Robust Bayesian Estimation and Inference for Dynamic Stochastic General Equilibrium Models

Yizhou Kuang†

†Department of Economics, Cornell University

Last updated: January 12^{th}, 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

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

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

Motivation - Estimation

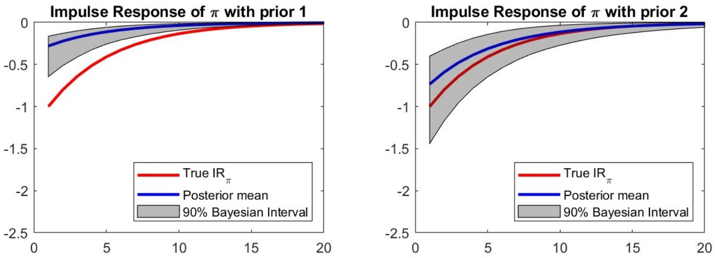
Table: Prior and Posterior Distribution of Structural Parameters

	True value	Prior distribution			Posterior distribution			
		Distr.	Mean	St. Dev.	Mode	Mean	5 percent	95 percent
σ_ϵ	1	Uniform	4	2.02	5.82	4.43	1.95	6.77
ϕ_π	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 $\mathrm{IR}_\pi(t, s, 1)$ 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 $(\rho, \frac{\sigma_e}{\phi_\pi - \rho})$

Motivation - Policy Analysis

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

$$y_t = \mathbb{E}_t[y_{t+1}] - \frac{1}{\sigma}(i_t - \mathbb{E}_t[\pi_{t+1}]) + g_t - \mathbb{E}_t[g_{t+1}]$$

$$\pi_t = \beta \mathbb{E}_t[\pi_{t+1}] + \kappa(y_t - g_t) + u_t$$

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

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

is trying to use the estimated parameter (history)

$(\sigma, \beta, \kappa, \psi_\pi, \psi_y, \rho_R, \rho_g, \rho_u, \sigma_R, \sigma_g, \sigma_u)$, to choose a policy rule (ψ'_π, ψ'_y) for

$$i^*_t = \psi'_\pi \pi_t + \psi'_y(y_t - g_t)$$

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 $(\psi_\pi, \psi_y) = (1.5, 0)$, and $(1.5, 0.125)$

Table: Policy Comparison under Different Distributions and Weights

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

- 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

- **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 (2015), Liao and Simoni (2019), Giacomini and Kitagawa (2021), Ke, Montiel Olea, and Nesbit (2022), Giacomini, Kitagawa, and Read (2022)
- **Sensitivity/weak identification in DSGE**: Müller (2012), Guerron-Quintana, Inoue, and Kilian (2013), Andrews and Mikusheva (2015), Ho (2022)

Literature and Contributions

- **Frequentist inference for set-identified 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-identified 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” identification rather than identification 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): $\pi_t = \beta E_t \{\pi_{t+1}\} + \lambda \left(\sigma + \frac{\varphi+\alpha}{1-\alpha}\right) \tilde{y}_t$

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 θ.

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

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.

$$
\begin{aligned}
S_t &= A(\theta)S_{t-1} + B(\theta)\varepsilon_t \\
Y_t &= C(\theta)S_{t-1} + D(\theta)\varepsilon_t
\end{aligned}
$$

S 4. Use a generic filter to compute the likelihood $p(y \mid \theta)$ through the state-space model.

S 5. Start from a prior distribtuion π_θ, update by MCMC methods using the likelihood and obtain the posterior distribution of θ, $\pi_{\theta|y}$.

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 y, if and only if $K(\theta) = K(\bar{\theta})$.

Algorithm

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

S.2* Optimize over the observationally equivalent set of parameters of this draw, find 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 identified 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

Define $N = APC' + B\Sigma D'$, where $P = E(S_t S_t')$,

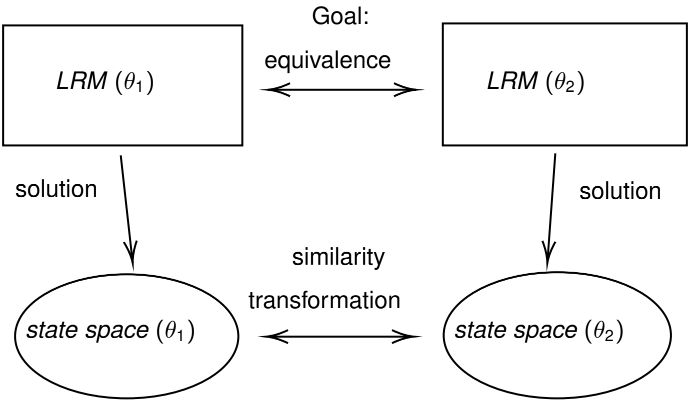
Assumption (Stability)

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

Assumption (Stochastic Minimality)

For every $\theta \in \Theta$, matrices $\mathcal{O}$ have full column rank and $\mathcal{C}$ have full row rank, i.e. $\text{rank}(\mathcal{O}) = \text{rank}(\mathcal{C}) = n_S$. Where $\mathcal{O} \equiv (C' \quad A'C' \quad \cdots \quad A'^{n_S-1}C')$, $\mathcal{C} \equiv (N \quad AN \quad \cdots \quad A^{n_S-1}N)$.

OE characterization



OE characterization

Theorem (KK21)

Let stability and stochastic minimality assumptions 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.*

OE characterization

Brief

- In order to use KK21, given a parameter $\bar{\theta}$, we need to link it to the solutions.
- Attain the solution, $S_t = \bar{A}(\theta)S_{t-1} + \bar{B}(\theta)\varepsilon_t$ and $P_t = \bar{F}(\theta)S_{t-1} + \bar{G}(\theta)\varepsilon_t$, plug in LRM, equate coefficients on both sides in terms of S_{t-1}, and ϵ_t.

$$\bar{\Gamma}_0^s\bar{A} + \bar{\Gamma}_0^p\bar{F} - \bar{\Gamma}_1^s(\bar{A})^2 - \bar{\Gamma}_1^p\bar{F}\bar{A} = \bar{\Gamma}_2$$

$$\bar{\Gamma}_1^s\bar{A}\bar{B} + \bar{\Gamma}_1^p\bar{F}\bar{B} - \bar{\Gamma}_0^s\bar{B} + \bar{\Gamma}_3 = \bar{\Gamma}_0^p\bar{G}$$

OE characterization

Brief

Therefore, we can solve for all observationally equivalent $\bar{\theta}$s following the procedure

S.1 Given θ, solve for state-space coefficients.

S.2 Characterize $\bar{\theta}$ by KK21 and the previous two equations, unknowns include (not limit to) $\bar{\theta}$.

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 θ_0.*

(2) *Parameters of interest $\eta : \Theta \to \mathbb{R}^n$ is continuous.*

Assumption (Regularity)

Let the prior of deep parameters θ, π_θ, be absolutely continuous with respect to a σ-finite measure on $(\Theta, \mathcal{A})$, and $\pi_\theta(\Theta) = 1$.

Multiple priors

- Define the class of all priors that the marginal distribution for K coincides with the given π_K, i.e.,

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

* The class of priors have the same push-forward measure π_K (by definition).
* The class of priors induce the same prior predictive distribution $p(y) = \int p(y \mid \theta) d\pi_\theta$.
* This class can be uniquely pinned down either by π_K or any element π_θ in it.

Robust Distributions

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.

Robust Distributions

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 Model

Consider the full model

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

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

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

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

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

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

The impulse response function is also not identified.

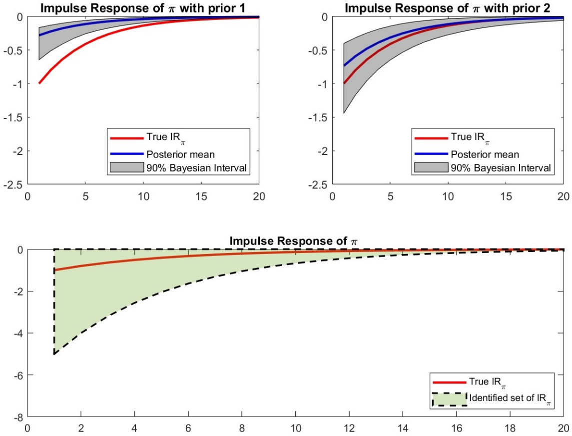
Example 1: Inference

Table: Estimated Identified Set

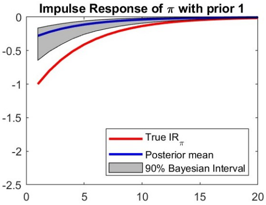
	True value	Identified set	Range of posterior mean
σ_e	1	$(0.2, \infty)$	$(0.21, \infty)$
ϕ_π	1.8	$(1, \infty)$	$(1.00, \infty)$
ρ	0.8	0.8	0.80

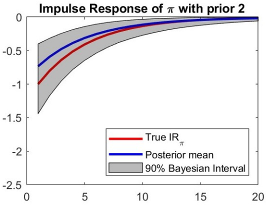
- Estimation of range of posterior means approximates the identified set well.

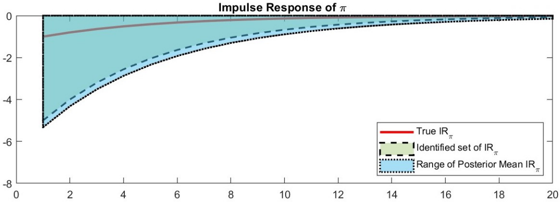
Example 1: Inference



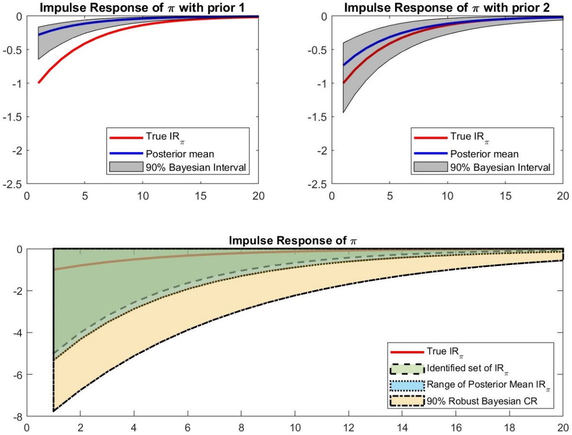
Example 1: Inference







Example 1: Inference



Example 2: An and Shorfheide (2007)

$$y_t = \mathbb{E}_t\left[y_{t+1}\right] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t\left[\pi_{t+1}\right] + \mathbb{E}_t\left[z_{t+1}\right]\right) + g_t - \mathbb{E}_t\left[g_{t+1}\right]$$

$$\pi_t = \beta\mathbb{E}_t\left[\pi_{t+1}\right] + \kappa\left(y_t - g_t\right)$$

$$i_t = \rho_R i_{t-1} + \left(1-\rho_R\right)\psi_\pi\pi_t + \left(1-\rho_R\right)\psi_y\left(y_t - g_t\right) + \varepsilon_{R,t}$$

$$z_t = \rho_z z_{t-1} + \varepsilon_{z,t}, \quad g_t = \rho_g g_{t-1} + \varepsilon_{g,t}.$$

- $(\psi_\pi, \psi_y, \rho_R, \sigma_R)$ are not identified.
- All the shocks, either has no effect on π_t or y_t, or affect π_t and y_t in the same direction.

Example 2: A Cost-push Shock Model

To generate meaningful trade-off between output gap and inflation,

$$y_t = \mathbb{E}_t[y_{t+1}] - \frac{1}{\sigma}(i_t - \mathbb{E}_t[\pi_{t+1}]) + g_t - \mathbb{E}_t[g_{t+1}]$$

$$\pi_t = \beta \mathbb{E}_t[\pi_{t+1}] + \kappa(y_t - g_t) + u_t$$

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

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

- Positive cost-push shock $u \longrightarrow y \downarrow, \pi \uparrow$
- Positive monetary policy shock $\epsilon_R \longrightarrow y \downarrow, \pi \downarrow$

Example 2: Policy

Table: Policy Comparison under Different Distributions and Weights

	$\lambda = 1$		$\lambda = 3$		$\lambda = 10$	
(ψ_π, ψ_y)	post 1	post 2	post 1	post 2	post 1	post 2
(1.5, 0)			✓		✓	✓
(1.5, 0.125)	✓	✓		✓		
(1.5, 1)	✓	✓				
(5, 0)			✓	✓	✓	✓

- Policy choices can be robust to prior choices.

Conclusion

In this paper, I attack the following problems:

- Estimation results of set-identified DSGE models are sensitive to 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.

Likelihood when T=1000,000

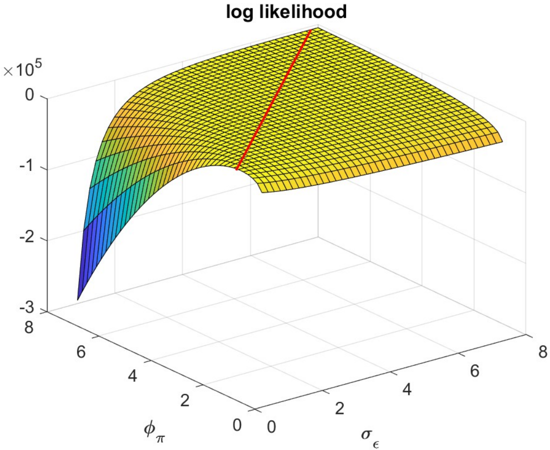


Figure: Likelihood function while fix $\rho = 0.8$

- Flat ridge along the $\sigma_\epsilon = \phi_\pi - 0.8$ line

Prior Sensitivity

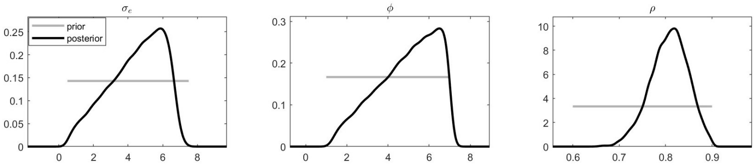


Figure: Cochrane model prior/posterior distribution with uniform priors

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

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

has a reduced Gröbner basis

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

- 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