Robust Bayesian Estimation and Inference for Dynamic Stochastic General Equilibrium Models

Yizhou Kuang[†]

[†]University of Manchester

Last updated: June 5^{th}, 2023

Motivation

- **DSGE models are widely used:**
  - U.S. Fed, Bank of Canada, Sveriges Riksbank etc.
- **Conclusions based on the models can be misleading because of 'identification':**
  - DSGE models are micro-founded, rich with parameters.
  - Multiple parameter vectors may yield same data generating process.
  - Standard Bayesian methods can be sensitive to prior choices.

Motivation

A monetary policy model (Cochrane 2011, JPE). Solved in its AR(1) form

$$\pi_t = \rho \pi_{t-1} + \frac{1}{\phi - \rho} \epsilon_t, \quad \phi > 1, |\rho| < 1, \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

parameter vector $(\phi, \sigma_\epsilon, \rho)$, Taylor rule parameter ϕ, monetary policy disturbance coefficient ρ, its standard error σ_ϵ. Inflation rate π_t is observed.

- Simulation: generate π_t, estimate $(\phi, \sigma_\epsilon, \rho)$ with different priors

Identification Failure

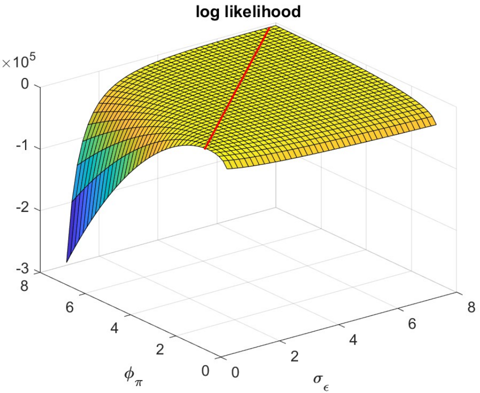
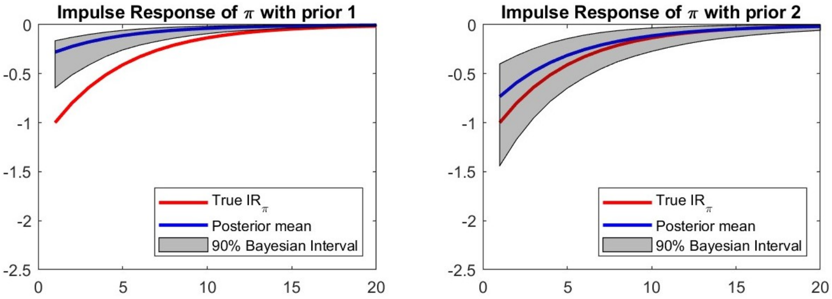


Figure: Likelihood function while fix $\rho = 0.8$ and sample size = 1 million

- Flat maxima along the $\sigma_\epsilon = \phi - 0.8$ line.
- Prior sensitivity becomes obvious.

Prior Sensitivity



- 1-unit monetary policy disturbance shock on inflation.
- Impulse response with two different priors (but has the same distribution over $(\rho, \frac{\sigma_\epsilon}{\phi-\rho})$).

Prior Sensitivity

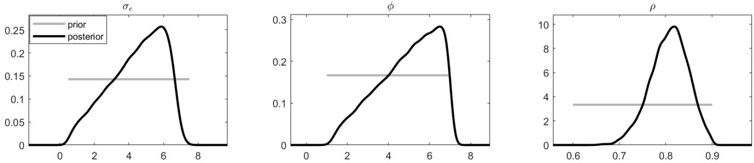


Figure: Cochrane model prior/posterior distribution with uniform priors

- The posterior of σ_ϵ and ϕ are extremely informative even if only $\frac{\sigma_\epsilon}{\phi - 0.8}$ is identified.
- Reason? Joint likelihood density more concentrated on areas with higher values of ϕ and σ_ϵ.

Research Question

- Given a DSGE model and observed data.
  - ∗ Identification/Sensitivity : Is there an identification issue?
    Is there a way to provide the entire set of parameters of interest, robust of priors?
  - ∗ Policy implications: Is it possible the identified set of parameters will agree on a single policy? What if it is not?

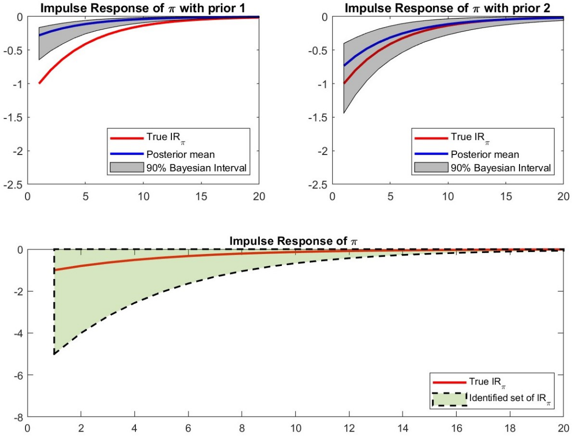
- Overview of the algorithm:

S.1 Run standard Bayesian estimation, get posterior draws of θ from a given prior $p(\theta)$.

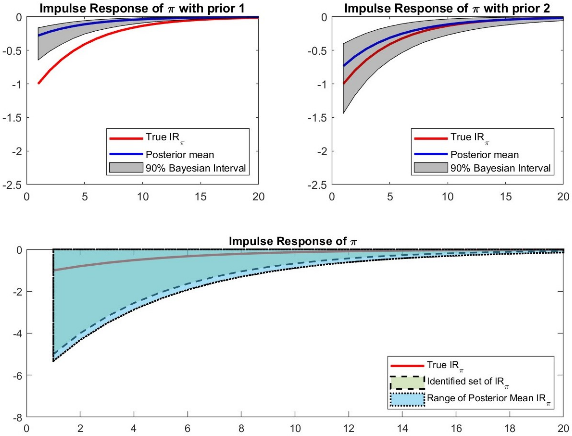
S.2* Optimize over the observationally equivalent set of parameters of each draw, find the lower and upper bounds of parameters of interest.

S.3 Average the lower/upper bounds for means and quantiles.

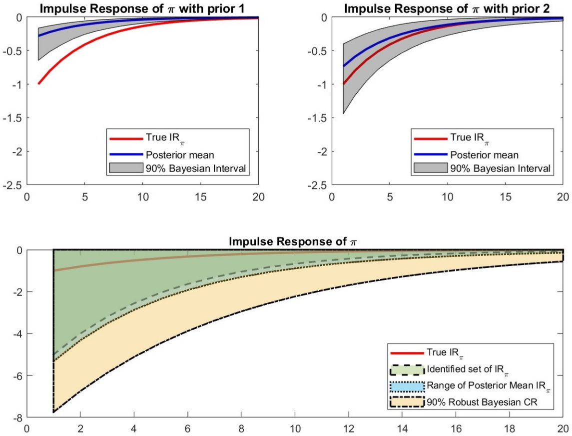
Preview of Results - Identification



Preview of Results - Inference



Preview of Results - Inference



Literature and Contributions

- **Identification in DSGE models**: Canova and Sala (2009), Iskrev (2010), Komunjer and Ng (2011), Qu and Tkachenko (2012), Qu and Tkachenko (2017), Kociecki and Kolasa (2018), Kociecki and Kolasa (2021)
- **Robust Bayesian analysis**: Berger et al. (1994), Berger, Insua, and Ruggeri (2000), Gustafson (2009), Giacomini and Kitagawa (2021), Ke, Montiel Olea, and Nesbit (2022), Giacomini, Kitagawa, and Read (2022)
- **Contribution of this paper**:
  - A robust Bayesian algorithm for DSGE models that is easy to implement and has a theory foundation.
  - "Global" identification rather than identification at a given parameter value (KK21).
  - DSGE model, which has some additional complications (GK21).
  - Exact identification rather than weak identification (Muller 2011, Andrews and Mikusheva 2015, Ho 2022, etc.)

Model Assumptions

Assumption (Linearity)

I work with linearized DSGE model with Gaussian shocks (first order perturbation).

Assumption (Determinacy)

Solution to the LREM is unique, i.e. no indeterminacy.

Assumption (Polynomial)

Structural parameters enter LREM in algebraic operations (addition, subtraction, multiplication, division and exponentiation by a rational number).

- e.g. NKPC in Gali (2015): $\pi_t = \beta E_t \{\pi_{t+1}\} + \lambda \left(\sigma + \frac{\varphi+\alpha}{1-\alpha}\right) \tilde{y}_t$

Remark: Baseline DSGE models like An& Schorfheide (2007) and Smets and Wouters (2007) satisfies these conditions. Some HANK models also do.

Extension and limitations

- The method works as long as the model can be rewritten as a state-space form.
- Violation of non-linearity/Gaussianity
  - Andreason, Fernández-Villaverde and Rubio-Ramírez (2017)
  - $\Rightarrow$ A superset of identified parameters.
- Can deal with much more sophisticated DSGE models than Smets and Wouters (2007)
  - Number of partially-identified parameters needs to be small, or the partially identified sets can be expressed as cross product of low dimensional sets.

Definitions

Definition (OE)

Parameter $\bar{\theta}$ is observationally equivalent to θ if they have the same likelihood $p(y \mid \theta)$ for all data realization y.

- A property independent of data

Definition (Identification)

θ is identified if it has no observationally equivalent parameters.

- Define the equivalence mapping $K : \Theta \to 2^{\Theta}$, that is, $p(y \mid \theta) = p(y \mid \bar{\theta})$ for all data y, if and only if $K(\theta) = K(\bar{\theta})$.

Definition (Reduced-form)

A C^1 function $\psi : \mathbb{R}^n \to \mathbb{R}^q$ is called a reduced-form parameter if it is identified. Consider $K(\theta)$ a generalized 'reduced-form' parameter.

Parameter Space

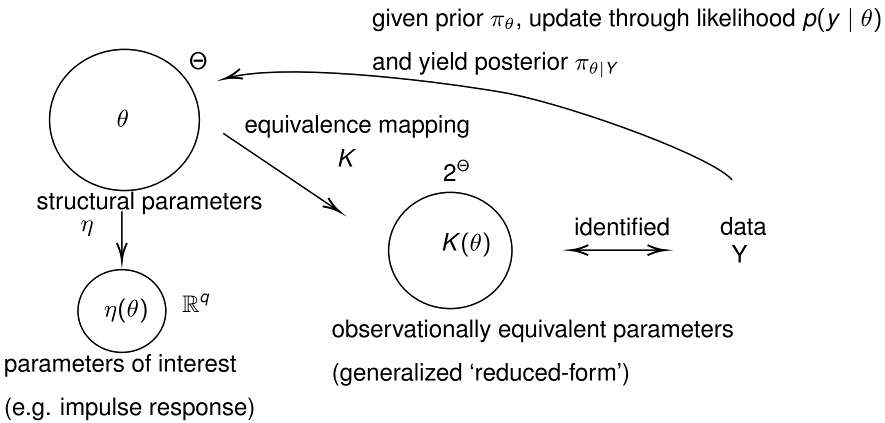


Figure: Connections between parameters

Example: Cochrane Model

Consider the full model

$$x_t = \rho x_{t-1} + \epsilon_t, \quad |\rho| < 1, \epsilon_t \sim N(0, \sigma_\epsilon)$$

$$i_t = r + E_t \pi_{t+1}$$

$$i_t = r + \phi \pi_t + x_t, \quad \phi > 1$$

Structural parameters are $\theta = (\rho, \phi, \sigma_\epsilon)$. The solution is equivalent to a AR(1) setting

$$\pi_t = \rho \pi_{t-1} + \frac{1}{\phi - \rho} \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

with reduced form parameters $\psi = (\rho, \frac{\sigma_\epsilon}{\phi - \rho})$, (ϕ, σ_ϵ) not jointly identifiable.

$K(\rho_0, \phi_0, \sigma_{0\epsilon}) = \{\rho_0, \phi_0, \sigma_{0\epsilon} \mid \rho = \rho_0, \frac{\sigma_\epsilon}{\phi - \rho} = \frac{\sigma_{0\epsilon}}{\phi_0 - \rho_0}\}$

Algorithm

Revisiting the algorithm

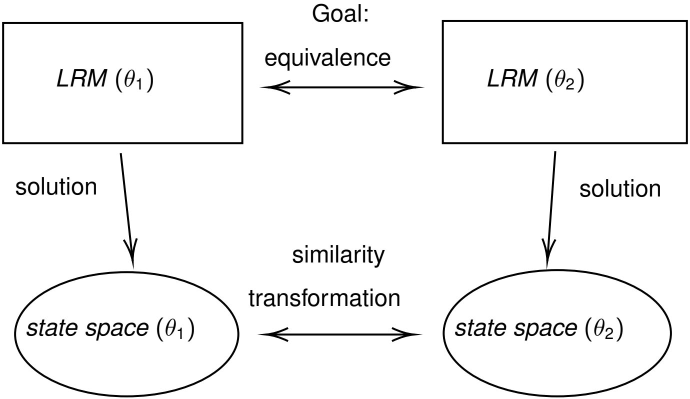
S.1 Run standard Bayesian estimation, get posterior draws of θ from a given prior $p(\theta)$.

S.2* Optimize over the observationally equivalent set of parameters of each draw, find the lower and upper bounds of parameters of interest.

- Finding the observationally equivalent set of a given parameter involves solving a polynomial system.

S.3 Average the lower/upper bounds for means and quantiles.

OE characterization (KN11, KK21)



Theoretical Results

Theorem (Posterior Mean)

For a given π_θ, let measurability and regularity assumptions hold, that is, given a prior π_θ absolutely continuous with respect to a σ-finite measure, we have a push-forward measure π_K of π_θ under K that is also absolutely continuous. 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 $\Pi_\theta(\pi_K)$ for a given π_K.

Theoretical results

Theorem (Consistency)

Let, in addition continuity assumption hold, assume further that induced prior π_K leads to a consistent posterior, and $\Theta \subset \mathbb{R}^p$, $H \subset \mathbb{R}^q$ for some $p, q < \infty$ are compact spaces. Then the Hausdorff distance between the set of posterior means and the convex hull of true identified set goes to zero almost surely as T increases, i.e.,

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

Example 1: Cochrane (2011)

Table: Estimated Identified Set for Cochrane Model

	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)

- Estimation of the range of posterior means closely approximates the identified set.
- Still working on the Smets & Wouters (2007) example.
- What is known about SW07 for now: not point-identified, and can be handled computationally.

Example 2: An& Schorfheide(2007)

Table: Estimated Identified Set 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)

Conclusion

In this paper, I address the following problems:

- The estimation results of the set-identified DSGE models are sensitive to the choice of priors (Identification)
  - ∗ Use a robust Bayesian algorithm, I can pick any 'reasonable' prior and obtain robust results.
  - ∗ I also prove it asymptotically finds the frequentist identified set.
- Researchers are silent about non-identified DSGE models (Inference)
  - ∗ The collection of posterior means of parameters of interest is given.
  - ∗ One may still have nontrivial conclusions even when the model suffers identification problems.
  - ∗ Robust Bayesian decision making from SDT can be useful.