Bayesian Sensitivity Analysis for Structural Models*

Yizhou Kuang

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

Abstract

This paper proposes a new algorithm to conduct robust Bayesian analysis for set-identified structural models. It combines standard Bayesian procedure with a characterization of observationally equivalent parameters. The algorithm finds the range of posterior means and the Bayesian credible region of both the structural parameters and any parameters of interest, ensuring robustness against the selection of priors within a class that produces identical marginal likelihoods. I provide theoretical support for this algorithm and apply the method to several monetary policy models, to show its relevance in policy analysis. The methodology finds that, in set-identified models, parameters of primary interest like impulse responses can have very different implications based on the prior, even within the same class. Additionally, optimal monetary policy rules could vary with the choice of prior within that class, particularly when historical policy parameters are not identified.

KEYWORDS: DSGE models, Bayesian inference, set identification, informative priors, policy analysis

*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, Raffaella Giacomini, Kyle Herkenhoff, Toru Kitagawa, Marcin Kolasa, Paul Levine, Arthur Lewbel, 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

There are several factors that make Bayesian methods particularly attractive to macroeconomists for the estimation of structural models. A key reason is that the Bayesian framework allows researchers to incorporate their prior knowledge about parameters directly into the estimation process of the model. This is particularly useful in macroeconomics, where the available sample sizes tend to be shorter. With a given prior, researchers can draw from a posterior distribution 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. Consequently, the results obtained from the standard Bayesian approach may be prone to inconsistency: the posterior mean can fail to converge to the true value, and its probability limit is sensitive to the choice of prior even asymptotically. Giacomini et al. (2019) also argue that a precise specification of a prior could be a poor representation of the belief that the researcher possesses in a given application. Given that the conditional (on the reduced-form parameters) prior distribution does not update with data, it is important to understand its impact through sensitivity analysis.

In this paper, I propose a robust Bayesian algorithm designed to identify the range of posterior means across a spectrum of priors that yield the same marginal likelihood. Additionally, it finds a robust Bayesian credible interval that guarantees a credibility level of at least $1-\alpha$ over the same class. This approach obviates the need for researchers to firmly commit to a prior; instead, they can begin with any "reasonable" prior that has a positive density over the (possibly constrained) parameter space. For each parameter drawn from the posterior distribution, researchers can obtain the observationally equivalent set of this parameter and use it as a constraint to solve an optimization problem for the upper and lower bounds of the parameters of interest η. The final step is to find the mean and the quantile of the bounds attained across posterior draws. This algorithm can be employed complementarily with standard Bayesian estimation methods. Moreover, researchers using this method do not need to determine a priori whether or not the model is identified. In fact, when the model is point-identified, the outcome will be identical to that obtained through the standard Bayesian approach.

On the theory side, I show that the proposed Bayesian sampling scheme is valid for all set-identified structural models. Specifically, the expected set obtained from this algorithm aligns with the collection of all posterior expectations generated by an arbitrary choice of prior within the aforementioned class. Therefore, researchers can apply the algorithm to conduct robust Bayesian inference on the η without the need to run estimation for every possible prior in this class. I also show that, under some regularity conditions, the estimated range of posterior means will asymptotically converge to the convex hull of the frequentist identified set.

From the practical perspective, however, while it is known that estimation may exhibit inconsistency in cases of set-identification, the implications of such discrepancies for policy making remain less understood. In Sections 2 and 4, I show that when identification fails, policy-related measures, like impulse response functions, and the formulation of monetary policy decisions can be sensitive to prior choices. I apply this algorithm to multiple models to show it is useful not only for conducting inference, also for policy-making. I start with a model in Cochrane (2011), which is simple enough to be analytically tractable. Using simulated data, I can compare the theoretically identified set and the estimated range of posterior means. I show that the true impulse response may not fall within the Bayesian credible region for some priors within the class even asymptotically. In contrast, the proposed algorithm performs well in estimating the frequentist identified set, even with small sample size. 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 based on 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, albeit similar in structure, have very different policy implications. First, in the presence of temporary shocks, parameters may remain unidentified while impulse responses are clearly identifiable. With serially correlated shocks, local identification fails but does not affect the optimal monetary policy. Lastly, in the variant featuring a cost-push shock, the uncertainty of parameters leads to ambiguity in the optimal policy. I also apply my method to Smets and Wouters (2007) to show its capability of dealing with larger models.

Literature: This paper is most closely related to the literature on identification in DSGE models.[1] 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

[1] There is a body of research exploring inference in weakly-identified DSGE models (Guerron-Quintana et al. (2013), Qu (2014), and Andrews and Mikusheva (2015)), alongside a distinct yet related line of work on sensitivity analysis in point-identified DSGE models (Müller (2012) and Ho (2022)).

(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. Andrews and Mikusheva (2015), 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 through numerical approximation methods.

Kociecki and Kolasa (2018) offer an alternative theoretical framework 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 parameter value. In this paper, I propose an easy-to-implement Bayesian algorithm for finding the identified set of parameters consistent with the data. In situations where identification failure is well-understood—a strong assumption from the outset—researchers can opt to revise the model by either incorporating more shocks or calibrating specific parameters to achieve point-identification. However, such practices can be somewhat ad hoc and may deviate from the original theoretical framework. 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.

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 many 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).[2] At the same time, there is an increasing literature on (quasi-) Bayesian analysis for set-identified models, e.g., Moon and Schorfheide (2012), Norets and Tang (2014), Baumeister and Hamilton (2015), Kline and Tamer (2016), Chen et al. (2018), Liao and Simoni (2019), Florens and Simoni (2021). The idea of incorporating uncertainty within the prior distribution traces its roots to works like Good (1950) and Robbins (1951) in the field of statistics.[3] This concept has been recently adopted in economics to conduct Bayesian inference in set-identified models (Giacomini et al. (2019), Giacomini and Kitagawa (2021), Ke et al. (2022), Giacomini et al. (2022), and Bacchiocchi and Kitagawa (2022)). While the robust Bayesian literature explores various aspects of robustness, my focus in this paper is on priors that yield identical marginal likelihoods, rendering them indistinguishable based solely on the model and input data. Moreover, additional constraints can be imposed to refine the class of priors, providing greater flexibility in the sensitivity analysis.

A key contribution of this paper is that it introduces a new Bayesian estimation technique and a tool for sensitivity analysis to structural models, applicable even when the links between structural parameters θ and reduced-form parameters[4] or the data are intractable. It has both the computational advantage of Bayesian methods, allowing researchers to guide the estimation with their prior knowledge, and a frequentist interpretation asymptotically. In the applications of this paper, I borrow the observational equivalence characterization from Kocięcki and Kolasa (2023) to perform robust Bayesian estimation in DSGE models. However, this estimation method is applicable to any structural model, whether point or set-identified, provided that the observationally equivalent set of a given parameter can be computed. When the mapping between θ and the reduced-form parameter is analytical, as is the case in the context of SVAR models, Giacomini and Kitagawa (2021) suggest proposing a prior distribution over the reduced-form parameters and then obtaining a range for the parameters of interest η from the inverse mapping of the posterior draws. However, for models lacking a clear reduced-form parameterization, such as DSGE models,[5] this method is not feasible. It is because the highly structural nature of DSGE models often precludes the possibility of deriving a straightforward, closed-form likelihood function based on structural parameters. Instead, I sample from the posterior of θ and directly obtain observationally equivalent

[2] See Molinari (2020) for a review.
[3] See Berger (1990) for a review.
[4] Here reduced-form parameters are by definition always identified, see Section 3 for a formal definition.
[5] In DSGE models, the spectral density of observed variables or the set of observationally equivalent θs might serve as the point-identified parameter, which can be infinite-dimensional.

sets of θ from those draws. I then optimize within these sets to establish the bounds for η. I then show that averaging these sets, based on a specific prior, captures the entire range of posterior means of η for a class of priors defined by that prior. The estimated set will converge to the frequentist identified set of η consistent with the data. Another contribution of this paper is it provides new insight for policy sensitivity analysis in DSGE models using robust estimation results, while in finite sample still respect the given prior up to its data implication. Therefore, the intent of this paper is not to join the debate over choosing between single or multiple priors for set-identified models; rather, it is to provide a new Bayesian inference methodology that performs well in finite samples, and is consistent with frequentist approaches asymptotically, while also providing a robust tool for sensitivity analysis conducive to effective policymaking.

The rest of this paper is organized as follows: Section 2 presents the motivations for this study, showing that the parameter estimates, impulse response functions, and optimal policies can all be sensitive to the choice of priors regardless of sample size in set-identified structural 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 Motivation

Estimation and inference in structural models that are rich in parameters and have complicated model structures can be challenging. While the Bayesian approach offers a partial solution by applying regularization with priors and using integration instead of optimization, little is known about the sensitivity of the estimation results when the model is not identified. Furthermore, the impact of non-identification on inference drawn from these estimations and its implications for policy analysis is yet to be fully understood.

The following example illustrates an identification failure, with a formal definition provided 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 identified from the data (i.e., the reduced-form parameter) 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 pick any $\bar{D} = D_0 Q$, with Q an orthonormal matrix, and still have the same covariance $\bar{D}\bar{D}' = D_0 D_0'$. Therefore, D is not identified.

It will become clearer later that using an arbitrarily informative prior distribution on D,[6] the posterior mean or mode of D can lead to a substantial divergence of D's posterior mean or mode from the actual D_0, especially when the prior mean or mode is considerably distant.

In this section, I examine two simple examples to show that in set-identified models, estimation, inference, and policies made based on standard Bayesian results can be sensitive to the choice of priors, regardless of sample size. To address this issue, I propose an algorithm to perform Bayesian estimation and inference that is robust to the unrevisable component of the prior.

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 the Bayesian alternative that has a preferable finite-sample interpretation (see Theorem 2).

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 following AR(1) form:

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

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 $\left(\rho, \frac{\sigma_e}{\phi_\pi - \rho}\right)$.

In the rest of this section and in Section 4, 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

[6] This holds true even when a Haar prior, covering all possible rotations, is applied.

a sample of 200 periods,[7] and save the generated observed variables. Following this, I estimate the parameters from the generated data, undertake a standard Bayesian analysis, and then evaluate these results in comparison to the predefined "true" values.

I run a standard Bayesian estimation of the parameters σ_e, ϕ_π, and ρ in *Dynare* (Adjemian et al. (2024)), using the "uninformative prior" such that the priors of the parameters are uniform and independent. 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.

Intriguingly, the posterior mean of (σ_e, ϕ_π) does not align with the actual parameter values, as illustrated in 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. [8]

Figure 1 shows that the marginal posterior distributions for ϕ_π and σ_e deviate notably from their prior distributions, which might be mistakenly interpreted as they are separately identified. However, a careful inspection of the likelihood function in terms of σ_e and ϕ_π (fixing ρ at its true value ρ_0) shows that the likelihood achieves its maximum near the line $\sigma_e = \phi_\pi - \rho_0$, and it is flatter near the maximum region when both values are high (see Figure 4 in Appendix B). Indeed, this scenario highlights the limitation of relying on a comparison of marginal priors and posteriors for gaining insights into parameter identification. The phenomenon of data-based learning about the identified parameters can "spill over" onto the unidentified parameters is also documented in Koop and Poirier (1997), Koop et al. (2013).

In this particular example, the identification failure is stark and straightforward to detect. Yet, with an increase in the number of estimated parameters and the model becoming more complex to solve, examining the likelihood function for model identification becomes impractical. This also renders the visual inspection or "eyeballing" approach ineffective for identifying the set of structural parameters.

As is mentioned above, the results are robust to the sample size, the number of posterior draws and replications. I also perform an exercise to explore the behavior of the posterior mode,

[7] Assuming monthly data, this equates to approximately 16 years of observations, which is considered a reasonable sample size for DSGE model estimations. Additionally, I conducted the same exercise with 500 periods and 1000 periods and the findings and conclusions drawn from these larger datasets are similar to those using 200 periods. These additional results are documented in the supplementary materials.

[8] For this study, I generate Markov chains with a length of 100,000, but the key attributes remain unchanged even when extending the chain length to 1 million or more.

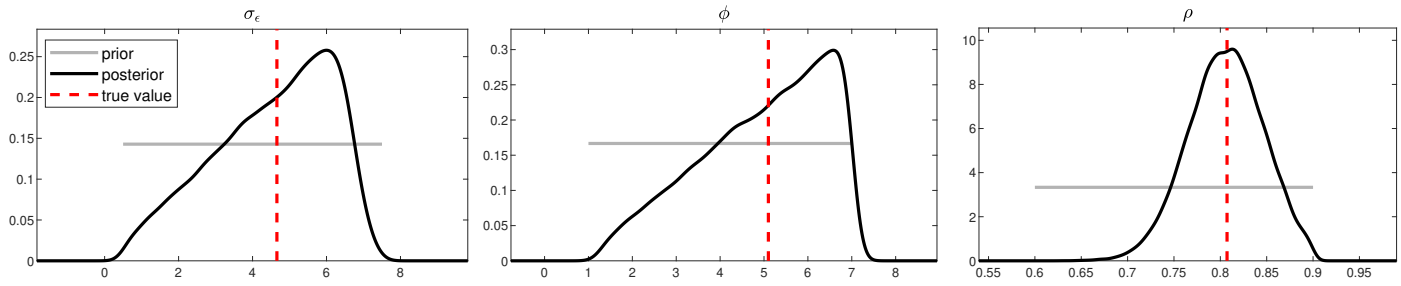


Figure 1: Prior and posterior for the Cochrane model. The red dashed line represents the true values; ρ is the correlation coefficient, ϕ_π is the Taylor rule parameter, and σ_e is the standard error of the monetary policy shock.

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

minimizing the negative log likelihood, penalized by the prior, with 1000 replications and a sample size ranging from 100 to 1000 observations.[9] As can be seen in Figure 5 and Figure 6 in Appendix B, the pattern illustrated above using MCMC is not Bayesian-specific.

2.2 The Impulse Responses

Following the discussion of estimation sensitivity in the previous section, a natural question arises: Could this sensitivity impact our policy analysis? The answer is affirmative. For instance, impulse response analysis, a tool frequently used in macroeconomic policy analysis, may also be vulnerable to such estimation sensitivities. Impulse responses can be generally defined without reference to the data-generating process as the following function (Koop et al. (1996), Potter (2000)):

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

[9]Because the objective function always has infinite minimizers, I use the particle swarm optimizer embedded in MATLAB with a uniform randomization of the starting points of particles around the prior mean, the minimizer reported is also random. Moreover, the reported numerical minimizers "picked" by the algorithm with many runs should reflect the shape of the objective 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[10] 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, denoted as $p_1(\theta)$. The other prior $p_2(\theta)$ is constructed in a way such that it always has the marginal likelihood regardless of the realization of $\{\pi_t\}_{t=1}^T$.[11] In Figure 2, I plot the true impulse response function of 20 periods with a 1-unit shock of ε at time 0. The Bayesian credible region, or the 90% highest-posterior-density interval, does cover the true impulse response function using $p_2(\theta)$ but not $p_1(\theta)$, when the parameters are not identified. This example also demonstrates how, in addition to parameter estimation, standard Bayesian inference for parameters of interest can be misleading in set-identified models.

2.3 Monetary Policy

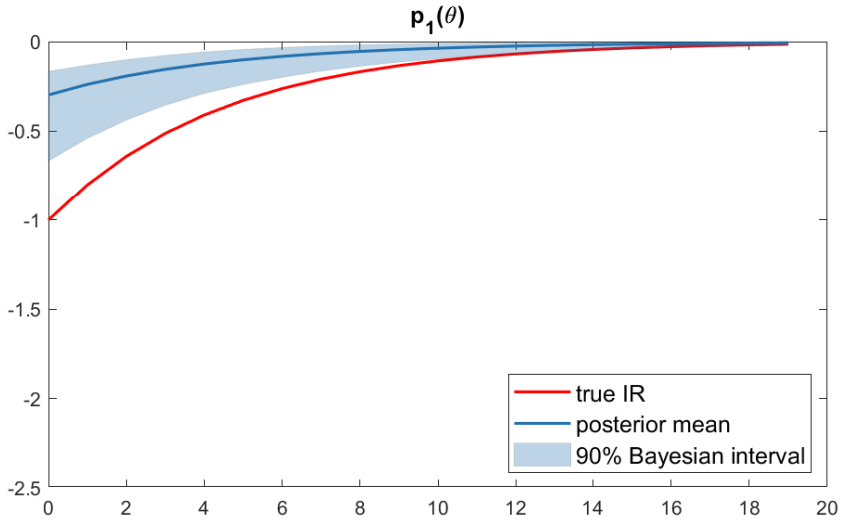
The next example, adapted from An and Schorfheide (2007), modifies only the total factor productivity shock to a cost-push shock, akin to the scenario described in (Galí, 2015, Chap. 5). As discussed in Section 4, introducing a cost-push shock breaks the divine coincidence, thereby rendering the policy response sensitive to estimates.

$$
\begin{aligned}
y_t &= \mathbb{E}_t\left[y_{t+1}\right] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t\left[\pi_{t+1}\right]\right) + g_t - \mathbb{E}_t\left[g_{t+1}\right] && (3)\\
\pi_t &= \beta\mathbb{E}_t\left[\pi_{t+1}\right] + \kappa\left(y_t - g_t\right) + u_t && (4)\\
i_t &= \rho_R i_{t-1} + \left(1-\rho_R\right)\psi_\pi \pi_t + \left(1-\rho_R\right)\psi_y\left(y_t - g_t\right) + \varepsilon_{R,t} && (5)\\
u_t &= \rho_u u_{t-1} + \varepsilon_{u,t} && (6)\\
g_t &= \rho_g g_{t-1} + \varepsilon_{g,t}. && (7)
\end{aligned}
$$

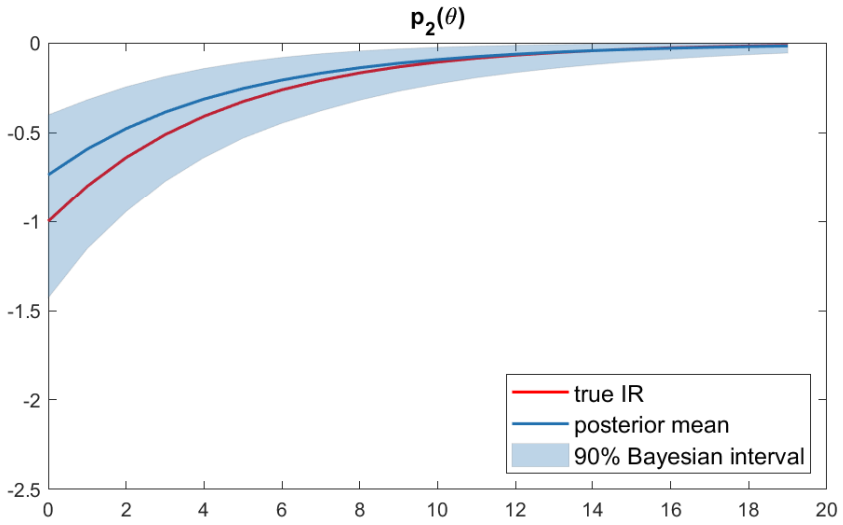
Here π_t is the inflation rate, i_t is the nominal interest rate, y_t is the output gap, g_t is the

[10] In the literature, it is more common to apply a shock that equals one standard deviation of the innovation term for impulse response analysis. However, for the purpose of illustrating the identification problem, I use a normalized unit shock in this simple example.

[11] The marginal likelihood for $\{\pi_t\}_{t=1}^T$ given prior $p(\theta)$ is $p(\pi) = \int p(\pi|\theta)p(\theta)d\theta$. It is also sometimes referred to as the prior predictive distribution.



(a) Impulse responses using prior setup 1



(b) Impulse responses using prior setup 2

Figure 2: The impulse response functions (IR) of inflation π to a one-unit shock in the Cochrane model; 20 periods of impulse responses are plotted. (a) uses the uniform prior, (b) uses prior 2 that has the same marginal likelihood.

	$\lambda = \frac{1}{3}$		$\lambda = 1$		$\lambda = 3$		$\lambda = 10$		$\lambda = 30$	
(ψ_π, ψ_y)	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$
(1.5, 0)					✓		✓	✓	✓	✓
(1.5, 0.125)	✓	✓	✓	✓		✓				

Table 2: Policy comparison under different prior distributions and weights. The check mark indicates that, for each prior distribution $p_i(\theta)$, where $i = 1, 2$, the policy associated with it results in a lower welfare loss according to its posterior.

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 $\sigma_R, \sigma_g, \sigma_u$. 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. As in the previous 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. Table **??** shows the optimal policy choices under different priors across different weights.

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.[12] That is, a researcher with $p_1(\theta)$ will disagree with a researcher who uses $p_2(\theta)$ in optimal policy choices, even though their models are the same and their priors induce the same marginal likelihood.

3 Theory

In this section, I first lay out the theoretical framework for set-identified structural models and introduce the corresponding algorithm. Next, I define the class of priors that underpins the

[12] 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 approximately between 2 and 6.

sensitivity analysis and discuss various properties of this class. In the final subsection, I show that the algorithm gives interpretable results in finite samples and describe its asymptotic properties.

3.1 Framework

Let $(\mathbf{Y}, \mathcal{Y})$ be the measurable space of a sample of observables $Y \equiv \{Y_t\}_{t=0}^{T} \subset \mathbf{Y}$, and let $(\Theta, \mathcal{A})$ be the measurable space of a structural parameter vector $\theta \in \Theta \subset \mathbb{R}^d$. There exists a regular conditional distribution $F(y \mid \theta)$, and density $p(y \mid \theta)$ of Y 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 that possesses reflexivity, symmetry, and transitivity. It partitions the space Θ into equivalent classes, independent of the data.

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.[13]

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,[14] there is no consensus regarding the definition of

[13] 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.

[14] 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.

reduced-form parameters some models, like DSGE. The agnostic nature of the mapping K caused by model complexity makes it hard to apply the methodology of Giacomini and Kitagawa (2021) here. With a slight 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.

Researchers are primarily interested in estimating the structural parameter vector θ, or transformations of θ, denoted as $\eta(\theta)$, through a measurable function $\eta : (\Theta, \mathcal{A}) \to (\mathcal{H}, \mathcal{D})$, $\mathcal{H} \subset \mathbb{R}^q$ for some $q < \infty$. 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

$$L_W(\theta, y, \pi) = \frac{1}{2}\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \left[a y_{t+s}^2 + b \pi_{t+s}^2 \right], \tag{8}$$

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 3 illustrates the graphical connection between these parameters.

Several challenges can arise in the estimation and inference of highly-structural models. Firstly, researchers often apply priors to structural parameters and derive likelihoods by solving the model, which can lead to non-trivial prior distributions over the parameters in the solution. Second, some commonly used structures used to generate the likelihood (e.g., the parameters of state-space structure, or the so-called ABC representation) may not identified. That disqualifies those parameters from being considered as reduced-form parameters. As a result, even with an injective mapping of the structural parameter vector θ to these coefficients, the Bayesian estimation of θ can still be sensitive to priors. Third, even if researchers have well-defined reduced-form parameters, it could still be challenging to back out all the θs that map into the same reduced-form, because the mapping from structural parameters to reduced-form parameters is model-specific and often numerical. Luckily, there is still a rich class of models that allows researchers to find the $K(\theta)$, even analytically. Therefore, for models that conform to this structure, and further assuming that $K(\theta)$ can be found for every given θ, I propose the following algorithm for estimation and inference:

Algorithm 1 (Robust Bayesian)

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

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

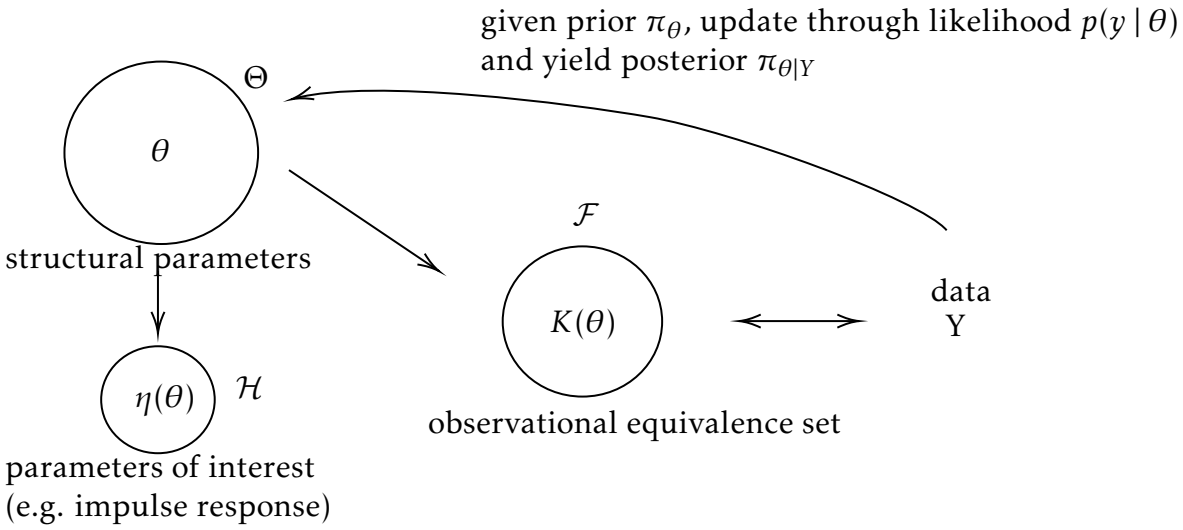


Figure 3: Connections between parameters

S.3 Optimize over $K(\theta^j)$ to find the minimum or maximum value of η, denote as $\underline{\eta}^j$, and $\overline{\eta}^j$.

S.4 Find the mean and the quantiles (e.g., 5% for $\underline{\eta}^j$ and 95% for $\overline{\eta}^j$) of $\underline{\eta}^j$ and $\overline{\eta}^j$ across draws. Report $\left[\frac{1}{M}\sum_{j=1}^{M}\underline{\eta}^j, \frac{1}{M}\sum_{j=1}^{M}\overline{\eta}^j\right]$ and $\left[\underline{q}^*_{0.05}, \overline{q}^*_{0.95}\right]$ as the set of posterior mean and robust Bayesian credible region.

The algorithm above is widely applicable so that the researchers can fit many structural models such as structural vector autoregressive models (Giacomini and Kitagawa (2021)), latent Dirichlet allocation (Ke et al. (2022)), SVMA models (Plagborg-Møller (2019)), DSGE models, and various other structural models into this framework.

3.2 Class of Priors

Several important questions remain to be addressed: What exactly is the class of priors that is being discussed so far, and why this class specifically? It is also unclear how the choice of priors affects the posterior mean and credible region under set-identification. In this subsection, I establish the need for multiple priors, justify the choice of this particular prior class, and show how the sensitivity of estimation is connected to identification failure of the model.

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 elicited information from micro evidence. When researchers have no knowledge about specific parameters, they typically adopt uninformative priors. However, 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 has argued, "Not knowing the chance of mutually exclusive events and knowing the chance to be equal are two quite different states of knowledge" (Syversveen (1998)). Such literature that justifies the use of classes of priors is extensive.

There are many reasons why a single prior is untenable. DeRoberts and Hartigan (1981) stated "In practice, prior knowledge is typically vague and any elicited prior distribution is only an approximation to the true one." 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. Berger et al. (1994) in their overview paper also argue that "In reality, it is very difficult to discriminate between, say, 0.10 and 0.15 as the subjective probability, $P(E)$, to assign to an event E, much less to discriminate between 0.10 and 0.100001. Yet standard Bayesian axiomatics assumes that the latter can (and will) be done."

Before introducing the class of priors considered in this paper, it is necessary to impose some structure on the mapped elements K. Specifically, I begin by assuming measurability of the set-valued function K, which maps Θ to the family of all closed subsets of Θ.

Assumption 1 *(i) 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 $F \in \mathcal{F}$.*

(ii) η is a continuous function of the structural parameter θ, mapping closed elements in the space $\mathcal{F}$ to $\mathcal{F}$.

Under Assumption 1 ((i)), K is called a random closed set in Θ. The corresponding prior π_K of $K(\theta)$ for a given π_θ (also denoted as the push-forward measure of π_θ, $K_*\pi_\theta$) can then be defined by

$$\pi_K(B) = \pi_\theta\left(\left\{\theta : K(\theta) \in B\right\}\right) \quad \text{for any } B \in \mathcal{B}(\mathcal{F}), \tag{9}$$

where $\mathcal{B}(\mathcal{F})$ is the σ-algebra generated by the Fell topology on $\mathcal{F}$ (Molchanov (2005), section 1.1).

The likelihood of θ is flat on $K(\theta)$ for any $y_{t=1}^T$, that is, $\theta \perp Y \mid K(\theta)$.[15] Therefore, observations of Y will not aid in determining θ given the knowledge of $K(\theta)$.

In general, the parameters spaces Θ, $\mathcal{H}$ can be subspaces of $\mathbb{R}^n$, and $K(\Theta)$ can be a subset

[15] However, that does not mean 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.

of a Baire space,[16] where the conditioned probability of $K(\theta)$ may be zero. To avoid the Borel-Kolmogorov paradox,[17] 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

$$\pi_{\theta|Y}(A) = \int_{\mathcal{F}} \pi_{\theta|K}(A)\, \mathrm{d}\pi_{K|Y}, \quad A \in \mathcal{A}. \tag{10}$$

From expression (10) it can also 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 $\pi_{\eta|Y}$. 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 following example clarifies the structure.

Example 2 *Suppose researchers observe only $Y_i \sim N(\theta_1 + \theta_2, 1)$, $i = 1, \ldots, t$. 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. It can be seen from (10) that only $\pi_{K|Y}$ 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 $\theta_1 + \theta_2$. In fact, in this example, the reduced-form parameter exists as the sum of two elements: $\phi = \theta_1 + \theta_2$.

When identification issues arise, the posterior of structural parameters integrates unrevisable prior belief over observationally equivalent parameters. Since this part of prior belief remains unaltered by the data, its influence persists even asymptotically.

In this paper, rather than requiring researchers to adhere to a single (conditional) prior, I consider the prior class in the following sense:

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

[17] 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.

Definition 3 (Conditional Prior Class) *Given a π_K supported only on $K(\Theta)$, the classes of conditional priors for θ given K are*

$$\Pi_{\theta|K} = \left\{ \pi_{\theta|K} : \pi_{\theta|K}\left(\left\{\theta : K(\theta) = K\right\}\right) = 1, \pi_K - \textit{almost surely} \right\}. \tag{11}$$

This class encompasses any conditional priors that assigns a probability one to the equivalent set, conditional on that same set. Given a prior distribution over the equivalent sets π_K, this condition prior class would then induce a class of priors for θ such that the marginal distribution for K coincides with π_K, i.e.,

$$\begin{aligned} \Pi_\theta(\pi_K) &\equiv \left\{ \int \pi_{\theta|K} d\pi_K : \pi_{\theta|K} \in \Pi_{\theta|K} \right\} \\ &= \left\{ \pi_\theta : \pi_\theta\left(\left\{\theta : K(\theta) \in B\right\}\right) = \pi_K(B), \text{ for } B \in \mathcal{B}(\mathcal{F}) \right\}. \end{aligned} \tag{12}$$

Moreover, each π_θ induces a π_K, thereby also uniquely defining a conditional prior class. In practice, When a reduced-form parameter exists and its closed-form mapping from structural parameters are known, along with a given reduced-form prior, one can follow the approach of Giacomini and Kitagawa (2021) by exploring the variations in (11). However, for more general situations where only structural parameter priors are available, (12) can be employed to identify the class of priors that yield the same marginal distribution over K.

There exists a variety of classes of multiple priors in the literature. For example, the ϵ-contaminated class of priors (Berger and Berliner (1984), Berger and Berliner (1986)), the class of priors with fixed location but different covariance matrices (Chamberlain and Leamer (1976), and the class of extended priors based on distributions near π_0, and a conditional distribution over its neighborhood $\pi_c(N(\theta)|\theta)$ (Lavine (1991)). In contrast to the aforementioned classes, this paper focuses on a prior class tailored for set-identified models, offering both theoretical and practical advantages. First, it naturally extends from a single prior to a prior class for set-identified models, and it simplifies to the single given prior $\Pi_\theta(\pi_K) = \{\pi_\theta\}$ in point-identified situations. This class is tractable, encompassing all conditional priors that assign probability 1 to the equivalence set $K(\theta)$. However, it also allows further refinements that can be imposed on the conditional priors, such as bounds on the parameter space. For practitioners, it simplifies the process, requiring just one prior over structural parameters without additional hyperparameters.[18] The defining prior

[18] Specifically, it eliminates the need to define the radius of the neighborhood around the benchmark prior.

of this class, π_K, is determined by the push-forward measure of π_θ, making the computation of π_K unnecessary. Moreover, any other prior $\tilde{\pi}_\theta$ within the same class can be sampled by redrawing from the observational equivalent set $K(\theta)$ of a given draw θ from prior π_θ. Lastly, the class has an economic implication of the class that generates the same prior/posterior predictive distribution (Geweke and Whiteman (2006), Geweke (2007), Del Negro and Schorfheide (2008), Weitzman (2009)). Therefore, it is impossible to evaluate one prior against another within the same class using only the model and its input. The following lemma formalizes this fact.

Lemma 1 (Prior Predictive Distribution) *For any given model and prior 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 lemma applies similarly to the posterior predictive distribution. Meanwhile, if researchers possess insights from economic theory or other sources that can further refine the conditional prior class, these refinements can be easily adopted in Algorithm 1 either by resampling from $K(\theta^j)$ or by narrowing the set in step S.2.

3.3 Robust Posterior Distributions

With the prior class defined, we can then explore the variation of posterior probabilities, means, and quantiles of the parameters of interest within this class.

Given the conditional prior class $\Pi_{\theta|K}$ and a posterior over K, $\pi_{K|Y}$, the class of posteriors for the parameters of interest follows

$$\Pi_{\eta|Y} = \left\{ \pi_{\eta|Y}(\cdot) = \int_{\mathcal{F}} \pi_{\theta|K}\left(\eta(\theta) \in \cdot\right) d\pi_{K|Y} : \pi_{\theta|K} \in \Pi_{\theta|K} \right\}. \tag{13}$$

For any $D \in \mathcal{D}$, the lowest and highest possible probabilities, denoted by $\underline{\pi}_{\eta|Y} : \mathcal{D} \to [0,1]$ and $\overline{\pi}_{\eta|Y} : \mathcal{D} \to [0,1]$ are

$$\underline{\pi}_{\eta|Y}(D) \equiv \inf_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}(D), \tag{14}$$

$$\overline{\pi}_{\eta|Y}(D) \equiv \sup_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}(D). \tag{15}$$

The lower and upper posterior probabilities outlined here serve as robust bounds on the

probability of a particular set, indicating the minimum and maximum probabilities of D attainable from $\Pi_{\theta|K}$ regardless of the chosen $\pi_{\theta|K}$. However, computing these probabilities using the given formulas is impractical due to the requirement of evaluating all conditional distributions within the class. Fortunately, there are alternative methods available to circumvent this necessity. I will start by outlining some necessary but weak regularity conditions.

Assumption 2 (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. This assumption, along with those detailed in the preceding paragraph, provide the basis for the subsequent theorem.

Theorem 1 (Lower and Upper Probabilities) *For any given π_θ, equivalence mapping K and continuous function η, let Assumptions 1 and 2 hold. For any $D \in \mathcal{D}$,*

$$\underline{\pi}_{\eta|Y}(D) = \pi_{\theta|Y}\left(\left\{\theta : \eta(K(\theta)) \subset D\right\}\right), \tag{16}$$

$$\overline{\pi}_{\eta|Y}(D) = \pi_{\theta|Y}\left(\left\{\theta : \eta(K(\theta)) \cap D \neq \emptyset\right\}\right). \tag{17}$$

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 1 provides a practical tool for computing the lower and upper probabilities. It bears resemblance to Theorem 1 in Giacomini and Kitagawa (2021), wherein the inverse mapping from the reduced-form parameter to the structural parameter is extended to the equivalent set. Assumption 1 and the continuity of η ensure that $K(\theta)$ and $\eta(K(\theta))$ are random closed sets. The proof can be found in Appendix A.

The robust probabilities (16) and (17) are called the containment functional and the capacity functional of a random set (Molchanov and Molinari (2018)), respectively. Morever, they correspond to special cases of a belief function and a plausibility function, respectively, in the imprecise probability literature (Walley (1991)).

3.3.1 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 1 has

a finite-sample interpretation: it reports the range of posterior means of the parameters of interest from a prior class.

Before introducing the main theorem, I present a lemma that establishes a connection between the distribution of the K-transformed posterior draws of θ and the draws from the posterior of $K_*\pi_\theta$.

Lemma 2 *The push-forward measure of $\pi_{\theta|Y}$ by a measurable multifunction $K : \Theta \to \mathcal{F}$, written as $\tilde{\pi}_{K|Y} = K_*\pi_{\theta|Y}$, coincides with the posterior distribution $\pi_{K|Y}$ of the push-forward measure $\pi_K = K_*\pi_\theta$.*

The proof can be found in Appendix A. This lemma, which generalizes the result in Ke et al. (2022) Appendix C without assuming a parametric structure of K, states that two distinct methods of obtaining posterior draws of equivalent sets yield identical results. For simplicity in notation and proof throughout the rest of the paper, we will regard the parameter of interest η as a scalar.[19] The main theorem of this paper is stated below.

Theorem 2 (Posterior Means of Scalar η) *For a given π_θ, let Assumptions 1 and 2 hold. 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

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

where $\Pi_{\theta|Y}$ collects the posteriors of Equation (12) with given $\pi_K = K_*\pi_\theta$.

This theorem states that if a researcher selects a prior π_θ, draws from its posterior, and computes the expected lower and upper bounds of η within the observationally equivalent θs, then it is tantamount to computing all posterior means within the same class of π_θ. The expectation of $\overline{\eta}^*(\theta)$ and $\underline{\eta}^*(\theta)$ under the distribution $\pi_{\theta|Y}$ is feasible because it needs draws only from one distribution.

Intuitively, Theorem 2 holds because π_θ serves solely to define the prior class. Even if π_θ contains some unrevised conditional prior of $\theta \mid K$, this information becomes irrelavant during the

[19] This still encompasses a wide range of interesting cases. For example, 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. The theoretical results can be easily extended to scenarios beyond scalar cases.

computation of the equivalence set $K(\theta)$ in step S.2 of Algorithm 1. The posterior of K is unaffected by this conditional prior and will be updated by data. As a result, knowing the observationally equivalent sets $K(\theta)$ and its posterior distribution is the same as knowing the class of all posterior distributions based on $\pi_{\theta|Y}$. Therefore, the posterior distribution of the identified set is the same.

3.3.2 Robust Credible Region

For the purpose of inference, this section introduces the robust Bayesian counterpart of the posterior mean and credible region in standard Bayesian inference. For $\alpha \in (0,1)$, let a robust Bayesian credible set be such that the posterior lower probability is greater or equal to $1-\alpha$. That is,

$$\pi_{\theta|Y}\left(\left\{\theta : \eta(K(\theta)) \subset C_\alpha\right\}\right) \geq 1-\alpha, \tag{18}$$

This definition aligns with those found in Moon and Schorfheide (2012), Norets and Tang (2014), and Kline and Tamer (2016), but is not uniquely determined as the volume-optimized credible set definition of Giacomini and Kitagawa (2021). In the scalar case for η, when the set C_α is further constrained to be convex, one approach to determining a robust credible set involves finding the smallest $\overline{q} \in \mathbb{R}$ such that

$$\inf_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}\left((-\infty, \overline{q}]\right) \geq 1 - \frac{\alpha}{2} \tag{19}$$

and the largest $\underline{q}$ where

$$\sup_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}\left((-\infty, \underline{q}]\right) \leq \frac{\alpha}{2}. \tag{20}$$

In fact, it can be shown that

$$\overline{q}^*_{1-\alpha/2} : \pi_{\theta|Y}\left(\overline{\eta}^*(\theta) \leq \overline{q}^*_{1-\alpha/2}\right) = 1 - \frac{\alpha}{2}, \tag{21}$$

and

$$\underline{q}^*_{\alpha/2} : \pi_{\theta|Y}\left(\underline{\eta}^*(\theta) \leq \underline{q}^*_{\alpha/2}\right) = \frac{\alpha}{2} \tag{22}$$

solve Equations (19) and (20). Consequently, a desired robust credible region is $[\underline{q}^*_{\alpha/2}, \overline{q}^*_{1-\alpha/2}]$, 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 3 *For any given π_θ, equivalence mapping K and continuous function η, let Assumptions 1 and 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 1. Details are also presented in Appendix A.

3.3.3 Asymptotic Properties

Transitioning from our discussion on finite sample interpretation in the previous section, we now turn our attention to asymptotic analysis. The results provided in this section allow us to connect our Bayesian results with frequentist perspectives.

In the scalar case, we can always truncate $\eta(K(\theta))$ by some upper and lower bounds, ensuring integrability without much loss of generality.

Let θ_0 be the true underlying value that generates the data. The data variable Y is superscripted with either the sample size T or ∞ to distinguish between finite and asymptotic behavior.

Assumption 3 *(i) Let* $\max\left\{|\underline{\eta}^*(\theta_0)|, |\overline{\eta}^*(\theta_0)|\right\} < \infty$, *and the equivalence mapping $K : \Theta \to \mathcal{F}$ is a correspondence continuous at θ_0.*

(ii) The posterior of $K(\theta)$ is consistent for $K(\theta_0)$, i.e.,

$$\pi_{\theta|Y^T}\left(\left\{\theta : d_H(K(\theta), K(\theta_0)) < \epsilon\right\}\right) \to 1 \text{ as } T \to \infty, \quad p\left(Y^\infty \mid \theta_0\right)\text{-a.s.}, \tag{23}$$

where $d_H(X, Y)$ is the Hausdorff distance[20] between X and Y.

(iii) $\underline{\eta}^*(\theta)$ and $\overline{\eta}^*(\theta)$ have finite posterior variance;

$$\mathbb{E}_{\theta|Y^T}\left(\sup_{\eta \in \eta(K(\theta))} |\eta|^2\right) < \infty, \quad p\left(Y^T \mid \theta_0\right)\text{-a.s.} \tag{24}$$

[20] 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 $\max\{|a - c|, |b - d|\}$.

The continuity of correspondences is defined as in Aliprantis and Kim (2006) Definition 17.2. Assumption 3 ((i)) allows the use of maximum theorem. Assumption 3 ((ii)) states that Bayesian estimation of the point-identified parameter gives standard consistency results. Assumption 3 ((iii)) ensures the almost-sure compactness of $\eta(K(\theta))$ and permits the application of Hölder's inequality to bound the distance between random sets.

Theorem 4 (Consistency of Posterior Mean) *Let Assumptions 1, 2, and 3 hold, and that $\Theta \subset \mathbb{R}^p$ for some $p < \infty$. Then the Hausdorff distance 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.,*

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

The proofs of Theorem 4 and Theorem 5 below follows directly from the proof of Theorem 3 and Theorem 4 in Giacomini and Kitagawa (2021), with the assumptions and conclusions tailored to the scalar case. Theorem 4 validates the application of the algorithm by asserting that the computed interval converges to the true identified set, provided that it is convex.

Next, I will state a set of high-level conditions under which the proposed robust credible set will asymptotically attains greater than $1-\alpha$ coverage for the true identified set, from a frequentist perspective.

Assumption 4 *(i) The true identified set $\eta(K(\theta_0))$ is closed and bounded, and $\eta(K(\theta))$ is π_θ-almost surely closed and bounded.*

(ii) The class of robust credible sets C_α consists of closed intervals in $\mathbb{R}$.

Assumption 5 *(i) Let $\hat{\theta}$ denote an element of the set of maximum likelihood estimators,*

$$\sqrt{T}\begin{pmatrix} \underline{\eta}^*(\theta) - \underline{\eta}^*(\hat{\theta}) \\ \overline{\eta}^*(\theta) - \overline{\eta}^*(\hat{\theta}) \end{pmatrix} | Y^T \quad \Rightarrow \quad \mathcal{N}(0,\Sigma), \text{ as } T\to\infty, p(Y^\infty \mid \theta_0)\text{-a.s.}, \tag{25}$$

$$\sqrt{T}\begin{pmatrix} \underline{\eta}^*(\hat{\theta}) - \underline{\eta}^*(\theta_0) \\ \overline{\eta}^*(\hat{\theta}) - \overline{\eta}^*(\theta_0) \end{pmatrix} | \theta_0 \quad \Rightarrow \quad \mathcal{N}(0,\Sigma), \text{ as } T\to\infty. \tag{26}$$

(ii) For the robust credible region $\left[\underline{q}^*_{\alpha/2}, \overline{q}^*_{1-\alpha/2}\right]$,

$$\hat{c}_T \equiv \sqrt{T}\begin{pmatrix} \underline{q}^*_{\alpha/2} - \underline{\eta}^*(\hat{\theta}) \\ \overline{q}^*_{1-\alpha/2} - \overline{\eta}^*(\hat{\theta}) \end{pmatrix} \xrightarrow{p} c \tag{27}$$

for some constant c *as* $T \to \infty$.

Here, $\Rightarrow$ denotes weak convergence, and $\xrightarrow{p}$ denotes convergence in probability. Assumption 4 reduces the event $\eta(K(\theta)) \subset C_\alpha$ to the inequalities $\underline{q}^*_{\alpha/2} < \underline{\eta}^*(\theta)$, and $\overline{\eta}^*(\theta) < \overline{q}^*_{1-\alpha/2}$. Assumption 5 ((i)) assumes the Bernstein-von Mises property for the support functions of the identified set in the scalar case, without requiring the existence of a reduced-form parameter. The convergence rate is derived from the central limit theorem for random sets (Molchanov (2005), Chapter 3, Theorem 2.1). Assumption 5 ((ii)) specifies that C_α should shrink towards $\left[\underline{\eta}^*(\hat{\theta}), \overline{\eta}^*(\hat{\theta})\right]$.

Theorem 5 *Under Assumptions 4 and 5,*

$$\liminf_{T\to\infty} P_{Y^T|\theta_0}\left(\eta(K(\theta_0)) \subset \left[\underline{q}^*_{\alpha/2}, \overline{q}^*_{1-\alpha/2}\right]\right) \geq 1-\alpha. \tag{28}$$

Theorem 5 provides an asymptotic basis for employing algorithm-generated averaged posterior quantiles as a robust credible set. It states that, similar to the Bernstein-von Mises theorem, the credible set asymptotically achieves coverage exceeding $1-\alpha$ for the frequentist true identified set.

3.4 DSGE Models

In this part, I shift from the general theory to a focused study of DSGE models. I start by sketching the structure within a DSGE context, then detail the approach for estimation and inference using a robust Bayesian methodology. The discussion on DSGE models covers their structure and outlines the standard procedure for their estimation, which involves solving linear rational expectation models and determining the likelihood. I further present the conditions of observational equivalence through relevant literature. In the application, I specify the class of priors utilized and examine the performance of the proposed algorithm. This part is intended to offer a comprehensive yet concise overview of applying the suggested method to DSGE model estimation.

3.4.1 Setup

A DSGE model, parametrized with structural parameter vector $\theta \in \Theta$, can be characterized by several Euler equations and market-clearing conditions at equilibrium. Upon linearizing these equilibrium conditions around the steady states, the model reduces to a linearized rational expectation (LRE) form following Kociecki and Kolasa (2018) and Kocięcki and Kolasa (2023),[21]

$$\Gamma_0(\theta)\begin{bmatrix} S_t \\ P_t \end{bmatrix} = \Gamma_1(\theta)\mathbb{E}_t\begin{bmatrix} S_{t+1} \\ P_{t+1} \end{bmatrix} + \Gamma_2(\theta)S_{t-1} + \Gamma_3(\theta)\varepsilon_t, \tag{29}$$

where $\Gamma_i(\theta), i = 0\ldots 3$ 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,[22] 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),

$$S_t = A(\theta)S_{t-1} + B(\theta)\varepsilon_t \tag{30}$$

$$Y_t = C(\theta)S_{t-1} + D(\theta)\varepsilon_t, \tag{31}$$

where $S_t \in \mathbb{R}^{n_S}$ is the state vector, $Y_t \in \mathbb{R}^{n_Y}$ is the observable vector, $\varepsilon_t \sim N(0, \Sigma(\theta))$ 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 (31) is derived from an additional solution equation $P_t = F(\theta)S_{t-1} + G(\theta)\varepsilon_t$[23] and a

[21] There are alternative representations 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 (29) because it allows researchers to operate under minimal state representation.

[22] 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.

[23] The coefficients F and G enter the likelihood of Y only through $C(\theta)$ and $D(\theta)$ and are thus irrelevant. In addition, in the literature people do not always separate the state variables from the policy variables. Therefore, their matrix A contains both values in A and F here.

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, this ABCD representation that researchers work directly on is not identified in general,[24] in the sense that different combinations of (A, B, C, D, Σ) could potentially have the same data generating process. Therefore, the lack of identification could result from the mapping from structural parameters to the state-space coefficients, or from the state-space parameters to the likelihood. Despite this, in practice, identification concerns are typically overlooked when estimating DSGE models because researchers can still achieve posterior distributions for structural parameters even without point-identification. However, as discussed in Section 3.3, in set-identified models, standard Bayesian estimation outcomes are sensitive to the choice of priors regardless of the sample size. This sensitivity indicates potential robustness issues in the estimation process and the policy conclusions derived from it.

A standard process for conducting Bayesian estimation in DSGE models can be found in Herbst and Schorfheide (2015). This type of algorithm, although widely used, would not detect lack of identification[25] and how the results are sensitive to the choice of priors. As shown in Section 2, relying on a prior for estimation can conceal underlying identification issues and potentially lead to misleading posterior distributions. Ideally, priors should serve as informed starting points that guide the estimation without dominating it. Researchers might be interested in all the inferential conclusions that are supported by the model and the data.Therefore, an estimation and inference method that maintains robustness against changes in priors within the class specified in (12) would be especially valuable.

To apply Algorithm 1 to DSGE models, I have adapted the second step (S.2) from the methods described in Kocięcki and Kolasa (2023). Further details are provided in the following section. In the case of point-identification, each set $K(\theta^j)$ should be a singleton, with the reported range consisting of a single value matching the standard Bayesian posterior mean. Therefore, there is no loss in sharpness. The multifunction K is a composition of the mapping from structural

[24] See Komunjer and Ng (2011) for a reference.

[25] It could be argued that identification failures might be observable, as illustrated in Figure 5, where the confidence bands for the MAP estimates remain wide for parameters that are not identified. However, this becomes less apparent when the range of the identified set is small or when the model is only locally, not globally, identified. Effective methods to detect global identification failure are covered in Qu and Tkachenko (2017) and Kociecki and Kolasa (2018).

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).[26]

3.4.2 Identification Conditions

A key step in Algorithm 1 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), Qu and Tkachenko (2017), Kociecki and Kolasa (2018), Kocięcki and Kolasa (2023)). Existing research on DSGE model identification typically focuses on checking at specific parameter vectors, and deriving necessary and sufficient conditions for point identification. They also offer valuable insights into characterizing the set of observationally equivalent parameters. This paper leverages these conditions to help identify the true set of parameters. For simplicity and when there is no ambiguity, I will omit the explicit dependence on the parameter θ for parameters A, B, C, D, Σ, and subsequent parameters. 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 6 (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 (30) and (31) to weakly stationary time series, where the eigenvalues of A remain inside the unit circle. Under Assumption 6, Wold decomposition applies; therefore, I can rewrite Y_t in Equations (30) and (31) in the form of a $VMA(\infty)$ process

$$Y_{t+1} = \left[C(I_{n_S} - AL)^{-1}BL + D\right]\varepsilon_{t+1}, \quad t = \cdots -1, 0, 1 \ldots, \tag{32}$$

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

$$IR_y(t, s, \delta) = \begin{cases} D\delta, & s = 0 \\ CA^{s-1}B\delta, & s = 1, 2, \ldots \end{cases}. \tag{33}$$

Define $P \equiv E(S_t S_t')$, which is also the unique (under Assumption 6) 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 $N = APC' + B\Sigma D'$.

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

Define the z-transform of $\{Y_t\}$ by

$$\Phi_Y(z;\theta) = \sum_{j=-\infty}^{+\infty} \Gamma_j^y z^{-j}. \tag{34}$$

The spectral density can be achieved by setting $z = e^{i\omega}$. 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.

Assumption 7 (Stochastic Minimality) *For every $\theta \in \Theta$, the matrices defined by*

$$\mathcal{O} \equiv (C' \quad A'C' \quad \cdots \quad A'^{n_S-1}C')$$

have full column rank and

$$\mathcal{C} \equiv (N \quad AN \quad \cdots \quad A^{n_S-1}N)$$

have full row rank, i.e., $\text{rank}(\mathcal{O}) = \text{rank}(\mathcal{C}) = n_S$.

Assumption 7 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[27] (see for example, Lindquist and Picci (2015)) is on (A, N) instead of (A, B), and Assumption 7 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. This assumption ensures 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), where it is also shown that if Assumption 7 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. Utilizing Gaussianity, I can characterize the observational equivalence of parameters based on their equivalence in the spectral density function. I assume zero intercepts in the measurement equation; however, incorporating non-zero intercepts

[27] 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. Here, the outputs are $\{Y_t\}$, the inputs are $\{\varepsilon_t\}$, and the states are $\{S_t\}$.

dependent on θ will give extra identification information in the first order. The subsequent lemma provides an alternative characterization of observational equivalence that is more convenient for analysis. This characterization is also the definition used in Komunjer and Ng (2011) and Qu and Tkachenko (2012).

Lemma 3 (Observational Equivalence) *In a linearized DSGE model with Gaussian shocks, if Assumption 6 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 can be found in Appendix A. Lemma 3, 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. Therefore, the spectral density $\Phi_Y(e^{i\omega};\theta)$ is always identified. However, as illustrated by the following theorem, the commonly used solution parametrization (A, B, C, D, Σ) people normally work with is not point-identified.

The simplest version of the theorem that characterizes observational equivalence across discrete-time linear state-space systems, namely the equivalence described by coordinate transformation $(A, B, C, D) \to (TAT^{-1}, TB, CT^{-1}, D)$, 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 a more general theorem below (Theorem 1 in Kocięcki and Kolasa (2023)) that accommodates both singular and non-singular state-space systems.

Theorem 6 (Observational Equivalence) *Let Assumptions 6 and 7 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 6 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, it is necessary to establish connections between the structural parameters θ and their respective state-space parameters, as well as between $\bar{\theta}$ and its associated state-space parameters, in order to construct an equation system characterizing their equivalence.

To do that, first substitute model solution expressed by Equation (30) to Equation (29), 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 (35)–(44). 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 6 together with the undetermined coefficient method, allows me to characterize the identified set by a system of equations. The following equation system is obtained:

$$\begin{aligned}
\bar{\Gamma}_0^s \bar{A} + \bar{\Gamma}_0^p \bar{F} - \bar{\Gamma}_1^s (\bar{A})^2 - \bar{\Gamma}_1^p \bar{F} \bar{A} &= \bar{\Gamma}_2 && (35)\\
\bar{\Gamma}_1^s \bar{A} \bar{B} + \bar{\Gamma}_1^p \bar{F} \bar{B} - \bar{\Gamma}_0^s \bar{B} + \bar{\Gamma}_3 &= \bar{\Gamma}_0^p \bar{G} && (36)\\
\bar{C} &= \bar{L}^s \bar{A} + \bar{L}^p \bar{F} && (37)\\
\bar{D} &= \bar{L}^s \bar{B} + \bar{L}^p \bar{G} + \bar{J} && (38)\\
\bar{A} &= T A T^{-1} && (39)\\
\bar{C} &= C T^{-1} && (40)\\
A Q A' - Q &= -B \Sigma B' + T^{-1} \bar{B} \bar{\Sigma} \bar{B}' \left(T^{-1}\right)' && (41)\\
A Q C' &= T^{-1} \bar{B} \bar{\Sigma} \bar{D}' - B \Sigma D' && (42)\\
C Q C' &= \bar{D} \bar{\Sigma} \bar{D}' - D \Sigma D' && (43)\\
Q &= Q', && (44)
\end{aligned}$$

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. Assumption 3 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 challenging to show this assumption holds in general when there are infinite solutions, even if the "discriminant locus" is excluded.

To solve equation system (35)–(44), analytical methods developed in computational algebraic geometry can be of use when it can be rewritten as a system of polynomials in unknowns.[28] That is, $\bar{\theta}$ appears in polynomial form within $\bar{\Gamma}_i$, $\bar{\Sigma}$, $\bar{L}$ and $\bar{J}$,[29] and (35)–(44) is a polynomial system.

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

[29] This is not a restrictive assumption. For instance, when parameters enter as a fraction, researchers can simply multiply both sides by the denominator to maintain a polynomial form. In cases where parameters

With the method 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).[30] The idea is to find solutions of system (35)–(44). Using theorems regarding Gröbner basis in algebraic geometry[31] allows for the computation of all solutions satisfying the polynomial system by simplifying the system. The procedure outlined by Kocięcki and Kolasa (2023) for finding $K(\theta)$ can be described as below.

Algorithm 2 (Observationally Equivalent Sets in DSGE)

S.1 *Given a structural parameter vector θ, where identification is checked, solve the model numerically and obtain a state-space representation with parameters (A, B, C, D, Σ).*

S.2 *Calculate the reduced Gröbner basis for the identification conditions, which are formulated as equations in (35)–(44) and parameterized by (A, B, C, D, Σ). This forms a polynomial system involving variables $(\bar{\theta}, \bar{B}, \bar{D}, \bar{F}, \bar{G}, T, Q)$.*

S.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.*

S.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 $K(\theta)$, either as a finite collection of points or a simple equation representing the roots of the polynomial system. Both the processes of solving an LRE model and reducing to the Gröbner basis can be executed within *SINGULAR* with arbitrary precision. In the applications discussed in Section 4, I configure the first part to operate with 600 digits and the latter with 20 digits, providing a higher level of accuracy compared to the default precision setting of *MATLAB*, which utilizes 16 digits.

enter the equations in more complex forms (e.g., one is an exponent of another), defining auxiliary parameters may be necessary, either by adding to the original vector or replacing some parameters to retain a polynomial structure. The relationship between the removed structural parameters and the newly defined auxiliary parameters is purely mathematical and known. Kocięcki and Kolasa (2023) call this new parameter vector semi-structural. In the applications discussed in Section 4, I do not distinguish structural and semi-structural parameters, except in Section 4.3 for the model from Smets and Wouters (2007).

[30] 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/.

[31] For comprehensive details, please see Cox et al. (2013).

4 Applications

In the DSGE framework, the standard approach to obtaining the posterior distribution of η does not involve estimating reduced-form parameters; instead, it entails evaluating likelihoods based on the parameters of the state-space representation. Priors are typically selected at the θ-level. For a given prior distribution, robust Bayesian estimation can then be performed by Algorithm 1, with second step executed through Algorithm 2.

In this section, I revisit all the models briefly discussed in Section 2, as well as the model from Smets and Wouters (2007), with a focused discussion on the three-equation New Keynesian models. For each model, I start with a specification and a predetermined set of true parameter values, simulate the model, and use the generated data for robust Bayesian estimation. Then I conduct sensitivity analysis 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:

$$\begin{aligned} x_t &= \rho x_{t-1} + \varepsilon_t, \quad |\rho| < 1, \varepsilon_t \sim N(0, \sigma_e) \\ i_t &= r + \mathbb{E}_t \pi_{t+1} \\ i_t &= r + \phi_\pi \pi_t + x_t, \quad \phi_\pi > 1, \end{aligned} \tag{45}$$

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 state-space 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

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

	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)

Table 3: Estimated Identified Set for Cochrane (2011) Model. Because of machine precision in *MATLAB*, I round numbers greater than 10^6 to infinity. Same in Table 5, 9, and 10.

In this case, the structural parameter vector is $\theta = (\rho, \phi_\pi, \sigma_e)$. However, methods such as simple regression can only identify $\left(\rho, \frac{\sigma_e}{\phi_\pi - \rho}\right)$. Invoking Theorem 6 gives the same result that (ϕ_π, σ_e) are not identified separately. In fact, $K(\theta) = \left\{(\bar\rho, \bar\phi_\pi, \bar\sigma_e) : \bar\rho = \rho, \frac{\bar\sigma_e}{\bar\phi_\pi - \rho} = \frac{\sigma_e}{\phi_\pi - \rho}\right\}$.

The true identified set for the impulse responses is given by $IR_0(t,s,1) = \frac{\rho_0^s}{\phi_\pi - \rho_0} \cdot Q$, where Q, as defined in Theorem 6, is a scalar that takes value in $\left(0, \frac{\phi_\pi - \rho_0}{1-\rho_0}\right)$ in this example. Therefore, $IR_0(t,s,1) = \left(\min\left\{0, -\frac{\rho_0^s}{1-\rho_0}\right\}, \max\left\{0, -\frac{\rho_0^s}{1-\rho_0}\right\}\right)$ for all horizon s.[32] If $\rho_0 > 0$, indicating a positive serial correlation among monetary shocks, the identified set of the inflation reaction coefficient is $(1, \infty)$, suggesting that while the policy is proactive in addressing inflation, the precise level of aggressiveness is completely undetermined by the available data.

As discussed in Section 2, the standard Bayesian estimation results from Table 1 and impulse response analysis in Figure 2 shows the sensitivity of the posterior to the choice of priors.[33]

Given the analytical result of the identified set $K(\theta_0)$ and $IR_0(t,s,1)$, I then compare them with the numerical results obtained by applying Algorithm 1.

Let the true parameter values be $\rho_0 = 0.8, \phi_{\pi 0} = 1.8, \sigma_{e0} = 1$. The identified set of ϕ_π and σ_e can be computed accordingly. The last two columns in Table **??** shows the estimation results from the algorithm.[34] The averaged values of the estimated identified sets match the theoretical values quite well.

I then proceed to compute the impulse responses for inflation π_t. First, I find the pointwise minimum and maximum impulse responses within each equivalence set associated with a posterior draw. Next, I compute the average or quantile of these bounds and determine the range of posterior means and the robust Bayesian credible region for the impulse responses. The lower panel of Figure

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

[33] 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 $\left(\rho, \frac{\sigma_e}{\phi_\pi - \rho}\right)$ and, therefore, the same marginal likelihood and posterior predictive distribution.

[34] Here I start with the prior setup $p_1(\theta)$ for the algorithm. Same for the lower panel of Figure 7.

7 shows that the region of model-consistent impulse response functions $IR_0(t,s,1)$ is of similar size to the estimated range of posterior means. Moreover, the 90% robust Bayesian credible region is much larger, as it should be.

4.2 Three-equation New Keynesian Models

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 parameter estimates. In the last variant, I introducing a cost-push shock (Clarida et al., 1999; Woodford, 2003a,b; Galí, 2015) to the second model. In that case, 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.

4.2.1 Temporary Shocks

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

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

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 (29) results in

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

Even this simple three-equation model is complex enough to obscure clear analytical insights, resembling a black box in terms of identification. However, it is obvious that τ is not identified in general.[35] 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 and τ are not separately identified. 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 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.

Based on data generated from the true values presented in Table 4, I perform a naive Bayesian estimation using the prior specification within the same table. *DYNARE* reports identification checks at the prior mean, and, not surprisingly, (ψ, τ) is pairwise unidentified at that point. 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

[35] For example, If I fix β, for each $\tau < 1$ there is always an observationally equivalent $\tau > 1$ such that $\frac{(1-\tau)(1-\beta\tau)}{\tau}$ remains unchanged. The latter can be excluded by restriction of support being $[0, 1]$ because τ stands for price stickiness.

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)

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 true identified set: for ψ, it is $(0,+\infty)$, and for τ, it is $(0.67,1)$. As discussed above, if $(\phi_\pi,\phi_y,\sigma,\beta,\kappa)$ are identified, the impulse response functions are also identified (see also 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.

4.2.2 An and Schorfheide (2007)

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

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)

equations:

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

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 example in 4.2.1, ν and ψ enter the model only through the ratio $\sigma\frac{1-\nu}{\nu\pi^2\psi}$. Because ν and ψ are not separately identifiable, they are replaced by $\kappa = \sigma\frac{1-\nu}{\nu\pi^2\psi}$ in estimation. However, even with this replacement, it will still be insufficient to achieve point identification for all parameters. As Komunjer and Ng (2011) and 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.

In the model, I set $\psi_\pi > 1$ and $\psi_y > 0$ to ensure the Blanchard-Kahn conditions are satisfied. The true values of the set-identified parameters do not always fall in the range of the posterior mean because of the sampling error. 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 range of values these non-identified parameters can take in the estimation result 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.

4.2.3 A Cost-push Shock Model

Just as sensitivity in estimates does not always translate to sensitivity in impulse responses, it remains uncertain whether and when the sensitivity of all these estimates might affect optimal policies, given their shared marginal likelihood.

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, known as 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:

$$
\begin{aligned}
y_t &= \mathbb{E}_t\left[y_{t+1}\right] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t\left[\pi_{t+1}\right]\right) + g_t - \mathbb{E}_t\left[g_{t+1}\right] && (46)\\
\pi_t &= \beta\mathbb{E}_t\left[\pi_{t+1}\right] + \kappa\left(y_t - g_t\right) + u_t && (47)\\
i_t &= \rho_R i_{t-1} + \left(1-\rho_R\right)\psi_\pi \pi_t + \left(1-\rho_R\right)\psi_y\left(y_t - g_t\right) + \varepsilon_{R,t} && (48)\\
u_t &= \rho_u u_{t-1} + \varepsilon_{u,t} && (49)
\end{aligned}
$$

$$g_t = \rho_g g_{t-1} + \varepsilon_{g,t}. \tag{50}$$

Here the parameters to be estimated are $\theta = (\sigma, \beta, \kappa, \psi_\pi, \psi_y, \rho_R, \rho_g, \rho_u, \sigma_R, \sigma_g, \sigma_u)$. u_t is the cost-push shock. A positive shock in u_t would lead to an increase in the concurrent inflation rate while simultaneously decreasing 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. For each given parameter combination, the welfare losses experienced by a representative household are second-order approximated, proportional to

$$E_0\left\{\sum_{t=0}^{\infty}\beta^t\left(\lambda\pi_t^2 + y_t^2\right)\right\}, \tag{51}$$

where $\lambda = \frac{1}{\nu\kappa}$.[36] However, 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.[37] 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 (51) 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).

First, I select a prior using the same hyperparameters as An and Schorfheide (2007), with the modification of substituting ρ_g for ρ_u, and σ_g for σ_u. Next, I choose a different posterior that yields the same posterior predictive distribution, employing a strategy similar to that used in the Cochrane model. I categorize four pairs of policies (ψ_π, ψ_y) into two groups[38] and determine the optimal policy based on various weights and posterior distributions. The findings are summarized in Table **??**.

Beyond what is already shown in Table **??**, Table **??** 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.

[36] The computation details can be found in Woodford (2003a), Galí (2015), Davig (2016).

[37] The weight choices are scattered to include most calibration choices in the literature and the rule of thumb choice $\frac{1}{\nu\kappa} = 1$.

[38] The policies I compare are from Galí (2015) Table 4.1.

	$\frac{1}{\nu\kappa}=\frac{1}{3}$		$\frac{1}{\nu\kappa}=1$		$\frac{1}{\nu\kappa}=3$		$\frac{1}{\nu\kappa}=10$		$\frac{1}{\nu\kappa}=30$	
(ψ_π,ψ_y)	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$
(1.5, 0)					✓		✓	✓	✓	✓
(1.5, 0.125)	✓	✓	✓	✓		✓				
(ψ_π,ψ_y)	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$
(1.5, 1)	✓	✓	✓	✓						
(5, 0)					✓	✓	✓	✓	✓	✓

Table 7: Policy Comparison under Different Distributions and Weights. The check mark indicates that, for each prior distribution $p_i(\theta)$, where $i = 1, 2$, the policy associated with it results in a lower welfare loss according to its posterior.

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	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$	$p_1(\theta)$	$p_2(\theta)$
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

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

$$\begin{aligned}&\arg\min_{\psi_\pi,\psi_y}\int L_W\left(\pi_t,y_t\mid\theta_{-(\psi_\pi,\psi_y)},\psi_\pi,\psi_y\right)d\pi_{\theta|Y}\\&\text{subject to Equations (46)}-\text{(50)}\end{aligned},\tag{52}$$

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.

Results in Table 8 reveal that while the welfare loss appears relatively consistent across the two prior distributions, the optimal Taylor rule parameters exhibit notable sensitivity to this selection. Indeed, the sensitivity of these policy-related parameters to variations in prior choices becomes more pronounced when considering a broader range of priors within the same class. To see that, I first estimate each parameter's range of posterior means, using Algorithm 1. The results are shown in Table 9. The range of ψ_π is not at all informative but the parameters ψ_y, ρ_R, σ_R have identified sets that are a proper subsets of their support. Subsequently, I compute the optimal Taylor rule policy parameters using values of ρ_R and σ_R from the true identified set, as reported in Table 10.

The range of optimal Taylor rule parameters that are consistent with the true identified set of structural parameters 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

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)

Table 10: Range of Optimal Policy Parameters within the True Identified Set

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	$(3.01, +\infty)$	$(1.00, +\infty)$	(1.60, 1490.36)	$(4.14, +\infty)$	$(9.95, +\infty)$
ψ_y	$(10.97, +\infty)$	$(0.00, +\infty)$	(0.00, 285.72)	$(0.00, +\infty)$	$(0.00, +\infty)$
$10^4\times$loss	(2.51, 2.53)	(2.26, 3.36)	(4.22, 5.07)	(5.41, 11.80)	(6.85, 35.56)

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 to start with, another prior with the same predictive distribution may result in a completely different optimal policy suggestion. Facing policy sensitivity, policy-makers can still make point-wise recommendations under certain statistical decision criteria (e.g., a min-max rule with respect to some ϵ-contaminated neighborhood of a given prior; see Berger (2013), Yata (2021), Giacomini et al. (2019), 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 example.

In this paper, I do not attempt to establish a rule for selecting the single optimal policy based on robust Bayesian outputs. Instead, I focus on presenting a method a method that allows for sensitivity analysis. As Giacomini and Kitagawa (2021) argue in their work, 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. This tool is also valuable in its own right for

separating 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. 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), but in a *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 to ease the computation burden. The mapping between semi-structural parameters and structural parameters used 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 constructed semi-structural parameters are point identified. This observation, in conjunction with the results presented in Table 14 and 15, suggests that all structural parameters become point identified if ε_p and ε_w are fixed. The estimated 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.

From the examples discussed in this section, we observe that in linearized DSGE models, identification issues typically arise within a small-dimensional subvector. This is partly because researchers generally have a solid understanding of their model and tend to avoid excessive parameterization. Additionally, although the computational cost of solving the model and reducing the polynomial system to its Gröbner basis escalates with the size of the model, the cost of sampling from observationally equivalent sets does not increase, provided the number of free parameters remains constant.

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

	True value	Posterior mean	(Robust) Bayesian CR
$\alpha_1 = \frac{(\gamma-1+\delta)\alpha}{\beta^{-1}\gamma^{\sigma_c}-1+\delta}$	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]

4.4 Discussions

4.4.1 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), Andreasen (2013), Morris (2014), Ivashchenko (2014), Herbst and Schorfheide (2015), Aruoba et al. (2017), Andreasen et al. (2018)). Since higher-order approximations retain more information from the nonlinear function, one might expect these models to have extra identification power that are not present when only linearized. 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.

Schmitt-Grohé and Uribe (2004) introduced the perturbation method, which uses second-order Taylor expansions around the steady state to approximate the solution. While Taylor expansions are relatively easy to compute, they can lead to explosive or non-stationary processes. To address these issues, Kim et al. (2008) developed the pruning method, which has since become widely adopted. The basic idea of pruning is to eliminate the terms from the policy functions that have higher-order effects than the approximation order.[39]

In addition, (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), and can be "minimized" into a minimal ABCD system. Consequently, Algorithm 1 remains applicable. However, the loss of Gaussianity means that the observationally equivalent sets found by the algorithm will be a superset of the actual $K(\theta)$.

4.4.2 Indeterminacy

It has been well-established that linear rational expectation models can exhibit multiple solutions with plausible parameter values (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

[39] See Andreasen et al. (2018) for a complete analysis.

and indeterminacy from a frequency domain perspective.

Kocięcki and Kolasa (2023) argue that Theorem 6 can handle indeterminate parametrization when a sufficient number of expectation errors are redefined as new fundamentals. However, a fixed structure of ABCD representation is needed, i.e., identification analysis needs to be done within either the determinate or the indeterminate part of the parameter space.

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. It's often assumed that this continuity property holds, but verifying this assumption can be exceedingly challenging.

5 Conclusions

The sensitivity of standard Bayesian results in set-identified models is well-known; however, it had not been investigated much in highly structural models such as the 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 set-identified structural models, I developed a novel robust Bayesian method to perform sensitivity analysis. The algorithm 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 structural models can suffer from other important issues, such as weak identification in DSGE models Canova and Sala (2009), Müller (2012), Guerron-Quintana et al. (2013), Andrews and Mikusheva (2015), Ho (2022), I do not attempt to address those issues in this paper. Although there's always the option to adjust the model (e.g., by introducing additional shocks) when identification fails, the method presented in this paper proves especially valuable in several scenarios:

1. When researchers want to make minimal adjustments or calibrations to the model.

2. When researchers possess confidence in their model setup and seek to assess the implications for estimates, even in situations where the model lacks point identification.

3. When policy-makers want to make robust decisions based on priors of equal predictive power.

The applications of the algorithm presented in this paper 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.

A Proofs

Proof for Lemma 1: 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 $\pi_\theta, \bar{\pi}_\theta \in \Pi_\theta(\pi_K)$,

$$\int_\Theta p(y \mid \theta) d\pi_\theta = \int_{\mathcal{F}} p(y \mid K) d\pi_K = \int_\Theta p(y \mid \theta) d\bar{\pi}_\theta \tag{53}$$

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

First, under Assumption 1, $K(\theta)$ 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 $A \in \mathcal{A}$,

$$\mathbf{1}_{\{K \subset A\}} \leq \pi_{\theta|K}(A), \quad \pi_K - a.s. \tag{54}$$

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 (54) is equivalent to showing

$$\int_B \mathbf{1}_{K_1^A} d\pi_K \leq \int_B \pi_{\theta|K}(A) d\pi_K \tag{55}$$

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

$$\begin{aligned}
\int_B \pi_{\theta|K}(A) d\pi_K &\geq \int_{B \cap K_1^A} \pi_{\theta|K}(A) d\pi_K && (56)\\
&= \pi_\theta\Big(A \cap \big\{\theta : K(\theta) \in B, K(\theta) \subset A\big\}\Big) && (57)\\
&= \pi_\theta\Big(\big\{\theta : K(\theta) \in B, K(\theta) \subset A\big\}\Big) && (58)\\
&= \pi_K\left(B \cap K_1^A\right) && (59)\\
&= \int_B \mathbf{1}_{K_1^A} d\pi_K && (60)
\end{aligned}$$

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 $\underline{\pi}^A_{\theta|K} \in \Pi_{\theta|K}$ that achieves the lower bound of $\pi_{\theta|K}$ obtained in (54), π_K almost surely. Consider the following three subsets of $\mathcal{F}$,

$$\begin{aligned}
K_0^A &= \big\{K \in K(\Theta) : K \cap A = \emptyset\big\}, && (61)\\
K_1^A &= \big\{K \in K(\Theta) : K \subset A\big\}, && (62)\\
K_2^A &= \big\{K \in K(\Theta) : K \cap A \neq \emptyset \text{ and } K \cap A^c \neq \emptyset\big\} && (63)
\end{aligned}$$

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

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 $\pi^A_{\theta|K} \in \Pi_{\theta|K}$, for any $\tilde{A} \in \mathcal{A}$

$$\underline{\pi}^A_{\theta|K}(\tilde{A}) = \begin{cases} \pi^A_{\theta|K}(\tilde{A}), & \text{for } K \in K_0^A \cup K_1^A, \\ \mathbf{1}_{\{S^A(K) \in \tilde{A}\}}(K), & \text{for } K \in K_2^A. \end{cases} \tag{64}$$

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

$$\int_B \underline{\pi}^A_{\theta|K}(A) d\pi_K = \int_B \underline{\pi}^A_{\theta|K}(A \cap K) d\pi_K \tag{65}$$

$$= \int_{B\cap K_0^A} \underline{\pi}^A_{\theta|K}(A\cap K)d\pi_K + \int_{B\cap K_1^A} \underline{\pi}^A_{\theta|K}(A\cap K)d\pi_K \tag{66}$$

$$+ \int_{B\cap K_2^A} \underline{\pi}^A_{\theta|K}(A\cap K)d\pi_K \tag{67}$$

$$= 0 + \int_{B\cap K_1^A} \underline{\pi}^A_{\theta|K}(A\cap K)d\pi_K + 0 \tag{68}$$

$$= \int_B \mathbf{1}_{K_1^A} d\pi_K, \tag{69}$$

where the first equality is from $\underline{\pi}^A_{\theta|K} \in \Pi_{\theta|K}$, the third equality follows because $A\cap K = \emptyset$ 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 1. We first show the special case of $\eta(\theta) = \theta$. Expanding the integral we have

$$\inf_{\pi_{\theta|Y}\in\Pi_{\theta|Y}} \pi_{\theta|Y}(A) = \inf_{\pi_{\theta|K}\in\Pi_{\theta|K}} \int_{\mathcal{F}} \pi_{\theta|K}(A)\, d\pi_{K|Y}, \quad A \in \mathcal{A} \tag{70}$$

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,

$$\inf_{\pi_{\theta|Y}\in\Pi_{\theta|Y}} \pi_{\theta|Y}(A) = \inf_{\pi_{\theta|K}\in\Pi_{\theta|K}} \int_{\mathcal{F}} \pi_{\theta|K}(A)\, d\pi_{K|Y} \tag{71}$$

$$= \inf_{\pi_{\theta|K}\in\Pi_{\theta|K}} \int_{\mathcal{F}} \mathbf{1}_{\{K\subset A\}}(K) d\pi_{K|Y} \tag{72}$$

$$= \pi_{K|Y}(\{K\subset A\}) \tag{73}$$

$$= \pi_{\theta|Y}\left(\left\{\theta : K(\theta)\subset A\right\}\right) \tag{74}$$

The last equality comes from equation (9).

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

$$\overline{\pi}_{\theta|Y}(A) = 1 - \underline{\pi}_{\theta|Y}(A^c) \tag{75}$$

$$= 1 - \pi_{\theta|Y}\left(\left\{\theta : K(\theta)\subset A^c\right\}\right) \tag{76}$$

$$= 1 - \pi_{\theta|Y}\left(\left\{\theta : K(\theta)\cap A = \emptyset\right\}\right) \tag{77}$$

$$= \pi_{\theta|Y}\left(\left\{\theta : K(\theta)\cap A \neq \emptyset\right\}\right) \tag{78}$$

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 $\tilde{A} \in \mathcal{A}$,

$$\overline{\pi}^{A}_{\theta|K}(\tilde{A}) = \begin{cases} \pi^{A}_{\theta|K}(\tilde{A}), & \text{for } K \in K^{A}_{0} \cup K^{A}_{1}, \\ \mathbf{1}_{\{S^{A^c}(K) \in \tilde{A}\}}(K), & \text{for } K \in K^{A}_{2}. \end{cases} \tag{79}$$

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 $\mathcal{A}$.

$$\begin{aligned} \inf_{\pi_{\eta|Y} \in \Pi_{\eta|Y}} \pi_{\eta|Y}(D) &= \inf_{\pi_{\theta|Y} \in \Pi_{\theta|Y}} \pi_{\theta|Y}\left(\eta^{-1}(D)\right) && (80) \\ &= \inf_{\pi_{\theta|K} \in \Pi_{\theta|K}} \int_{\mathcal{F}} \mathbf{1}_{\{K \subset \eta^{-1}(D)\}}(K) d\pi_{K|Y} && (81) \\ &= \pi_{K|Y}\left(\left\{K \subset \eta^{-1}(D)\right\}\right) && (82) \\ &= \pi_{\theta|Y}\left(\left\{\theta : \eta(K(\theta)) \subset D\right\}\right) && (83) \end{aligned}$$

□

Proof for Lemma 2:

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),

$$\pi_{\theta|Y}(A) = \frac{\int_A p(y \mid \theta) d\pi_\theta}{\int p(y \mid \theta) d\pi_\theta}, \quad \text{for any } A \in \mathcal{A}. \tag{84}$$

Plug in this formula

$$\pi^{*}_{K|Y}(B) = \pi_{\theta|Y}(K^{-1}(B)) = \pi_{\theta|Y}(\{\theta : K(\theta) \in B\}) = \frac{\int_{\{\theta : K(\theta) \in B\}} p(y \mid \theta) \mathrm{d}\pi_\theta}{\int p(y \mid \theta) \mathrm{d}\pi_\theta}. \tag{85}$$

And this is equal to

$$\pi_{K|Y}(B) = \frac{\int_B p(y \mid K)\mathrm{d}\pi_K}{\int p(y \mid K)\mathrm{d}\pi_K} = \frac{\int_{\{\theta:K(\theta)\in B\}} \tilde{p}(y \mid K(\theta))\mathrm{d}\pi_\theta}{\int \tilde{p}(y \mid K(\theta))\mathrm{d}\pi_\theta} = \frac{\int_{\{\theta:K(\theta)\in B\}} p(y \mid \theta)\mathrm{d}\pi_\theta}{\int p(y \mid \theta)\mathrm{d}\pi_\theta}, \tag{86}$$

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*

$$\pi_\theta = \int_{\mathcal{F}} \pi_{\theta|K} d\pi_K, \quad \pi_\theta\left(\left\{\theta : K(\theta) \in \cdot\right\}\right) = \pi_K(\cdot), \tag{87}$$

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

Proof for Theorem 2:

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 2. Infimum case is similar.
Pick any two priors $\pi_\theta, \tilde{\pi}_\theta \in \Pi_\theta$,

$$\mathbb{E}_{\tilde{\pi}_\theta}\left[\eta(\theta)\right] \le \mathbb{E}_{\tilde{\pi}_\theta}\left[\overline{\eta}^*(\theta)\right] = \mathbb{E}_{\pi_\theta}\left[\overline{\eta}^*(\theta)\right] \tag{88}$$

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

$$\tilde{\Pi}_\theta = \left\{\tilde{\pi}_\theta \in \Pi_\theta : \mathbb{E}_{\tilde{\pi}_{\theta|K}}\left[\eta(\theta)\right] \ge \sup_{\theta' \in K} \eta(\theta) - \epsilon, \pi_K - \text{ almost surely }\right\} \tag{89}$$

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

$$\mathbb{E}_{\tilde{\pi}_\theta}\left[\eta(\theta)\right] = \mathbb{E}_{\pi_K}\left[\mathbb{E}_{\tilde{\pi}_{\theta|K}}\left(\eta(\theta)\right)\right] \ge \mathbb{E}_{\pi_K}\left[\sup_{\theta' \in K} \eta(\theta) - \epsilon\right] = \mathbb{E}_{\pi_\theta}\left[\overline{\eta}^*(\theta)\right] - \epsilon \tag{90}$$

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 3: From Theorem 1,

$$\inf_{\pi_{\eta|Y}\in\Pi_{\eta|Y}} \pi_{\eta|Y}\left((-\infty,q]\right) = \pi_{\theta|Y}\left(\left\{\theta : \eta\left(K(\theta)\right)\subset(-\infty,q]\right\}\right) \tag{91}$$

$$= \pi_{\theta|Y}\left(\left\{\theta : \overline{\eta}^{*}(\theta)\leq q\right\}\right) \tag{92}$$

whereas

$$\sup_{\pi_{\eta|Y}\in\Pi_{\eta|Y}} \pi_{\eta|Y}\left((-\infty,q]\right) = \pi_{\theta|Y}\left(\left\{\theta : \eta\left(K(\theta)\right)\cap(-\infty,q]\neq\varnothing\right\}\right) \tag{93}$$

$$= \pi_{\theta|Y}\left(\left\{\theta : \underline{\eta}^{*}(\theta)\leq q\right\}\right) \tag{94}$$

□

Proof for Lemma 3:

Under a linearized DSGE model with Gaussian shocks, with stability assumption 6, 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. In other words, it guarantees the same information contained in $\Phi_Y(z;\theta)$ and the likelihood $p(y \mid \theta)$.

□

B Supplementary Tables and Graphs

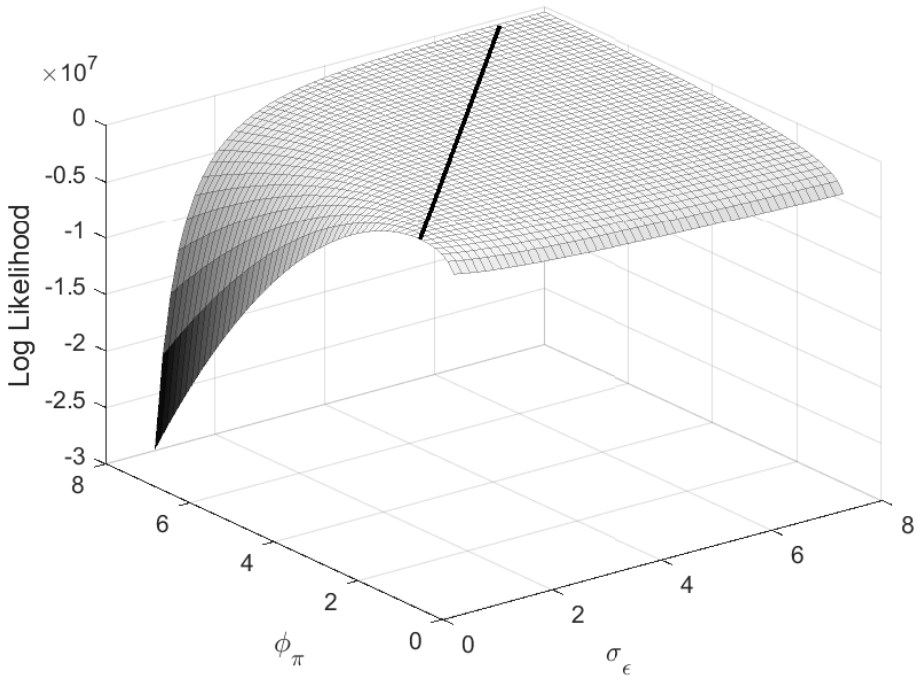


Figure 4: Log-likelihood of the Cochrane model when the correlation coefficient ρ is fixed at 0.8 and sample size $T = 1,000,000$. The thick line is the maxima ridge of the likelihood function.

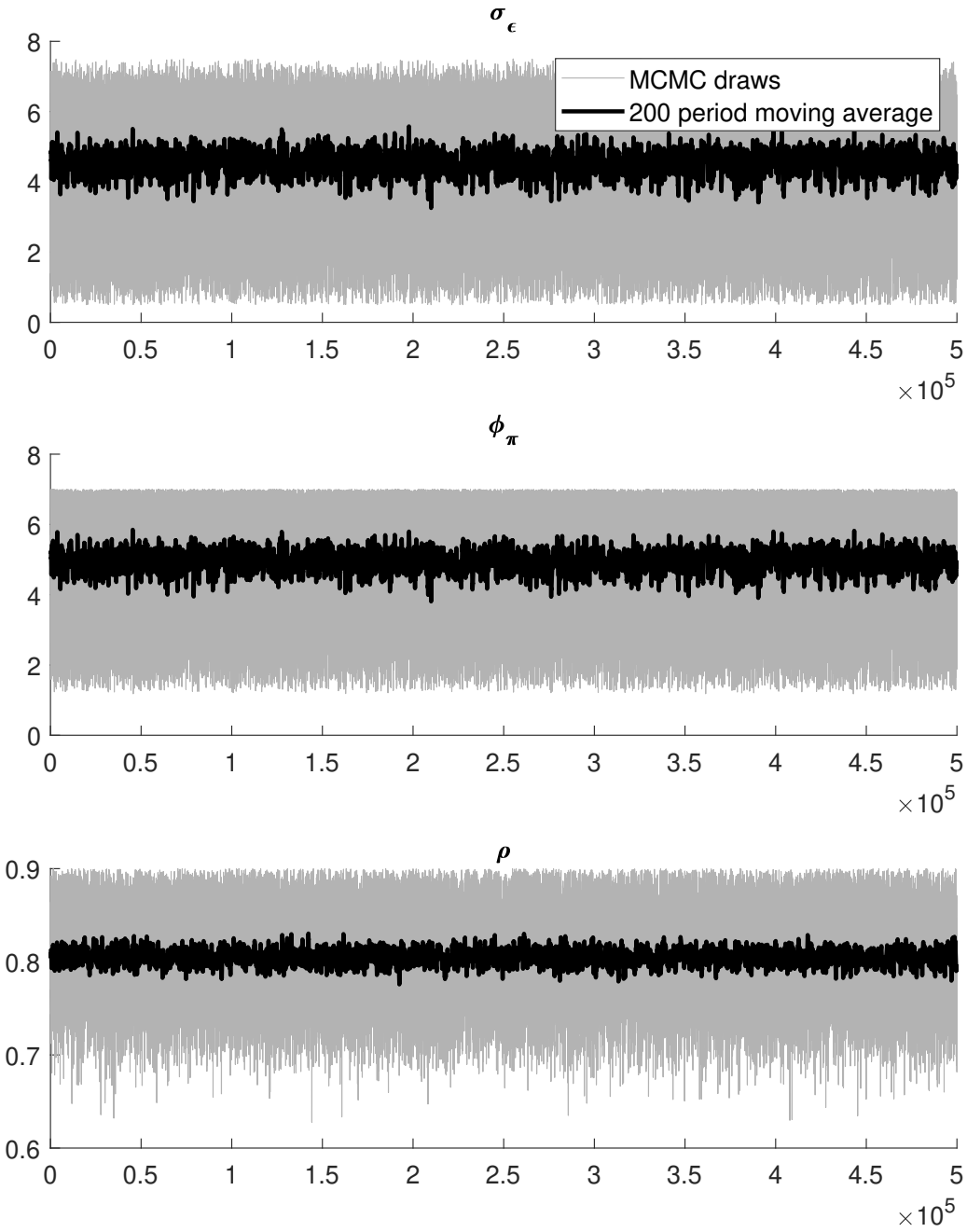


Figure 5: Trace plot of MCMC draws parameters

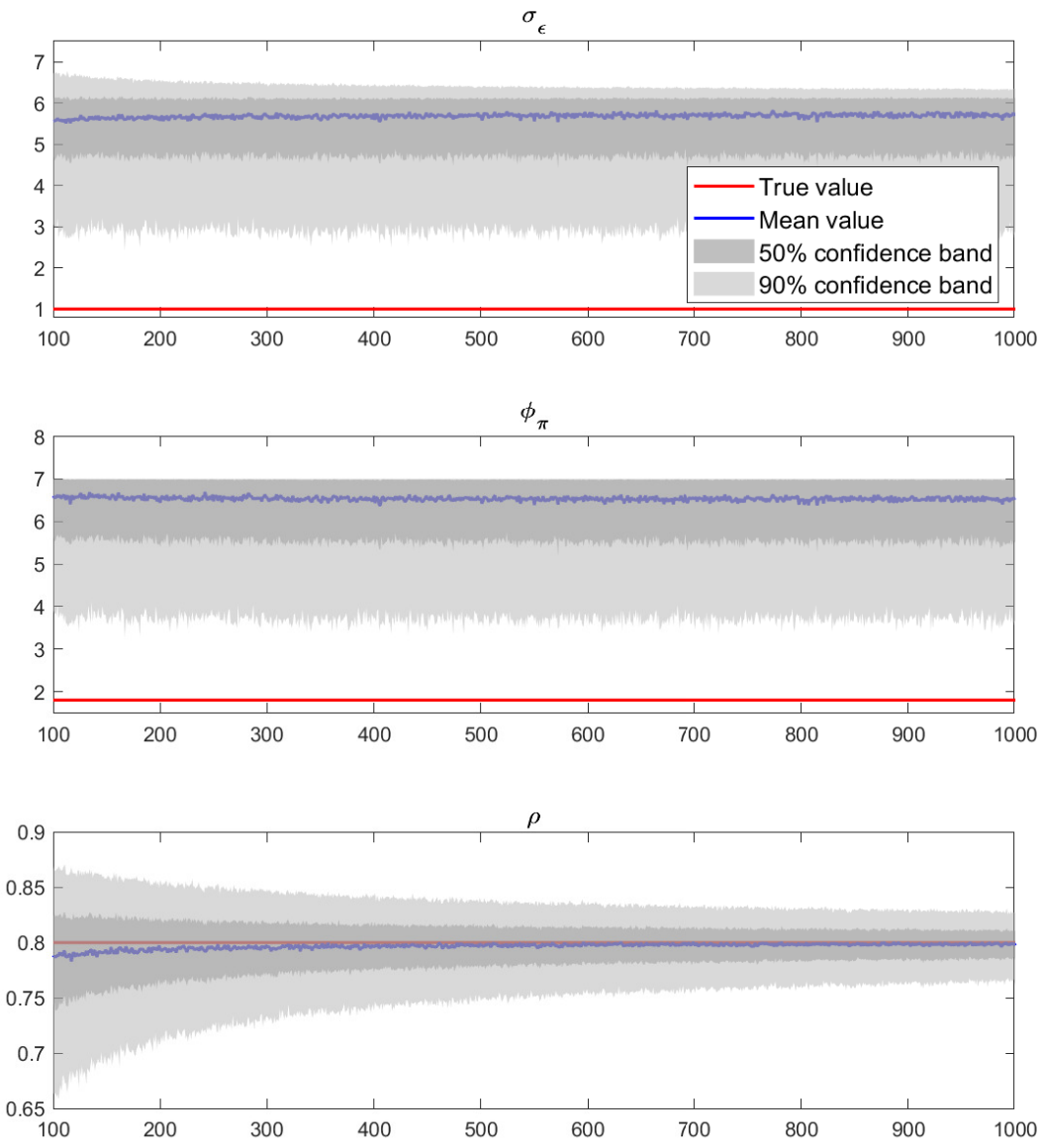


Figure 6: Maximum A Posteriori (MAP) Estimates of structural parameters with 1000 replications

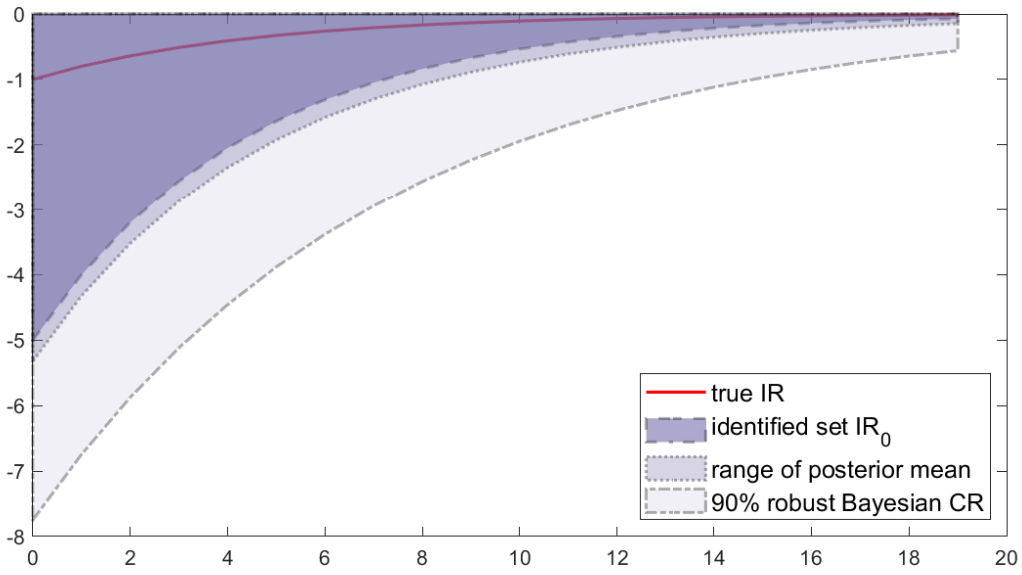


Figure 7: IRF in the Cochrane model. The red line is the impulse response with true parameter values. IR_0 is the theoretical identified set of IR. Range of posterior mean and the robust Bay

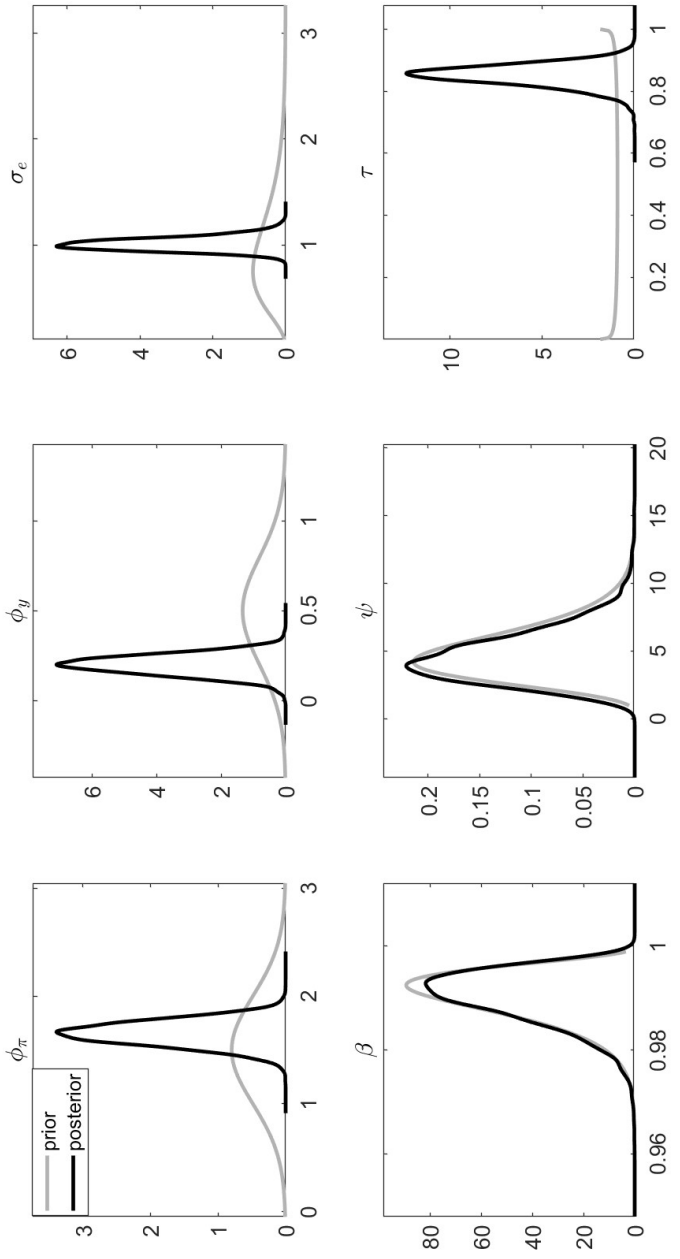


Figure 8: Prior and posterior for the 3-equation model

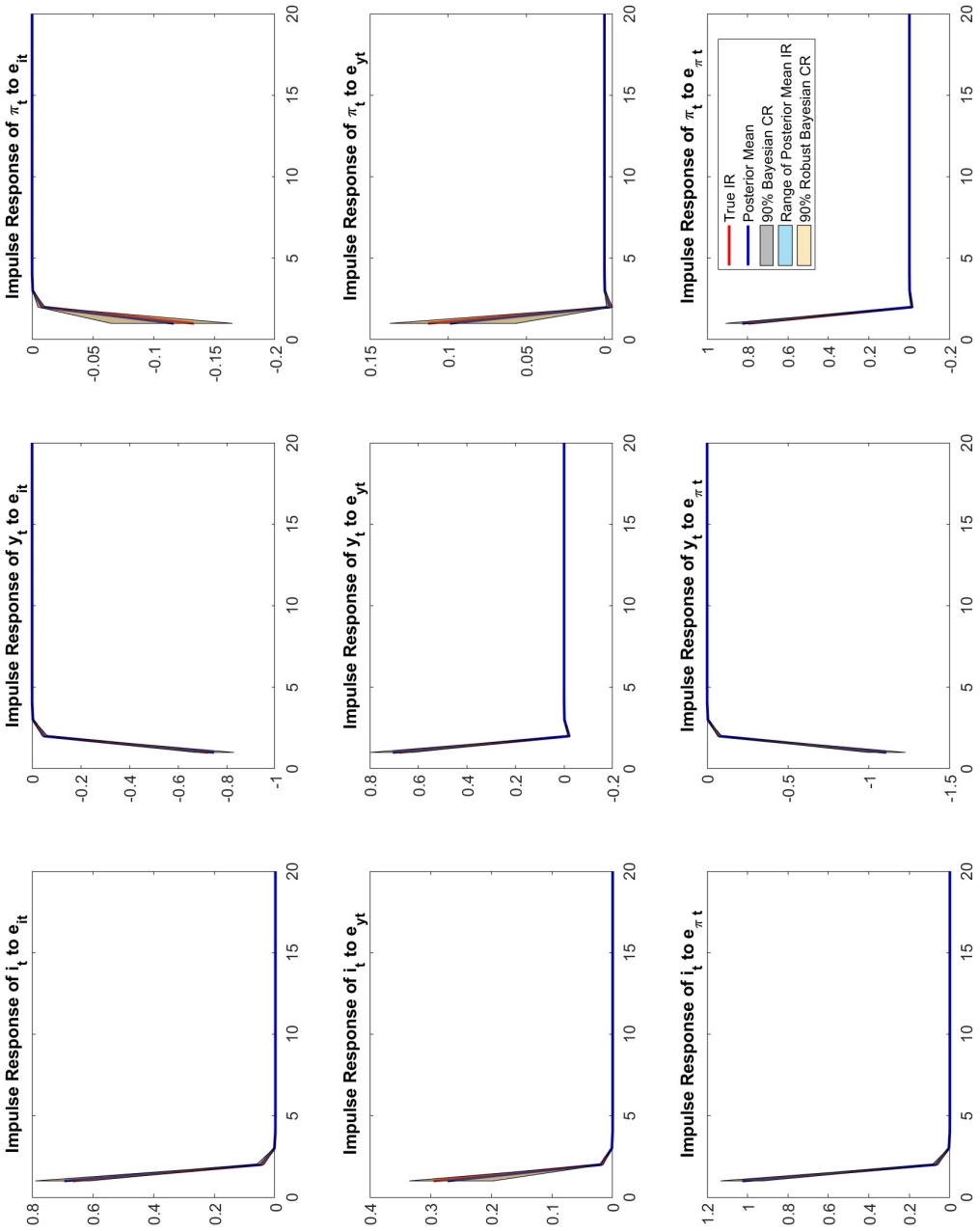


Figure 9: IRF in the three-equation NK model

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

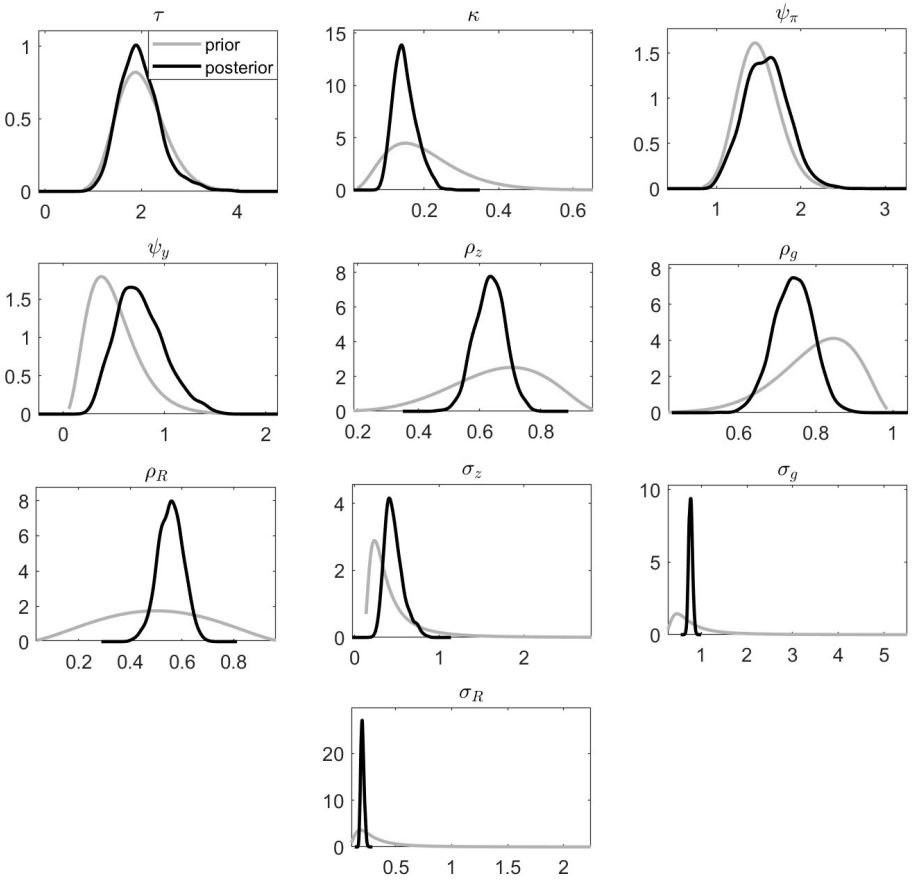


Figure 10: Prior and posterior for the AS model

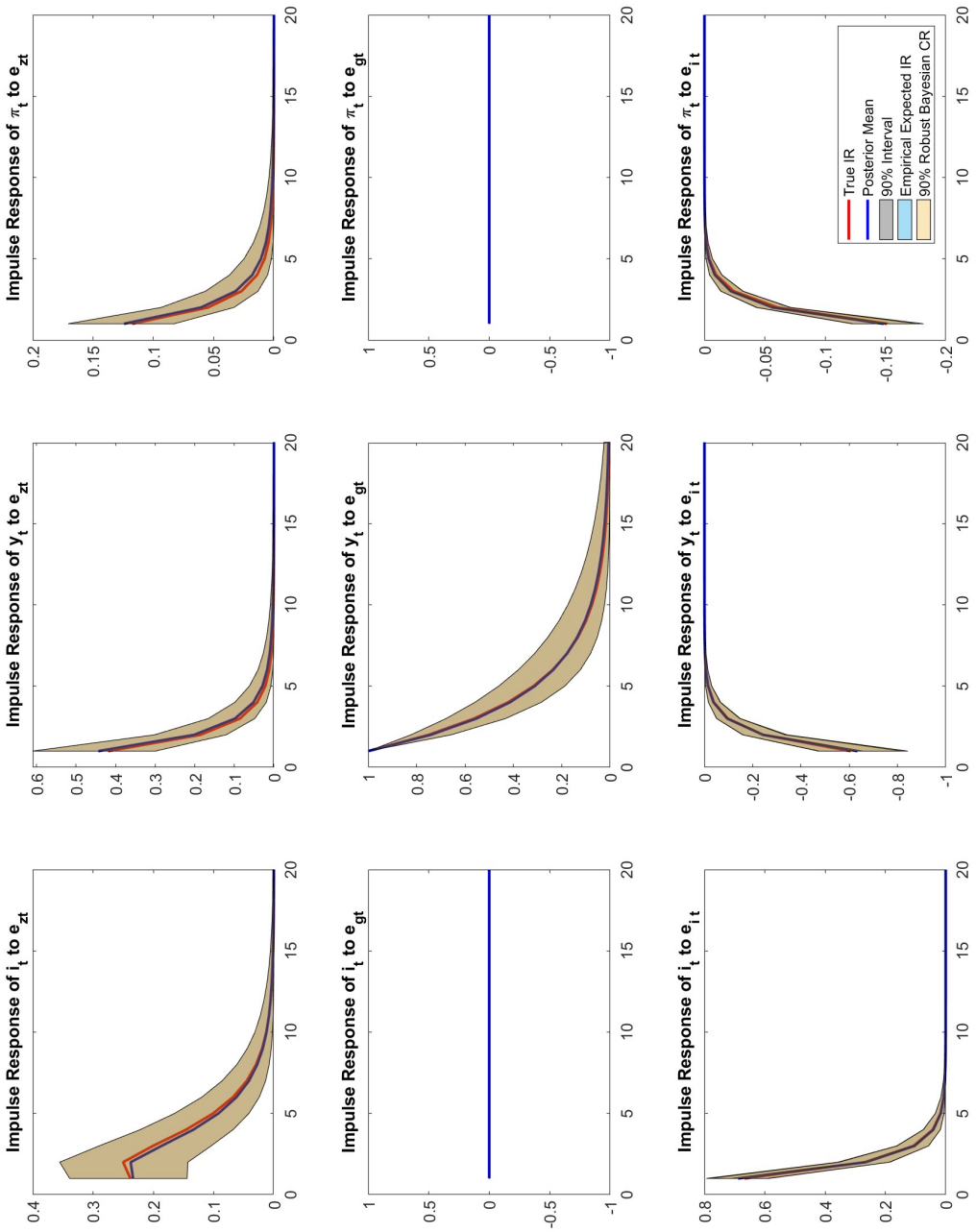


Figure 11: IRF in the AS model

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

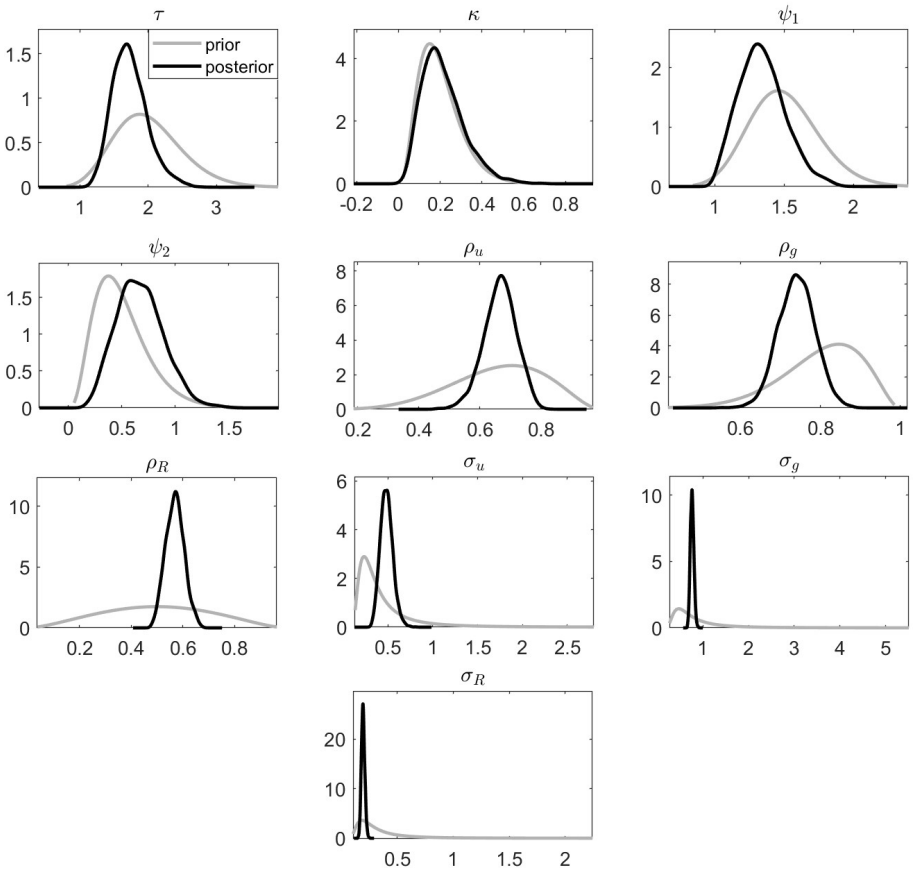


Figure 12: Prior and posterior for the cost-push shock model

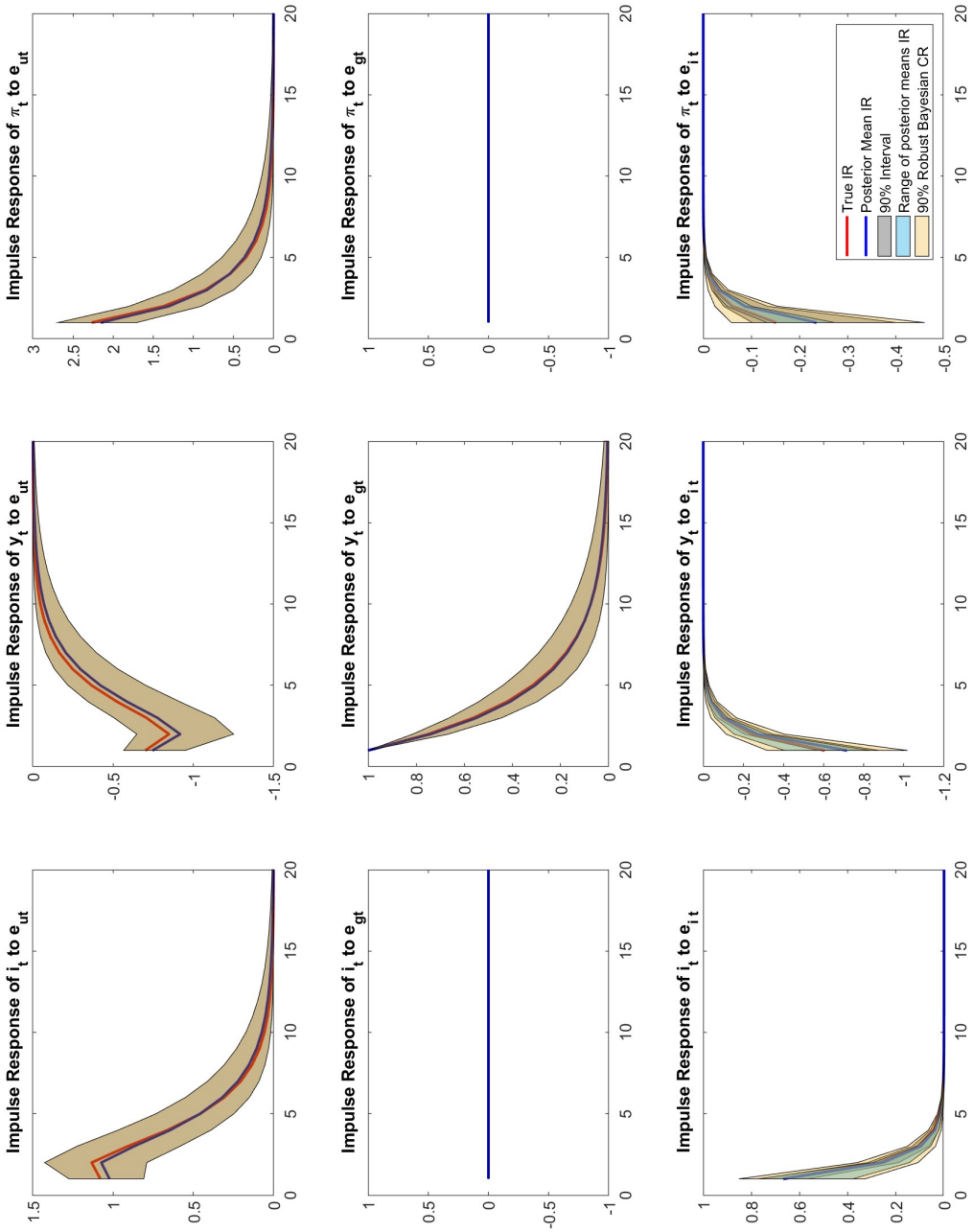


Figure 13: IRF in the Cost-push model

Table 14: Estimated Identified Set of Untransformed Structural Parameters for the Smets-Wouters Model

	True value	Posterior mean	(Robust) Bayesian CR
$100(\bar{\gamma}-1)$	0.43	0.43	[0.43,0.43]
$\bar{\pi}$	0.70	0.75	[0.69, 0.80]
$\bar{l}$	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]

Table 15: Estimated Identified Set of Parameters of Shock Processes for the Smets-Wouters Model

	True value	Identified set	(Robust) Bayesian CR
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]

References

Adjemian, Stéphane, Michel Juillard, Fréderic Karamé, Willi Mutschler, Johannes Pfeifer, Marco Ratto, Normann Rion, and Sébastien Villemot, "Dynare: Reference Manual, Version 6," Dynare Working Papers 80, CEPREMAP February 2024.

Al-Sadoon, Majid M 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," *Technical notes*, 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.

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.

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

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 and L Mark Berliner, "Bayesian input in Stein estimation and a new minimax empirical Bayes estimator," *Journal of Econometrics*, 1984, *25* (1-2), 87–108.

__ **and** __ , "Robust Bayes and empirical Bayes analysis with ε-contaminated priors," *The Annals of Statistics*, 1986, pp. 461–486.

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.

__ **, Elías Moreno, Luis Raul Pericchi, M Jesús Bayarri, José M Bernardo, Juan A Cano, Julián De la Horra, Jacinto Martín, David Ríos-Insúa, Bruno Betrò et al.**, "An overview of robust Bayesian analysis," *Test*, 1994, *3* (1), 5–124.

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.

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.

Chamberlain, Gary and Edward E Leamer, "Matrix weighted averages and posterior bounds," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1976, *38* (1), 73–84.

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," *Econometrica*, 2007, *75* (5), 1243–1284.

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.

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.

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.

Giacomini, Raffaella, Toru Kitagawa, and Harald Uhlig, "Estimation under ambiguity," Technical Report, Cemmap working paper 2019.

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

Good, Isidore Jacob, "Probability and the Weighing of Evidence," 1950.

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.

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

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.

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.

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.

Lavine, Michael, "Sensitivity in Bayesian statistics: the prior and the likelihood," *Journal of the American Statistical Association*, 1991, *86* (414), 396–399.

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

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.

Norets, Andriy and Xun Tang, "Semiparametric inference in dynamic binary choice models," *Review of Economic Studies*, 2014, *81* (3), 1229–1262.

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.

Plagborg-Møller, Mikkel, "Bayesian inference on structural impulse response functions," *Quantitative Economics*, 2019, *10* (1), 145–184.

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, "Inference in dynamic stochastic general equilibrium models with possible weak identification," *Quantitative Economics*, 2014, *5* (2), 457–494.

__ **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.

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

Robbins, Herbert, "Asymptotically subminimax solutions of compound statistical decision problems," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1951.

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.

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.

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.

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.

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

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.