Robust Bayesian Inference for Set-Identified Models

Contents

Section 1

quick overview of Bayesian statistics

Bayes rule

$$P(\text{Hypothesis}|\text{Data}) = \frac{P(\text{Data}|\text{Hypothesis})P(\text{Hypothesis})}{P(\text{data})}$$

- Prior: Our Hyptothesis (= parameters) before seeing the data, set by the researcher

- Likelihood: The probability of observing the data given the Hypothesis, our "model"

Bayes rule

$$P(\text{Hypothesis}|\text{Data}) = \frac{P(\text{Data}|\text{Hypothesis})P(\text{Hypothesis})}{P(\text{data})}$$

- Prior: Our Hyptothesis (= parameters) before seeing the data, set by the researcher

- Likelihood: The probability of observing the data given the Hypothesis, our "model"
- Posterior: Belief about the Hypothesis (= parameters) given the data, the point of Bayesian statistics

Bayes rule

$$P(\text{Hypothesis}|\text{Data}) = \frac{P(\text{Data}|\text{Hypothesis})P(\text{Hypothesis})}{P(\text{data})}$$

- Prior: Our Hyptothesis (= parameters) before seeing the data, set by the researcher

- Likelihood: The probability of observing the data given the Hypothesis, our "model"
- Posterior: Belief about the Hypothesis (= parameters) given the data, the point of Bayesian statistics
- Evidence: (Usually unobserved) normalising constant

Example model, Bayesian linear regression

Given we chose the model

$$y = \theta_0 + \theta_1 x + \epsilon \quad y, x, \theta_0, \theta_1 \in \mathbb{R} \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

We set relatively uninformative priors

$$\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix} \right) = \mathcal{N}(\mu_{prior}, \Sigma_{prior}), \quad \sigma^2 \sim \text{Inv-Gamr}$$

The likelihood defined by the model given that we have k samples follows the distribution

$$y|\beta_0, \beta_1 \sim \mathcal{N}_k(X\beta, diag(\sigma^2)), \quad X = \begin{pmatrix} 1 & x_1 \\ ... & ... \\ 1 & x_k \end{pmatrix}$$

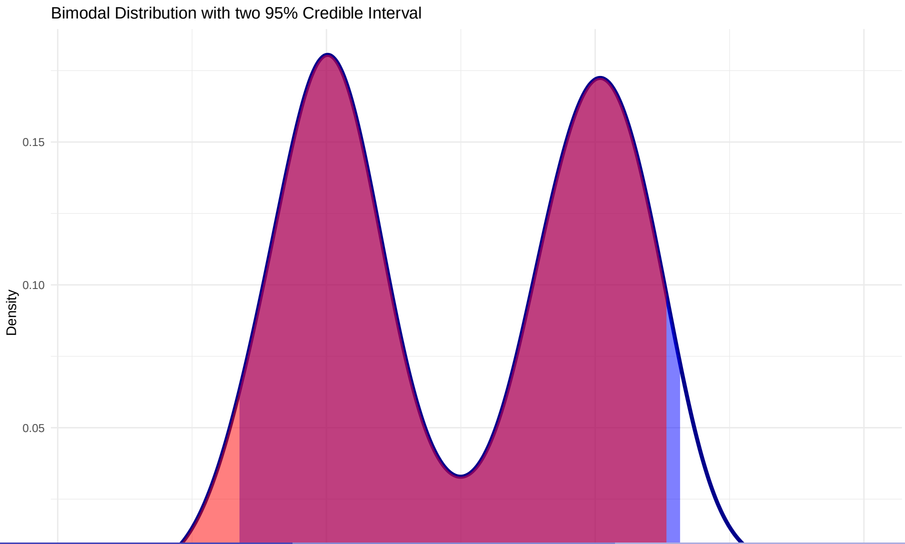
Therefore the posterior is

$$P((\theta_0, \theta_1)^T | y) \propto \mathcal{N}\left(\Sigma_{posterior} \frac{1}{\sigma^2} X^T y, (\Sigma_{prior}^{-1} + \frac{1}{\sigma^2} X^T X)^{-1} \right)$$

!!!! Note the above is false, do not bother with the full distribution

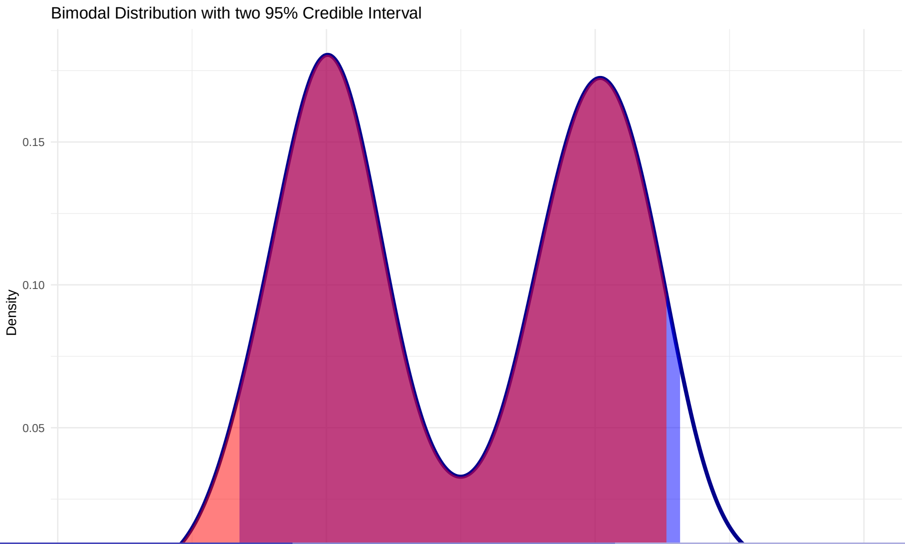
Inference: credible intervals

Say, given any sample $Y = (y_1, ..., y_n)$ the posterior $ P(| y) $ looks approximately like



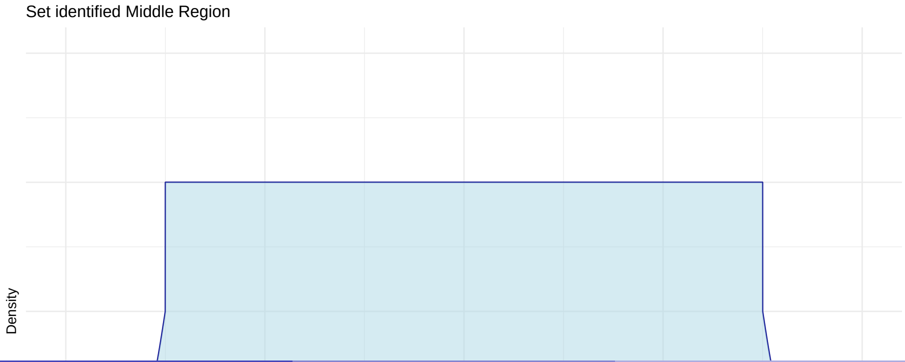
Inference: credible intervals

Say, given any sample $Y = (y_1, ..., y_n)$ the posterior \$ P(| y) \$ looks approximately like



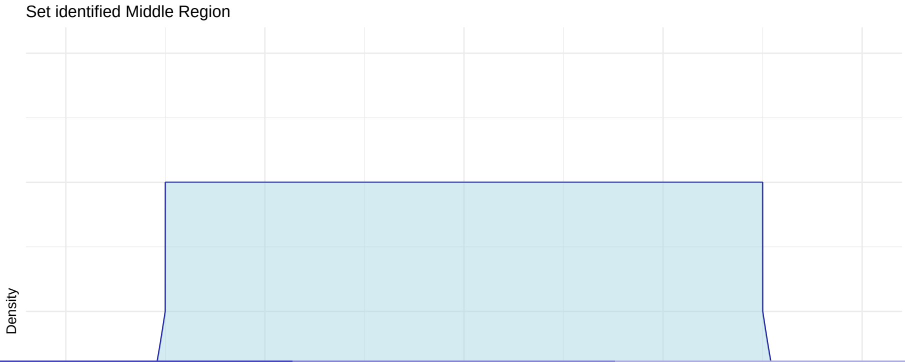
Set identification

- A parameter set Θ is said to be (point) identified if $\Theta = \{\theta\}$



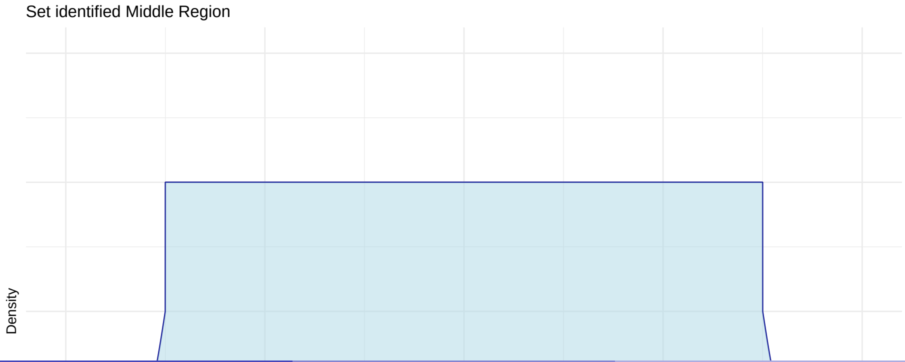
Set identification

- A parameter set Θ is said to be (point) identified if $\Theta = \{\theta\}$
- A parameter set Θ is said to be set identified if $\exists \theta_0, \theta_1 \in \Theta s.t. \theta_0 \neq \theta_1$



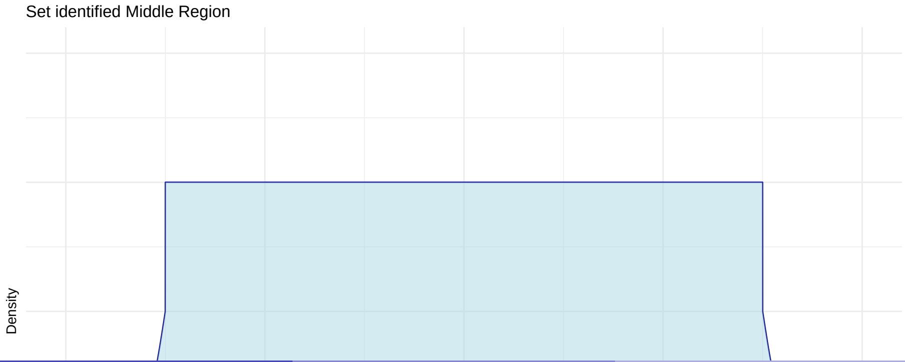
Set identification

- A parameter set Θ is said to be (point) identified if $\Theta = \{\theta\}$
- A parameter set Θ is said to be set identified if $\exists \theta_0, \theta_1 \in \Theta s.t. \theta_0 \neq \theta_1$
- In a Bayesian setting parameters are set identified if $\exists \theta_0, \theta_1 \in \Theta \; s.t. \theta_0 \neq \theta_1, \quad p(y|\theta_0) = p(y|\theta_1) \; \forall y \in Y$



Set identification

- A parameter set Θ is said to be (point) identified if $\Theta = \{\theta\}$
- A parameter set Θ is said to be set identified if $\exists \theta_0, \theta_1 \in \Theta s.t. \theta_0 \neq \theta_1$
- In a Bayesian setting parameters are set identified if $\exists \theta_0, \theta_1 \in \Theta \ s.t. \theta_0 \neq \theta_1, \quad p(y|\theta_0) = p(y|\theta_1) \ \forall y \in Y$
- Note this is for any draw of the data and the priors must be equal. Not necessarily the posteriors

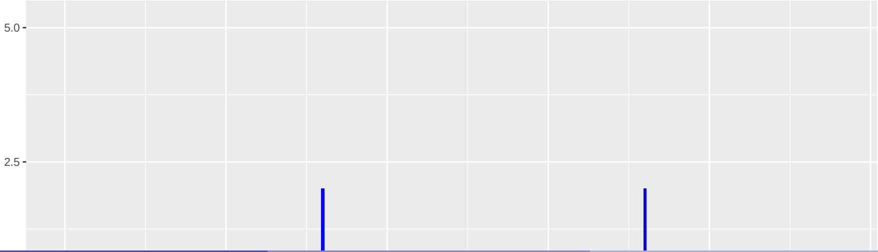


Section 2

The paper

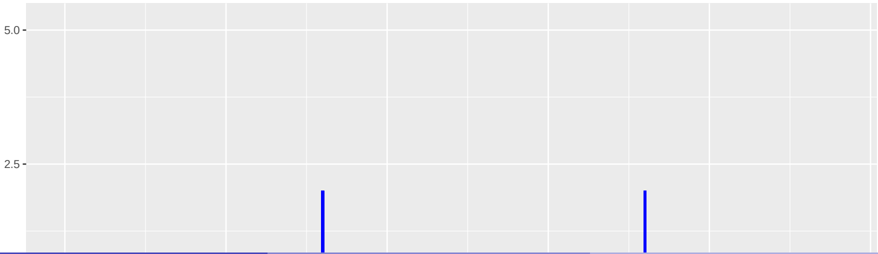
The problem with set identified models

- Let $\mathbf{Y} \subseteq \mathbb{R}^d$ be the set of samples Y and $\Theta \subseteq \mathbb{R}^d$ be the set of possible parameter values θ.



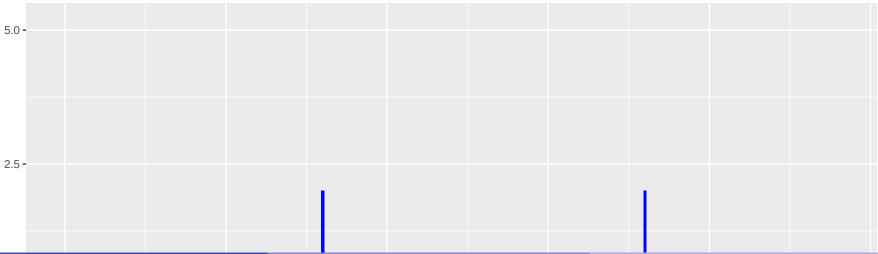
The problem with set identified models

- Let $\mathbf{Y} \subseteq \mathbb{R}^d$ be the set of samples Y and $\Theta \subseteq \mathbb{R}^d$ be the set of possible parameter values θ.
- Suppose we only observe a reduced for parameter $\phi = g(\theta)$, where $g : \Theta \to \Phi, \quad g(\theta) = g(\theta^{'})$ iff $p(Y|\theta) = p(Y|\theta^{'}) \forall Y \in \mathbf{Y}$. Note in set identification g is not injective.



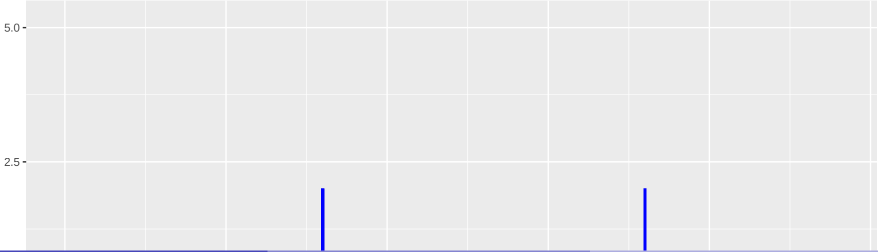
The problem with set identified models

- Let $\mathbf{Y} \subseteq \mathbb{R}^d$ be the set of samples Y and $\Theta \subseteq \mathbb{R}^d$ be the set of possible parameter values θ.
- Suppose we only observe a reduced for parameter $\phi = g(\theta)$, where $g : \Theta \to \Phi, \quad g(\theta) = g(\theta^{'})$ iff $p(Y|\theta) = p(Y|\theta^{'}) \forall Y \in \mathbf{Y}$. Note in set identification g is not injective.
- Define the identified set $IS_\theta(\phi) = \{\theta \in \Theta | g(\theta) = \phi\}$



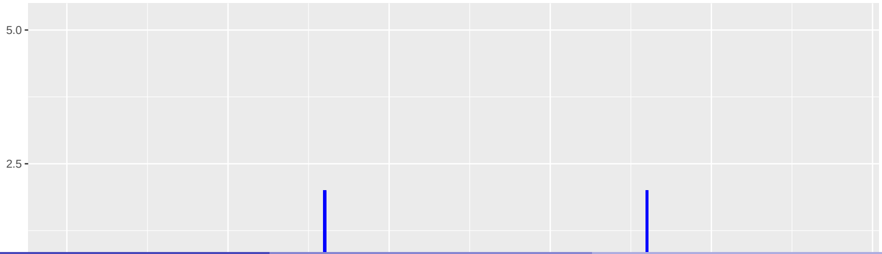
The problem with set identified models

- Let $\mathbf{Y} \subseteq \mathbb{R}^d$ be the set of samples Y and $\Theta \subseteq \mathbb{R}^d$ be the set of possible parameter values θ.
- Suppose we only observe a reduced for parameter $\phi = g(\theta)$, where $g : \Theta \to \Phi, \quad g(\theta) = g(\theta^{'})$ iff $p(Y|\theta) = p(Y|\theta^{'}) \forall Y \in \mathbf{Y}$. Note in set identification g is not injective.
- Define the identified set $IS_\theta(\phi) = \{\theta \in \Theta | g(\theta) = \phi\}$
- In set identified structural models the prior $\pi_{\theta|Y}$ can be decomposed as $\pi_{\theta|Y}(\cdot) = \int_\Phi \pi_{\theta|\phi}(\cdot) d\pi_{\phi|Y}$.



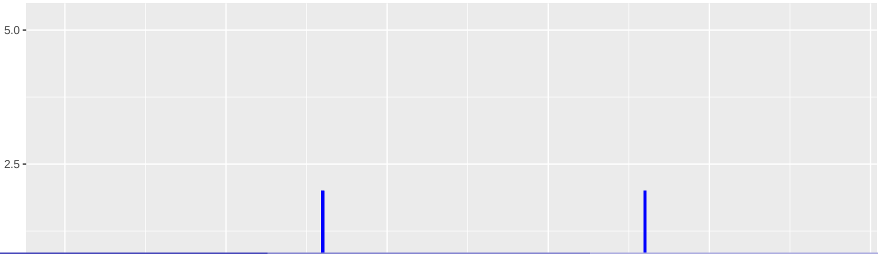
The problem with set identified models

- Let $\mathbf{Y} \subseteq \mathbb{R}^d$ be the set of samples Y and $\Theta \subseteq \mathbb{R}^d$ be the set of possible parameter values θ.
- Suppose we only observe a reduced for parameter $\phi = g(\theta)$, where $g : \Theta \rightarrow \Phi, \quad g(\theta) = g(\theta^{'})$ iff $p(Y|\theta) = p(Y|\theta^{'}) \forall Y \in \mathbf{Y}$. Note in set identification g is not injective.
- Define the identified set $IS_\theta(\phi) = \{\theta \in \Theta | g(\theta) = \phi\}$
- In set identified structural models the prior $\pi_{\theta|Y}$ can be decomposed as $\pi_{\theta|Y}(\cdot) = \int_\Phi \pi_{\theta|\phi}(\cdot) d\pi_{\phi|Y}$.
- $\pi_{\phi|Y}$ is updates as data comes in, $\pi_{\theta|\phi}$ is not.



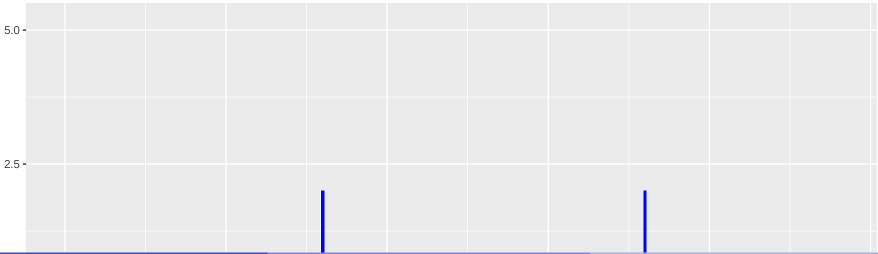
The problem with set identified models

- Let $\mathbf{Y} \subseteq \mathbb{R}^d$ be the set of samples Y and $\Theta \subseteq \mathbb{R}^d$ be the set of possible parameter values θ.
- Suppose we only observe a reduced for parameter $\phi = g(\theta)$, where $g : \Theta \to \Phi, \quad g(\theta) = g(\theta^{'})$ iff $p(Y|\theta) = p(Y|\theta^{'}) \forall Y \in \mathbf{Y}$. Note in set identification g is not injective.
- Define the identified set $IS_\theta(\phi) = \{\theta \in \Theta | g(\theta) = \phi\}$
- In set identified structural models the prior $\pi_{\theta|Y}$ can be decomposed as $\pi_{\theta|Y}(\cdot) = \int_\Phi \pi_{\theta|\phi}(\cdot) d\pi_{\phi|Y}$.
- $\pi_{\phi|Y}$ is updates as data comes in, $\pi_{\theta|\phi}$ is not.
-



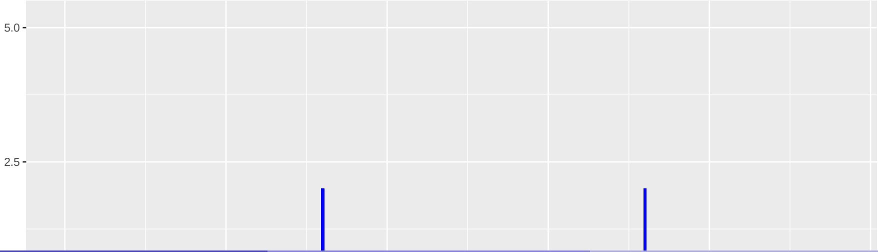
The problem with set identified models

- Let $\mathbf{Y} \subseteq \mathbb{R}^d$ be the set of samples Y and $\Theta \subseteq \mathbb{R}^d$ be the set of possible parameter values θ.
- Suppose we only observe a reduced for parameter $\phi = g(\theta)$, where $g : \Theta \to \Phi, \quad g(\theta) = g(\theta^{'})$ iff $p(Y|\theta) = p(Y|\theta^{'}) \forall Y \in \mathbf{Y}$. Note in set identification g is not injective.
- Define the identified set $IS_\theta(\phi) = \{\theta \in \Theta | g(\theta) = \phi\}$
- In set identified structural models the prior $\pi_{\theta|Y}$ can be decomposed as $\pi_{\theta|Y}(\cdot) = \int_\Phi \pi_{\theta|\phi}(\cdot) d\pi_{\phi|Y}$.
- $\pi_{\phi|Y}$ is updates as data comes in, $\pi_{\theta|\phi}$ is not.
-



The problem with set identified models

- Let $\mathbf{Y} \subseteq \mathbb{R}^d$ be the set of samples Y and $\Theta \subseteq \mathbb{R}^d$ be the set of possible parameter values θ.
- Suppose we only observe a reduced for parameter $\phi = g(\theta)$, where $g : \Theta \to \Phi, \quad g(\theta) = g(\theta^{'})$ iff $p(Y|\theta) = p(Y|\theta^{'}) \forall Y \in \mathbf{Y}$. Note in set identification g is not injective.
- Define the identified set $IS_{\theta}(\phi) = \{\theta \in \Theta | g(\theta) = \phi\}$
- In set identified structural models the prior $\pi_{\theta|Y}$ can be decomposed as $\pi_{\theta|Y}(\cdot) = \int_{\Phi} \pi_{\theta|\phi}(\cdot) d\pi_{\phi|Y}$.
- $\pi_{\phi|Y}$ is updates as data comes in, $\pi_{\theta|\phi}$ is not.
-



A running example

- As a running example I will be using an SVAR(0) of the form

$$Ay_t = \epsilon_t, \quad y_t, \epsilon_t \in \mathbb{R}^2, A \in \mathcal{M}(2,2)$$

A running example

- As a running example I will be using an SVAR(0) of the form

$$Ay_t = \epsilon_t, \quad y_t, \epsilon_t \in \mathbb{R}^2, A \in \mathcal{M}(2,2)$$

- The reduced form VAR(0) is $y_t = A^{-1}\epsilon_t = \eta_t$ with $E[\eta_t \eta_t^T] = \Sigma = A^{-1}(A^{-1})^T = (A^T A)^{-1}$

A running example

- As a running example I will be using an SVAR(0) of the form

$$Ay_t = \epsilon_t, \quad y_t, \epsilon_t \in \mathbb{R}^2, A \in \mathcal{M}(2,2)$$

- The reduced form VAR(0) is $y_t = A^{-1}\epsilon_t = \eta_t$ with $E[\eta_t \eta_t^T] = \Sigma = A^{-1}(A^{-1})^T = (A^T A)^{-1}$
- Note that $\forall Q \in \mathcal{O}(2)$, we have that $\Sigma = (A^T Q Q^T A)^{-1}$, the model is set identified

Breaking down the priors in set identified models

Multiple priors, posterior lower and upper probabilities

Set of posterior means

Robust credible regions

Diagnostic tools

Asymptotic properties 1

Asymp ic properties 2

Thank you

Section 3

Graveyard