Robust Bayesian Estimation and Inference for Dynamic Stochastic General Equilibrium Models

Yizhou Kuang[†]

[†]Department of Economics, Cornell University

Last updated: January 12th, 2023

Outline of Talk

Motivation:

What happens if a DSGE model is not identified?

Literature:

My contribution

Model and structure setup:

Assumptions and robust Bayes

Proposed algorithm:

Specific steps

Theoretical results:

Finite sample and asymptotic properties

Simulation and application:

Policy implications from estimation results

Motivation

DSGE models are widely used:

U.S. Fed, Banque de France, Sveriges Riksbank, IMF etc.

They are taught in almost every Econ Ph.D. program.

Analysis of the models is challenging because of ‘identification’:

DSGE models are micro-founded, rich with parameters.

Multiple parameter vectors may yield the same data generating process.

Standard Bayesian methods can be sensitive to prior choices.

Motivation - Estimation

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

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

parameter vector (ϕ, σ, ρ), Taylor rule parameter ϕ, monetary policy disturbance coefficient ρ, its standard error σ. Inflation rate π_t is observed.

Motivation - Estimation

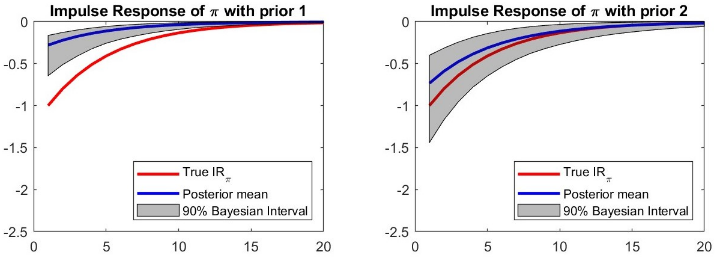
Table: Prior and Posterior Distribution of Structural Parameters

	True value	Prior distribution			Posterior distribution			
		Distr.	Mean	St. Dev.	Mode	Mean	5 percent	95 percent
σ	1	Uniform	4	2.02	5.82	4.43	1.95	6.77
ϕ	1.8	Uniform	4	1.73	6.49	4.91	2.78	7.00
ρ	0.8	Uniform	0.75	0.09	0.82	0.81	0.74	0.87

here

Motivation - Impulse Response

Figure: Impulse Responses from Standard Bayesian Estimation



The impulse response IR $(t, s, 1)$ here denotes the effect of an one-unit shock at time t on π_{t+s}

Prior 1 and prior 2 induce the same distribution over $(\rho, \frac{e}{})$

Motivation - Policy Analysis

Suppose a central bank using the following small-sized DSGE model

$$y_t = E_t[y_{t+1}] \quad \frac{1}{\sigma}(i_t \quad E_t[\pi_{t+1}]) + g_t \quad E_t[g_{t+1}]$$

$$\pi_t = \beta E_t[\pi_{t+1}] + \kappa(y_t \quad g_t) + u_t$$

$$i_t = \rho_R i_{t\;\;1} + (1 \quad \rho_R)\psi_{\;}\;\pi_t + (1 \quad \rho_R)\psi_y(y_t \quad g_t) + \varepsilon_{R;t}$$

$$u_t = \rho_u u_{t\;\;1} + \varepsilon_{u;t}, \quad g_t = \rho_g g_{t\;\;1} + \varepsilon_{g;t}.$$

is trying to use the estimated parameter (history)

$(\sigma, \beta, \kappa, \psi_{\;}, \psi_y, \rho_R, \rho_g, \rho_u, \sigma_R, \sigma_g, \sigma_u)$, to choose a policy rule $(\psi^0_{\;}, \psi^0_y)$ for

$$i_t = \psi^0_{\;}\pi_t + \psi^0_y(y_t \quad g_t)$$

that minimize expected welfare loss in the form of $\lambda\pi_t^2 + y_t^2$ in the future.

Motivation - Policy Analysis

Now consider two policies $(\psi\ \ , \psi_y) = (1.5, 0)$, and $(1.5, 0.125)$

Table: Policy Comparison under Different Distributions and Weights

	$\lambda = 1$		$\lambda = 3$		$\lambda = 10$	
$(\psi\ \ , \psi_y)$	post 1	post 2	post 1	post 2	post 1	post 2
(1.5, 0)			X		X	X
(1.5, 0.125)	X	X		X		

Policy choices are sensitive to prior choices as well.

Research Question

Given a DSGE model and observed data.

- Sensitivity analysis: How much can the posterior mean change as I change the prior (asymptotically)?
- Policy implications: Is it always possible to support a single policy rule robust of priors? If not, what is that range of policies?

Literature and Contributions

Identi cation 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 (2015), Liao and Simoni (2019), Giacomini and Kitagawa (2021), Ke, Montiel Olea, and Nesbit (2022), Giacomini, Kitagawa, and Read (2022)

Sensitivity/weak identi cation in DSGE : Müller (2012), Guerron-Quintana, Inoue, and Kilian (2013), Andrews and Mikusheva (2015), Ho (2022)

Literature and Contributions

Frequentist inference for set-identi ed models: Horowitz and Manski (2000), Manski (2003), Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Stoye (2009), Romano and Shaikh (2010), Kaido, Molinari, and Stoye (2019)

Bayesian inference for set-identi ed models: Baumeister and Hamilton (2015), Kline and Tamer (2016), Chen, Christensen, and Tamer (2018)

My contribution :

- A new robust Bayesian algorithm applied to DSGE models with theory.
- I work on “global” identi cation rather than identi cation at certain point (KK21).
- I study DSGE model, which has further complications (GK21).

Setup

Assumption

Linearized DSGE model with Gaussian shocks.

Linear state-space representation

Assumption

Solution to the linear rational expectation model (LRM) is unique, i.e., no indeterminacy.

Coefficients of the state-space model are uniquely determined.

Assumption

Deep parameters enter LRM in an algebraic expression way.

e.g., NKPC in Gali (2015): $_t = E_t \{ \; _{t+1} \} + \quad + \frac{' +}{1} \; y_t$

Estimate a Linearized DSGE model

Standard precedure

S 1. Summarize a macro model with equilibrium conditions, measurement equations, etc.

S 2. Log-linearization the equations around steady state, represent the model by a linear rational expectation model (LRM) with deep parameters .

$$ {}_0(\ )\begin{bmatrix} S_t \\ P_t \end{bmatrix} = {}_1(\ )E_t\begin{bmatrix} S_{t+1} \\ P_{t+1} \end{bmatrix} + {}_2(\ )S_{t\ 1} + {}_3(\ )"_t $$

S_t state variables, P_t policy variables.

Estimate a Linearized DSGE model

Standard precedure

S 3. Solve the LRM, combine with measurement equations and attain a state-space representation.

$$S_t = A(\ )S_{t\ \ 1} + B(\ )"_t$$

$$Y_t = C(\ )S_{t\ \ 1} + D(\ )"_t$$

S 4. Use a generic lter to compute the likelihood p(y j) through the state-space model.

S 5. Start from a prior distribtuion , update by MCMC methods using the likelihood and obtain the posterior distribution of , jy.

De nitions

De nition (OE)

Parameter is observationally equivalent to if they have the same likelihood p(y j) for all data realization y.

- A property independent of data

De nition (Identi cation)

is identi ed if it has no observationally equivalent parameters.

- De ne the equivalence mapping K : ! 2 , that is, p(y j) = p(y j) for all y, if and only if K( ) = K( ).

Algorithm

S.1 Run standard Bayesian estimation, get posterior draws of   from a given prior   .

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

Finding the OE set of a given parameter involves solving a polynomial system.

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

Remark: When the model is identi ed at all draws, the proposed algorithm gives the same result as the standard Bayesian method.

Comparison with Giacomini and Kitagawa (2021)

In GK21, the SVAR model

$$A_0 Y_t = a + \sum_{j=1}^{p} A_j Y_{t-j} + \varepsilon_t \text{ for } t = 1, \ldots, T$$

have explicit reduced-form parameters.

What they did: prior over reduced-form $\rightarrow$ structural parameters.

In GK21, the mapping between structural parameters and reduced-form coefficients is analytically tractable.

OE characterization

Assumptions

De ne $N = APC^{0} + B \quad D^{0}$, where $P = E(S_t S_t^{0})$,

Assumption (Stability)

For every 2 and for any z 2 C, $\det(zI_{n_S} \quad A) = 0$ implies $j\ z\ j< 1$.

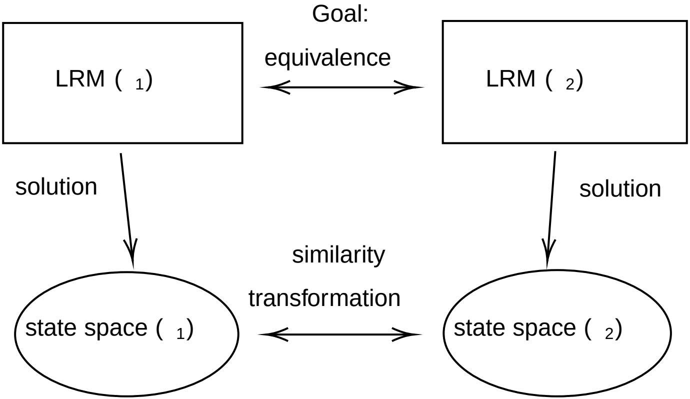
Assumption (Stochastic Minimality)

For every 2 , matrices O have full column rank and C have full row rank,

i.e. rank(O) = rank(C) = n_S. Where O $(C^{0} \quad A^{0}C^{0} \qquad A^{0 n_S \quad 1}C^{0})$,

C $(N \quad AN \qquad A^{n_S \quad 1}N)$.

OE characterization



OE characterization

Theorem (KK21)

Let stability and stochastic minimality assumptions hold. Then if and only if

1) $A = \ TAT \ ^{1}$,

2) $C = \ CT \ ^{1}$,

3) $AQA^{0} \quad Q = \ T \ ^{1}B \ \ B^{0}T^{0 \ 1} \quad B \ \ B^{0}$,

4) $CQC^{0} = \ D \ \ D^{0} \quad D \ \ D^{0}$,

5) $AQC^{0} = \ T \ ^{1}B \ \ D^{0} \quad B \ \ D^{0}$,

for some nonsingular n" n" matrix T and symmetric n" n" matrix Q. In addition, if then both T and Q are unique.

OE characterization

Brief

In order to use KK21, given a parameter , we need to link it to the solutions.

Attain the solution, $S_t = A(\ )S_{t\ \ 1} + B(\ )"_t$ and $P_t = F(\ )S_{t\ \ 1} + G(\ )"_t$, plug in LRM, equate coef cients on both sides in terms of $S_{t\ \ 1}$, and ${}_t$.

$$
{}^s_0 A + {}^p_0 F \qquad {}^s_1 (A)^2 \qquad {}^p_1 F A = \quad {}_2
$$

$$
{}^s_1 AB + {}^p_1 F B \qquad {}^s_0 B + \quad {}_3 = \quad {}^p_0 G
$$

OE characterization

Brief

Therefore, we can solve for all observationally equivalent s following the procedure

S.1 Given , solve for state-space coef cients.

S.2 Characterize by KK21 and the previous two equations, unknowns include (not limit to) .

S.3 Reduce a polynomial system to its reduced Grobner basis. here

Additional Assumptions

Assumption (Measurability)

The equivalence mapping K is measurable.

Assumption (Continuity)

(1) K is a continuous correspondence at $_0$.

(2) Parameters of interest : ! R^n is continuous.

Assumption (Regularity)

Let the prior of deep parameters , , be absolutely continuous with respect to a - nite measure on (; A), and ( ) = 1.

Multiple priors

De ne the class of all priors that the marginal distribution for K coincides with the given $_K$, i.e.,

$$(\ _K) = f \quad : \quad (f \ : K(\ ) \ 2 \ Bg) = \ _K(B); \ \text{for } B \ 2 \ B(F)g$$

The class of priors have the same push-forward measure $_K$ (by de nition).

The class of priors induce the same prior predictive distribution $p(y) = R \, p(y \, j \ ) d$.

This class can be uniquely pinned down either by $_K$ or any element in it.

Robust Distributions

Theorem (Posterior Mean)

For a given , let measurability and regularity assumptions hold, that is, given a prior absolutely continuous with respect to a - nite measure, we have a push-forward measure $_K$ of under K that is also absolutely continuous. De ne

$$\bar{\ } (\ ) = \sup_{0 \in K(\ )} (\ ^0); \quad \underline{\ } (\ ) = \inf_{0 \in K(\ )} (\ ^0)$$

Then, the set of posterior means is characterized by

$$\sup_{jY \in jY} E_{jY} [\ (\ )] = E_{jY} [\bar{\ } (\ )]; \qquad \inf_{jY \in jY} E_{jY} [\ (\ )] = E_{jY} \underline{\ } (\ )$$

where $_{jY}$ collects the posteriors of $(\ _K)$ for a given $_K$.

Robust Distributions

Theorem (Consistency)

Let, in addition continuity assumption hold, assume further that induced prior $_K$ leads to a consistent posterior, and $R^p; H$ R^q for some $p; q < 1$ are compact spaces. Then the Hausdorff distance between the set of posterior means and the convex hull of true identi ed set goes to zero almost surely as T increases, i.e.,

$$\lim_{T!1} d_H \quad E_{jY^T} \quad \_\,(\ ); \quad (\ ) \quad ; \quad h\,\_\,(\ _0); \quad (\ _0)\,i\,! \quad ! \quad 0; \quad p\,(Y^1 \ j \ _0)\text{-a.s.}$$

Example 1: Cochrane Model

Consider the full model

$$x_t = \quad x_{t \quad 1} + \quad _t; \quad j \quad j < 1; \quad _t \quad N(0; \quad)$$

$$i_t = r + E_t \quad _{t+1}$$

$$i_t = r + \qquad _t + x_t; \qquad > 1$$

Deep parameters are = (; ;). The solution is equivalent to a AR(1) setting

$$_t = \quad _{t \quad 1} + \frac{1}{\qquad} \, _t; \quad _t \quad N(0; \quad ^2)$$

with reduced form parameters = (; ——), (;) not jointly identi able.

The impulse response function is also not identi ed.

Example 1: Inference

Table: Estimated Identi ed Set

	True value	Identi ed set	Range of posterior mean
e	1	(0:2; 1)	(0:21; 1)
	1.8	(1; 1)	(1:00; 1)
	0.8	0.8	0.80

Estimation of range of posterior means approximates the identi ed set well.

Example 1: Inference

Example 1: Inference

Example 1: Inference

Example 2: An and Shorfheide (2007)

$$y_t = E_t[y_{t+1}] \quad \frac{1}{\ } (i_t \quad E_t[\ _{t+1}] + \color{red}{E_t[z_{t+1}]}) + g_t \quad E_t[g_{t+1}]$$

$$_t = \quad E_t[\ _{t+1}] + \quad (y_t \quad g_t)$$

$$i_t = \quad _R i_{t \ 1} + (1 \quad _R) \quad _t + (1 \quad _R) \ _y (y_t \quad g_t) + \text{"}_{R;t}$$

$$z_t = \quad _z z_{t \ 1} + \text{"}_{z;t}; \quad g_t = \quad _g g_{t \ 1} + \text{"}_{g;t}:$$

$(\ ; \ _y; \ _R; \ _R)$ are not identi ed.

All the shocks, either has no effect on $_t$ or y_t, or affect $_t$ and y_t in the same direction.

Example 2: A Cost-push Shock Model

To generate meaningful trade-off between output gap and in ation,

$$y_t = E_t[y_{t+1}] \quad \frac{1}{\ } (i_t \quad E_t[\ _{t+1}]) + g_t \quad E_t[g_{t+1}]$$

$$_t = \quad E_t[\ _{t+1}] + \quad (y_t \quad g_t) + u_t$$

$$i_t = \ _R i_{t \ 1} + (1 \quad _R) \quad _t + (1 \quad _R) \ _y (y_t \quad g_t) + "_{R;t}$$

$$u_t = \ _u u_{t \ 1} + "_{u;t}; \quad g_t = \ _g g_{t \ 1} + "_{g;t}:$$

Positive cost-push shock u ! y #; "

Positive monetary policy shock $_R$! y #; #

Example 2: Policy

Table: Policy Comparison under Different Distributions and Weights

	= 1		= 3		= 10	
(; y)	post 1	post 2	post 1	post 2	post 1	post 2
(1.5, 0)			X		X	X
(1.5, 0.125)	X	X		X		
(1.5, 1)	X	X				
(5, 0)			X	X	X	X

Policy choices can be robust to prior choices.

Conclusion

In this paper, I attack the following problems:

- Estimation results of set-identi ed DSGE models are sensitive to choice of priors (Identi cation)
  - Use a robust Bayesian algorithm, I can pick any `reasonable' prior and obtain robust results.
  - I also prove it asymptotically nds the frequentist identi ed set.
- Researchers are silent about non-identi ed 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 identi cation problems.

Likelihood when T=1000,000

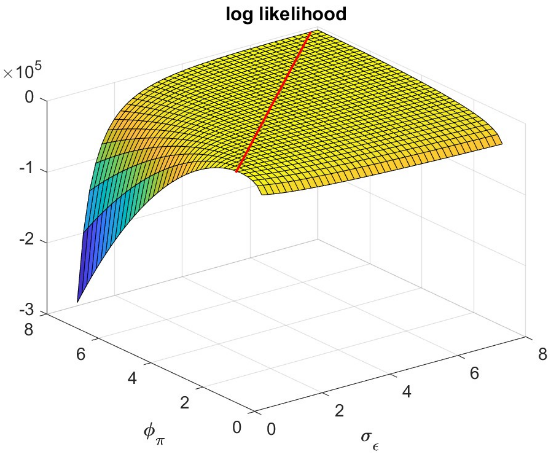


Figure: Likelihood function while fix $\rho = 0.8$

Flat ridge along the $\sigma \quad = \phi \qquad$ 0.8 line

Prior Sensitivity

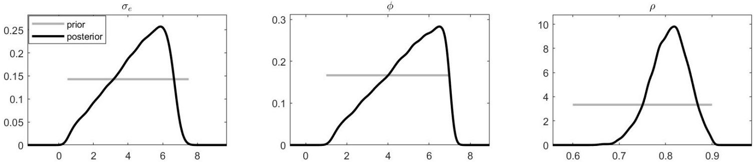


Figure: Cochrane model prior/posterior distribution with uniform priors

The posterior of σ and ϕ are extremely informative even if only 0:8 is identified.

Reason? Joint likelihood density more concentrated on areas with higher values of ϕ and σ.

Back to main

Gröbner Basis

A reduced Gröbner basis is a set of multivariate polynomials enjoying certain properties that allow simple algorithmic solutions. For example, the equations:

$$x^3 \quad 2xy, \quad x^2 \quad 2y^2 + x.$$

has a reduced Gröbner basis

$$x^2, \quad xy, \quad y^2 \quad \frac{x}{2}.$$

Any zero of a Gröbner basis is also a zero of the original system.

Reduced Gröbner bases are unique for any given set of polynomials and any monomial ordering. main