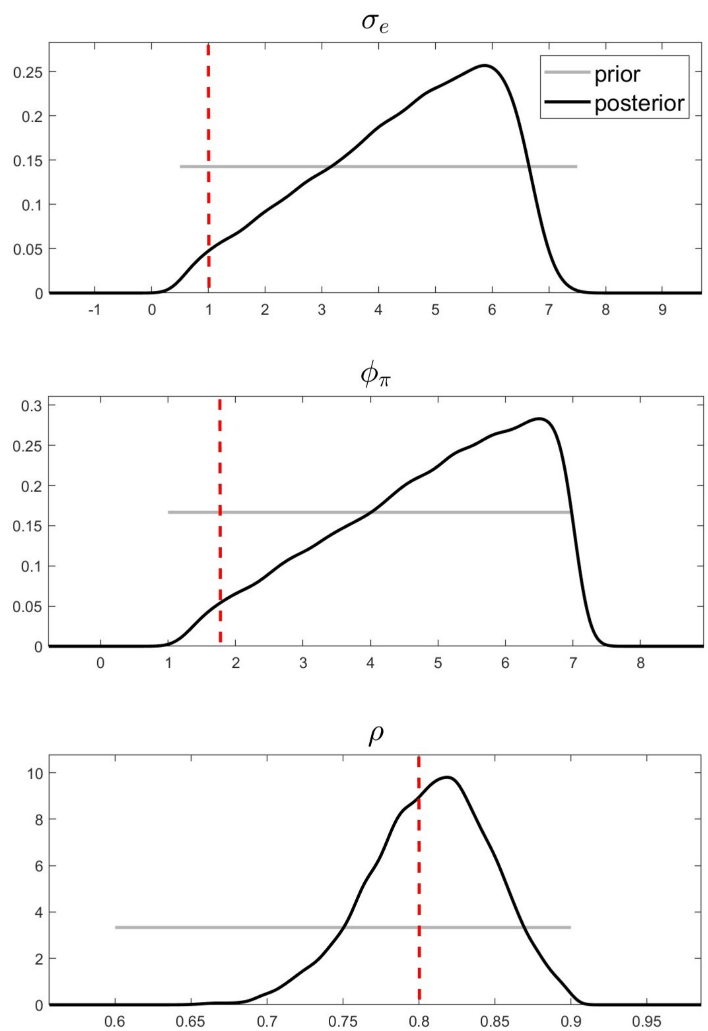


Figure 1: Prior and posterior for the Cochrane model. The red dashed line represents the true values; ρ is the correlation coefficient, ϕ_π is the Taylor rule parameter, and σ_e is the standard error of the monetary policy shock.

change in a macroeconomic framework. Impulse responses can be generally defined without reference to the data-generating process as the following function (Koop et al., 1996; Potter, 2000; Jordà, 2005):

$$IR(t,s,\delta) = \mathbb{E}(y_{t+s} \mid \varepsilon_t = \delta; \omega_{t-1}) - \mathbb{E}(y_{t+s} \mid \varepsilon_t = 0; \omega_{t-1}), \quad s = 0,1,2,\ldots, \tag{1}$$

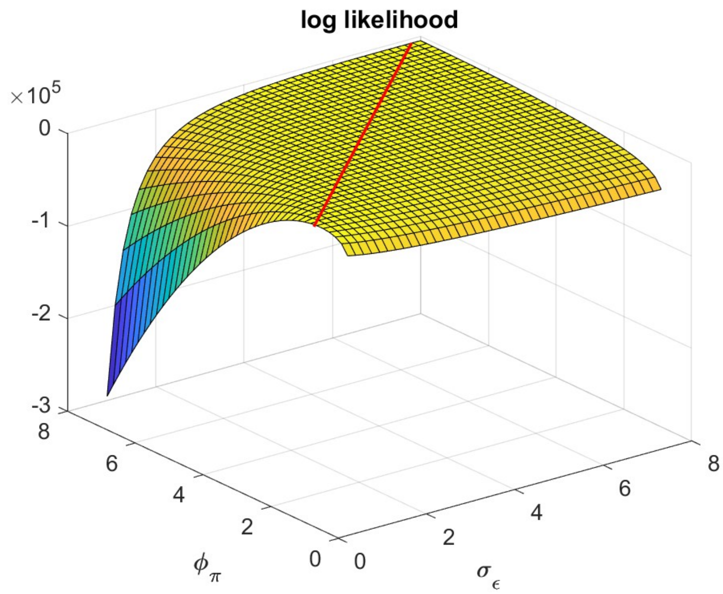


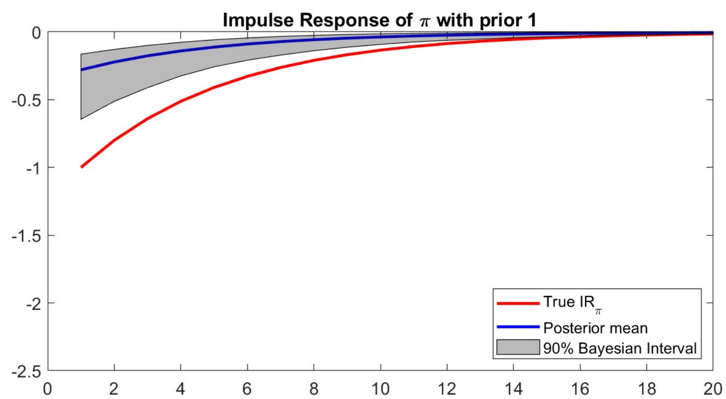
Figure 2: Log-likelihood of the Cochrane model when the correlation coefficient ρ is equal to 0.8. The red line is the maxima ridge of the likelihood function.

where y_{t+s} is the variable of interest at time $t+s$, ε_t is the exogenous shock, δ is the size of the shock with the same dimension as ε_t, and ω_{t-1} is a particular realization of the information set available up to time $t-1$, Ω_{t-1}.

In this exercise, I compute the impulse responses of inflation π_t to a 1-unit change in monetary policy shock. To do that, I first compute the posterior mean and standard Bayesian confidence interval of impulse responses using two priors. One is the uniform prior used above. The other is constructed in a way such that it always has the same prior and posterior predictive distribution over $\{\pi_t\}$ regardless of the data realization.[5] In Figure 3, I draw the true impulse response function of 20 periods with a 1-unit shock of ε at time 1. The Bayesian credible region, or the 90% highest-posterior-density interval, does not always cover the true impulse response function, denoted as $IR_\pi |_{\theta_0}$, when the parameters are not identified.

This is also an example to show that, even from a probabilistic point of view, inference based on estimation results can be misleading. That is, even when researchers explore the entirety of the posterior distribution on the impulse response function, they will find it unlikely the true impulse response function is actually true. Again, this result persists across 1000 replications of simulated samples; therefore, it cannot be explained as sampling errors.

[5]The prior predictive distribution is the distribution of a data point marginalized over its prior distribution. In this case, it is $p(\boldsymbol{\pi}) = \int p(\boldsymbol{\pi} \mid \theta) d\pi_\theta$, where π_θ is the prior distribution, and $p(\boldsymbol{\pi} \mid \theta)$ is the likelihood of $\{\pi_t\}$.



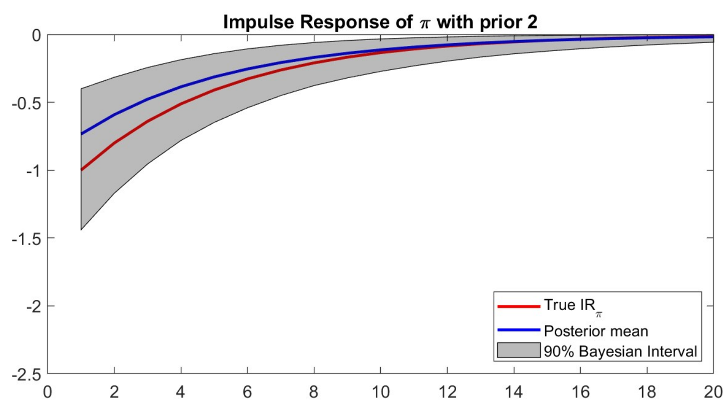


Figure 3: The impulse response functions (IR) of inflation π to a one-unit shock in the Cochrane model; 20 periods of impulse responses are plotted.

2.3 Policy Analysis

Considering the sensitivity results presented in the previous section, a natural question is, Does this sensitivity issue affect policy analysis? The short answer is yes, sometimes. A more detailed discussion is in section 5. The following example comes from An and Schorfheide (2007). Relative to their model, I change only the total factor productivity shock to a cost-push shock, similar to the setting in (Galí, 2015), Chap. 5. As will be discussed in section 4, the existence of a cost-push shock prevents the divine coincidence.

$$
\begin{aligned}
y_t =& \mathbb{E}_t\left[y_{t+1}\right] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t\left[\pi_{t+1}\right]\right) + g_t - \mathbb{E}_t\left[g_{t+1}\right] \\
\pi_t =& \beta \mathbb{E}_t\left[\pi_{t+1}\right] + \kappa\left(y_t - g_t\right) + u_t \\
i_t =& \rho_R i_{t-1} + \left(1-\rho_R\right)\psi_\pi \pi_t + \left(1-\rho_R\right)\psi_y\left(y_t - g_t\right) + \varepsilon_{R,t} \\
u_t =& \rho_u u_{t-1} + \varepsilon_{u,t} \\
g_t =& \rho_g g_{t-1} + \varepsilon_{g,t}.
\end{aligned}
\tag{2}
$$

Here π_t is the inflation rate, i_t is the nominal interest rate, y_t is the output gap, g_t is the government spending shock, $\varepsilon_{R,t}$ is the monetary policy shock, and u_t is the cost-push shock. The structural parameters to be estimated are the inverse intertemporal elasticity of substitution σ; composite parameter κ, which denotes the slope of the new-Keynesian Phillips curve; ψ_π and ψ_y, which are the strength of the interest rate response to deviations of inflation or the output gap from their target levels, respectively; autoregressive coefficients ρ_R, ρ_g, ρ_u; and standard deviations $\sigma_R, \sigma_g, \sigma_u$, respectively. They are all stacked into a vector $\theta = (\sigma, \kappa, \psi_\pi, \psi_y, \rho_R, \rho_g, \rho_u, \sigma_R, \sigma_g, \sigma_u)$, and discount factor β is calibrated at its true value. I use two different priors in this example. The first prior is the same as that of An and Schorfheide (2007), except the total-factor productivity shock's parameters are replaced by the cost-push shock. I obtain 50,000 posterior draws and drop 40,000 for the burn-in. As in the above example, I generate another posterior distribution with the same posterior predictive distribution as that of the first prior.

Imagine that the policy maker is trying to choose between two policy parameter combinations, $(\psi_\pi, \psi_y) = (1.5, 0)$ and $(\psi_\pi, \psi_y) = (1.5, 0.125)$, to determine whether the monetary policy should respond to the output gap to minimize the welfare loss of the form $\lambda \pi_t^2 + y_t^2$, where λ is the relative weight on the inflation that cannot be recovered from estimation. Table 2 shows the optimal policy choices under different priors across different weights.

Table 2: Policy comparison under different distributions and weights. The check mark denotes the policy with lower welfare loss.

	$\lambda = \frac{1}{3}$		$\lambda = 1$		$\lambda = 3$		$\lambda = 10$		$\lambda = 30$	
(ψ_π, ψ_y)	post 1	post 2	post 1	post 2	post 1	post 2	post 1	post 2	post 1	post 2
(1.5, 0)					✓		✓	✓	✓	✓
(1.5, 0.125)	✓	✓	✓	✓		✓				

Although the optimal parameters agree under both priors for most weights, they are different when the weight on inflation in the objective function is three times the output gap.[6] That is, a researcher with prior 1 will disagree with a researcher who uses prior 2 in optimal policy choices, even though their models are the same and their priors induce the same predictive distribution (hence, the same marginal data density).

3 DSGE and Robust Bayes

In this section, I first set up the problem under a DSGE model framework, characterize it in a robust Bayes setting, and propose an algorithm for robust estimation and inference. Then, I explain the structure of DSGE models and the robust Bayes framework. For the DSGE part, I start by demonstrating a standard way to estimate a DSGE model, which involves solving a linear rational expectation model and finding the likelihood. Each step could lead to a failure of identification. Finally, I show the theorems from previous literature to characterize observational equivalence. For the robust Bayes part, I define a prior class that I work with and show the proposed algorithm's finite sample and asymptotic properties.

3.1 Model Specification

A DSGE model with structural parameter vector $\theta \in \Theta$ is typically characterized by several Euler equations and market clearing conditions in equilibrium. After linearizing the equilibrium conditions around the steady states, I have a linearized rational expectation (LRE) model of the form (Kociecki and Kolasa, 2018; Kocięcki and Kolasa, 2023)[7]

$$\Gamma_0(\theta)\begin{bmatrix} S_t \\ P_t \end{bmatrix} = \Gamma_1(\theta)\mathbb{E}_t\begin{bmatrix} S_{t+1} \\ P_{t+1} \end{bmatrix} + \Gamma_2(\theta)S_{t-1} + \Gamma_3(\theta)\varepsilon_t, \tag{3}$$

[6]The feature that the optimal policy is sensitive to the choice of priors does not appear only at $\lambda = 3$; it holds for λ values between 2 and 6.

[7]Alternative representations are used in the literature. For example, Blanchard and Kahn (1980) use $\Gamma_0\mathbb{E}_t(S_{t+1}) = \Gamma_1 S_t + c + \Psi\varepsilon_t$. Sims (2002) introduces endogenous forecast error $\eta_t^y \equiv y_t - \mathbb{E}_{t-1}y_t$ so that $\Gamma_0(\theta)S_t = \Gamma_1(\theta)S_{t-1} + \Psi(\theta)\varepsilon_t + \Pi(\theta)\eta_t$, which is called the canonical form. Al-Sadoon and Zwiernik (2019) use $\sum_{i=-q}^{p} B_i\mathbb{E}_t(S_{t-i}) = \sum_{i=0}^{k} A_i\varepsilon_{t-i}$. These forms can be transformed easily from one to another. Although some may be the subclass of the other more general forms, they are general enough to represent almost all linear DSGE models. I work on the form of Equation (3) because it allows researchers to operate under minimal state representation.

where $\Gamma_i(\theta), i = 0\dots 3$ are matrices of coefficients that are also functions of structural parameters θ; S_t, P_t, ε_t and η_t contain the state variables, policy variables (non-state endogenous variables), structural shocks, and expectation errors respectively; ε_t can in general include sunspot shocks in the case of infinite stable solutions to LRE, which is called indeterminacy,[8] and measurement errors when they exist. The LRE models can be solved numerically to yield a linear state-space representation; the solution combined with measurement (without a constant term) is also known as the ABC(D) representation (Fernández-Villaverde and Rubio-Ramírez, 2007),

$$S_t = A(\theta)S_{t-1} + B(\theta)\varepsilon_t \tag{4}$$

$$Y_t = C(\theta)S_{t-1} + D(\theta)\varepsilon_t, \tag{5}$$

where $S_t \in \mathbb{R}^{n_S}$ is the state vector, $Y_t \in \mathbb{R}^{n_Y}$ is the observable vector, $\varepsilon_t \sim i.i.d.N(0, \Sigma(\theta))$ has dimension n_ε, $A(\theta)$ is an $n_S \times n_S$ matrix, $B(\theta)$ is an $n_S \times n_\varepsilon$ matrix, $C(\theta)$ is $n_Y \times n_S$, $D(\theta)$ is $n_Y \times n_\varepsilon$, and $\Sigma(\theta)$ is a positive definite, $n_\varepsilon \times n_\varepsilon$ matrix. Here A, B, C, and D are the coefficients of the solution. Representation (5) can be derived from an additional solution equation $P_t = F(\theta)S_{t-1} + G(\theta)\varepsilon_t$[9] and a measurement equation with policy variables

$$Y_t = L(\theta)\begin{bmatrix} S_t \\ P_t \end{bmatrix} + J(\theta)\varepsilon_t.$$

Different from structural vector autoregressive (SVAR) models, the ABCD representation that researchers work directly on is not identified in general (Komunjer and Ng, 2011) in the sense that different combinations of (A, B, C, D, Σ) could potentially have the same likelihood. Therefore, the failure of identification may come from the mapping from structural parameters to the state-space coefficients, or the mapping from the state-space parameters to the likelihood. However, in practice, researchers are not concerned about the identification issue when estimating a DSGE model because the posterior distributions of structural parameters can always be attained regardless of identification. As will become clear later, in set-identified models, standard Bayesian estimation results are sensitive to the choice of priors, no matter how big the sample size is.

[8] Lubik and Schorfheide (2003) have shown the stable solutions can be represented as $S_t = \Theta S_{t-1} + \Theta_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t$, where ε stands for structural shocks, and ϵ is the sunspot shocks. To keep things simple, I do not consider the possibility of indeterminacy for the main part of this paper, and put this discussion to section 6.

[9] The coefficients F and G enter the likelihood of Y only through $C(\theta)$ and $D(\theta)$ and are thus irrelevant. Another reason it is not called an ABCDFG representation is that in other papers the authors do not separate the state variables from the policy variables. Therefore, their matrix A contains both values in A and F here.

Next, we outline the standard process for conducting Bayesian estimation in DSGE models. This is presented in the form of a pseudo-algorithm.[10]

Algorithm 1 (Bayesian Estimation of Linearized DSGE Models)

(1) Write down DSGE as a constrained optimization problem. Find the optimal first-order conditions and steady states. Then linearize the model around its steady states to obtain its LRE form expressed by Equation (3).

(2) Solve the LRE model and obtain a state-space form Equations (4) and (5).

(3) Set a prior distribution π_θ for structural parameters θ.

(4) Use MCMC methods, such as random walk Metropolis–Hastings (Robert et al., 1999) or sequential Monte Carlo methods (Herbst and Schorfheide, 2014), to obtain draws from posterior $\pi_{\theta|Y}$, and call it θ_{j+1}.

- *Likelihood can be evaluated from the state-space form using the Kalman filter.*

(5) Repeat steps 2 and 3 M times. Obtain the posterior distribution of parameters of interest $\{\eta(\theta_j)\}_{j=1}^M$ by transforming the posterior draws of θ.

This type of algorithm, although widely used, would reveal only the parameter values supported by the model and the data when the parameters are identified. It is not clear from this procedure when there is lack of identification[11] and how the results are sensitive to the choice of priors. As shown in section 1, starting with a more or less ad hoc prior obscures the underlying identification problems and can result in misleading posteriors (see also Poirier 1998; Gustafson 2009; Moon and Schorfheide 2012; Morris 2014). Researchers might be interested in all the inferential conclusions that are supported by the model and the data. An estimation and inference method that is robust to the choice of priors would be particularly valuable in this case.

Before I introduce the proposed algorithm and jump to applications of robust Bayesian tools in DSGE models, I summarize the framework. I set up the structure by defining the key concepts used in this paper. Then, I present the algorithm and theorems needed for it to work.

[10]See Herbst and Schorfheide (2015) for an exhaustive description of each step.

[11]One might argue identification failure can be found from the MCMC trace plot using a uniform prior, as in Figure 5. However, when the range of the identified set is small, or when the model is locally but not globally identified, this feature becomes less obvious. Effective methods to detect global identification failure are covered in Qu and Tkachenko (2017) and Kociecki and Kolasa (2018).

Let $(\mathbf{Y}, \mathcal{Y})$ be the measurable space of a sample of observables $Y^T \equiv \{Y_t\}_{t=0}^T \subset \mathbf{Y}$, and let $(\Theta, \mathcal{A})$ be the measurable space of a parameter vector $\theta \in \Theta \subset \mathbb{R}^d$. I assume the existence of a regular conditional distribution $F(y \mid \theta)$, and density $p(y \mid \theta)$ of $Y_{1:T}$ given θ, which represents the likelihood function. The concepts of observational equivalence and identification are defined as follows (Rothenberg, 1971):

Definition 1 (Observational Equivalence) *Given a model with likelihood density $p(y \mid \theta)$, θ and $\bar{\theta}$ are observationally equivalent if $p(y \mid \theta) = p(y \mid \bar{\theta})$ for all observed data $y \in \mathbf{Y}$. It can also be written as $\theta \sim \bar{\theta}$.*

By definition, this observational equivalence is an equivalence relation in mathematics and therefore possesses reflexivity, symmetry, and transitivity. It partitions the space Θ into equivalent classes. For any data realization, each parameter vector has the same likelihood within the class.

Definition 2 (Identification) *Given a model, the parameters θ of the model are identified if there exists no other $\bar{\theta} \in \Theta$ observationally equivalent to θ.*

Definition 2 is sometimes referred to as global identification.[12]

For any given θ, I can define the observationally equivalent set by a mapping $K : \Theta \to \mathcal{F} \subset 2^{\Theta}$ such that $K(\theta) \equiv \{\theta' : \theta' \sim \theta\}$, and $\mathcal{F}$ is the family of closed subsets of Θ. Without further assumptions, K characterizes the indices that dictate the likelihood, i.e., $p(y \mid \theta) = p(y \mid \bar{\theta})$ for all $y \in \mathbf{Y}$ if and only if $K(\theta) = K(\bar{\theta})$ (see, e.g., Barankin et al. (1960)). It immediately follows that $\theta \sim \bar{\theta}$ if and only if $K(\theta) = K(\bar{\theta})$.

In contrast with Giacomini and Kitagawa (2021), who work with the reduced-form parameters or minimal-sufficient parameters directly,[13] there is no consensus regarding the definition of reduced-form parameters in DSGE models. The agnostic nature of the mapping K caused by model complexity makes it hard to apply the methodology of Giacomini and Kitagawa (2021) here. In an abuse of notation, I will use K to denote both the mapping and a generic mapped element in $\mathcal{F}$, which is a subset of Θ.

[12] If I replace "$\bar{\theta} \in \Theta$" with "$\bar{\theta} \in \mathcal{N}_r(\theta)$ for some neighborhood of θ", I have the notion of local identification.

[13] Let $\theta \in \Theta$ be the structural parameters. If there exists a continuously differentiable function $\phi = g(\theta)$ that maps a neighborhood of θ to a subset of $\mathbb{R}^r$ such that $\tilde{p}(y \mid \phi) = p(y \mid \theta)$ for all $y \in \mathbf{Y}$ and $\theta \in \Theta$ for some function $\tilde{p}(y \mid \phi)$ and if, in addition, ϕ is globally identified, then ϕ is called a reduced-form parameter. Identification analysis in econometrics normally proceeds as follows: first, find the reduced-form model representation where the parameters are always identified, and then disentangle the link between structural parameters and reduced-form parameters. See, for example, Koopmans (1949); Koopmans and Reiersol (1950); Barankin et al. (1960); Picci (1977); Dawid (1979); Florens and Simoni (2021); Giacomini and Kitagawa (2021) for more discussion.

Researchers are primarily interested in estimating the structural parameter vector θ. However, sometimes transformations of θ, namely $\eta(\theta)$, with a measurable function $\eta : (\Theta, \mathcal{A}) \to (\mathcal{H}, \mathcal{D})$, $\mathcal{H} \subset \mathbb{R}^q$ for some $q < \infty$ are more of interest. Examples include a particular policy parameter or a finite-period impulse response. More generally, η can be the optimal-choice parameters of a policy rule that minimizes some welfare loss. For example, in a basic New Keynesian model (Galí (2015)), an interest rate rule $i_t = r_t + \phi_\pi \pi_t + \phi_y y_t$ with natural rate r_t can be used to minimize welfare loss in terms of output gap and inflation of the form

$$L_W(\theta, y, \pi) = \frac{1}{2}\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \left[a y_{t+s}^2 + b \pi_{t+s}^2 \right],$$

and $\eta = \arg\min_{\phi_\pi, \phi_y} L_W$; η may also depend on the initial condition if the welfare loss is conditional. Detailed applications are shown in section 4. In that case, η is also a function of state variables. Figure 4 illustrates the graphical connection between these parameters.

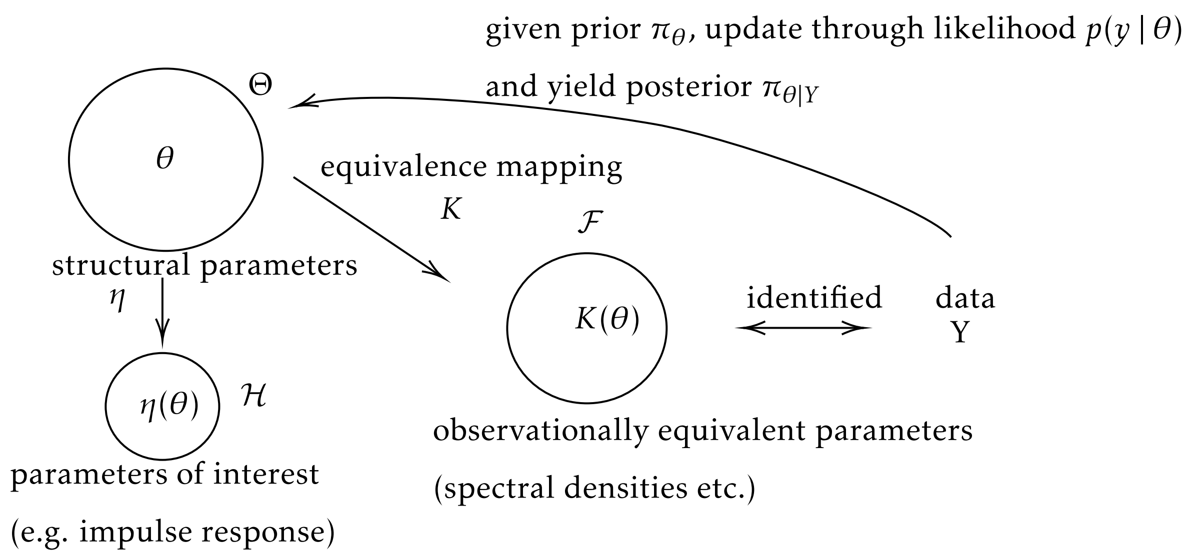


Figure 4: Connections between parameters

The specification above is model-free so that the researchers can fit Bayesian models such as structural vector autoregressive models (Giacomini and Kitagawa, 2021), latent Dirichlet allocation (Ke et al., 2022), and DSGE in this framework. However, because of the differences in model complexity and sources of non-identification, different approaches should be taken to perform robust Bayesian analysis. A comparison of the algorithms used in these models is in the appendix. A few things significantly complicate the problem under the DSGE settings. First, in practice, researchers do not start by estimating reduced-form parameters in DSGE models. Instead, researchers put a prior on structural parameters and generate the likelihoods based on a linear state-space representation. As shown below, even the parameters of state-space representation (or ABCD representation) in general are not identified. So, even if the mapping from structural parameter vector θ to the state-space coefficients is injective, the Bayesian estimation results of θ can still be sensitive to priors. That disqualifies the state-space form coefficients from being the so-called reduced-form parameters. Second, even if researchers have well-defined reduced-form parameters (e.g., spectral densities), it is still challenging to back out all the θs that map into the same spectral density, because the mapping from structural parameters to spectral densities is model-specific and often numerical. Luckily, because of the pioneering work done by Komunjer and Ng (2011), Kocięcki and Kolasa (2023), there exist handy tools to characterize observationally equivalent θ directly without computing the likelihood. For now, let us take that tool as given. An overview of the algorithm is presented here, with a more detailed version in section 5.

Algorithm 2 (Robust Posterior Mean in DSGE)

(1) Specify a prior π_θ. Use standard Bayesian methods to obtain $\pi_{\theta|Y}$.

(2) Draw from $\pi_{\theta|Y}$ M times. For each draw θ^j, compute its observationally equivalent set $K(\theta^j)$.

(3) Optimize over $K(\theta^j)$ to find the minimum or maximum value of $\eta(\theta)$, η^j_{min}, and η^j_{max}.

(4) Take the average of η^j_{min} and η^j_{max} over draws. Report $\left[\frac{1}{M}\sum_{j=1}^{M}\eta^j_{min}, \frac{1}{M}\sum_{j=1}^{M}\eta^j_{max}\right]$.

I add two steps to standard Bayesian estimation procedures, where step 2 is based on Kocięcki and Kolasa (2023). In the case of point-identification, each set $K(\theta^j)$ should be a singleton, and the reported range will be equal to the standard Bayesian posterior mean. Therefore, there is no loss in sharpness. Moreover, I will show in the next subsections that this reported set equals the range of posterior means.

3.2 Identification Conditions

A key step in Algorithm 2 is to find the observationally equivalent set of a given parameter vector. Theoretical results on conditions to characterize parameter identification in DSGE models have been widely studied (Iskrev, 2010; Komunjer and Ng, 2011; Qu and Tkachenko, 2012, 2017; Kociecki and Kolasa, 2018). While the existing research on checking DSGE identification is performed at a given parameter vector, sufficient and necessary conditions can be insightful to characterize the observationally equivalent set. Moreover, these results are used in this paper to provide tools for finding the true identified set. From this point on, I suppress the dependence on argument θ for parameters A, B, C, D, Σ, and other parameters to come, for brevity. Therefore, $\bar{A}$ denotes $A(\bar{\theta})$, and the same applies for the notations of $\bar{B}$, $\bar{C}$, $\bar{D}$, $\bar{F}$, $\bar{G}$ and $\bar{\Sigma}$.

Assumption 1 (Stability) *For every $\theta \in \Theta$ and for any $z \in \mathbb{C}$, $\det(zI_{n_S} - A) = 0$ implies $|z| < 1$.*

Assumption 1 restricts $\{S_t\}$ and $\{Y_t\}$ in Equations (4) and (5) to weakly stationary time series, where the eigenvalues of A remain inside the unit circle. Under Assumption 1, Wold decomposition applies; therefore, I can rewrite Y_t in Equations (4) and (5) in the form of a $VMA(\infty)$ process

$$Y_{t+1} = \left[C(I_{n_S} - AL)^{-1}BL + D\right]\varepsilon_{t+1}, \quad t = \cdots -1, 0, 1\ldots, \tag{6}$$

where L is the lag operator. The implied impulse response will be

$$IR_y(t, s, \delta) = \begin{cases} D\delta, & s = 0 \\ CA^{s-1}B\delta, & s = 1, 2, \ldots \end{cases}. \tag{7}$$

Define $P \equiv E(S_t S_t')$, which is also the unique (under Assumption 1) solution to the Lyapunov equation $P = APA' + B\Sigma B'$. The autocovariances of $\{Y_t\}$, $\Gamma_j^y = E(Y_t Y_{t-j}')$ can be expressed as $\Gamma_0^y = CPC' + D\Sigma D'$ and $\Gamma_j^y \equiv CA^{j-1}N$ for $j > 0$, where $N = APC' + B\Sigma D'$.

I define the z-transform of $\{Y_t\}$ by

$$\Phi_Y(z;\theta) = \sum_{j=-\infty}^{+\infty} \Gamma_j^y z^{-j}. \tag{8}$$

The spectral density can be achieved by setting $z = e^{i\omega}$, i.e. $\Phi_Y(e^{i\omega};\theta) = \Sigma_{j=-\infty}^{\infty}\Gamma_j^y e^{-ij\omega}$. It can also be written in terms of the ABCD representation parameters $\Phi_Y(z) = H(z)\Sigma H'(z^{-1})$, where $H(z) = C(zI_{n_S} - A)^{-1}B + D$ is the transfer function. Define $\mathcal{O} \equiv (C' \quad A'C' \quad \cdots \quad A'^{n_S - 1}C')$, $\mathcal{C} \equiv (N \quad AN \quad \cdots \quad A^{n_S-1}N)$ for the assumption below.

Assumption 2 (Stochastic Minimality) *For every $\theta \in \Theta$, matrices $\mathcal{O}$ have full column rank and $\mathcal{C}$ have full row rank, i.e.,* $\text{rank}(\mathcal{O}) = \text{rank}(\mathcal{C}) = n_S$.

Assumption 2 is the same as stochastic minimality in Kocięcki and Kolasa (2023) and autocovariance minimality in Komunjer and Zhu (2020). It differs from the minimality definition in

Komunjer and Ng (2011) in that the controllability[14] (see for example, Lindquist and Picci (2015)) is on (A, N) instead of (A, B), and Assumption 2 does not require the econometrician to observe ϵ_t. Intuitively, the rank conditions for $\mathcal{C}$ stand for the controllability of the innovations representation of the state-space system. The full column rank of $\mathcal{O}$ guarantees the observability of the model. The main purpose of this assumption is to ensure there exists no other state-space representation that has a lower-dimensional state-space but the same spectral density. The practicality of this assumption is also discussed in Kocięcki and Kolasa (2023), who show that if Assumption 2 holds at one θ, it holds almost everywhere in Θ.

From definition 1, two structures are said to be observationally equivalent if they imply the same probability distribution for the observables. With Gaussianity, I can then characterize the observational equivalence of parameters by their equivalence in the spectral density function: In my state-space representation, I have only zero intercepts in the measurement equation; allowing for non-zero intercepts that depend on θ will give extra identification information in the first order. The following lemma gives an alternative characterization of observational equivalence that is easier to work with. It is also the definition of observational equivalence in Komunjer and Ng (2011) and Qu and Tkachenko (2012).

Lemma 1 (Observational Equivalence) *In a linearized DSGE model with Gaussian shocks, if Assumption 1 is satisfied, two structural parameter vectors θ_0 and θ_1 are observationally equivalent if and only if $\Phi_Y(z;\theta_0) = \Phi_Y(z;\theta_1)$ for all $z \in \mathbb{C}$.*

The proof of all lemmas and theorems is in the appendix. Lemma 1, when combined with Definition 2, states that the structural parameter θ can be identified if and only if there is no other structural parameter that generates the same spectral density. Intuitively speaking, this is because the spectral density contains the same information as serial dependence of Y_t.

By definition, when researchers observe $\{Y_t\}$, its spectral density $\Phi_Y(e^{i\omega};\theta)$ is identified. However, as the following theorem will clarify, the mapping from (A, B, C, D, Σ) to Φ_Y is not injective.

The simplest version of the theorem that characterizes observational equivalence across discrete-time linear state-space systems, namely the equivalence described by coordinate trans-

[14]The matrix pair (A, B) is said to be controllable if the matrix $(B \quad AB \quad \cdots \quad A^{n_S} B)$ has rank n_S. The pair (A, C) is called observable if (A', C') is controllable. In a deterministic system, controllability means that for any initial state, it is always possible to achieve any final state from any initial state by admissible shocks (inputs). Observability means that it is always possible to reconstruct the initial state by observing the output trajectory, given the evolution of the shocks. In my notation, the outputs are $\{Y_t\}$, the inputs are $\{\varepsilon_t\}$, and the states are $\{S_t\}$.

formation $(A,B,C,D) \to (TAT^{-1}, TB, CT^{-1}, D)$, is well-documented in linear system literature. Komunjer and Ng (2011) extend it to accommodate the case $\Sigma \neq I_{n_\varepsilon}$, and discuss the singular and non-singular cases separately because of the need to formulate a minimal system. Alternatively, I use the following more general theorem (Theorem 1 in Kocięcki and Kolasa (2023)) that accommodates both singular and non-singular state-space systems.

Theorem 1 (Observational Equivalence) *Let Assumptions 1 and 2 hold. Then $\theta \sim \bar{\theta}$ if and only if (1) $\bar{A} = TAT^{-1}$, (2) $\bar{C} = CT^{-1}$, (3) $AQA' - Q = T^{-1}\bar{B}\bar{\Sigma}\bar{B}'T'^{-1} - B\Sigma B'$,(4) $CQC' = \bar{D}\bar{\Sigma}\bar{D}' - D\Sigma D'$, (5)$AQC' = T^{-1}\bar{B}\bar{\Sigma}\bar{D}' - B\Sigma D'$, for some nonsingular $n_\varepsilon \times n_\varepsilon$ matrix T and symmetric $n_\varepsilon \times n_\varepsilon$ matrix Q. In addition, if $\theta \sim \bar{\theta}$, then both T and Q are unique.*

Theorem 1 is an adapted version of Corollary 4.5 in Glover (1973), reformulated and proved by Kocięcki and Kolasa (2023) to fit the discrete case. It states that two state-space representations are observationally equivalent up to some similarity transformation. However, I still need to connect the structural parameters θ and $\bar{\theta}$ with their state-space parameters to form an equation system with unknown $\bar{\theta}$.

To do that, I first substitute model solution expressed by Equation (4) to Equation (3), and impose $\mathbb{E}_t\varepsilon_{t+1} = 0$. This leads to a equation system with variables S_{t-1} and ϵ_t. Then, I use the undetermined coefficient method, letting the coefficients of S_{t-1} and ϵ_t be zero, to obtain the first four equations of system (9). This method is based on the fact that the state-space system should always conform with the linear rational expectation model, regardless of the state realizations. Theorem 1 together with the undetermined coefficient method, allows me to characterize the identified set by a system of equations. The following equation system is obtained:

$$
\begin{aligned}
\bar{\Gamma}_0^s \bar{A} + \bar{\Gamma}_0^p \bar{F} - \bar{\Gamma}_1^s (\bar{A})^2 - \bar{\Gamma}_1^p \bar{F}\bar{A} &= \bar{\Gamma}_2 \\
\bar{\Gamma}_1^s \bar{A}\bar{B} + \bar{\Gamma}_1^p \bar{F}\bar{B} - \bar{\Gamma}_0^s \bar{B} + \bar{\Gamma}_3 &= \bar{\Gamma}_0^p \bar{G} \\
\bar{C} &= \bar{L}^s \bar{A} + \bar{L}^p \bar{F} \\
\bar{D} &= \bar{L}^s \bar{B} + \bar{L}^p \bar{G} + \bar{J} \\
\bar{A} &= TAT^{-1} \\
\bar{C} &= CT^{-1} \\
AQA' - Q &= -B\Sigma B' + T^{-1}\bar{B}\bar{\Sigma}\bar{B}'\left(T^{-1}\right)' \\
AQC' &= T^{-1}\bar{B}\bar{\Sigma}\bar{D}' - B\Sigma D' \\
CQC' &= \bar{D}\bar{\Sigma}\bar{D}' - D\Sigma D' \\
Q &= Q',
\end{aligned}
\tag{9}
$$

where $\Gamma_0 = \begin{bmatrix}\Gamma_0^s & \Gamma_0^p\end{bmatrix}$, $\Gamma_1 = \begin{bmatrix}\Gamma_1^s & \Gamma_1^p\end{bmatrix}$, and $L = [L^s \quad L^p]$; each superscript corresponds to either the state or policy component. The unknowns in this system are $\bar{\theta}$ (as which $\bar{\Gamma}_i$ are explicitly expressed), the elements of $\bar{B}$, $\bar{D}$, $\bar{F}$, $\bar{G}$, T, and Q. Others can be canceled via substitution or reparametrization. The parameters A, B, C, D, and Σ are known when researchers check identification at a fixed θ point.

To solve equation system (9), analytical methods developed in computational algebraic geometry can be of use if equation system (9) can be rewritten as a system of polynomials in unknowns.[15] That is, $\bar{\theta}$ appears in the form of polynomial variables in $\bar{\Gamma}_i$, $\bar{\Sigma}$, $\bar{L}$ and $\bar{J}$. This is not a restrictive assumption. For example, if a fraction of two parameters enter the system, researchers can simply multiply both sides by the denominator and still obtain a polynomial. When they enter in a more complicated form (e.g., one is an exponent of another), it may be necessary to define auxiliary parameters, adding to the original vector or replacing some of them to keep a polynomial form. The relation between the structural parameters and the newly defined auxiliary parameters is purely mathematical. Therefore, researchers can easily convert the new parameter vector back to the original structural parameters. Kocięcki and Kolasa (2023) call this new parameter vector semi-structural. To avoid unnecessary complexity, without any loss of generality, I do not distinguish structural and semi-structural parameters unless necessary. That is, I assume that θ shown in Γ_0, Γ_1, Γ_2, Γ_3, Σ, and L, J in a rational functional form, and (9) is a polynomial system.

[15] Polynomial systems are well-studied in mathematics, especially when the number of solutions is finite.

With the algorithm proposed by Kocięcki and Kolasa (2023), it is possible to find the complete equivalent set $K(\theta)$ with any given θ. All the computations of finding solutions of these systems are done using *SINGULAR* (Decker et al., 2022).[16] Their idea is to transform the identification conditions provided by Theorem 1 into a polynomial system, characterized by equations such as Equation (9). Theorems about Gröbner basis in algebraic geometry[17] then allow me to find all the solutions satisfying the polynomial system by reducing the system to its simplest form. The steps Kocięcki and Kolasa (2023) taken to find $K(\theta)$ are described below.

Algorithm 3 (Observationally Equivalent Sets)

(1) Start with a structural parameter vector θ chosen by the researcher, where identification is checked. Solve numerically the model and have a state-space representation with known (A, B, C, D, Σ).

(2) Calculate the reduced Gröbner basis associated with identification conditions satisfying Equation (9) based on (A, B, C, D, Σ). This is a polynomial system in $(\bar{\theta}, \bar{B}, \bar{D}, \bar{F}, \bar{G}, T, Q)$.

(3) If the Gröbner basis has multiple roots of $(\bar{\theta}, \bar{B}, \bar{D}, \bar{F}, \bar{G}, T, Q)$, rule out the ones that violate model constraints.

(4) If more than one solution remains, return the equivalent sets $K(\theta)$.

The algorithm takes an initial parameter value θ as input, solves an LRE model and a polynomial system, and outputs $K(\theta)$. Both the procedure of solving an LRE model and the reduction to the Gröbner basis can be done within *SINGULAR* with arbitrary precision. In the applications in section 5, I set the first part to have 600 digits and the latter to have 20 digits, which is more accurate than the default setting of *MATLAB*, which uses 16 digits of precision.

3.3 Multiple Priors and Random Sets

The next step after finding the observationally equivalent sets is to show thatalgorithm 2 indeed provides an estimator of (the convex hull of) the identified set. However, it is still unclear how the choice of priors affects the posterior mean and credible region under set-identification. When identification issues arise, the posterior of structural parameters incorporates non-revisable prior

[16] It is a computer algebra system for polynomial computations, with special emphasis on commutative and non-commutative algebra, algebraic geometry, and singularity theory. It is free and open-source under the GNU General Public Licence. See more information at https://www.singular.uni-kl.de/.

[17] See the appendix for a brief introduction about the Gröbner basis, or Cox et al. (2013) for more details.

knowledge over observationally equivalent parameters. In this section I show how the sensitivity of estimation is connected to identification failure of the model. Then I set up the framework for robust Bayesian inference with multiple priors.

In a Bayesian world, the unknown structural parameters θ are assumed to be $(\Theta, \mathcal{A})$-valued random variables, defined on a probability space. Let π_θ be a prior distribution of θ, which can be a belief of the researcher or information from micro evidence. I also need the measurability of the set-valued function K that maps Θ to the family of all closed subsets of Θ.

Assumption 3 *The equivalence mapping $K : \Theta \to \mathcal{F}$ is Effros measurable, that is, $K^{-}(F) \equiv \{\theta : K(\theta) \cap F \neq \emptyset\} \in \mathcal{A}$ for each $F \in \mathcal{F}$.*

Under Assumption 3, K is called a random closed set in 2^{Θ}. The multifunction K is a composition of the mapping from structural parameters θ to DSGE state-space solutions (A, B, C, D, Σ), with some algebraic transformation, and the mapping from the coefficients of a polynomial system to its solutions (also called a variety).[18] The corresponding prior π_K of $K(\theta)$ with given π_θ can then be taken as given, defined by π_θ such that

$$\pi_K(B) = \pi_\theta\left(\left\{\theta : K(\theta) \in B\right\}\right) \quad \text{for any } B \in \mathcal{B}(\mathcal{F}), \tag{10}$$

where $\mathcal{B}(\mathcal{F})$ is the σ-algebra generated by the Fell topology on $\mathcal{F}$ (Molchanov (2005), section 1.1). This π_K can also be denoted by $K_*(\pi_\theta)$ to explicitly show its dependence on π_θ.

The likelihood of θ is flat on $K(\theta)$ for any $y_{t=1}^{T}$, that is, $\theta \perp Y \mid K(\theta)$. That does not mean, however, that the induced likelihood of $\eta(\theta)$ is flat on $K(\theta)$ as well. In the most extreme case, a parameter that is set-identified may map into a point-identified parameter of interest.

In addition, the fact that researchers cannot discriminate one θ from another does not mean the two points have equal prior probability. In that sense, a flat prior does not equal non-informativeness. As Fisher would argue, “Not knowing the chance of mutually exclusive events and knowing the chance to be equal are two quite different states of knowledge” (Syversveen, 1998). Pericchi and Walley (1991) contend that there is not a single distribution that can model ignorance satisfactorily, hence an examination of a class of priors is necessary. If that is the case, researchers should care more about the credible regions of parameters, instead of their respective posterior probability. This motivates the application of robust Bayesian methods.

[18] Because polynomials are continuous functions in their variables, the set of roots must be closed.

In general, the parameters spaces Θ, $\mathcal{H}$ can be subspaces of $\mathbb{R}^n$, and $K(\Theta)$ can be a subset of a Baire space,[19] where the conditioned probability of $K(\theta)$ may be zero. To avoid the Borel-Kolmogorov paradox,[20] in the rest of this paper, I assume the conditional distributions $\pi_{\theta|K}$ are regular and defined based on conditional expectations. That is, $\pi_{\theta|K}(A) \equiv \mathbb{E}[\mathbf{1}_A(\theta) \mid K]$, similarly for $\pi_{\eta|K}$. The posterior of θ, $\pi_{\theta|Y}$, can be expressed as

$$\pi_{\theta|Y}(A) = \int_{\mathcal{F}} \pi_{\theta|K}(A) \, \mathrm{d}\pi_{K|Y}, \quad A \in \mathcal{A}. \tag{11}$$

From expression (11) it can be seen that the conditional prior of θ given $K(\theta)$ cannot be updated by the data. The same argument also holds for the posteriors $\pi_{\eta|Y}$. The following example might be helpful in understanding how to connect these conditional priors and why set-identification could happen.

Example 2 *Let the structural parameter be $\theta = (\theta_1, \theta_2)$, and the equivalent set K is defined as $K(\theta) = \left\{\bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2) \mid \theta_1 + \theta_2 = \bar{\theta}_1 + \bar{\theta}_2\right\}$. Consider two prior distributions on θ:*

$$\pi_\theta : \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \quad \tilde{\pi}_\theta : \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left(\begin{bmatrix} a \\ -a \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right),$$

where a is a non-zero constant, and π_θ and $\tilde{\pi}_\theta$ induce the same prior distribution on K. Suppose researchers observe only $Y_i \sim N(\theta_1 + \theta_2, 1)$, $i = 1, \dots, t$. Then, in the posterior distribution of θ (expression (11)) $\pi_{K|Y}$ gets updated by $\{Y_i\}$, but the conditional distributions $\pi_{\theta|K}$ and $\tilde{\pi}_{\theta|K}$ remain unchanged because the data do not contain any more information than $\theta_1 + \theta_2$.

I refer to π_K as revisable prior knowledge, and $\pi_{\theta|K}$ as unrevisable prior knowledge, à la Giacomini and Kitagawa (2021) (see also Poirier (1998)). The robust Bayesian analysis does not require researchers to commit to a single (conditional) prior; instead, any prior satisfying the following condition is allowed:

Definition 3 (Multiple-Prior Class) *Given a π_K supported only on $K(\Theta)$, the classes of conditional*

[19] This space is still a Polish space equipped with a Borel σ-algebra. Hence there exists a regular conditional distribution.

[20] In probability theory, the Borel–Kolmogorov paradox suggests that conditional probability with respect to an event of probability zero can be indeterminate or ill-posed.

priors for θ given K are

$$\Pi_{\theta|K} = \left\{ \pi_{\theta|K} : \pi_{\theta|K} \left(\left\{ \theta : K(\theta) = K \right\} \right) = 1, \pi_K - \textit{almost surely} \right\}. \tag{12}$$

This condition prior class would then induce a class of proper priors for θ that also coincides with the prior class defined in Ke et al. (2022), the class of all priors that the marginal distribution for K coincides, given π_K, i.e.,

$$\begin{aligned} \Pi_\theta(\pi_K) &\equiv \left\{ \int \pi_{\theta|K} d\pi_K : \pi_{\theta|K} \in \Pi_{\theta|K} \right\} \\ &= \left\{ \pi_\theta : \pi_\theta \left(\left\{ \theta : K(\theta) \in B \right\} \right) = \pi_K(B), \text{ for } B \in \mathcal{B}(\mathcal{F}) \right\} \end{aligned}. \tag{13}$$

Given π_K, the equivalence of having $\Pi_\theta(\pi_K)$ and $\Pi_{\theta|K}$ is shown in appendix. I will use these two notations of multiple prior classes interchangeably in proofs, to avoid unnecessary complication.

This is not the only way to define a multiple-prior class. However, this specific case has the advantage of being both tractable, because it collects all the prior distributions that assign probability 1 to the equivalence set $K(\theta)$, and convenient, as it generates the same prior predictive distribution, the distribution of observations expected before observing any data (Geweke and Whiteman, 2006; Geweke, 2007; Del Negro and Schorfheide, 2008; Weitzman, 2009). Therefore, any two priors that does not belong to the same multiple-prior class will result to different predictions over data. It is possible to evaluate one prior against another using data-based criteria, but this method is not applicable if they are in the same class. The following lemma formalizes this fact.

Lemma 2 (Prior Predictive Distribution) *For any given π_K, the prior predictive distribution defined as*

$$p(y) = \int_\Theta p(y \mid \theta) d\pi_\theta$$

is constant across $\pi_\theta \in \Pi_\theta(\pi_K)$ for all y.

The same result also holds for the posterior predictive distribution.

Moreover, under some regularity conditions, the range of posterior means is close (in probability) to the frequentist estimator of the identified set (Ke et al., 2022). The connections between parameters can now be characterized by Figure 4. Given $(\Pi_{\theta|K}, \pi_{K|Y})$, I can then define the class of posteriors for the parameters of interest as

$$\Pi_{\eta|Y} = \left\{ \pi_{\eta|Y}(\cdot) = \int_{\mathcal{F}} \pi_{\theta|K}\left(\eta(\theta) \in \cdot\right) d\pi_{K|Y} : \pi_{\theta|K} \in \Pi_{\theta|K} \right\}. \tag{14}$$

If researchers can put a prior on K, and draw from $\Pi_{\theta|K}$, the methods in Giacomini and Kitagawa (2021) can be applied to find the robust distribution $\pi_{\eta|Y}$.

However, in practice, researchers are supposed to specify a prior π_θ for θ, not $K(\theta)$, before estimating a DSGE model. The challenging part of working directly with distributions on $K(\theta)$ is mentioned in section 3. Therefore, for convenience, I work with the class $\Pi_\theta(\pi_K)$ of distributions on θ. Firstly, the π_K that characterizes this class is pinned down by the push-forward measure of π_θ, and there is no need to compute π_K. Secondly, any other prior $\tilde{\pi}_\theta$ within the same class can be obtained by redrawing from the observational equivalent set $K(\theta)$ of a given draw θ from prior π_θ.

3.4 Robust Distributions

From the previous section, it becomes clear the posterior distributions of both θ and η can be vulnerable to the choice of unrevisable conditional priors $\pi_{\theta|K}$. I am now fully equipped to show the main theorem of this paper. The class $\Pi_{\theta|K}$ is often hard to define explicitly from the model because each $\pi_{\theta|K}$ can have different support with different K. Moreover, $\pi_{K|Y}$ is not what is being estimated directly in practice. In this section, I will define the robust probabilities $\underline{\pi}_{\eta|Y}$ and $\overline{\pi}_{\eta|Y}$ and show a more practical way to estimate these probabilities. To invoke theories on random closed sets, I need to put some structure on η.

Assumption 4 *The parameter of interest $\eta : \Theta \to \mathcal{H}$ is a continuous function of the structural parameter θ.*

I first characterize the posterior class $\Pi_{\eta|Y}$ by its lowest and highest possible probabilities on different sets, denoted by $\underline{\pi}_{\eta|Y} : \mathcal{D} \to [0,1]$ and $\overline{\pi}_{\eta|Y} : \mathcal{D} \to [0,1]$, respectively, and defined as

$$\underline{\pi}_{\eta|Y}(D) \equiv \inf_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}(D)$$

$$\overline{\pi}_{\eta|Y}(D) \equiv \sup_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}(D).$$

The lower and upper posterior probabilities defined here can be considered robust probability bounds on a specific set. In other words, they measure the lowest and highest probability of D that can be obtained from $\Pi_{\theta|K}$ regardless of what prior distribution $\pi_{\theta|K}$ is used. Although computing the probabilities from these formulas is easy to understand and tempting to apply directly, this method is unrealistic because, besides the above-mentioned reasons, it requires exhausting all conditional distributions within the class. Several regularity conditions are needed to derive analytical results for $\underline{\pi}_{\eta|Y}$ and $\overline{\pi}_{\eta|Y}$. I begin with some weak but necessary regularity conditions.

Assumption 5 (Regularity) *Let the prior of structural parameters θ, π_θ be non-atomic on $(\Theta, \mathcal{A})$, and $\pi_\theta(\Theta) = 1$.*

In the case where $\Theta = \mathbb{R}^n$, π_θ being absolutely continuous with respect to the Lebesgue measure is sufficient to guarantee non-atomicity. With this property, we can leverage the theorems elsewhere (Ke et al., 2022; Giacomini and Kitagawa, 2021) and apply them to the DSGE framework.

Theorem 2 (Lower and Upper Probabilities) *For any given π_θ, equivalence mapping K and function η, , under Assumptions 3, 4 and 5, and $\eta(K(\theta))$ closed for π_θ-almost every θ, let $D \in \mathcal{D}$,*

$$\underline{\pi}_{\eta|Y}(D) = \pi_{\theta|Y}\left(\left\{\theta : \eta(K(\theta)) \subset D\right\}\right), \tag{15}$$

$$\overline{\pi}_{\eta|Y}(D) = \pi_{\theta|Y}\left(\left\{\theta : \eta(K(\theta)) \cap D \neq \varnothing\right\}\right). \tag{16}$$

Moreover, $\left\{\pi_{\eta|Y}(D) : \pi_{\eta|Y} \in \Pi_{\eta|Y}\right\}$ is a connected interval, i.e., each point between the lower and upper bound is attainable.

Theorem 2 is similar to Theorem 1 in Giacomini and Kitagawa (2021), with inverse mapping from the reduced-form parameter to the structural parameter replaced by the equivalent set. Assumption 3 and 4 guarantee $K(\theta)$ and $\eta(K(\theta))$ are random closed sets. Proof can be found in the appendix. The robust probabilities (15) and (16) are called the containment functional and the capacity functional of a random set (Molchanov and Molinari (2018)), respectively. They are also special cases of a belief function and a plausibility function, respectively, in the imprecise probability literature (Walley (1991)).

This theorem allows me to compute the lower and upper probabilities. For example, when η is a subvector of θ, the lower probability can be obtained by the following algorithm.

Algorithm 4 (Robust Probabilities)

(1) Set J=0.

(2) Draw from posterior $\pi_{\theta|Y}$ for a given prior defined over Θ.

(3) Compute its equivalent set $K(\theta)$.

(4) If $\eta(K(\theta)) \subset D$, set $J = J + 1$.

(5) Repeat steps 1–3 M times; the lower probability of D is therefore $\frac{J}{M}$.

The steps to compute the upper probability are similar.

3.4.1 Robust Posterior Mean

In standard Bayesian analysis, researchers are particularly interested in the expectation of parameters of interest and its credible region. In this section, I show that the results from Algorithm 2 has a finite-sample interpretation: it reports the range of posterior means of the parameters of interest from a prior class.

Before I show the main theorem, I need a lemma that connects the K–transformation of the posterior draws of θ to draws from the posterior of $K_*\pi_\theta$, the push-forward prior of K.

Lemma 3 *The push-forward measure of $\pi_{\theta|Y}$ by a measurable multifunction $K : \Theta \to \mathcal{F}$, written as $\pi^*_{K|Y} = K_*\pi_{\theta|Y}$, coincides with the posterior distribution $\pi_{K|Y}$ of the push-forward measure $\pi_K = K_*\pi_\theta$.*

This lemma, which generalizes the result in Ke et al. (2022) Appendix C without assuming a parametric structure of K, states that two different methods of achieving posterior draws of equivalent sets are identical. The main theorem of this paper is stated below.

Theorem 3 (Posterior Means of Scalar η) *For a given π_θ, let Assumptions 3, 4, and 5 hold, that is, given a non-atomic prior π_θ, there is a push-forward measure π_K of π_θ under K that is also non-atomic. In addition, let the parameter of interest η be a scalar. Define*

$$\overline{\eta}^*(\theta) = \sup_{\theta' \in K(\theta)} \eta(\theta'), \quad \underline{\eta}^*(\theta) = \inf_{\theta' \in K(\theta)} \eta(\theta').$$

Then, the set of posterior means is characterized by

$$\sup_{\pi_{\theta|Y} \in \Pi_{\theta|Y}} \mathbb{E}_{\theta|Y}\Big[\eta(\theta)\Big] = \mathbb{E}_{\theta|Y}\Big[\overline{\eta}^*(\theta)\Big], \quad \inf_{\pi_{\theta|Y} \in \Pi_{\theta|Y}} \mathbb{E}_{\theta|Y}\Big[\eta(\theta)\Big] = \mathbb{E}_{\theta|Y}\Big[\underline{\eta}^*(\theta)\Big],$$

where $\Pi_{\theta|Y}$ collects the posteriors of Equation (13) with given $\pi_K = K_*\pi_\theta$.

This theorem states that if the researcher picks a prior π_θ, draws from its posterior, and finds the lower and upper bounds of η within the observationally equivalent θs, then it is as if the researcher knew all the priors that generate the same prior on K and collected the range of the posterior means. The expectation of $\overline{\eta}^*(\theta)$ and $\underline{\eta}^*(\theta)$ under distribution $\pi_{\theta|Y}$ is effective, because it computes the bounds of $\mathbb{E}_{\eta|Y}(\eta)$ robust to the choice of conditional priors. It is also attainable via numerical methods because it needs draws only from one distribution.

Intuitively, Theorem 3 holds because even if the posterior $\pi_{\theta|Y}$ that the researcher draws from absorbs some arbitrarily specified unrevised conditional prior of $\theta \mid K$, the posterior of K will not be affected by this conditional prior if it is possible to collapse the sampled space from Θ to $K(\Theta)$. In other words, the observationally equivalent sets $K(\theta)$ contain the same information as the class of all conditional priors. Therefore, the posterior distribution of the identified set is the same.

3.4.2 Robust Credible Region

This section introduces the robust Bayesian counterpart of the posterior mean and credible region in standard Bayesian inference. For $\alpha \in (0,1)$, consider a subset $C_\alpha \subset \mathcal{H}$ such that the posterior lower probability $\underline{\pi}_{\eta|Y}(C_\alpha)$ is greater than or equal to α :

$$\underline{\pi}_{\eta|Y}(C_\alpha) = \pi_{\theta|Y}\left(\left\{\theta : \eta(K(\theta)) \subset C_\alpha\right\}\right) \geq \alpha.$$

As has been mentioned in Giacomini and Kitagawa (2021), such set will not be unique unless some extra minimality condition is imposed. Let

$$\begin{aligned} C^*_\alpha &\in \arg\min_{D\in\mathcal{D}} Leb(D) \\ s.t. \quad & \pi_{\theta|Y}\left(\{\theta : \eta(K(\theta)) \subset C_\alpha\}\right) \geq \alpha, \end{aligned}$$

where $Leb(D)$ is the Lebesgue measure of D. This optimization problem is still intimidating to solve, because of the curse of dimensionality. However, if the researcher focuses on the scalar case for η,[21] and further constrains C^*_α to be convex, the problem of finding the robust credible region becomes

[21] Although impulse responses can be, in principle, infinite-dimensional, researchers are more often than not interested in the pointwise coverage probability at each time period than in the overall coverage of an IRF.

finding the smallest $\overline{q} \in \mathbb{R}$ such that

$$\inf_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}\left((-\infty, \overline{q}]\right) \geq 1 - \alpha \tag{17}$$

and the largest $\underline{q}$ such that

$$\sup_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}\left((-\infty, \underline{q}]\right) \leq \alpha. \tag{18}$$

If it can be shown that $\overline{q}^*_{1-\alpha/2} : \pi_{\theta|Y}\left(\overline{\eta}^*(\theta) \leq \overline{q}^*_{1-\alpha/2}\right) = 1 - \frac{\alpha}{2}$ and $\underline{q}^*_{\alpha/2} : \pi_{\theta|Y}\left(\underline{\eta}^*(\theta) \leq \underline{q}^*_{\alpha/2}\right) = \frac{\alpha}{2}$ solve Equations (17) and (18), then the researcher can obtain the desired robust credible region as $[\underline{q}^*_{\alpha/2}, \overline{q}^*_{1-\alpha/2}]$, with

$$\inf_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}\left(\left[\underline{q}^*_{\alpha/2}, \overline{q}^*_{1-\alpha/2}\right]\right) \geq 1 - \alpha.$$

This result is shown in the following theorem.

Theorem 4 *Let the assumptions in Theorem 2 hold. For any given $q \in \mathbb{R}$ and given prior π_θ,*

$$\inf_{\pi_{\theta|Y} \in \Pi_{\theta|Y}} \pi_{\theta|Y}\left(\eta(\theta) \leq q\right) = \pi_{\theta|Y}\left(\overline{\eta}^*(\theta) \leq q\right)$$

and

$$\sup_{\pi_{\theta|Y} \in \Pi_{\theta|Y}} \pi_{\theta|Y}\left(\eta(\theta) \leq q\right) = \pi_{\theta|Y}\left(\underline{\eta}^*(\theta) \leq q\right).$$

The proof of this theorem is a direct result of Theorem 2. Details are also presented in the appendix.

The next theorem is a direct application of Theorem 3 in Giacomini and Kitagawa (2021) to DSGE models. Without loss of generality, it is possible to truncate the spaces Θ and H to their compact subspaces to always allow integrability. Let θ_0 denote the underlying true value that generates the data.

Assumption 6 *The equivalence mapping $K : \Theta \to \mathcal{F}$ is a continuous correspondence at θ_0.*

The continuity of correspondences is defined as in Aliprantis and Kim (2006) Definition 17.2. Assumption 6 is easy to verify when the zero set of a polynomial system is finite, the structural parameter space is complex, or both (Alexanderian, 2013). However, it is not easy to show the assumption holds in general when there are infinite solutions, even if the "discriminant locus" is excluded.

Theorem 5 (Consistency of Posterior Mean) *Let Assumptions 3, 5, and 6 hold, and assume further that the induced prior π_K leads to a consistent posterior*[22] *and that $\Theta \subset \mathbb{R}^p, H \subset \mathbb{R}^q$ for some $p,q < \infty$ are compact spaces. Then the Hausdorff distance*[23] *between the set of posterior means and the convex hull of the true identified set goes to zero almost surely as T increases, i.e.,*

$$\lim_{T\to\infty} d_H\left(\mathbb{E}_{\theta|Y^T}\left(\left[\underline{\eta}^*(\theta), \overline{\eta}^*(\theta)\right]\right), \left[\underline{\eta}^*(\theta_0), \overline{\eta}^*(\theta_0)\right]\right) \to 0, \quad p\left(Y^\infty \mid \theta_0\right)\text{-a.s.}$$

The proof of Theorem 5 follows directly from Theorem 3 in Giacomini and Kitagawa (2021). It provides a justification for using the algorithm-generated posterior means as a consistent estimator of the convex hull of the identified set. The theorem also implies that the range of posterior means will converge to the true identified set.

4 Applications

As stated in the previous sections, in the DSGE framework, the go-to procedure for obtaining the posterior distribution of η is not through the estimation of some reduced-form parameters, but through the evaluation of likelihoods in terms of the parameters of the state-space representation. The priors are typically chosen at θ-level. Therefore, it is very challenging to proceed as in Giacomini and Kitagawa (2021): first estimate the reduced-form parameter, and then draw from unrevised priors subject to constraints. However, it is still possible to find the identified set for the parameters of interest thanks to work done by Kociecki and Kolasa (2018). Theorem 3 states that it is possible to circumvent the trouble of drawing from a class of unrevised priors by finding the complete observationally equivalent set of θ. Here I show the main algorithm again, but with more details than provided for Algorithm 2.

Algorithm 5 (Robust Bayesian Mean and Credible Region)

(1) Perform the standard MCMC exercise:

- *(a) take any prior on θ with full support on the parameter space;*
- *(b) based on the data available, get posterior draws θ_j subject to $\pi_{\theta|Y}$ from the standard Bayesian DSGE sampler.*

[22] That is, for any neighborhood V_0 of $K(\theta_0)$, $\pi_{K|Y^T}(V_0) \xrightarrow{p} 1$ as $T \to \infty$.

[23] The Hausdorff distance is defined as $d_H(X,Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\right\}$. In the one-dimensional η case, the Hausdorff distance between $[a,b]$ and $[c,d]$ is $\max\{|a-c|,|b-d|\}$.

(2) For each posterior draw θ_j,

(a) using Algorithm 3 proposed by Kocięcki and Kolasa (2023), find the equivalent class of θ_j, $K(\theta_j)$ characterized by a reduced Gröbner basis;

(b) optimize over the Gröbner basis constraints to find the identified upper and lower bounds of each element of θ or the object of interest $\eta(\theta)$.

(3) Draw another θ from the posterior and repeat M times step 2.

(4) Compute the estimated range of posterior mean or quantiles by averaging over the mean (or quantiles) of the minimum or maximum obtained in steps 2 and 3.

Theorem 3 gives a theoretical foundation for this algorithm. It is similar to Theorem 2 in Giacomini and Kitagawa (2021) except I deal with posterior $\pi_{\theta|Y}$. A detailed coding strategy for the examples used in this section can be found in the appendix.

The exercises in this section are done in the following way. I start with an analytically tractable, toy model to apply the identification theorems in section 2. Then I perform analysis based on simulation results. First, I start with a set of true values and model specification, simulate the model, and use the simulated data to run the algorithm proposed in this paper. Then I conduct inference based on algorithm-generated results.

4.1 A Taylor-rule Model

Consider a simple model introduced in Cochrane (2011) that consists of a monetary policy shock transition, a Fisher equation, and a monetary policy rule:

$$\begin{aligned}
x_t &= \rho x_{t-1} + \varepsilon_t, \quad |\rho| < 1, \varepsilon_t \sim N(0, \sigma_e) \\
i_t &= r + \mathbb{E}_t \pi_{t+1} \\
i_t &= r + \phi_\pi \pi_t + x_t, \quad \phi_\pi > 1,
\end{aligned}$$

where x_t is the monetary policy shock, i_t is the nominal interest rate, r is the constant real rate, and π_t is the inflation rate. Only π_t is assumed to be observed. This system is not minimal without further simplification; therefore, it is necessary to first minimize the system by keeping only x_t as the state variable. The solution yields

$$A = \rho, B = 1, C = \frac{\rho}{\rho - \phi_\pi}, D = \frac{1}{\rho - \phi_\pi}, \Sigma = \sigma_e^2,$$

which is equivalent to an AR(1) setting

$$\pi_t = \rho\pi_{t-1} - \frac{1}{\phi_\pi - \rho}\varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_e^2).$$

Here the structural parameter vector is $\theta = (\rho, \phi_\pi, \sigma_e)$, and with simple regression it is possible to identify $(\rho, \frac{\sigma_e}{\phi_\pi - \rho})$. Invoking Theorem 1 gives the same result, that (ϕ_π, σ_e) are not jointly identifiable. Therefore, the impulse response function is also not identified. Because ρ is identified, $\frac{\sigma_e}{\phi_\pi - \rho_0} = \frac{\sigma_{e0}}{\phi_{\pi 0} - \rho_0}$ for any pair of (ϕ_π, σ_e) that is observationally equivalent to $(\phi_{\pi 0}, \sigma_{e0})$. The identified set of impulse responses is $IR(t,s,1)\,|_{K(\theta_0)} = -\frac{\rho_0^s}{\phi_\pi - \rho_0} \cdot Q$, denoting the impulse responses evaluated at all points in $K(\theta_0)$, where Q is a scalar from Theorem 1 can take values between $(0, \frac{\phi_\pi - \rho_0}{1-\rho_0})$. Therefore, $IR(t,s,1)\,|_{K(\theta_0)} = (\min\{0, -\frac{\rho_0^s}{1-\rho_0})\}, \max\{0, -\frac{\rho^s}{1-\rho})\})$ for all s.[24] If $\rho_0 > 0$, $\phi_\pi\,|_{K(\theta_0)} = (1, \infty)$ and $\sigma_e\,|_{K(\theta_0)} = (1-\rho_0, \infty)$, $IR_\pi(t,s,1)\,|_{K(\theta_0)} = (-\frac{\rho_0^s}{1-\rho_0}, 0)$.

Given that I have the analytical result of the identified set $K(\theta_0)$ and $IR(t,s,1)\,|_{K(\theta_0)}$, where θ indicates the structural parameters σ_e, ρ, and ϕ_π, I then run Algorithm 5 using *MATLAB* and *SINGULAR* to see if the results match. I start with simulated data of 200 periods, and then run an MCMC estimation of the parameters $(\sigma_e, \phi_\pi, \rho)$ in *DYNARE* (Adjemian et al., 2011). The local identification results from embedded methods based on Iskrev (2010), Komunjer and Ng (2011), and Qu and Tkachenko (2012) confirm that (σ_e, ϕ_π) are pairwise collinear. In the MCMC exercise, as has been mentioned in section 2, I pick the uniform prior as the first prior. Then I use a hierarchical scheme to draw the second posterior based on the first posterior draws and their observationally equivalent sets so that these two posteriors always induce the same posterior distribution over $(\rho, \frac{\sigma_e}{\phi_\pi - \rho})$ and, therefore, the same posterior predictive distribution. The numerical details of the hyperparameters for redrawing are shown in the appendix. The standard Bayesian result from Table 1 shows the sensitivity of the posterior to the choice of priors.

Table 3: Estimated Identified Set for Cochrane (2011) Model[25]

	True value	Identified set	Range of post mean	Robust Bayesian credible region
σ_e	1	$(0.2, \infty)$	$(0.21, \infty)$	$(0.14, \infty)$
ϕ_π	1.8	$(1, \infty)$	$(1.00, \infty)$	$(1.00, \infty)$
ρ	0.8	0.8	0.80	(0.74, 0.87)

[24] Some may argue that if normalizing ε to a standard Gaussian is allowed, identification can be achieved. However, because the ε_t here can include expectation errors and sunspot shocks, there is no reason to assume unit variance in addition to zero conditional expectations.

Then I use Algorithm 2 proposed in this paper to estimate the identified set for $\theta = (\rho, \phi_\pi, \sigma_e)$. For the given true parameter values ($\rho_0 = 0.8, \phi_{\pi 0} = 1.8, \sigma_{e0} = 1$), it is possible to apply the theoretical results above. It is known that ϕ_π can take values $(1, \infty)$, and the fact that only $\frac{\sigma_{e0}}{\phi_{\pi 0} - \rho_0} = 1$ is identified means σ_e is bounded below by 0.2. The last column in Table 3 shows the empirical results from my algorithm.[26] The averaged values of the estimated identified sets match the theoretical values quite well.

I then proceed to computing the impulse responses for inflation π_t. The first part is shown in Figure 3 and also in the upper panel of Figure 7. Using the strategy mentioned in Algorithm 2, I find the (pointwise) minimum and maximum impulse response within each equivalence set attached to a posterior draw. I average the ranges of impulse responses, and estimate the range of posterior means and the robust Bayesian credible region for impulse responses. The lower panel of Figure 7 shows that the region of model-consistent impulse response functions $IR_\pi |_{K(\theta_0)}$ is much larger than the standard Bayesian confidence interval. The estimated range of posterior means for IR_π is of similar size and covers the theoretical $IR_\pi |_{K(\theta_0)}$, whereas the 90% robust Bayesian credible region is much larger, as it should be.

4.2 Three-equation New Keynesian Model

In this section I present applications to three different variants of a baseline New Keynesian model, also known as the three-equation NK models. They are well-studied small-scale New Keynesian DSGE models that consist of final-goods producing firms, intermediate-goods producing firms, households, a central bank, and a fiscal authority (Goodfriend and King, 1997; Clarida et al., 2000; King, 2000; Woodford, 2003a; Lubik and Schorfheide, 2004; An and Schorfheide, 2007; Galí, 2015; Herbst and Schorfheide, 2015). The first variant has only temporary shocks, where the (local) identification failure is within semi-structural parameter κ. The second example is the An and Schorfheide (2007) model, where local identification fails non-trivially, but the optimal policy does not depend on model parameterization. In the last variant, I make modifications to the model in An and Schorfheide (2007) by introducing a cost-push shock (Clarida et al., 1999; Woodford, 2003a,b; Galí, 2015). By doing that, policy-makers face a trade-off between the output gap and inflation when nominal rates are set, and the optimal monetary policy will depend on estimation results.

[25]Because of machine precision in *MATLAB*, I round numbers greater than 10^6 to infinity. Same in Table 5, 9, and 10.

[26]Here I use the prior setup 1 for the algorithm. Same for the lower panel of Figure 7.

4.2.1 Temporary Shocks

Consider a benchmark three-equation New Keynesian model similar to Galí and Gertler (1999). Instead of having AR(1) shocks, I remove the auto-correlation and make them i.i.d.:

$$\begin{aligned} y_t &= \mathbb{E}_t y_{t+1} - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t \pi_{t+1}\right) + \varepsilon_{yt} \\ \pi_t &= \beta \mathbb{E}_t \pi_{t+1} + \kappa y_t + \varepsilon_{\pi t} \\ i_t &= \rho i_{t-1} + (1-\rho)\left(\phi_\pi \pi_t + \phi_y y_t\right) + \varepsilon_{Rt} \\ \varepsilon_{jt} &\sim N(0,1); \quad j = y, \pi, R \end{aligned},$$

where π_t is the inflation, y_t is the output gap, i_t is the nominal interest rate, ϵ_{yt} is the demand shock, $\epsilon_{\pi t}$ is the supply shock, and ϵ_{Rt} is the monetary policy shock. The equations are referred to as the dynamic IS equation, New Keynesian Phillips curve, and an interest rate feedback rule with output gap rule specification, respectively. Here the structural parameters are the inverse intertemporal elasticity of substitution σ; the Calvo price stickiness τ; the elasticity of marginal disutility with respect to labor ψ; the influence of inflation rate and the output gap in the interest rate rule ψ_π and ψ_y, respectively; $\kappa = \frac{(1-\tau)(1-\beta\tau)}{\tau}(\sigma+\psi)$ is the slope of the Phillips curve. The structural parameter vector is $\theta = (\sigma, \beta, \tau, \psi, \phi_\pi, \phi_y)$, and monetary policy adjustment rate ρ is calibrated to be 0.1. Putting the above equations in a standard LRE form, i.e., Equation (3) results in

$$\Gamma_0 = \begin{pmatrix} 1 & -(1-\rho)\phi_y & -(1-\rho)\phi_\pi \\ 1 & \sigma & 0 \\ 0 & -(1-\tau)(1-\beta\tau)(\sigma+\psi) & 0 \end{pmatrix}, \quad \Gamma_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma & 1 \\ 0 & 0 & \beta\tau \end{pmatrix}, \quad \Gamma_2 = \begin{pmatrix} \rho \\ 0 \\ 0 \end{pmatrix}, \quad \Gamma_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \tau \end{pmatrix}.$$

Even this simple three-equation model is too complex to be analytically enlightening. It is hard to tell the identifiability of all parameters just by examining the model. However, it is obvious that τ is not identified in general.[27] Moreover, even if identification failure of this kind is excluded, by rescaling ψ and τ continuously, it is possible to achieve the same κ. Because (ψ, τ) enter the equation system only via κ, ψ can be jointly unidentified with τ. This means the structural parameters of the three-equation model, when micro-founded, is neither locally nor globally identified; that is, if the researcher tries to estimate $(\sigma, \beta, \tau, \psi)$ instead of just (σ, β, κ) for calibration or policy analysis purposes. In terms of identification of impulse responses, however, because the model dynamics

[27] If I fix $\beta = \frac{4}{5}$, $\tau = \frac{9}{10}$ is always observationally equivalent to $\tau = \frac{45}{31}$. The latter case can be excluded by restricting the support to $[0,1]$ because τ stands for price stickiness.

will be affected only through κ, they should be identified if $(\phi_\pi, \phi_y, \sigma, \beta, \kappa)$ are identified. The Blanchard-Kahn condition (Blanchard and Kahn, 1980), which guarantees determinacy, will not be restrictive here.

Table 4: Three-equation Model Prior and Posterior Distribution of Structural Parameters

	True value	Prior distribution			Posterior distribution			
		Distr.	Mean	St. Dev.	Mode	Mean	5 percent	95 percent
ϕ_π	1.7	Normal	1.5	0.5	1.67	1.65	1.46	1.82
ϕ_y	0.2	Normal	0.5	0.3	0.21	0.19	0.11	0.29
σ	1	Gamma	1	0.5	0.99	1.01	0.92	1.11
β	0.99	Beta	0.9	0.005	0.99	0.99	0.98	1.00
ψ	1	Gamma	5	2	4.23	4.60	1.46	7.36
τ	0.75	Beta	0.5	0.3	0.84	0.85	0.80	0.91

Table 5: Estimated Identified Set of Structural Parameters for Three-equation Model

	True value	Identified set	Range of post mean	Robust Bayesian CR
ϕ_π	1.7	1.7	1.65	(1.46, 1.83)
ϕ_y	0.2	0.2	0.20	(0.11, 0.29)
σ	1	1	1.01	(0.92, 1.12)
β	0.99	0.99	0.99	(0.98, 1.00)
ψ	1	$(0, +\infty)$	$(0, +\infty)$	$(0, +\infty)$
τ	0.75	$(0.67, 1)$	$(0.69, 1.00)$	(0.63, 1.00)

Based on data generated from the true values presented in Table 4, I perform a naive Bayesian estimation using the prior distributions within the same table. *DYNARE* reports identification checks at the prior mean, and, not surprisingly, (ψ, τ) is pairwise unidentified. The solutions to the Gröbner basis also show the same identification results for each MCMC draw. This knowledge helps to reduce the identification problem to finding the (ψ, τ) pairs that lead to the same κ, when combined with parameter bounds and the Blanchard-Kahn condition, reduced to

$$\frac{(1-\tau)(1-0.99\tau)}{\tau}(1+\psi) = \frac{103}{600}, \quad \psi > 0, \quad 0 < \tau < 1,$$

which then provides the identified set: for ψ, it is $\psi\,|_{K(\theta_0)}= (0, +\infty)$, and for τ, it is $\tau\,|_{K(\theta_0)}= (0.67, 1)$. As discussed in section 4, if $(\phi_\pi, \phi_y, \sigma, \beta, \kappa)$ are identified, the impulse response functions are also identified (see Figure 9). The true impulse response is very close to the posterior mean, the standard Bayesian credible region is tight, and the 90% robust Bayesian credible region coincides with the 90% standard credible region when the impulse responses are identified. This result shows that the algorithm proposed in this paper does not cause a loss (excluding computation time) when the parameters of interest are identified.

One may argue that application of the robust Bayesian method is unnecessary in this case, when the identification failure is obvious and all one needs to do is imposing an additional normalization restriction. However, in general, researchers do not know if the parameters $(\phi_\pi, \phi_y, \sigma, \beta, \kappa)$ are identified, because the DSGE models are not always analytically tractable, and the identification failure can be imperceptible. Since there is no loss in the robust Bayes method, estimation using Algorithm 5 is always recommended.

4.2.2 An and Schorfheide (2007)

A more economically meaningful example would be to have nontrivial identification failures (i.e., identification issues of a less-mechanical nature). The following example is also very similar to the exercises in Herbst and Schorfheide (2015), where the authors allow for correlation between productivity growth and government spending. The equilibrium is characterized by the following linearized equations:

$$
\begin{aligned}
y_t =& \mathbb{E}_t\left[y_{t+1}\right] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t\left[\pi_{t+1}\right] - \mathbb{E}_t\left[z_{t+1}\right]\right) + g_t - \mathbb{E}_t\left[g_{t+1}\right] \\
\pi_t =& \beta\mathbb{E}_t\left[\pi_{t+1}\right] + \sigma\frac{1-v}{v\pi^2\psi}\left(y_t - g_t\right) \\
i_t =& \rho_R i_{t-1} + \left(1-\rho_R\right)\psi_\pi\pi_t + \left(1-\rho_R\right)\psi_y\left(y_t - g_t\right) + \varepsilon_{R,t} \\
z_t =& \rho_z z_{t-1} + \varepsilon_{z,t} \\
g_t =& \rho_g g_{t-1} + \varepsilon_{g,t}.
\end{aligned}
$$

Here the parameters are $\theta = (\sigma, \beta, \nu, \psi, \psi_\pi, \psi_y, \rho_R, \rho_g, \rho_z, \sigma_R, \sigma_g, \sigma_z)$, including the inverse elasticity of substitution σ; the elasticity of demand for each intermediate good $\frac{1}{\nu}$; and the quadratic loss in price adjustment ψ. The endogenous variables are $\left(y_t, \pi_t, i_t, g_t, z_t\right)'$, where $S_t = \left(z_t, g_t, i_t\right)'$, $P_t = \left(\pi_t, y_t\right)'$. In the original model there is no measurement error. I therefore drop a few steady state parameters and assume without loss of generality that the output gap y_t, inflation rate π_t, and nominal interest rate i_t are directly observed. Similar to the three-equation model, it is not hard to see (ν, ψ) enter the model only through the ratio $\sigma \frac{1-\nu}{\nu\pi^2\psi}$. Because (ν, ψ) are not jointly identifiable, they are replaced by $\kappa = \sigma \frac{1-\nu}{\nu\pi^2\psi}$ in estimation. However, this replacement will still not be enough to generate point-identification. As Komunjer and Ng (2011) or Qu and Tkachenko (2012) show, the monetary policy parameters $(\psi_\pi, \psi_y, \rho_R, \sigma_R)$ cannot be identified in this output gap rule specification, although under the output growth specification these parameters are locally identifiable (Ivashchenko and Mutschler, 2020). The robust Bayesian estimation results are reported in Table 6.

Table 6: Estimated Identified Set of Structural Parameters for AS Model

	True value	Identified Set	Range of Posterior Mean	Robust Bayesian CR
τ	2	2.00	1.97	(1.36, 2.76)
κ	0.15	0.15	0.15	(0.10, 0.21)
ψ_π	1.5	(1.00, 4.87)	(1.00, 4.11)	(1.00, 5.36)
ψ_y	1	(0.00, 1.15)	(0.00, 0.94)	(0.00, 1.44)
ρ_z	0.65	0.65	0.63	(0.56, 0.71)
ρ_g	0.75	0.75	0.74	(0.66, 0.82)
ρ_R	0.6	(0.58, 0.60)	(0.54, 0.56)	(0.45, 1.00)
$100\sigma_z$	0.45	0.45	0.47	(0.31, 0.67)
$100\sigma_g$	0.8	0.80	0.77	(0.70, 0.84)
$100\sigma_R$	0.2	(0.19, 0.20)	(0.20, 0.21)	(0.18, 0.23)

In the model, I impose $\psi_\pi > 1$ and $\psi_y > 0$ to guarantee the Blanchard-Kahn condition holds. The true values of the set-identified parameters do not always fall in the range of the posterior mean, because of the finite-sample estimation error, just like the true values of the point-identified parameters do not always equal the posterior mean. The "Identified set" column in Table 6 is computed by finding the observationally equivalent set of the true values using the method of Kocięcki and Kolasa (2023). The result shows that the range of values these non-identified parameters can take is a proper subset of the support. In fact, the identified set can be parametrized by only one free variable changing continuously within an interval (see the appendix for more details).

4.2.3 A Cost-push Shock Model

Just like sensitivity in estimates does not always cause sensitivity in impulse responses, one might want to know whether and when optimal policies can be affected by sensitivity of estimates. In the model of An and Schorfheide (2007), the three shocks either have no impact on the output gap or inflation, or they shift them in the same direction. That feature, which is called a divine coincidence (Blanchard and Galí, 2007; Galí, 2015), makes policy analysis a trivial problem, because policy-makers will maximize the response of the interest rate to dampen the effect of shocks, and this maximization fully stabilizes both the inflation rate and the output gap at the same time. In other words, there is no trade-off between stabilizing inflation and the output gap (Alves, 2014). However, in practice, most central banks still perceive this trade-off. To address this issue, the literature extends the standard New Keynesian model with additional frictions that allow the gap between efficient output and output under flexible prices to vary over time (Erceg et al., 2000; Woodford, 2003a; Benigno and Woodford, 2005; Ravenna and Walsh, 2006; Blanchard and Galí, 2007). In this section, I assume this gap is exogenous, and add to the Phillips curve a cost-push shock to capture the gap and generate opposite dynamics for inflation and output (see, for example, Clarida et al. (1999); Galí (2002); Woodford (2003a); Blanchard and Galí (2007)). To keep the number of shocks unchanged, I drop the total-factor productivity shock z_t in An and Schorfheide (2007):

$$
\begin{aligned}
y_t =& \mathbb{E}_t\left[y_{t+1}\right] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t\left[\pi_{t+1}\right]\right) + g_t - \mathbb{E}_t\left[g_{t+1}\right] \\
\pi_t =& \beta \mathbb{E}_t\left[\pi_{t+1}\right] + \kappa\left(y_t - g_t\right) + u_t \\
i_t =& \rho_R i_{t-1} + \left(1-\rho_R\right)\psi_\pi \pi_t + \left(1-\rho_R\right)\psi_y\left(y_t - g_t\right) + \varepsilon_{R,t} \\
u_t =& \rho_u u_{t-1} + \varepsilon_{u,t} \\
g_t =& \rho_g g_{t-1} + \varepsilon_{g,t}.
\end{aligned}
\tag{19}
$$

Here the estimated parameters are $\theta = (\sigma, \beta, \kappa, \psi_\pi, \psi_y, \rho_R, \rho_g, \rho_u, \sigma_R, \sigma_g, \sigma_u)$. Moreover, u_t is the cost-push shock. A positive shock in u_t would increase the concurrent inflation rate and decrease the output gap. The state variable vector is $S_t = \left(u_t, g_t, i_t\right)'$, and the vector of policy variables is $P_t = \left(\pi_t, y_t\right)'$. As in the previous example, $(\psi_\pi, \psi_y, \rho_R, \sigma_R)$ are not identified. Using Algorithm 5, I can again attain a range of posterior means for these parameters. For each given parameter combination within the range, the welfare losses experienced by a representative household are second-order approximated, proportional to

$$E_0\left\{\sum_{t=0}^{\infty}\beta^t\left(\lambda\pi_t^2+y_t^2\right)\right\}, \tag{20}$$

where $\lambda = \frac{1}{\nu\kappa}$.[28] However, as discussed in the previous examples, the structural parameter ν in the semi-structural parameter κ cannot be identified, which makes the weight on π_t in the objective function agnostic. Therefore, the exercises I perform next are for multiple weight choices.[29] The central banks can pursue either a policy characterized by a period-by-period optimization to minimize $\lambda\pi_t^2 + y_t^2$ or a state-contingent sequence of $\{y_t, \pi_t\}$ that minimizes expression (20) directly. Whereas the former policy, called the optimal policy under discretion, does not need a central bank to commit itself to any future actions, the latter requires the central banks to be able to commit with full credibility to a policy plan (Taylor, 1993; Woodford, 2001; Taylor, 2007).

This exercise is also related to policy analysis in DSGE models under parameter uncertainty (see, for example, Wieland (2000); Kimura and Kurozumi (2007); Edge et al. (2010)), but under the Bayesian framework.

First, I pick a prior, using the same hyperparameters as used in An and Schorfheide (2007), except I substitute ρ_g for ρ_u, and σ_g for σ_u. *DYNARE* returns 10,000 draws from the posterior after a 40,000 burn-in period. Then, I pick another posterior that generates the same posterior predictive distribution using a similar strategy as in the Cochrane model. I divide four pairs of policies (ψ_π, ψ_y) into two groups[30] and select the better policy under different weights and posterior distributions. The results are in Table 7.

Beyond what is already shown in Table 2, Table 7 shows that choices between polarized policies are more robust to the choice of priors. That is to say, when the alternative policies are polarized, researchers will have to hold a polarized prior belief on structural parameters (but still within the same prior class) to disagree with each other's policy choices.

[28]The computation details can be found in Woodford (2003a); Galí (2015); Davig (2016).
[29]The weight choices are scattered to cover most calibration choices and the rule of thumb choice $\frac{1}{\nu\kappa} = 1$.
[30]The policies I compare are from Galí (2015) Table 4.1.

Table 7: Policy Comparison under Different Distributions and Weights

	$\frac{1}{\nu\kappa}=\frac{1}{3}$		$\frac{1}{\nu\kappa}=1$		$\frac{1}{\nu\kappa}=3$		$\frac{1}{\nu\kappa}=10$		$\frac{1}{\nu\kappa}=30$	
(ψ_π, ψ_y)	post 1	post 2	post 1	post 2	post 1	post 2	post 1	post 2	post 1	post 2
(1.5, 0)					✓		✓	✓	✓	✓
(1.5, 0.125)	✓	✓	✓	✓		✓				
(1.5, 1)	✓	✓	✓	✓						
(5, 0)					✓	✓	✓	✓	✓	✓

Optimal discretionary policy parameters are therefore chosen by solving the following optimization:

$$\begin{aligned} &\arg\min_{\psi_\pi,\psi_y} \int L_W\left(\pi_t, y_t \mid \theta_{-(\psi_\pi,\psi_y)}, \psi_\pi, \psi_y\right) d\pi_{\theta|Y}, \\ &\text{subject to Equation (19)} \end{aligned} \tag{21}$$

where $L_W(\pi_t, y_t \mid \theta_{-(\psi_\pi,\psi_y)}, \psi_\pi, \psi_y) = \mathbb{E}_{t-1}\left(\frac{1}{\nu\kappa}\pi_t^2 + y_t^2\right)$ denotes the conditional expected loss of welfare under given θ except the choice of ψ_π and ψ_y.

Table 8: Optimal Policy under Different Distributions and Weights

	$\frac{1}{\nu\kappa}=\frac{1}{3}$		$\frac{1}{\nu\kappa}=1$		$\frac{1}{\nu\kappa}=3$		$\frac{1}{\nu\kappa}=10$		$\frac{1}{\nu\kappa}=30$	
optimal policy	post. 1	post. 2	post. 1	post. 2	post. 1	post. 2	post. 1	post. 2	post. 1	post. 2
ψ_π	2.35	10.39	1.36	4.78	1.76	4.56	3.90	6.17	9.21	14.85
ψ_y	3.48	26.7	0.23	3.38	0.00	0.85	0.00	0.02	0.00	0.00
$10^4\times$loss	2.49	2.50	3.13	3.20	4.17	4.30	5.81	5.83	7.51	7.41

The estimation results show that although the welfare loss does not seem to vary too much with the choice of posterior (hence the prior) distribution, the optimal Taylor rule parameters can be susceptible to this choice. This again shows the importance of robust Bayesian estimation and inference. In fact, these policy-related parameters' sensitivity to priors' choices can be even more significant. To see that, I first estimate each parameter's range of posterior means, using Algorithm 5. The results are shown in Table 9. Then I compute the optimal Taylor rule policy parameters using the grid search between the lower and upper bounds of ρ_R and σ_R,[31] reported in Table 10.

[31] Because the range of the grid search is determined by the lower and upper bounds of the elements of a parameter vector, each pair (ρ_R, σ_R) in the grid search might not necessarily correspond to an element of the estimated range of posterior means. However, sampling from the set of posterior means is challenging, as it is an average of subsets of manifolds.

Table 9: Estimated Identified Set of Structural Parameters for Cost-push Shock Model

	True value	Identified Set	Range of Posterior Mean	Robust Bayesian CR
τ	2	2.00	1.73	(1.34, 2.19)
κ	0.15	0.15	0.21	(0.08, 0.40)
ψ_π	1.5	$(1, +\infty)$	$(1.00, +\infty)$	$(1.00, +\infty)$
ψ_y	1	$(0.22, +\infty)$	$(0.20, +\infty)$	$(0.15, +\infty)$
ρ_u	0.65	0.65	0.67	(0.58, 0.75)
ρ_g	0.75	0.75	0.74	(0.67, 0.81)
ρ_R	0.6	(0.49, 1.00)	(0.49, 1.00)	(0.47, 1.00)
$100\sigma_u$	0.45	0.45	0.49	(0.37, 0.61)
$100\sigma_g$	0.8	0.80	0.77	(0.70, 0.83)
$100\sigma_R$	0.2	(0.16, 0.33)	(0.17, 0.34)	(0.15, 0.38)

In Table 9, the point-identified parameters have the same values as their posterior means. This is because for each posterior draw, the only values these parameters can take in the observationally equivalent set are themselves. However, the range of ψ_π is not informative because $\psi_\pi > 1$ and $\psi_y > 0$ have been restricted to guarantee determinacy. On the other hand, the parameters $(\psi_y, \rho_R, \sigma_R)$ (numerically) have an identified set that is a proper subset of their support.

Table 10: Range of Optimal Policy Parameters

weight	$\frac{1}{\nu\kappa} = \frac{1}{3}$	$\frac{1}{\nu\kappa} = 1$	$\frac{1}{\nu\kappa} = 3$	$\frac{1}{\nu\kappa} = 10$	$\frac{1}{\nu\kappa} = 30$
ψ_π	(1.62, 281.32)	(1.10, 190.72)	(1.97, 216.62)	$(4.93, +\infty)$	$(12.72, +\infty)$
ψ_y	(1.81, 631.40)	(0.00, 123.85)	(0.00, 34.74)	$(0.00, +\infty)$	$(0.00, +\infty)$
$10^4\times$loss	(0.75, 2.50)	(1.33, 3.18)	(3.09, 4.13)	(4.54, 12.62)	(5.36, 35.83)

The range of optimal Taylor rule parameters that are consistent with the estimated range of parameter posterior means is also wide. As ρ_R gets closer to 1, the optimal values of ψ_π and ψ_y increase dramatically. This result is almost mechanical because as the auto-correlation of monetary policy increases, the central bank has to increase the value of (ψ_π, ψ_y) to maintain the same reaction strength to inflation and output gap. From Table 10, it can be seen that even if a researcher has a good sense of what the weight should be and has a credible prior for parameters, another prior with the same predictive distribution may result in a completely different optimal policy suggestion. Facing parameter ambiguity, policy-makers can still make recommendations under certain statistical decision criteria (e.g., a min-max rule with respect to some $\epsilon-$ contaminated neighborhood of a given prior; see Berger (2013); Yata (2021); Manski (2021) for examples of recent literature), or additional assumptions. For example, a normalization of the value ρ_R identifies all the parameters.

In this paper, however, I do not attempt to provide a rule to pick the single optimal policy based on robust Bayesian outputs. Rather, I provide a method to report the robust Bayesian output. As Giacomini and Kitagawa (2021) argue, from the output, one can learn what inferential conclusions can be supported by the model-imposed restrictions and the data. Manski (2013) concludes that "everyone concerned with policy making should keep in mind several dangers of policy analysis with incredible certitude". By comparing the output across different sets of identification restrictions, researchers can better understand each set's identification power and choose upon needs. Lastly, it is a valuable tool to separate the information contained in the data from any prior input that is not revised by the data.

4.3 Smets and Wouters (2007)

The Smets-Wouters model have become a modern workhorse and benchmark model for analyzing monetary and fiscal policy in European central banks, and policy institutions in the US as well. Beyond its theoretical contributions, the model demonstrated practical applicability by fitting it to US economic data. This empirical validation showed that DSGE models could be directly applied to real-world economic analysis and forecasting, bridging the gap between theoretical research and policy implementation. In the original paper, they estimate a fully specified, medium-sized new Keynesian model with many frictions and rigidities, and extracting a full set of implied shocks from those estimates. To compare my identification result with Kocięcki and Kolasa (2023) in the 'global' sense, I adopted their setup in my application. In contrast to the initial approach in Smets and Wouters (2007), where the output gap is viewed as the disparity between actual output and its potential in the absence of nominal rigidities and markup shocks, Kocięcki and Kolasa (2023) considers it as the deviation of output from its deterministic trend. The transformation can be seen in Table 11. Given $\gamma = 100(\bar{\gamma}-1)$, we can directly back out the identified structural parameters λ from α_2, σ_c from α_4, β from α_5, δ from α_7. This further allows identification of ϕ_w from α_3, ψ from α_6, ι_p from α_{11} and ι_w from α_{14}. Notice that parameters within the pairs (ξ_p, ε_p) and (ξ_w, ε_w) enter the model through α_{12} and α_{16} respectively, and are identified jointly rather than being identifiable individually. Therefore, I fix ε_p and ε_w in the estimation procedure.

Table 11 shows that all the semi-structural parameters are point identified, which in turn means all structural parameters are point identified if ε_p and ε_w are fixed. The identified set is therefore a singleton equal to the posterior mean. The robust Bayesian credible region is the usual Bayesian credible region in this case.

Table 11: Estimated Identified Set of (Semi-)structural Parameters for the Smets-Wouters Model

	True value	Identified set	(Robust) Bayesian CR
Transformed Structural Parameters			
$\alpha_1 = \frac{(\gamma-1+\delta)\alpha}{\beta^{-1}\gamma^{\sigma_c}-1+\delta}$	0.17	0.17	[0.16,0.18]
$\alpha_2 = \frac{\lambda\gamma^{-1}}{1+\lambda\gamma^{-1}}$	0.41	0.41	[0.41,0.42]
$\alpha_3 = \frac{(1-\alpha)(\sigma_c-1)}{\phi_w\sigma_c(1+\lambda\gamma^{-1})(1-\alpha_1-g_y)}$	0.13	0.13	[0.12,0.14]
$\alpha_4 = \frac{1-\lambda\gamma^{-1}}{(1+\lambda\gamma^{-1})\sigma_c}$	0.12	0.13	[0.13,0.13]
$\alpha_5 = \frac{1}{1+\beta\gamma^{1-\sigma_c}}$	0.50	0.50	[0.50,0.50]
$\alpha_6 = \frac{1}{(1+\beta\gamma^{1-\sigma_c})\varphi\gamma^2}$	0.09	0.09	[0.08,0.10]
$\alpha_7 = \beta\gamma^{-\sigma_c}(1-\delta)$	0.97	0.97	[0.97,0.97]
$\alpha_8 = (1-\delta)\gamma^{-1}$	0.97	0.97	[0.97,0.97]
$\alpha_9 = (1-\alpha_8)(1+\beta\gamma^{1-\sigma_c})\varphi\gamma^2$	0.29	0.31	[0.28,0.34]
$\alpha_{10} = \frac{\iota_p}{1+\beta\gamma^{1-\sigma_c}\iota_p}$	0.19	0.16	[0.14,0.19]
$\alpha_{11} = \frac{\beta\gamma^{1-\sigma_c}}{1+\beta\gamma^{1-\sigma_c}\iota_p}$	0.80	0.83	[0.81,0.86]
$\alpha_{12} = \frac{(1-\beta\gamma^{1-\sigma_c}\xi_p)(1-\xi_p)}{(1+\beta\gamma^{1-\sigma_c}\iota_p)\xi_p[(\phi_p-1)\varepsilon_p+1]}$	0.02	0.02	[0.02,0.02]
$\alpha_{13} = \frac{1}{1-\lambda\gamma^{-1}}$	3.41	3.37	[3.29,3.44]
$\alpha_{14} = \frac{\iota_w}{1+\beta\gamma^{1-\sigma_c}}$	0.29	0.29	[0.27,0.31]
$\alpha_{15} = \frac{1+\beta\gamma^{1-\sigma_c}\iota_w}{1+\beta\gamma^{1-\delta_c}}$	0.79	0.79	[0.77,0.81]
$\alpha_{16} = \frac{(1-\beta\gamma^{1-\sigma_c}\xi_w)(1-\xi_w)}{(1+\beta\gamma^{1-\sigma_c})\xi_w[(\phi_w-1)\varepsilon_w+1]}$	0.00	0.01	[0.01,0.01]
Untransformed Structural Parameters			
$100(\bar{\gamma}-1)$	0.43	0.43	[0.43,0.43]
$\bar{\pi}$	0.70	0.75	[0.69, 0.80]

(Continued next page)

	True value	Identified set	(Robust) Bayesian CR
$\bar{l}$	0	-0.26	[-0.61, 0.11]
g_y	0.18	0.17	[0.14,0.19]
ϕ_p	1.60	1.57	[1.53,1.60]
α	0.19	0.19	[0.19,0.20]
ψ	0.54	0.53	[0.51,0.54]
σ_l	1.83	1.99	[1.64,2.39]
ρ	0.81	0.80	[0.79,0.80]
r_π	2.04	1.97	[1.91,2.03]
r_y	0.08	0.07	[0.07,0.08]
$r_{\Delta y}$	0.22	0.22	[0.21,0.23]
Parameters of Shock Processes			
ρ_a	0.95	0.95	[0.95,0.95]
ρ_b	0.22	0.23	[0.20,0.25]
ρ_g	0.97	0.97	[0.97,0.97]
ρ_i	0.71	0.70	[0.69,0.71]
ρ_r	0.15	0.15	[0.15,0.16]
ρ_p	0.89	0.87	[0.85,0.89]
ρ_w	0.96	0.96	[0.96,0.97]
ρ_{ga}	0.52	0.52	[0.49,0.54]
μ_p	0.69	0.64	[0.59,0.69]
μ_w	0.84	0.85	[0.83,0.86]
σ_a	0.45	0.45	[0.44,0.46]
σ_b	0.23	0.21	[0.20,0.22]
σ_g	0.53	0.52	[0.51,0.53]
σ_i	0.45	0.45	[0.44,0.46]
σ_r	0.24	0.24	[0.23,0.24]
σ_p	0.14	0.14	[0.14,0.15]
σ_w	0.24	0.24	[0.23,0.25]

4.4 Discussions

4.4.1 Diagnostic Tools

Although the steps used in Algorithm 5 to find the observationally equivalent parameters can be, in theory, arbitrarily accurate in solving the models and the polynomial systems, it is always good to have a handy tool to verify that they are indeed equivalent. While the inverse mapping from the spectral density to structural parameters is impossible to achieve, the mapping from θ to spectral densities is more approachable. One thing researchers can do is to check if the spectral densities generated from the equivalent class $K(\theta_0)$ from Algorithm 5 are actually the same. The state-space model, together with Gaussian assumptions,[32] allows reducing the cost of computing the likelihood function from $O(T^3)$ to $O(T \log T)$ in each evaluation using Whittle's approximation method (Whittle, 1951, 1953; Pawitan and O'sullivan, 1994), also called the frequency domain quasi-maximum likelihood in other publications (Qu and Tkachenko, 2012, 2017). See details of (penalized) Whittle's likelihood approximation in the appendix.

The following algorithm is an application of such method to verify that the $K(\theta)$ obtained from Algorithm 5 is an equivalent set.

Algorithm 6 (Sanity Check)

(1) For each $K(\theta_j)$ obtained from Algorithm 5, pick n draws $\theta_1,\ldots,\theta_n$.

(2) Compute the Whittle likelihood function

$$L_T(f) = \frac{1}{T}\sum_{k=[-T/2]+1}^{[T/2]}\Big[\log\det(f_\theta(\omega_k)) + tr(f_\theta^{-1}(\omega_k))I(\omega_k)\Big],$$

where $f_\theta(\omega_k)$ is the spectral density evaluated at $\omega_k = 2\pi\frac{k}{T}$; $I(\omega_k)$ is the periodogram

$$I(\omega_k) = w(\omega_k)w(\omega_k)^*, \quad w(\omega_k) = \frac{1}{\sqrt{2\pi T}}\sum_t Y_t \exp(-i\omega_k t)$$

evaluated at the same point ω_k.

(3) Compare the likelihood between draws from step 1; if the difference is smaller than some tolerance level ε, admit the achieved $K(\theta)$.

[32] In general, if the researcher allows for non-Gaussian processes, the Whittle likelihood can be understood as a quasi-likelihood of the data based on the asymptotic distribution of the discrete Fourier transforms of the data.

While Algorithm 6 checks the validity of the identified set only from model structure and independently of realized data, sometimes stylized facts, data unused in estimation, or prior knowledge of the researchers allows them to further narrow these identified sets a posteriori, if not achieve point-identification.

There is a vast literature on the applications of this type of knowledge. For example, to identify the effect of macroeconomic shocks, researchers can use contemporaneous restrictions (Blanchard and Perotti, 2002), narrative methods (Friedman and Schwartz, 2008), proxy SVAR (Stock and Watson, 2008), long-run restrictions (Shapiro and Watson, 1988), sign restrictions (Enders et al., 2021), and factor-augmented VARs (Bernanke et al., 2005).[33]

To be more specific, take example 1, and let Y_t and ε_t be scalars. Assume further the true coefficient $D_0 = 1$. Even in this simplest case D is not identified, $D\,|_{K(D_0)} = \pm 1$. However, if in addition the sign of the impact of ε_t on Y_t is known to be positive, point-identification can be achieved. Moreover, these kind of restrictions can be imposed only a posteriori on θ when the model structure becomes complicated.

4.4.2 Non-linearity

There has been a growing literature on the estimation of nonlinear DSGE models (Schmitt-Grohé and Uribe, 2004; Fernández-Villaverde and Rubio-Ramírez, 2007; Andreasen, 2011, 2013; Morris, 2014; Ivashchenko, 2014; Herbst and Schorfheide, 2015; Aruoba et al., 2017; Andreasen et al., 2018). The perturbation method was proposed by Schmitt-Grohé and Uribe (2004) to approximate the nonlinear model with higher-order Taylor expansions around the steady state. Although the Taylor expansions are straightforward to compute, they could generate explosive or non-stationary processes. The pruning method was proposed by Kim et al. (2008) to resolve this concern, and has been well adapted since then.

The basic idea of pruning is to eliminate the terms from the policy functions that have higher-order effects than the approximation order.[34] Because higher-order approximation preserves more information from the nonlinear function, one may expect to gain extra identification power in some non-identified models. Mutschler (2015) showed that the models in Kim (2003) and An and Schorfheide (2007), which are known to have the issue of identification in their linearized Gaussian form, are identifiable with a second-order approximation.

[33] See Ramey (2016) for a more detailed but non-exhaustive list of applications.
[34] See Andreasen et al. (2018) for a complete analysis.

Fortunately, however, Algorithm 5 can still be easily extended to accommodate the non-Gaussian innovations, even if they are white noise. Morris (2014) (Chapter 3) showed that under some fairly modest assumptions, the deviations-from-means of the pruned state-space of second-order approximation can be reparametrized to ABCD representation. He also showed that the errors, although they become non-Gaussian, are white noise processes with finite covariance. This reparametrized form satisfies assumptions 1–2 in Komunjer and Ng (2011). In addition, after being "minimalized", the reparametrized form becomes a minimal ABCD system. Theorem 1 can be invoked using similar arguments if equivalences between linear state-space systems with white noises can be established.

4.4.3 Indeterminacy

It has become a well-established fact that linear rational expectation models can have more than one solution under realistic parameter choices (Sims, 2002; Lubik and Schorfheide, 2003; Farmer et al., 2015; Funovits, 2017; Bianchi and Nicolò, 2021). However, identification exercises taking into account indeterminacy remain rare in the literature. Qu and Tkachenko (2017) are the first to propose a framework to check global identification in linearized DSGE models that allow both determinacy and indeterminacy from a frequency domain perspective.

Kocięcki and Kolasa (2023) have shown that Theorem 1 can handle indeterminate parametrization when a sufficient number of expectation errors are redefined as new fundamentals (Farmer et al., 2015). A fixed structure of ABCD representation is needed, i.e., identification analysis needs to be done within the determinate or indeterminate parameter subspace.

One direct result of allowing for indeterminacy is the possible failure of the continuity property. There are examples in the literature of discontinuity of solutions to linear rational expectation models (e.g., (Al-Sadoon and Zwiernik, 2019)). Nevertheless, researchers need only the continuity of solutions around the true value to apply the proposed algorithm in this paper. This continuity property is conjectured to be true (almost surely), but more work is needed to confirm it.

5 Conclusions

The sensitivity of standard Bayesian results in set-identified models is well-known; however, it had not been investigated in DSGE settings. I showed in this paper that not only parameter estimates but also inference based on estimation results, such as impulse response functions and optimal policies, can be sensitive to the choice of priors.

To provide insight into partially identified DSGE models, I developed a new algorithm to find the complete identified set of parameters in linearized DSGE models. Although Bayesian estimation results of partially identified models are sensitive to the choice of priors, the framework proposed in this paper can be used to conduct robust Bayesian inference on the parameters of interest without the need to exhaust all possible priors.

While DSGE models can suffer from other important issues, such as weak identification (Canova and Sala, 2009; Müller, 2012; Guerron-Quintana et al., 2013; Andrews and Mikusheva, 2015; Ho, 2022), I do not address those issues in this paper. Although there always exists a choice to modify the model (e.g., add more shocks) whenever identification fails, the method in this paper is particularly useful if researchers have some confidence in their model setup and want to know the implications for estimates and optimal policies even when the model is not point-identified.

The applications of the algorithm developed here are based on assumptions of linearized models with Gaussian shocks under determinacy, but they cover a wide range of DSGE models, and a numerical extension to non-linearity and indeterminacy is promising. At the same time, since reparametrization of the Gröbner basis becomes expensive to solve when set-identified structural parameters are high-dimensional, better optimization tools with constraints defined by polynomials may significantly reduce the computational burden.

6 Appendix

6.1 Proofs

Proof for Lemma 1:

Under a linearized DSGE model with Gaussian shocks, with stability assumption 1, Y_t is a weakly stationary time series.

Moreover, the expectation of Y_t is 0. Therefore, the distribution of Y_t is fully characterized by its second moments, $\Gamma(j)$, $j = -\infty, \ldots, \infty$. In other words, the vector of second moments is a sufficient statistic for θ.

Since $\Phi_Y(z;\theta)$ is a z-transform of second moments Y_t, which is a one-to-one mapping between the second moments $\Gamma(j)$ and $\Phi_Y(z)$ (Hannan (2009), p. 46), $\Phi_Y(z)$ must also be a sufficient statistic. Intuitively, it guarantees the same information contained in $\Phi_Y(z;\theta)$ and the likelihood $p(y \mid \theta)$. $\square$

Proof for Lemma 2: Note $\Pi_\theta(\pi_K) = \left\{\pi_\theta : \pi_\theta\left(\left\{\theta : K(\theta) \in B\right\}\right) = \pi_K(B), \text{ for } B \in \mathcal{B}(\mathcal{F})\right\}$. For any $\pi_\theta, \bar{\pi}_\theta \in \Pi_\theta(\pi_K)$,

$$\int_\Theta p(y \mid \theta) d\pi_\theta = \int_{\mathcal{F}} p(y \mid K) d\pi_K = \int_\Theta p(y \mid \theta) d\bar{\pi}_\theta \tag{22}$$

where the equalities comes from change-of variables formula (see for example Stroock (1994); Folland (1999)) and $K_*(\pi_\theta) = K_*(\bar{\pi}_\theta)$. □

Proof for Lemma 3:

The proof of this Lemma is a simple generalization of the same proof in Ke et al. (2022) online appendix C.

Let $p(y \mid \theta)$ be the likelihood of y conditional on θ being the structural parameter. Since the likelihood depends on θ only through $K(\theta)$, we have $\tilde{p}(y \mid K(\theta)) = p(y \mid \theta)$. The data Y updates π_θ to $\pi_{\theta|Y}$ in the following sense (see Ghosal and Van der Vaart (2017) formula 1.1),

$$\pi_{\theta|Y}(A) = \frac{\int_A p(y \mid \theta) d\pi_\theta}{\int p(y \mid \theta) d\pi_\theta}, \quad \text{for any } A \in \mathcal{A}.$$

Plug in this formula

$$\pi^*_{K|Y}(B) = \pi_{\theta|Y}(K^{-1}(B)) = \pi_{\theta|Y}(\{\theta : K(\theta) \in B\}) = \frac{\int_{\{\theta:K(\theta)\in B\}} p(y \mid \theta) \mathrm{d}\pi_\theta}{\int p(y \mid \theta) \mathrm{d}\pi_\theta}.$$

And this is equal to

$$\pi_{K|Y}(B) = \frac{\int_B p(y \mid K) \mathrm{d}\pi_K}{\int p(y \mid K) \mathrm{d}\pi_K} = \frac{\int_{\{\theta:K(\theta)\in B\}} \tilde{p}(y \mid K(\theta)) \mathrm{d}\pi_\theta}{\int \tilde{p}(y \mid K(\theta)) \mathrm{d}\pi_\theta} = \frac{\int_{\{\theta:K(\theta)\in B\}} p(y \mid \theta) \mathrm{d}\pi_\theta}{\int p(y \mid \theta) \mathrm{d}\pi_\theta},$$

where the second equality comes from change of variable (see for example Lemma 5.0.1 from Stroock (1994)). □

Lemma 4 *Given $\Pi_\theta(\pi_K)$ and its corresponding measurable function $K : \Theta \to \mathcal{F}$, there is a unique pair $(\Pi_{\theta|K}, \pi_K)$ up to a measure zero set such that for any $\pi_\theta \in \Pi_\theta(\pi_K)$, there exists a $\pi_{\theta|K} \in \Pi_{\theta|K}$ and π_K such that*

$$\pi_\theta = \int_{\mathcal{F}} \pi_{\theta|K} d\pi_K, \quad \pi_\theta\left(\left\{\theta : K(\theta) \in \cdot\right\}\right) = \pi_K(\cdot), \tag{23}$$

and, conversely, Π_θ *is uniquely determined by* $(\Pi_{\theta|K}, \pi_K)$.

Proof for Lemma 4:

($\Rightarrow$) For any $\pi_\theta \in \Pi_\theta$ let π_K be defined by 10. From disintegration theorem we know there exists a regular conditional probability as a function $\kappa : \mathcal{F} \times \mathcal{A} \to [0,1]$, i.e., a Markov kernel, such that

1. For every $F \in \mathcal{F}$, $\kappa(F, \cdot)$ is a probability measure on $\mathcal{A}$.
2. For all $A \in \mathcal{A}$, $\kappa(\cdot, A)$ is $\mathcal{B}(\mathcal{F})$-measurable.
3. For all $A \in \mathcal{A}$ and $B \in \mathcal{B}(\mathcal{F})$,

$$\pi_\theta(A \cap K^{-1}(B)) = \int_B \kappa(F, A)\pi_K(\mathrm{d}F)$$

$\kappa(K, A)$ is therefore our desired $\pi_{\theta|K}(A)$.

$\pi_{\theta|K'}\left(\left\{\theta : K(\theta) = K'\right\}\right) = 1$ follows directly from definition bullet 3. Hence, $\pi_{\theta|K} \in \Pi_{\theta|K}$.

($\Leftarrow$) On the other hand, for any selected $\pi_{\theta|K} \in \Pi_{\theta|K}$ and π_K, by the tower rule the conditional distribution $\pi_{\theta|\phi}$ can be constructed as $\pi_\theta = \int_{\mathcal{F}} \pi_{\theta|K} d\pi_K$. □

Proof for Theorem 2: This proof, while taking into account the differences in topological structure, mirrors the approach taken in Theorem 1 of Giacomini and Kitagawa (2021), which can be divided into four steps.

First, under Assumption 3, $K(\theta)$ is a random closed set induced by some probability measure on $(\Theta, \mathcal{A})$. Since $\eta : \Theta \to \mathcal{H}$ is a continuous function, for any closed set D, $\{\theta : \eta(K(\theta)) \cap D \neq \emptyset\} = \{\theta : K(\theta) \cap \eta^{-1}(D) \neq \emptyset\} \in \mathcal{A}$ by Effros-measurability of K and the fact that $\eta^{-1}(D)$ is a closed subset of Θ. Therefore $\eta(K(\theta))$ is also a random closed set.

In the second step, I show that for any $\pi_{\theta|K} \in \Pi_{\theta|K}$ and $A \in \mathcal{A}$,

$$\mathbf{1}_{\{K \subset A\}} \leq \pi_{\theta|K}(A), \quad \pi_K - a.s. \tag{24}$$

Note that the set $\{\theta : K(\theta) \subset A\} = \{\theta : K(\theta) \cap A^c \neq \emptyset\}^c$ is also measurable. Denote $K_1^A = \{K \in K(\Theta) : K \subset A\}$. Showing (24) is equivalent to showing

$$\int_B \mathbf{1}_{K_1^A} d\pi_K \leq \int_B \pi_{\theta|K}(A) d\pi_K \tag{25}$$

for every $\pi_{\theta|K} \in \Pi_{\theta|K}$ and $B \in \mathcal{B}(\mathcal{F})$. It then goes

$$\begin{aligned}
\int_B \pi_{\theta|K}(A) d\pi_K &\geq \int_{B\cap K_1^A} \pi_{\theta|K}(A) d\pi_K \\
&= \pi_\theta\left(A \cap \left\{\theta : K(\theta) \in B, K(\theta) \subset A\right\}\right) \\
&= \pi_\theta\left(\left\{\theta : K(\theta) \in B, K(\theta) \subset A\right\}\right) \\
&= \pi_K\left(B \cap K_1^A\right) \\
&= \int_B \mathbf{1}_{K_1^A} d\pi_K
\end{aligned}$$

where the first equality comes from the definition of the conditional distribution.

In the third step, I show that, for each $A \in \mathcal{A}$, there exists $\underline{\pi}_{\theta|K}^A \in \Pi_{\theta|K}$ that achieves the lower bound of $\pi_{\theta|K}$ obtained in (24), π_K almost surely. Consider the following three subsets of $\mathcal{F}$,

$$K_0^A = \left\{K \in K(\Theta) : K \cap A = \emptyset\right\}, \quad K_1^A = \left\{K \in K(\Theta) : K \subset A\right\}, \quad K_2^A = \left\{K \in K(\Theta) : K \cap A \neq \emptyset \text{ and } K \cap A^c \neq \emptyset\right\}$$

They are measurable, mutually disjoint and forms a partition of $K(\Theta)$. Next, I will construct a $\Theta-$valued selection $S^A(K)$ on K_2^A, and the conditional probability distribution $\underline{\pi}_{\theta|K}^A$ that achieves the lower bound given $S^A(K)$ and an arbitrary conditional distribution $\pi_{\theta|K}^A \in \Pi_{\theta|K}$.

Let $S^A(K) = \arg\max_{\theta \in A^\epsilon \cap K} d(\theta, A)$, where $d(\theta, A) = \inf_{\theta' \in A} \parallel \theta - \theta' \parallel$, $A^\epsilon = \{\theta : d(\theta, A) \leq \epsilon\}$. Note that $S^A(K) \in A^c$ by construction if defined on a nonempty K_2^A. Theorem 2.27, as presented in Chapter 1 of Molchanov (2005), establishes that $S^A(K)$ is a random variable. Given $S^A(K)$, and an arbitrary $\pi_{\theta|K}^A \in \Pi_{\theta|K}$, for any $\tilde{A} \in \mathcal{A}$

$$\underline{\pi}_{\theta|K}^A(\tilde{A}) = \begin{cases} \pi_{\theta|K}^A(\tilde{A}), & \text{for } K \in K_0^A \cup K_1^A, \\ \mathbf{1}_{\{S^A(K) \in \tilde{A}\}}(K), & \text{for } K \in K_2^A. \end{cases} \tag{26}$$

It can be verified that $\underline{\pi}_{\theta|K}^A$ is a probability measure in $(\Theta, \mathcal{A})$ and $\underline{\pi}_{\theta|K}^A \in \Pi_{\theta|K}$. For any $B \in \mathcal{B}(\mathcal{F})$, it then follows

$$\begin{aligned}
\int_B \underline{\pi}_{\theta|K}^A(A) d\pi_K &= \int_B \underline{\pi}_{\theta|K}^A(A \cap K) d\pi_K \\
&= \int_{B\cap K_0^A} \underline{\pi}_{\theta|K}^A(A \cap K) d\pi_K + \int_{B\cap K_1^A} \underline{\pi}_{\theta|K}^A(A \cap K) d\pi_K + \int_{B\cap K_2^A} \underline{\pi}_{\theta|K}^A(A \cap K) d\pi_K
\end{aligned}$$

$$= 0 + \int_{B\cap K_1^A} \underline{\pi}^A_{\theta|K}(A\cap K)d\pi_K + 0$$
$$= \int_B \mathbf{1}_{K_1^A} d\pi_K,$$

where the first equality is from $\underline{\pi}^A_{\theta|K} \in \Pi_{\theta|K}$, the third equality follows because $A \cap K = \varnothing$ for $K \in K_0^A$, and by construction $\underline{\pi}^A_{\theta|K}(A\cap K) = \mathbf{1}_{\{S^A(K)\in A\cap K\}}(K) = 0$ for $K \in K_2^A$. Given that this equality applies for all $B \in \mathcal{B}(\mathcal{F})$, it follows that $\underline{\pi}^A_{\theta|K}(A) = \mathbf{1}_{K_1^A}$, and the lower bound is always attainable.

We are now in position to show Theorem 2. We first show the special case of $\eta(\theta) = \theta$. Expanding the integral we have

$$\inf_{\pi_{\theta|Y}\in\Pi_{\theta|Y}} \pi_{\theta|Y}(A) = \inf_{\pi_{\theta|K}\in\Pi_{\theta|K}} \int_{\mathcal{F}} \pi_{\theta|K}(A)\, d\pi_{K|Y}, \quad A \in \mathcal{A}$$

The lower bound of $\pi_{\theta|Y}(A)$ is minimized over the class $\Pi_{\theta|Y}$ by plugging in the attainable pointwise lower bound of $\pi_{\theta|K}$, which is $\mathbf{1}_{\{K\subset A\}}$. Therefore,

$$\begin{aligned}
\inf_{\pi_{\theta|Y}\in\Pi_{\theta|Y}} \pi_{\theta|Y}(A) &= \inf_{\pi_{\theta|K}\in\Pi_{\theta|K}} \int_{\mathcal{F}} \pi_{\theta|K}(A)\, d\pi_{K|Y} \\
&= \inf_{\pi_{\theta|K}\in\Pi_{\theta|K}} \int_{\mathcal{F}} \mathbf{1}_{\{K\subset A\}}(K) d\pi_{K|Y} \\
&= \pi_{K|Y}(\{K \subset A\}) \\
&= \pi_{\theta|Y}\left(\left\{\theta : K(\theta) \subset A\right\}\right)
\end{aligned}$$

The last equality comes from equation (10).

The expression of the posterior upper probability follows directly from its conjugacy with the lower probability.

$$\begin{aligned}
\overline{\pi}_{\theta|Y}(A) &= 1 - \underline{\pi}_{\theta|Y}(A^c) \\
&= 1 - \pi_{\theta|Y}\left(\left\{\theta : K(\theta) \subset A^c\right\}\right) \\
&= 1 - \pi_{\theta|Y}\left(\left\{\theta : K(\theta) \cap A = \varnothing\right\}\right) \\
&= \pi_{\theta|Y}\left(\left\{\theta : K(\theta) \cap A \neq \varnothing\right\}\right)
\end{aligned}$$

To show $\left\{\pi_{\theta|Y}(A) : \pi_{\eta|Y} \in \Pi_{\theta|Y}\right\}$ is a connected interval, we use similar construction as the second step, for any $\tilde{A} \in \mathcal{A}$,

$$\overline{\pi}^{A}_{\theta|K}(\tilde{A}) = \begin{cases} \pi^{A}_{\theta|K}(\tilde{A}), & \text{for } K \in K^{A}_{0} \cup K^{A}_{1}, \\ \mathbf{1}_{\{S^{A^c}(K)\in\tilde{A}\}}(K), & \text{for } K \in K^{A}_{2}. \end{cases} \tag{27}$$

Consider a mixture of these two conditional priors, $\pi^{\lambda}_{\theta|K} \equiv \lambda\underline{\pi}^{A}_{\theta|K} + (1-\lambda)\overline{\pi}^{A}_{\theta|K}$. Note that $\pi^{\lambda}_{\theta|K} \in \Pi_{\eta|Y}$ for any $\lambda \in [0,1]$. Since λ can be chosen arbitrarily, $\left\{\pi_{\theta|Y}(A) : \pi_{\theta|Y} \in \Pi_{\eta|Y}\right\}$ is connected.

Now, for more general forms of η, the same argument follows, by replacing set A above by $\eta^{-1}(D)$, which is also measurable with respect to $\mathcal{A}$.

$$\begin{aligned}
\inf_{\pi_{\eta|Y}\in\Pi_{\eta|Y}} \pi_{\eta|Y}(D) &= \inf_{\pi_{\theta|Y}\in\Pi_{\theta|Y}} \pi_{\theta|Y}\left(\eta^{-1}(D)\right) \\
&= \inf_{\pi_{\theta|K}\in\Pi_{\theta|K}} \int_{\mathcal{F}} \mathbf{1}_{\{K\subset\eta^{-1}(D)\}}(K)\, d\pi_{K|Y} \\
&= \pi_{K|Y}\left(\left\{K \subset \eta^{-1}(D)\right\}\right) \\
&= \pi_{\theta|Y}\left(\left\{\theta : \eta(K(\theta)) \subset D\right\}\right)
\end{aligned}$$

□

Proof for Theorem 3:

It is enough to show the equality for the prior class for the supremum case, i.e., $\sup_{\pi_\theta\in\Pi_\theta} \mathbb{E}_{\pi_\theta}\left[\eta(\theta)\right] = \mathbb{E}_{\pi_\theta}\left[\overline{\eta}^{*}(\theta)\right]$. Equality in posterior follows immediately from Lemma 3. Infimum case is similar.

Pick any two priors $\pi_\theta, \tilde{\pi}_\theta \in \Pi_\theta$,

$$\mathbb{E}_{\tilde{\pi}_\theta}\left[\eta(\theta)\right] \leq \mathbb{E}_{\tilde{\pi}_\theta}\left[\overline{\eta}^{*}(\theta)\right] = \mathbb{E}_{\pi_\theta}\left[\overline{\eta}^{*}(\theta)\right]$$

where the second equality follows from the definition of $\overline{\eta}^{*}(\theta)$ and $\pi_\theta, \tilde{\pi}_\theta$ have the property that they induce the same π_K.

To show the reverse inequality, choose any $\epsilon > 0$ and from Lemma 4 we can define

$$\tilde{\Pi}_\theta = \left\{ \tilde{\pi}_\theta \in \Pi_\theta : \mathbb{E}_{\tilde{\pi}_{\theta|K}}\left[\eta(\theta)\right] \geq \sup_{\theta'\in K} \eta(\theta) - \epsilon, \pi_K - \text{ almost surely} \right\}$$

and this set is nonempty. Then, for any $\tilde{\pi}_\theta \in \tilde{\Pi}_\theta$,

$$\mathbb{E}_{\tilde{\pi}_\theta}\left[\eta(\theta)\right] = \mathbb{E}_{\pi_K}\left[\mathbb{E}_{\tilde{\pi}_{\theta|K}}\left(\eta(\theta)\right)\right] \geq \mathbb{E}_{\pi_K}\left[\sup_{\theta' \in K} \eta(\theta) - \epsilon\right] = \mathbb{E}_{\pi_\theta}\left[\overline{\eta}^*(\theta)\right] - \epsilon$$

Let ϵ decrease to 0. The proof can also be done using random set theory, see Theorem 2.18 in Molchanov and Molinari (2018). □

Proof for Theorem 4: From Theorem 2,

$$\begin{aligned}\inf_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}\left((-\infty, q]\right) &= \pi_{\theta|Y}\left(\left\{\theta : \eta\left(K(\theta)\right) \subset (-\infty, q]\right\}\right) \\ &= \pi_{\theta|Y}\left(\left\{\theta : \overline{\eta}^*(\theta) \leq q\right\}\right)\end{aligned}$$

whereas

$$\begin{aligned}\sup_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}\left((-\infty, q]\right) &= \pi_{\theta|Y}\left(\left\{\theta : \eta\left(K(\theta)\right) \cap (-\infty, q] \neq \varnothing\right\}\right) \\ &= \pi_{\theta|Y}\left(\left\{\theta : \underline{\eta}^*(\theta) \leq q\right\}\right)\end{aligned}$$

□

6.2 Supplementary Tables and Graphs

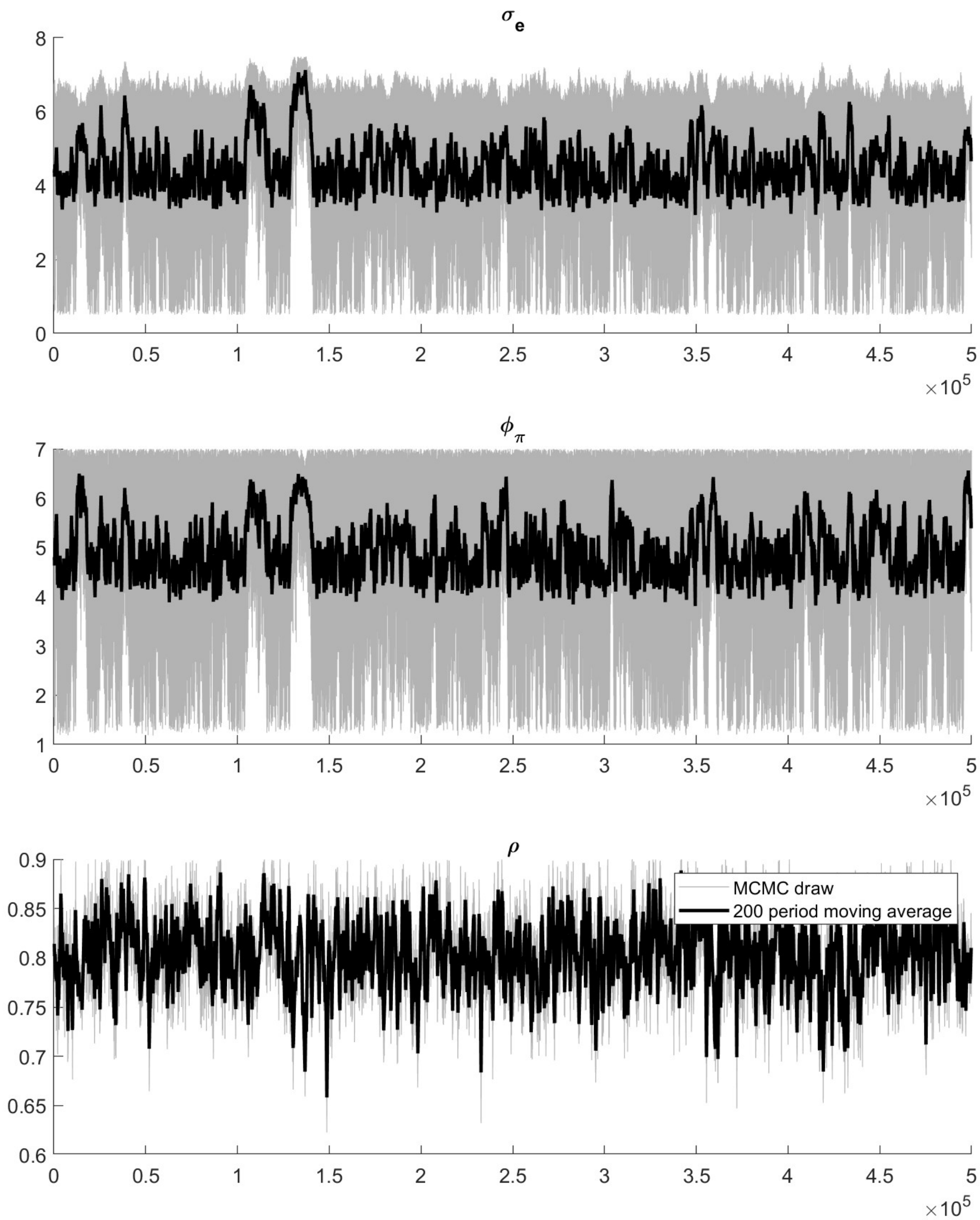


Figure 5: Trace plot of MCMC draws parameters

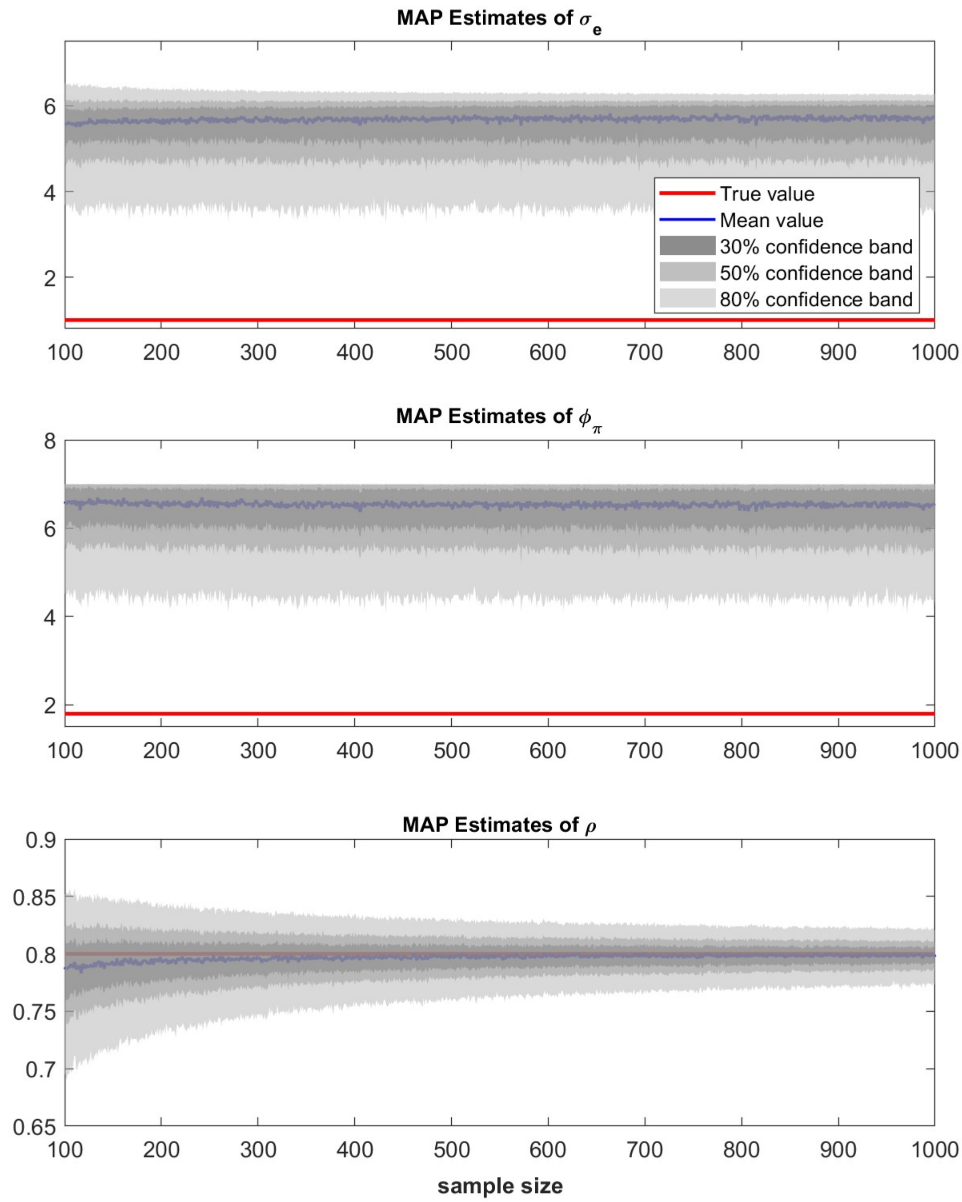


Figure 6: Estimates of structural parameters with 1000 replications

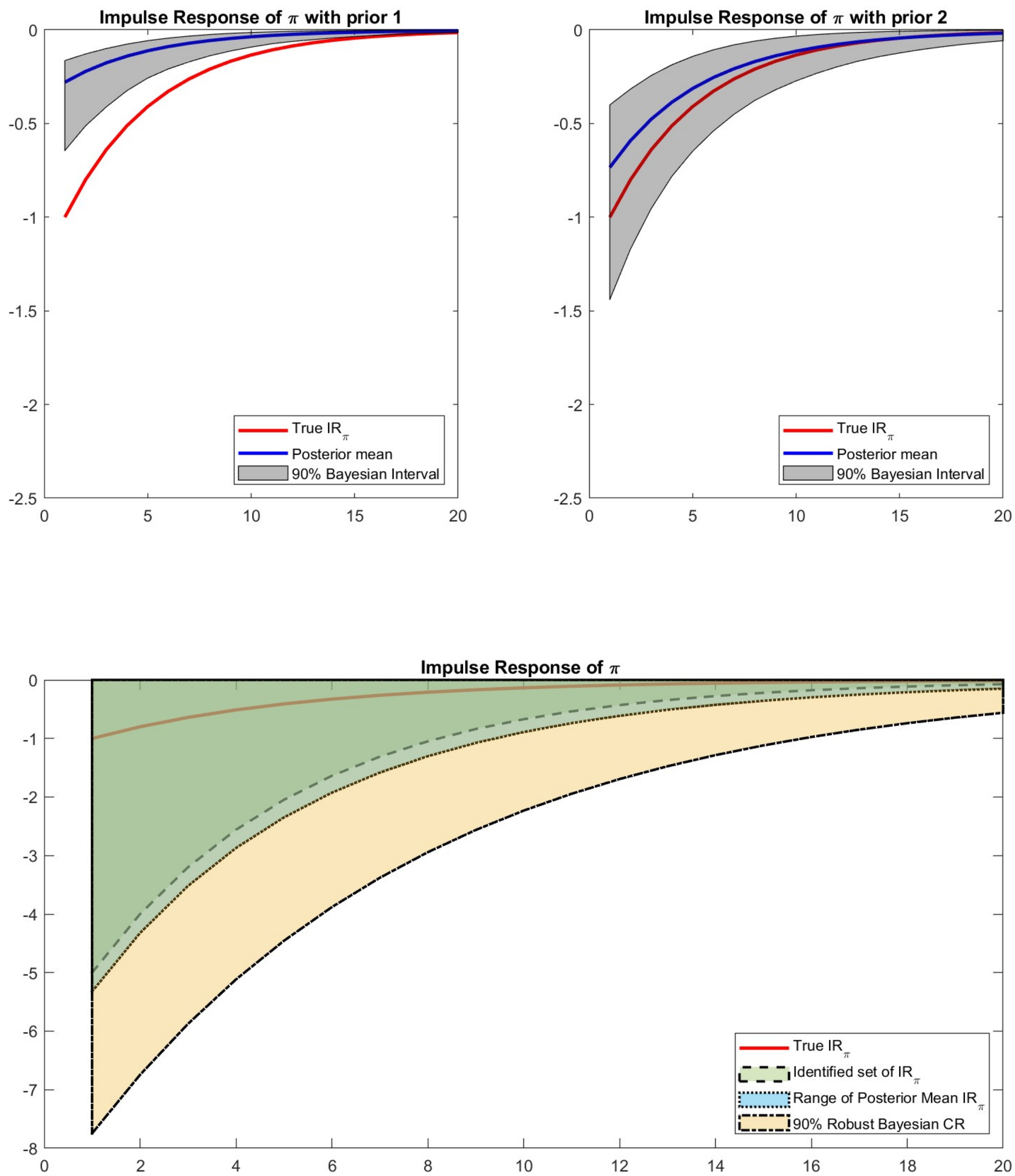


Figure 7: IRF in the Cochrane model

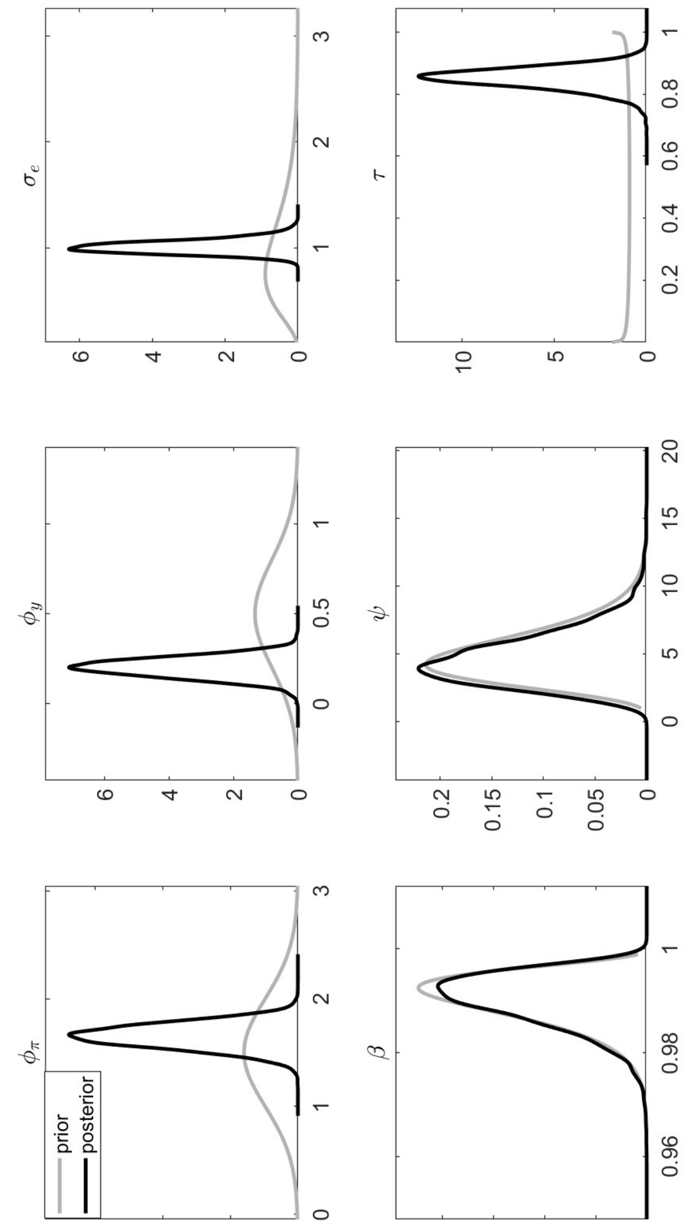


Figure 8: Prior and posterior for the 3-equation model

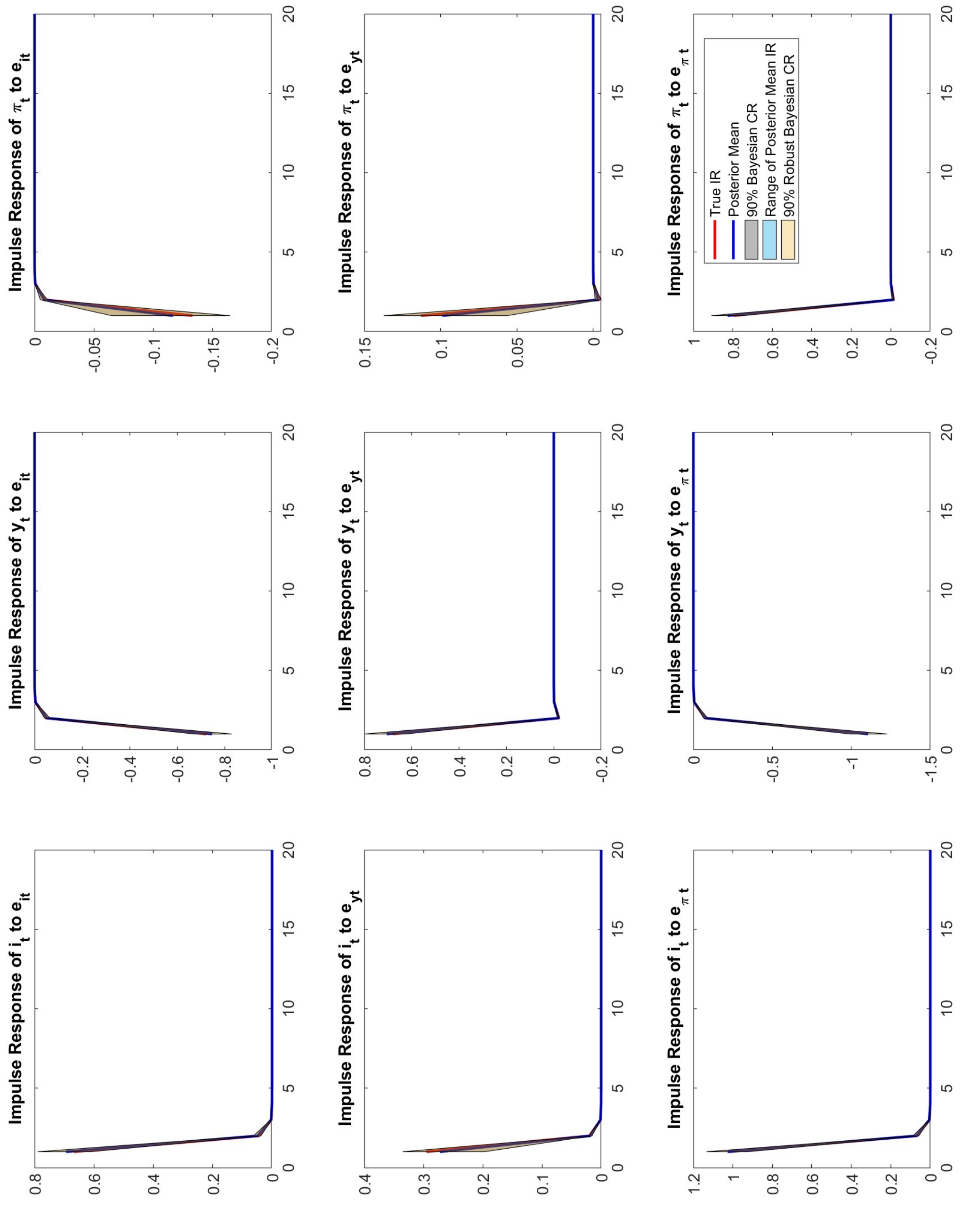


Figure 9: IRF in the three-equation NK model

Table 12: AS Model Prior and Posterior Distribution of Structural Parameters

	True value	Prior distribution			Posterior distribution			
		Distr.	Mean	St. Dev.	Mode	Mean	5 percent	95 percent
τ	2	Gamma	2	0.5	1.79	1.97	1.29	2.61
κ	0.15	Gamma	0.2	0.1	0.13	0.15	0.10	0.20
ϕ_π	1.5	Gamma	1.5	0.25	1.52	1.59	1.16	1.99
ϕ_y	1	Gamma	0.5	0.25	0.67	0.76	0.35	1.11
ρ_z	0.65	Beta	0.66	0.15	0.64	0.63	0.55	0.71
ρ_g	0.75	Beta	0.8	0.1	0.74	0.74	0.66	0.82
ρ_R	0.6	Beta	0.5	0.2	0.56	0.56	0.47	0.63
$100\sigma_z$	0.45	Inv Gamma	0.5	4	0.43	0.46	0.30	0.64
$100\sigma_g$	0.8	Inv Gamma	1	4	0.76	0.77	0.70	0.83
$100\sigma_R$	0.2	Inv Gamma	0.4	4	0.20	0.20	0.18	0.23

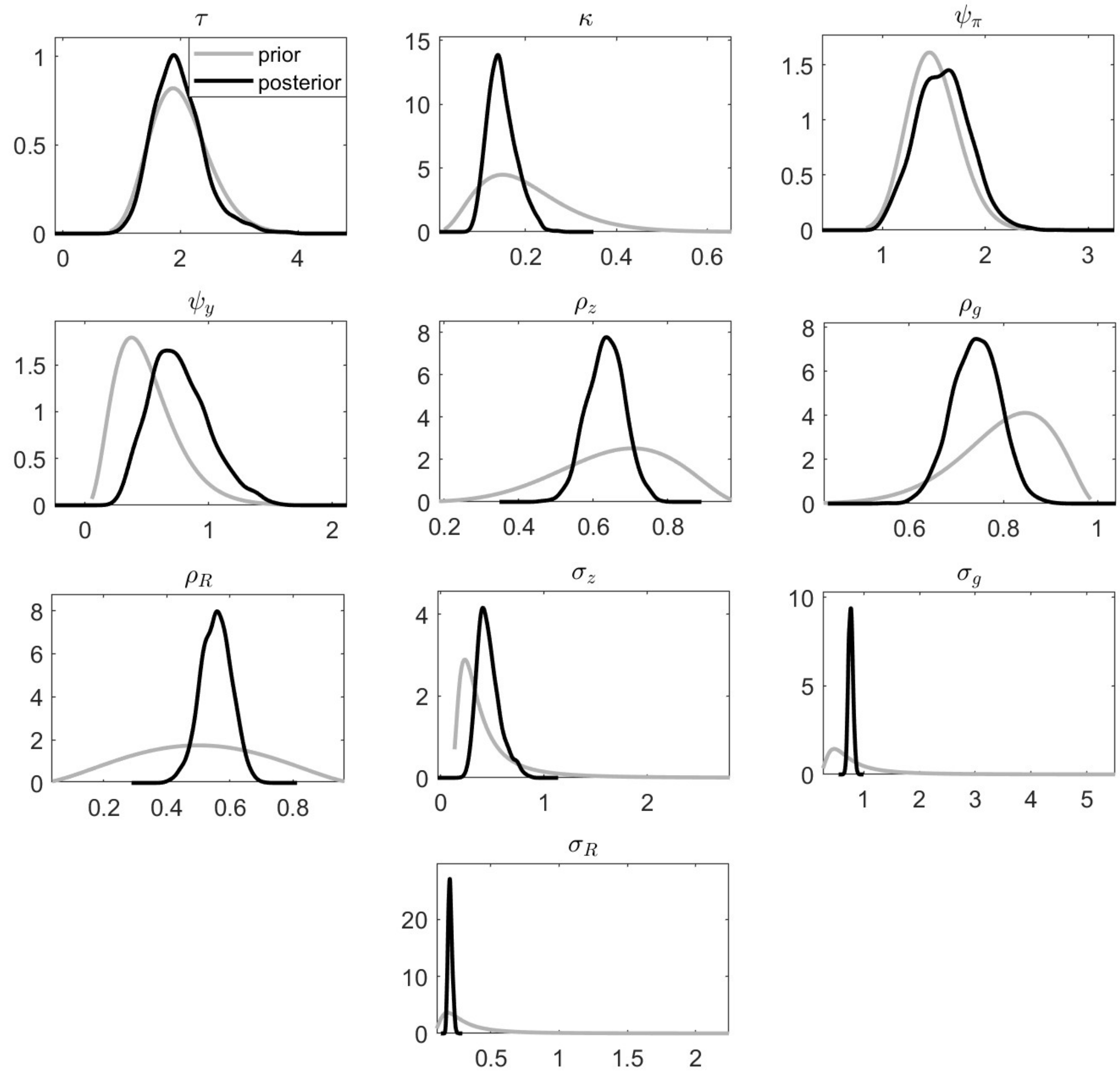


Figure 10: Prior and posterior for the AS model

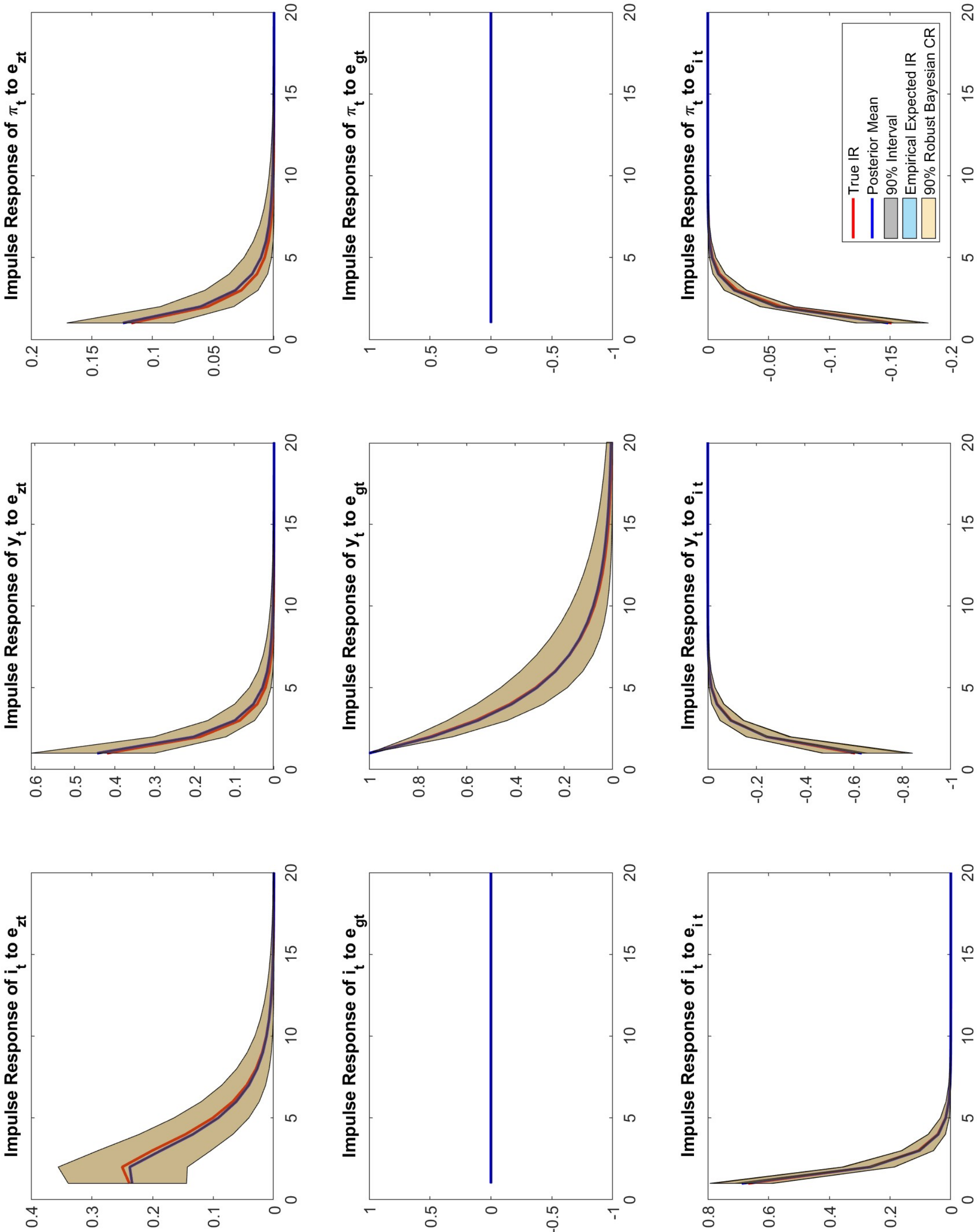


Figure 11: IRF in the AS model

Table 13: Cost-push Model Prior and Posterior Distribution of Structural Parameters

	True value	Prior distribution			Posterior distribution			
		Distr.	Mean	St.Dev.	Mode	Mean	5 percen	95 percent
τ	2	Gamma	2	0.5	1.68	1.73	1.31	2.12
κ	0.15	Gamma	0.2	0.1	0.16	0.21	0.06	0.37
ψ_π	1.5	Gamma	1.5	0.25	1.27	1.34	1.05	1.58
ψ_y	1	Gamma	0.5	0.25	0.56	0.67	0.30	1.00
ρ_u	0.65	Beta	0.66	0.15	0.67	0.67	0.58	0.75
ρ_g	0.75	Beta	0.8	0.1	0.74	0.74	0.67	0.82
ρ_R	0.6	Beta	0.5	0.2	0.56	0.57	0.51	0.62
$100\sigma_u$	0.45	Inv Gamma	0.5	4	0.45	0.48	0.36	0.59
$100\sigma_g$	0.8	Inv Gamma	1	4	0.75	0.76	0.70	0.83
$100\sigma_R$	0.20	Inv Gamma	0.4	4	0.19	0.20	0.17	0.22

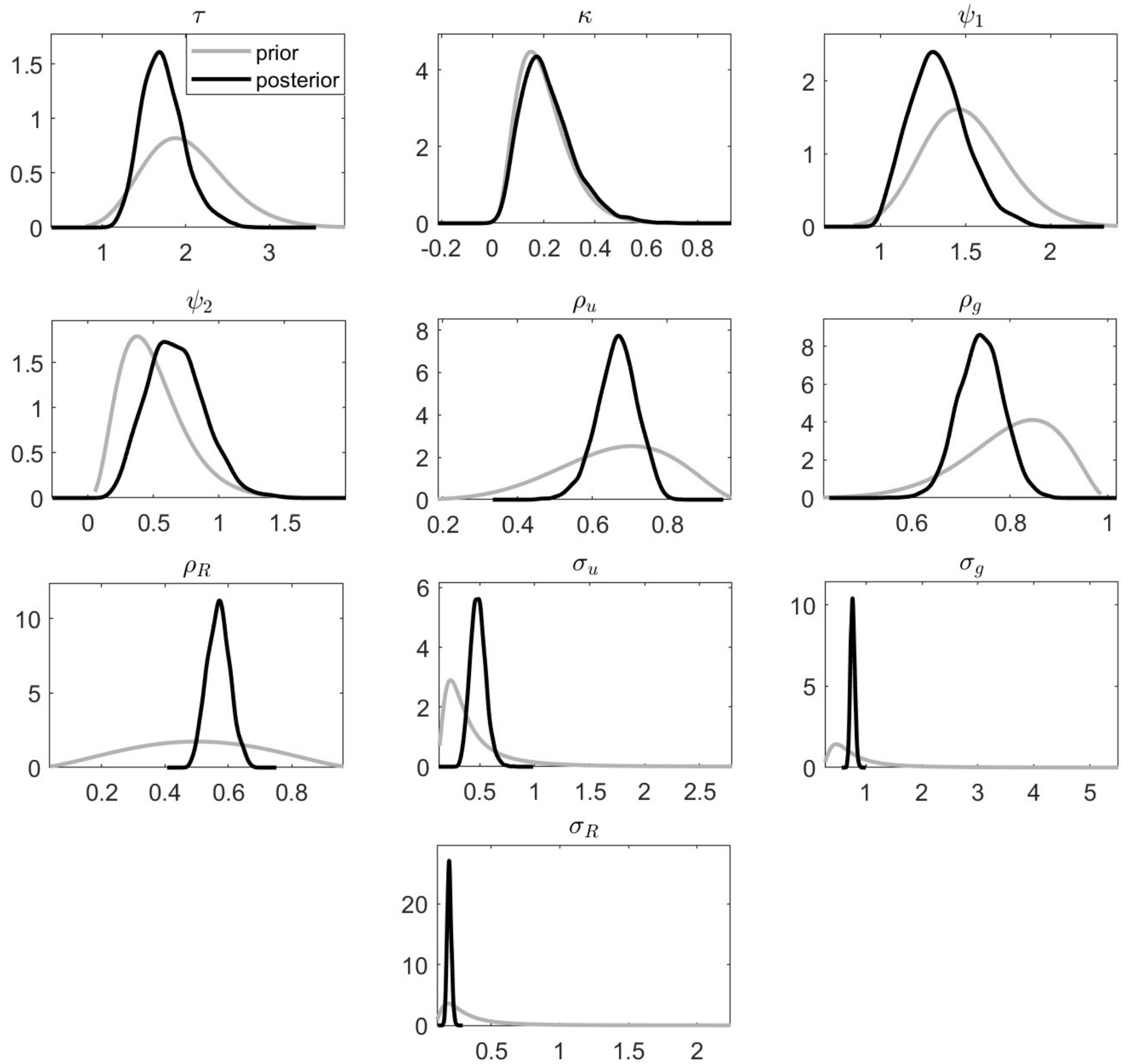


Figure 12: Prior and posterior for the cost-push shock model

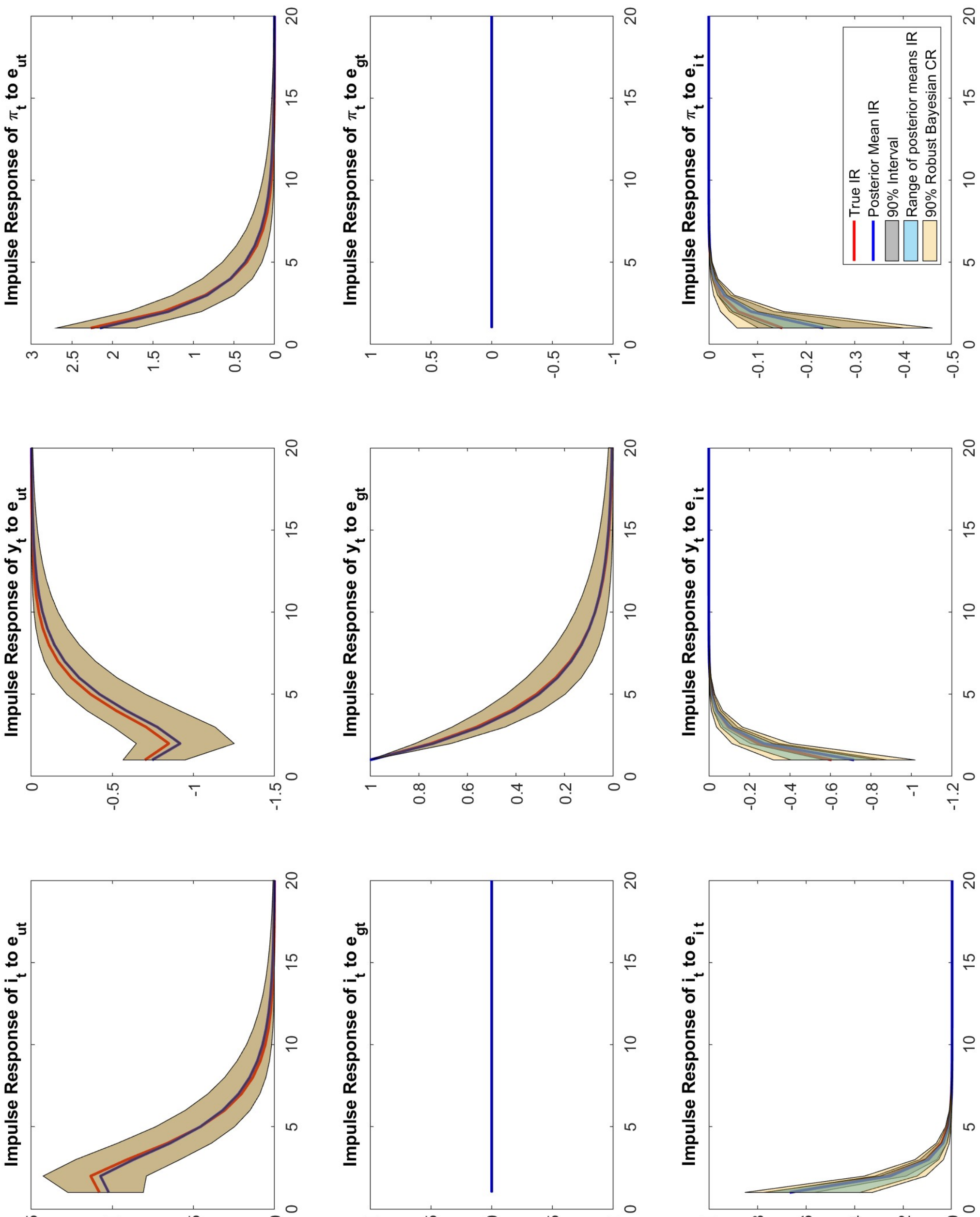


Figure 13: IRF in the Cost-push model

6.3 Application Details

The main steps has been introduced in Algorithm 5. In this section I will go into more detail about how each model is treated differently while solving symbolic Gröbner basis terms and computing the bounds.

6.3.1 Cochrane Model

1. In *Dynare* model write-up, write down the model, calibrate the parameter values and then run *stoch_simul* to simulate a 200 period dataset with variables x_t, π_t, r_t.

2. Check Identification at prior mean using *Dynare* embeded methods.

3. Estimate the model with a prior such that the calibrated value (i.e, the true value) is given nontrivial density, using the only observable π_t. Save the M posterior draws of θ excluding the burn-in period.

4. In *SINGULAR* script, inputs all the draws of θ and output for each θ the Gröbner basis reduced from 9. Ring is set to be rational field. Most of the polynomials are in terms of transformation matrices T, Q and state-space parameters and redundant. A typical result from one draw is presented here. Note given a θ, the computed solution and reduction below are exact, not an approximation.

$$
\begin{aligned}
0 =& 1306669315376553981827097\!6 \cdot Q^2 + 95877633354222629606266441 \cdot Q \\
0 =& 10116030142747 \cdot T \cdot Q + 4186785337110 \cdot Q \\
0 =& G \cdot Q \\
0 =& 18759706557158656509172154380576 3 \cdot G \cdot T + 12459679923307361666013197809920 \cdot Q \\
& + 9142363789675221734790733348334 5 \\
0 =& 10116030142747 \cdot \bar{F} \cdot T + 4186785337110 \\
0 =& D - G \\
0 =& 91423637896752217347907333483345 \cdot T^{-1} \cdot \bar{B} + 187597065571586565091721543805763 \cdot G \\
& - 30104838802210930638489585120384 \cdot Q \\
0 =& 1853999779646306024554490449467702854272471867461696 \cdot \bar{\sigma}_e \\
& - 9811548045114026857241967637688790821764961999743969 \cdot T^2
\end{aligned}
$$

$$
\begin{aligned}
&+ 88530008575628738450385530629619625869205349612800 \cdot Q \\
0 =& 4929950657305 \cdot \bar{\phi}_\pi - 10116030142747 \cdot T - 4186785337110 \\
0 =& 5170129 \cdot \bar{\rho} - 4390758
\end{aligned}
$$

5. Although look like a lot, they are very simple polynomials of scalar variables in the Cochrane case. Impose symbolically the upper lower bounds of parameters. Reparametrize in *MATLAB* the M Gröbner basis to expressions of free parameters by symbolic function *solve* with *ReturnConditions= true* and keeping only the structural parameters. The Gröbner basis above will become

$$
\begin{aligned}
\bar{\sigma}_e =& (1221027570018689850472910384549454312224 \cdot x^2)/ \\
& 834270390849036833577588029975794152 37 \\
& - (160204484624566949989110628032952841 40 \cdot x)/ \\
& 641746454499259102751990792289072424 9 \\
& + 1050978598846472222807755026384334575/ \\
& 987302237691167850387678141983188346 \\
\bar{\phi}_\pi =& x \\
\bar{\rho} =& 515645/604628 \\
parameters :& \ x \\
conditions :& \ 1 < x \quad \text{and} \quad x \neq 515645/604628
\end{aligned}
$$

6. It is obvious ρ is identified in this case. Translate the returned symbolic conditions of $\bar{\sigma}_e$ and $\bar{\phi}_\pi$ to optimization constraints. Set optimization function, and run constrained optimization M times, store the upper and lower bounds of parameters of interest.

7. Take average of the bounds for a robust Bayesian posterior mean, quantiles for robust Bayesian credible region.

6.3.2 Three-equation NK Model

1. The first two steps are the same as estimation in Cochrane model, except with a different model setup.

2. Estimate using i_t, π_t and y_t as observables. Store both the posterior draws of θ and the state-space parameters computed in *Dynare* from algorithm *gensys* based on Sims (2002) corresponding to each θ. These parameters are used as initial values for *SINGULAR* to solve.

3. In *SINGULAR* script, define the first ring to be real numbers with 500 digits precision. Take the parameters θ^j and state-space values $(A^j, B^j, C^j, D^j, \Sigma^j)$, solve for $(A(\theta^j), B(\theta^j), C(\theta^j), D(\theta^j), \Sigma(\theta^j))$ that is accurate up to its 500 digits, which should be very close to the initial value$(A^j, B^j, C^j, D^j, \Sigma^j)$.

4. Define a second ring in *SINGULAR* that is accurate up to 10 digits. Solve again for the Gröbner basis. Solutions from our first MCMC draw result looks like below.

$$0 = Q_{1,1}$$
$$0 = T - 1$$
$$0 = G_{2,3}^2 - 0.75155924653409373544$$
$$0 = G_{2,2} - 0.07951992942099192501 2$$
$$0 = G_{2,1} + 0.08791825703800079447 1$$
$$0 = G_{1,3} + 1.33748131551708435491 \cdot G_{2,3}$$
$$0 = G_{1,2} - 0.76809465922214837765$$
$$0 = G_{1,1} + 0.76604181621551468463$$
$$0 = (\bar{F} \cdot T)_{2,1} + 0.00879182570380007944 71$$
$$0 = (\bar{F} \cdot T)_{1,1} + 0.07660418162155146846 3$$
$$0 = D_{3,3} - G_{2,3}$$
$$0 = D_{3,2} - 0.07951992942099192501 2$$
$$0 = D_{3,1} + 0.08791825703800079447 1$$
$$0 = D_{2,3} + 1.33748131551708435491 \cdot G_{2,3}$$
$$0 = D_{2,2} - 0.76809465922214837765$$
$$0 = D_{2,1} + 0.76604181621551468463$$
$$0 = D_{1,3} - 1.32358664224738610 36 \cdot G_{2,3}$$
$$0 = D_{1,2} - 0.22949614903646230548$$
$$0 = D_{1,1} - 0.75808364840879226454$$
$$0 = (T^{-1} \cdot \bar{B})_{1,3} - 1.32358664224738610 36 \cdot G_{2,3}$$

$$0 = (T^{-1} \cdot \bar{B})_{1,2} - 0.22949614903646230548$$

$$0 = (T^{-1} \cdot \bar{B})_{1,1} - 0.75808364840879226454$$

$$0 = \bar{\tau} \cdot G_{2,3} - 0.86692516778214123747 \cdot \bar{\tau}$$

$$0 = \bar{\psi} \cdot G_{2,3} - 0.86692516778214123747 \cdot \bar{\psi} + 1.0802 \cdot G_{2,3} - 0.93645256623826896472$$

$$0 = \bar{\psi} \cdot \bar{\tau}^2 - 2.00770897364841033909 \cdot \bar{\psi} \cdot \bar{\tau} + 1.00770897364841033909 \cdot \bar{\psi} + 1.0802 \cdot \bar{\tau}^2$$
$$- 2.27568102557546129216 \cdot \bar{\tau} + 1.08852723333501284829$$

$$0 = \bar{\beta} \cdot \bar{\tau} - 0.99235 \cdot \bar{\tau}$$

$$0 = \bar{\beta} \cdot \bar{\psi} + 1.0802 \cdot \bar{\beta} - 0.99235 \cdot \bar{\psi} - 1.07193647$$

$$0 = \bar{\sigma} - 1.0802$$

$$0 = \bar{\phi}_y - 0.15787$$

$$0 = \bar{\phi}_\pi - 1.6818$$

5. Solve in *MATLAB* using *solve* for a reparametrized equation system. Do that in a two-step flavor. First solve for the uniquely determined variables. Then remove these variables from equations by plug-in their values, and keeping only the non-determinant structural parameters in the solver's equations and solve again. The results looks like

$$\begin{aligned}
\bar{\psi}: &\quad -(1080200 \cdot z^2 - 2275680 \cdot z + 1088530)/(1000000 \cdot z^2 - 2007710 \cdot z + 1007710) \\
\bar{\tau}: &\quad z \\
parameters: &\quad z \\
conditions: &\quad z \neq 1.0 \text{ and } z \neq 1.00771 \text{ and } ((1.00771 < z \text{ and } z < 1.37252) \text{ or } (z < 1.0 \text{ and } 0.734205 < z))
\end{aligned}$$

6. The last two steps are similar to Cochrane model, except now optimize with $\bar{\psi}$, $\bar{\tau}$ and their conditions.

6.3.3 The An&Schorfheide Model

Everything is the same in step 1-3. Except here I used the Theorem 1-S in Komunjer and Ng (2011) for simpler equations. 4. Reduced Gröbner basis from the first posterior draw is,

$$0 = T_{3,3} - 1$$

$$0 = T_{3,2}$$

$$0 = T_{3,1}$$

$$0 = T_{2,3}$$

$$0 = T_{2,2} - 1$$

$$0 = T_{2,1}$$

$$0 = T_{1,3}^2 - 17.25315337 \cdot T_{1,3}$$

$$0 = T_{1,2}$$

$$0 = T_{1,1} + 0.6369912784 \cdot T_{1,3} - 1$$

$$0 = U_{3,3} \cdot T_{1,3}$$

$$0 = U_{3,2}$$

$$0 = U_{3,1}$$

$$0 = U_{2,3}$$

$$0 = U_{2,2} - 1$$

$$0 = U_{2,1}$$

$$0 = U_{1,3} - 0.6984103625 \cdot T_{1,3}$$

$$0 = U_{1,2}$$

$$0 = U_{1,1} + 0.3927596209 \cdot T_{1,3} - 1$$

$$0 = \bar{G}_{2,3} + 0.1525963868$$

$$0 = \bar{G}_{2,2}$$

$$0 = \bar{G}_{2,1} - 0.1409624802$$

$$0 = \bar{G}_{1,3} + 0.8983721938$$

$$0 = \bar{G}_{1,2} - 1$$

$$0 = \bar{G}_{1,1} - 0.6578842935$$

$$0 = (\bar{F} \cdot T)_{2,3} + 0.08132929628$$

$$0 = (\bar{F} \cdot T)_{2,2}$$

$$0 = (\bar{F} \cdot T)_{2,1} - 0.09074741585$$

$$0 = (\bar{F} \cdot T)_{1,3} + 0.4788054281$$

$$0 = (\bar{F} \cdot T)_{1,2} - 0.70811$$

$$0 = (\bar{F} \cdot T)_{1,1} - 0.4235261716$$

$$0 = \bar{\sigma}_R^2 - 0.0405458496 \cdot U_{3,3}^2$$

$$0 = \bar{\sigma}_g^2 - 0.5232930921$$

$$0 = \bar{\sigma}_z^2 - 0.6186927788 \cdot T_{1,3} - 0.1479094681$$

$$0 = \bar{\rho}_R - 0.53297 \cdot U_{3,3}$$

$$0 = \bar{\rho}_g - 0.70811$$

$$0 = \bar{\rho}_z + 0.01573846956 \cdot T_{1,3} - 0.64377$$

$$0 = \bar{\psi}_y \cdot T_{1,3} + 5.171016239 \cdot T_{1,3}$$

$$0 = \bar{\psi}_y \cdot U_{3,3} - 1.876278215 \cdot \bar{\psi}_y + 10.0771457 \cdot U_{3,3} - 9.702265116$$

$$0 = \bar{\psi}_\pi + 4.313937786 \cdot \bar{\psi}_y - 3.558745724$$

$$0 = \bar{\kappa} - 0.10679$$

$$0 = \bar{\tau} + 0.4609655598 \cdot T_{1,3} - 1.3391$$

The parametrized solutions look like

$$\begin{aligned}
\bar{\psi}_\pi :&\quad (59 \cdot (8083902471764131279771571541312 \cdot x - 8341942821160251303963513600493))/\\
&\quad (2251799813685248 \cdot (4503599627370496 \cdot x - 8450005869917379))\\
\bar{\psi}_y :&\quad -(45383429619478200 \cdot x - 43695117561067360)/\\
&\quad (4503599627370496 \cdot x - 8450005869917379)\\
\bar{\rho}_R :&\quad (53297 \cdot x)/100000\\
\bar{\sigma}_R^2 :&\quad (2921636370399683 \cdot x^2)/72057594037927936\\
U_{3,3} :&\quad x\\
parameters :&\quad x\\
conditions :&\quad x \neq 8450005869917379/4503599627370496\\
&\quad \text{and } 1092377939026684/1134585740486955 < x\\
&\quad \text{and } x < 9462938096098712419076960626 0819/9336180820645158205890981905 8880
\end{aligned}$$

6.3.4 The Cost-push Shock Model

The steps and solution forms look very similar to An& Schorfheide case, here's the observational equivalence set for the first posterior draw.

$$
\begin{aligned}
\bar{\psi}_\pi: &\quad -(18567476663545134437002172836354 \cdot x + 2725175536675775611678249887487)/ \\
&(9007199254740992 \cdot (2251799813685248 \cdot x - 3967300010722725)) \\
\bar{\psi}_y: &\quad -(11419659612338326 \cdot x - 8975929578081457)/ \\
&(4503599627370496 \cdot x - 7934600021445450) \\
\bar{\rho}_R: &\quad (56759 \cdot x)/100000 \\
\bar{\sigma}_R^2: &\quad (2859004378036277 \cdot x^2)/72057594037927936 \\
U_{3,3}: &\quad x \\
parameters: &\quad x \\
conditions: &\quad x < 3967300010722725/2251799813685248 \\
&\text{and } x \neq 3967300010722725/2251799813685248 \\
&\text{and } 33009086163239882578714353555713/38849886267196804860949424122370 < x
\end{aligned}
$$

It is worth noting that although the coding I show above is model specific in *SINGULAR* and MATLAB. This algorithm can be in principle, generalized without going to (almost manually) decipher the solutions of reduced Gröbner basis. This unfortunately, will ask more coding techniques than what the author is capable of.

6.4 Methodology

6.4.1 LREMs: Solutions and Indeterminacy

Linear rational expectation models can be seen as a generalization of classical linear systems, where the state of the system depends only on past and present values of the state variables and shocks. The state in LREMs can depend on information available to form expectations about future states (Al-Sadoon, 2018). In other words, LREMs are both backward-looking and forward-looking. The literature regarding how to solve or regularize a LREM is huge. Sims (2002) derived the existence and uniqueness conditions to solve a LREM. This paper is also known for deriving a widely-used algorithm to solve LREMs numerically called *gensys*. Lubik and Schorfheide (2003) characterize the complete set of solutions to LREMs with indeterminacies. While all these papers focus on the numerical side, Al-Sadoon (2018) and Al-Sadoon and Zwiernik (2019) rigorously define the solution space, the solution concept, as well as existence and uniqueness under both linear system approach and the spectral approach. In this paper, the numerical results are based on Sims (2002) using QZ decomposition (generalized Schur decomposition) whereas theoretical steps are derived using results from Al-Sadoon and Zwiernik (2019). Therefore I briefly discuss this numerical method allowing for indeterminacy.

Write the LREM in a canonical form

$$\Gamma_0 S_t = \Gamma_1 S_{t-1} + \Psi \varepsilon_t + \Pi \eta_t$$

A stable solution of this form exists if there exist expectation errors η_t as a function of the exogenous shocks ε_t such that the explosive components of y_t will be offset.

Assume for some ξ,

$$\mathbb{E}_t\left(\xi^{-h} S_{t+h}\right) \overset{h\to\infty}{\longrightarrow} 0, \quad \xi > 1$$

Perform a QZ decomposition (generalized Schur decomposition) [35] to Γ_0 and Γ_1. There exist matrices Q, Z, Λ, and Ω, such that $Q'\Lambda Z' = \Gamma_0, Q'\Omega Z' = \Gamma_1, QQ' = ZZ' = I_{n\times n}$, and both Λ and Ω are upper-triangular. Although the QZ decomposition is not unique, the resulting generalized eigenvalues ω_{ii}/λ_{ii} are, where ω_{ii} and λ_{ii} are diagonal element of Ω and Λ. Let $w_t = Z'y_t$, pre-multiply the canonical form by Q to obtain

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ 0 & \Omega_{22} \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_{1\cdot} \\ Q_{2\cdot} \end{bmatrix} \left(\Psi \varepsilon_t + \Pi \eta_t\right) \tag{28}$$

Where the first set of equations has absolute generalized eigenvalues smaller than threshold ξ and is therefore non-explosive. The second set of equations can be rewritten as

$$w_{2,t} = \Lambda_{22}^{-1} \Omega_{22} w_{2,t-1} + \Lambda_{22}^{-1} Q_{2.} \left(\Psi \varepsilon_t + \Pi \eta_t\right)$$

A stable solution exists if and only if $w_{2,0} = 0$ and the column space of $Q_{2\cdot}\Psi$ is contained in the

[35] See Golub and Van Loan (2013) for a reference

column space of $Q_{2.}\Pi$, i.e.,

$$\text{span}\left(Q_{2.}\Psi\right) \subset \text{span}\left(Q_{2.}\Pi\right), \quad \text{or} \quad Q_{2.}\Psi = Q_{2.}\Pi\lambda \quad \text{for some matrix } \lambda$$

The solution is unique if and only if the row space of $Q_1\Pi$ is contained in the row space of $Q_{2.}\Pi$, i.e.,

$$Q_{1.}\Pi = \Phi Q_{2.}\Pi, \quad \text{for some matrix } \Phi$$

When there are multiple solutions, use SVD decomposition to get rid of the linearly dependent rows,

$$Q_{2.}\Pi = \begin{bmatrix} U_{.1} & U_{.2} \end{bmatrix} \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{.1}^{*} \\ V_{.2}^{*} \end{bmatrix} = U_{.1} D_{11} V_{.1}^{*},$$

where V^* is the conjugate transpose of V. $[U_{.1}\ U_{.2}]$ and $[V_{.1}\ V_{.2}]$ are unitary matrices. Solving for forecast errors η_t is then reduced to solving

$$U_{.1} D_{11} V_{.1}^{*} \eta_t = -Q_{2.}\Psi\varepsilon_t \text{ for all } t > 0$$

The solution presented below are in the the same form as Qu and Tkachenko (2017) result, equivalent to expression (17) in Lubik and Schorfheide (2003), only they use a different decomposition strategy by incorporate the last two terms, which sum up to an element in the null space of $V_{.2}$. The full solution of forecast errors can be characterized by

$$\eta_t = -V_{.1} D_{11}^{-1} U_{.1}^{*} Q_{2.}\Psi\varepsilon_t + V_{.2}\epsilon_t \text{ with } E_{t-1}\epsilon_t = 0.$$

for any ϵ_t that is conformable with $V_{.2}$.

Premultiply (28) by $\begin{bmatrix} I & \Phi \\ 0 & I \end{bmatrix}$ to get

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} - \Phi\Lambda_{22} \\ 0 & I \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} - \Phi\Omega_{22} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_{1.} - \Phi Q_{2.} \\ 0 \end{bmatrix} \left(\Psi\varepsilon_t + \Pi\eta_t\right)$$

Plug in the solutions of η_t, and then preply again by $Z \begin{bmatrix} \Lambda_{11} & \Lambda_{12} - \Phi\Lambda_{22} \\ 0 & I \end{bmatrix}^{-1}$, we end up with our solutions of the form $S_t = \Theta_S S_{t-1} + \Theta_\varepsilon \varepsilon_t + \Theta_\epsilon \epsilon_t$, where $\Theta_1 = Z_{.1}\Lambda_{11}^{-1}[\Omega_{11}\Omega_{12} - \Phi\Omega_{22}]Z^*$, $\Theta_\epsilon =$

$Z_{\cdot 1}\Lambda_{11}^{-1}(Q_{1\cdot} - \Phi Q_{2\cdot})\Psi$, and $\Theta_{\varepsilon} = Z_{\cdot 1}\Lambda_{11}^{-1}Q_{1\cdot}\Pi\left(I - V_{\cdot 1}V_{\cdot 1}^{*}\right)V_{\cdot 2}$.

6.4.2 Whittle's Approximation

To solve the problem of estimation the spectral denstiy and other parameters in times series models, Whittle (1951, 1953) introduced an approximate likelihood function under Gaussian settings. The Whittle likelihood is a frequency-domain approximation to the exact likelihood and is considered a standard method in parametric spectral analysis on account of its $O(T\log T)$ computational efficiency. The one-time computational burden is insignificant with a sample size of three digits. However, with multiple draws of parameters to compare, it is still extremely beneficial to reduce the cost. While bias can exist in the finite sample, when one try to approximate MLE of a continuous times series with discrete observation, or when the data is multivariate, researchers have derived all kinds of variations for the Whittle likelihood (Pawitan and O'sullivan, 1994; Choudhuri et al., 2004; Krafty and Collinge, 2013; Jesus and Chandler, 2017; Sykulski et al., 2019). We only present here the original Whittle likelihood for the sake of compactness.

The likelihood function for the parameters θ conditional on observations of Y_t in a SSM (state-space model) can be represented at

$$L(\theta \mid Y) = P(Y_{1:T} \mid \theta) = \Pi_{t=1}^{T} p(y_t \mid Y_{1:t-1}, \theta)$$

which is not directly computable because of unobserved S_t. We then compute the density $p(y_t \mid Y_{1:t-1}, \theta)$ using

$$\begin{aligned} p\left(s_t \mid Y_{1:t-1}, \theta\right) &= \int p\left(s_t \mid s_{t-1}, Y_{1:t-1}, \theta\right) p\left(s_{t-1} \mid Y_{1:t-1}, \theta\right) d s_{t-1} \\ p\left(y_t \mid Y_{1:t-1}, \theta\right) &= \int p\left(y_t \mid s_t, Y_{1:t-1}, \theta\right) p\left(s_t \mid Y_{1:t-1}, \theta\right) d s_t \end{aligned}$$

where $p\left(s_t \mid s_{t-1}, Y_{1:t-1}, \theta\right)$ and $p\left(y_t \mid s_t, Y_{1:t-1}, \theta\right)$ are directly parametrized by vec(A, B, C, D, Σ). This algorithm, although conceptually easy to understand, will involve inversion of covariance matrices, which largely increases computational cost. An alternative approximation to the log likelihood function $-2\log L(\theta \mid Y)$ is defined as

$$-2\log L(\theta \mid Y) \approx 2n\log 2\pi + \sum_{j=0}^{n-1}\left[\log f_{\theta}\left(\omega_j\right) + I_j / f_{\theta}\left(\omega_j\right)\right]$$

where $f_{\theta}(\omega_j)$ is the spectral density evaluated at $\omega_j = \frac{j}{T}$. I_j is the periodogram formed from the distribution of the Fourier transformed data $I_j = X_j' X_j$, evaluated at the same point ω_k. $X_j = T^{-1/2} \sum_{t=1}^{T} Y_t \exp\left(-2\pi i \omega_j t\right), j = 1, \ldots, J$. Fox and Taqqu (1986) have shown that the $\hat{\theta}_W$ that minimizes the Wittle likelihood, is asymptotically efficient, therefore asymptotically equivalent to MLE in ARFIMA models, indicating the efficiency for ARMA models.

References

Adjemian, Stéphane, Houtan Bastani, Michel Juillard, Ferhat Mihoubi, George Perendia, Marco Ratto, and Sébastien Villemot, "Dynare: Reference manual, version 4," 2011.

Adolfson, Malin, Stefan Laséen, Jesper Lindé, and Lars EO Svensson, "Optimal monetary policy in an operational medium-Sized DSGE Model," *Journal of Money, credit and Banking*, 2011, *43* (7), 1287–1331.

Al-Sadoon, Majid M, "The linear systems approach to linear rational expectations models," *Econometric Theory*, 2018, *34* (3), 628–658.

_ **and Piotr Zwiernik**, "The identification problem for linear rational expectations models," *arXiv preprint arXiv:1908.09617*, 2019.

Alexanderian, Alen, "On continuous dependence of roots of polynomials on coefficients," *The University of Texas at Austin, USA*, 2013.

Aliprantis, Charalambos and Border Kim, *Infinite dimensional analysis: A Hitchhiker's Guide*, Springer, 2006.

Alves, Sergio Afonso Lago, "Lack of divine coincidence in New Keynesian models," *Journal of Monetary Economics*, 2014, *67*, 33–46.

An, Sungbae and Frank Schorfheide, "Bayesian analysis of DSGE models," *Econometric reviews*, 2007, *26* (2-4), 113–172.

Andreasen, Martin M, "Non-linear DSGE models and the optimized central difference particle filter," *Journal of Economic Dynamics and Control*, 2011, *35* (10), 1671–1695.

_ , "Non-linear DSGE models and the central difference Kalman filter," *Journal of Applied Econometrics*, 2013, *28* (6), 929–955.

— , **Jesús Fernández-Villaverde, and Juan F Rubio-Ramírez**, "The pruned state-space system for non-linear DSGE models: Theory and empirical applications," *The Review of Economic Studies*, 2018, *85* (1), 1–49.

Andrews, Donald WK and Gustavo Soares, "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 2010, *78* (1), 119–157.

Andrews, Isaiah and Anna Mikusheva, "Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models," *Quantitative Economics*, 2015, *6* (1), 123–152.

Antsaklis, Panos J and Anthony N Michel, *Linear systems*, Vol. 8, Springer, 1997.

Aruoba, S Borağan, Luigi Bocola, and Frank Schorfheide, "Assessing DSGE model nonlinearities," *Journal of Economic Dynamics and Control*, 2017, *83*, 34–54.

Bacchiocchi, Emanuele and Toru Kitagawa, "Locally-but not globally-identified SVARs," *Quaderni-Working Paper DSE*, 2022.

Barankin, Edward et al., "Sufficient parameters: solution of the minimal dimensionality problem," *Annals of the Institute of Statistical Mathematics*, 1960, *12* (2), 91–118.

Benigno, Pierpaolo and Michael Woodford, "Inflation stabilization and welfare: The case of a distorted steady state," *Journal of the European Economic Association*, 2005, *3* (6), 1185–1236.

Beresteanu, Arie and Francesca Molinari, "Asymptotic properties for a class of partially identified models," *Econometrica*, 2008, *76* (4), 763–814.

— , **Ilya Molchanov, and Francesca Molinari**, "Sharp identification regions in models with convex moment predictions," *Econometrica*, 2011, *79* (6), 1785–1821.

Berger, James O, "Robust Bayesian analysis: sensitivity to the prior," *Journal of statistical planning and inference*, 1990, *25* (3), 303–328.

— , *Statistical decision theory and Bayesian analysis*, Springer Science & Business Media, 2013.

Bernanke, Ben S, Jean Boivin, and Piotr Eliasz, "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach," *The Quarterly journal of economics*, 2005, *120* (1), 387–422.

Bianchi, Francesco and Giovanni Nicolò, "A generalized approach to indeterminacy in linear rational expectations models," *Quantitative Economics*, 2021, *12* (3), 843–868.

Blanchard, Olivier and Jordi Galí, "Real wage rigidities and the New Keynesian model," *Journal of money, credit and banking*, 2007, *39*, 35–65.

__ **and Roberto Perotti**, "An empirical characterization of the dynamic effects of changes in government spending and taxes on output," *the Quarterly Journal of economics*, 2002, *117* (4), 1329–1368.

Blanchard, Olivier Jean and Charles M Kahn, "The solution of linear difference models under rational expectations," *Econometrica: Journal of the Econometric Society*, 1980, pp. 1305–1311.

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

Chen, Xiaohong, Timothy M Christensen, and Elie Tamer, "Monte Carlo confidence sets for identified sets," *Econometrica*, 2018, *86* (6), 1965–2018.

Chernozhukov, Victor, Han Hong, and Elie Tamer, "Estimation and confidence regions for parameter sets in econometric models 1," *Econometrica*, 2007, *75* (5), 1243–1284.

Choudhuri, Nidhan, Subhashis Ghosal, and Anindya Roy, "Contiguity of the Whittle measure for a Gaussian time series," *Biometrika*, 2004, *91* (1), 211–218.

Christiano, Lawrence J, Martin S Eichenbaum, and Mathias Trabandt, "On DSGE models," *Journal of Economic Perspectives*, 2018, *32* (3), 113–40.

__ **, Mathias Trabandt, and Karl Walentin**, "DSGE models for monetary policy analysis," in "Handbook of monetary economics," Vol. 3, Elsevier, 2010, pp. 285–367.

Clarida, Richard, Jordi Gali, and Mark Gertler, "The science of monetary policy: a new Keynesian perspective," *Journal of economic literature*, 1999, *37* (4), 1661–1707.

__ **,** __ **, and** __ , "Monetary policy rules and macroeconomic stability: evidence and some theory," *The Quarterly journal of economics*, 2000, *115* (1), 147–180.

Cochrane, John H, "Determinacy and identification with Taylor rules," *Journal of Political economy*, 2011, *119* (3), 565–615.

Cox, David, John Little, and Donal OShea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, Springer Science & Business Media, 2013.

Davig, Troy, "Phillips curve instability and optimal monetary policy," *Journal of Money, Credit and Banking*, 2016, *48* (1), 233–246.

Dawid, A Philip, "Conditional independence in statistical theory," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1979, *41* (1), 1–15.

Decker, Wolfram, Gert-Martin Greuel, Gerhard Pfister, and Hans Schönemann, "Singular 4-3-0 — A computer algebra system for polynomial computations," http://www.singular.uni-kl.de 2022.

DeRoberts, Lorraine and JA Hartigan, "Bayesian inference using intervals of measures," *The Annals of Statistics*, 1981, pp. 235–244.

Edge, Rochelle M, Thomas Laubach, and John C Williams, "Welfare-maximizing monetary policy under parameter uncertainty," *Journal of Applied Econometrics*, 2010, *25* (1), 129–143.

Enders, Zeno, Michael Kleemann, and Gernot J Müller, "Growth expectations, undue optimism, and short-run fluctuations," *Review of Economics and Statistics*, 2021, *103* (5), 905–921.

Erceg, Christopher J, Dale W Henderson, and Andrew T Levin, "Optimal monetary policy with staggered wage and price contracts," *Journal of monetary Economics*, 2000, *46* (2), 281–313.

Farmer, Roger EA, Vadim Khramov, and Giovanni Nicolò, "Solving and estimating indeterminate DSGE models," *Journal of Economic Dynamics and Control*, 2015, *54*, 17–36.

Fernández-Villaverde, Jesús and Juan F Rubio-Ramírez, "Estimating macroeconomic models: A likelihood approach," *The Review of Economic Studies*, 2007, *74* (4), 1059–1087.

Florens, Jean-Pierre and Anna Simoni, "Revisiting identification concepts in Bayesian analysis," *Annals of Economics and Statistics*, 2021, (144), 1–38.

Folland, Gerald B, *Real analysis: modern techniques and their applications*, Vol. 40, John Wiley & Sons, 1999.

Fox, Robert and Murad S Taqqu, "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *The Annals of Statistics*, 1986, *14* (2), 517–532.

Friedman, Milton and Anna Jacobson Schwartz, *A monetary history of the United States, 1867-1960*, Vol. 16, Princeton University Press, 2008.

Funovits, Bernd, "The full set of solutions of linear rational expectations models," *Economics Letters*, 2017, *161*, 47–51.

Galí, Jordi, "New perspectives on monetary policy, inflation, and the business cycle," 2002.

— , *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press, 2015.

— **and Mark Gertler**, "Inflation dynamics: A structural econometric analysis," *Journal of monetary Economics*, 1999, *44* (2), 195–222.

Geweke, John, "Bayesian model comparison and validation," *American Economic Review*, 2007, *97* (2), 60–64.

— **and Charles Whiteman**, "Bayesian forecasting," *Handbook of economic forecasting*, 2006, *1*, 3–80.

Ghosal, Subhashis and Aad Van der Vaart, *Fundamentals of nonparametric Bayesian inference*, Vol. 44, Cambridge University Press, 2017.

Giacomini, R and T Kitagawa, "Robust Bayesian Inference for Set-Identified Models," *Econometrica*, 2021.

— **, — , and M Read**, "Robust Bayesian Inference in Proxy SVARs," *Journal of Econometrics*, 2022.

Glover, Keith, "Structural aspects of system identification." PhD dissertation, Massachusetts Institute of Technology 1973.

Golub, Gene H and Charles F Van Loan, *Matrix computations*, JHU press, 2013.

Goodfriend, Marvin and Robert G King, "The new neoclassical synthesis and the role of monetary policy," *NBER macroeconomics annual*, 1997, *12*, 231–283.

Guerron-Quintana, Pablo, Atsushi Inoue, and Lutz Kilian, "Frequentist inference in weakly identified dynamic stochastic general equilibrium models," *Quantitative Economics*, 2013, *4* (2), 197–229.

Gustafson, Paul, "What are the limits of posterior distributions arising from nonidentified models, and why should we care?," *Journal of the American Statistical Association*, 2009, *104* (488), 1682–1695.

__ , *Bayesian inference for partially identified models: Exploring the limits of limited data*, Vol. 140, CRC Press, 2015.

Hannan, Edward James, *Multiple time series*, John Wiley & Sons, 2009.

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

Herbst, Edward P and Frank Schorfheide, "Bayesian estimation of DSGE models," in "Bayesian Estimation of DSGE Models," Princeton University Press, 2015.

Ho, Paul, "Global robust Bayesian analysis in large models," *Journal of Econometrics*, 2022.

Horowitz, Joel L and Charles F Manski, "Nonparametric analysis of randomized experiments with missing covariate and outcome data," *Journal of the American statistical Association*, 2000, *95* (449), 77–84.

Hurwicz, Leonid, "Generalization of the concept of identification," *Statistical inference in dynamic economic models*, 1950, *10*, 245–57.

Imbens, Guido W and Charles F Manski, "Confidence intervals for partially identified parameters," *Econometrica*, 2004, *72* (6), 1845–1857.

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

Ivashchenko, Sergey, "DSGE model estimation on the basis of second-order approximation," *Computational Economics*, 2014, *43* (1), 71–82.

__ **and Willi Mutschler**, "The effect of observables, functional specifications, model features and shocks on identification in linearized DSGE models," *Economic Modelling*, 2020, *88*, 280–292.

Jesus, Joao and Richard E Chandler, "Inference with the Whittle likelihood: A tractable approach using estimating functions," *Journal of Time Series Analysis*, 2017, *38* (2), 204–224.

Jordà, Òscar, "Estimation and inference of impulse responses by local projections," *American economic review*, 2005, *95* (1), 161–182.

Justiniano, Alejandro and Bruce Preston, "Monetary policy and uncertainty in an empirical small open-economy model," *Journal of Applied Econometrics*, 2010, *25* (1), 93–128.

Kaido, Hiroaki, Francesca Molinari, and Jörg Stoye, "Confidence intervals for projections of partially identified parameters," *Econometrica*, 2019, *87* (4), 1397–1432.

Ke, Shikun, José Luis Montiel Olea, and James Nesbit, "A robust machine learning algorithm for text analysis," Technical Report, Working paper 2022.

Kim, Jinill, "Functional equivalence between intertemporal and multisectoral investment adjustment costs," *Journal of Economic Dynamics and control*, 2003, *27* (4), 533–549.

__ , Sunghyun Kim, Ernst Schaumburg, and Christopher A Sims, "Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models," *Journal of Economic Dynamics and Control*, 2008, *32* (11), 3397–3414.

Kimura, Takeshi and Takushi Kurozumi, "Optimal monetary policy in a micro-founded model with parameter uncertainty," *Journal of Economic Dynamics and Control*, 2007, *31* (2), 399–431.

King, Robert G, "The new IS-LM model: language, logic, and limits," *FRB Richmond Economic Quarterly*, 2000, *86* (3), 45–104.

Kline, Brendan and Elie Tamer, "Bayesian inference in a class of partially identified models," *Quantitative Economics*, 2016, *7* (2), 329–366.

Kociecki, Andrzej and Marcin Kolasa, "Global identification of linearized DSGE models," *Quantitative Economics*, 2018, *9* (3), 1243–1263.

Kocięcki, Andrzej and Marcin Kolasa, "A solution to the global identification problem in DSGE models," *Journal of Econometrics*, 2023, *236* (2), 105477.

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

— **and Yinchu Zhu**, "Likelihood ratio testing in linear state space models: An application to dynamic stochastic general equilibrium models," *Journal of Econometrics*, 2020, *218* (2), 561–586.

Koop, Gary and Dale J Poirier, "Learning about the across-regime correlation in switching regression models," *Journal of Econometrics*, 1997, *78* (2), 217–227.

— **, M Hashem Pesaran, and Ron P Smith**, "On identification of Bayesian DSGE models," *Journal of Business & Economic Statistics*, 2013, *31* (3), 300–314.

— **,** — **, and Simon M Potter**, "Impulse response analysis in nonlinear multivariate models," *Journal of econometrics*, 1996, *74* (1), 119–147.

Koopmans, Tjalling C, "Identification problems in economic model construction," *Econometrica, Journal of the Econometric Society*, 1949, pp. 125–144.

— **and Olav Reiersol**, "The identification of structural characteristics," *The Annals of Mathematical Statistics*, 1950, *21* (2), 165–181.

Krafty, Robert T and William O Collinge, "Penalized multivariate Whittle likelihood for power spectrum estimation," *Biometrika*, 2013, *100* (2), 447–458.

Liao, Yuan and Anna Simoni, "Bayesian inference for partially identified smooth convex models," *Journal of Econometrics*, 2019, *211* (2), 338–360.

Lindé, Jesper, "DSGE models: still useful in policy analysis?," *Oxford Review of Economic Policy*, 2018, *34* (1-2), 269–286.

Lindley, Dennis Victor, *Bayesian statistics: A review*, SIAM, 1972.

Lindquist, Anders and Giogio Picci, "Linear stochastic systems," *Series in Contemporary Mathematics*, 2015, *1.*

Lubik, Thomas A and Frank Schorfheide, "Computing sunspot equilibria in linear rational expectations models," *Journal of Economic dynamics and control*, 2003, *28* (2), 273–285.

— **and** — , "Testing for indeterminacy: An application to US monetary policy," *American Economic Review*, 2004, *94* (1), 190–217.

Manski, Charles F, *Identification problems in the social sciences*, Harvard University Press, 1995.

— , *Partial identification of probability distributions*, Vol. 5, Springer, 2003.

— , *Public policy in an uncertain world: analysis and decisions*, Harvard University Press, 2013.

— , "Econometrics for decision making: Building foundations sketched by haavelmo and wald," *Econometrica*, 2021, *89* (6), 2827–2853.

Molchanov, Ilya, *Theory of random sets*, Springer-Verlag, 2005.

— **and Francesca Molinari**, *Random sets in econometrics*, Vol. 60, Cambridge University Press, 2018.

Molinari, Francesca, "Microeconometrics with partial identification," *Handbook of econometrics*, 2020, *7*, 355–486.

Moon, Hyungsik Roger and Frank Schorfheide, "Bayesian and frequentist inference in partially identified models," *Econometrica*, 2012, *80* (2), 755–782.

Morris, Stephen David, "Identification of dynamic stochastic general equilibrium models." PhD dissertation, University of California, San Diego 2014.

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

Mutschler, Willi, "Identification of DSGE models—The effect of higher-order approximation and pruning," *Journal of Economic Dynamics and Control*, 2015, *56*, 34–54.

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

— **, Stefano Eusepi, Marc P Giannoni, Argia M Sbordone, Andrea Tambalotti, Matthew Cocci, Raiden Hasegawa, and M Linder**, "The frbny dsge model," *FRB of New York Staff Report*, 2013, (647).

Pawitan, Yudi and Finbarr O'sullivan, "Nonparametric spectral density estimation using penalized Whittle likelihood," *Journal of the American Statistical Association*, 1994, *89* (426), 600–610.

Pericchi, Luis Raúl and Peter Walley, "Robust Bayesian credible intervals and prior ignorance," *International Statistical Review/Revue Internationale de Statistique*, 1991, pp. 1–23.

Picci, Giorgio, "Some connections between the theory of sufficient statistics and the identifiability problem," *SIAM Journal on Applied Mathematics*, 1977, *33* (3), 383–398.

Poirier, Dale J, "Revising beliefs in nonidentified models," *Econometric theory*, 1998, *14* (4), 483–509.

Potter, Simon M, "Nonlinear impulse response functions," *Journal of Economic Dynamics and Control*, 2000, *24* (10), 1425–1446.

Qu, Zhongjun and Denis Tkachenko, "Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models," *Quantitative Economics*, 2012, *3* (1), 95–132.

— **and** — , "Global identification in DSGE models allowing for indeterminacy," *The Review of Economic Studies*, 2017, *84* (3), 1306–1345.

Ramey, Valerie A, "Macroeconomic shocks and their propagation," *Handbook of macroeconomics*, 2016, *2*, 71–162.

Ravenna, Federico and Carl E Walsh, "Optimal monetary policy with the cost channel," *Journal of Monetary Economics*, 2006, *53* (2), 199–216.

Robert, Christian P, George Casella, and George Casella, *Monte Carlo statistical methods*, Vol. 2, Springer, 1999.

Romano, Joseph P and Azeem M Shaikh, "Inference for the identified set in partially identified econometric models," *Econometrica*, 2010, *78* (1), 169–211.

Rothenberg, Thomas J, "Identification in parametric models," *Econometrica: Journal of the Econometric Society*, 1971, pp. 577–591.

Schmitt-Grohé, Stephanie and Martın Uribe, "Solving dynamic general equilibrium models using a second-order approximation to the policy function," *Journal of economic dynamics and control*, 2004, *28* (4), 755–775.

Shapiro, Matthew D and Mark W Watson, "Sources of business cycle fluctuations," *NBER Macroeconomics annual*, 1988, *3*, 111–148.

Sims, Christopher A, "Solving linear rational expectations models," *Computational economics*, 2002, *20* (1-2), 1.

Smets, Frank and Rafael Wouters, "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American economic review*, 2007, *97* (3), 586–606.

Stock, James H and Mark W Watson, "What's New in Econometrics–Time Series, Lecture 7: Structural VARs," *Cambridge, MA.: National Institute for Economic Research. URl: http://www. nber. org/minicourse_2008. html*, 2008.

Stoye, Jörg, "More on confidence intervals for partially identified parameters," *Econometrica*, 2009, *77* (4), 1299–1315.

Stroock, Daniel W, *A concise introduction to the theory of integration*, Birkhäuser Boston, MA, 1994.

Sykulski, Adam M, Sofia C Olhede, Arthur P Guillaumin, Jonathan M Lilly, and Jeffrey J Early, "The debiased Whittle likelihood," *Biometrika*, 2019, *106* (2), 251–266.

Syversveen, Anne Randi, "Noninformative bayesian priors. interpretation and problems with construction and applications," *Preprint statistics*, 1998, *3* (3), 1–11.

Taylor, John B, "Discretion versus policy rules in practice," in "Carnegie-Rochester conference series on public policy," Vol. 39 Elsevier 1993, pp. 195–214.

— , *Monetary policy rules*, Vol. 31, University of Chicago Press, 2007.

Walley, Peter, *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, 1991.

Wasserman, Larry Alan, "A robust Bayesian interpretation of likelihood regions," *The Annals of Statistics*, 1989, *17* (3), 1387–1393.

Weitzman, Martin L, "On modeling and interpreting the economics of catastrophic climate change," *The review of economics and statistics*, 2009, *91* (1), 1–19.

Whittle, Peter, *Hypothesis testing in time series analysis*, Vol. 4, Almqvist & Wiksells boktr., 1951.

— , "Estimation and information in stationary time series," *Arkiv för matematik*, 1953, *2* (5), 423–434.

Wieland, Volker, "Monetary policy, parameter uncertainty and optimal learning," *Journal of Monetary Economics*, 2000, *46* (1), 199–228.

Woodford, Michael, "The Taylor rule and optimal monetary policy," *American Economic Review*, 2001, *91* (2), 232–237.

—, "Interest and prices: Foundations of a Theory of Monetary Policy," 2003.

—, "Optimal interest-rate smoothing," *The Review of Economic Studies*, 2003, *70* (4), 861–886.

Yata, Kohei, "Optimal Decision Rules Under Partial Identification," *arXiv preprint arXiv:2111.04926*, 2021.