Bayesian Methods for DSGE models
Lecture 1
Macro models as data generating processes

Kristoffer Nimark
Cornell University

July 2, 2018

Bayesian Methods for DSGE models

Course admin

Lectures: Mon-Fri, 9.00-11.00, room 24.104

Practicas: Mon-Fri, 16.45-17.45 (with Sarah Zoi)

To contact me:

- pkn8@cornell.edu
- Office in Merce Rodoreda building

Bayesian Methods for DSGE models

Course overview

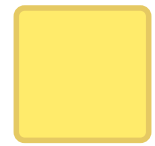
1. (i) Introduction to Bayesian analysis and (ii) Macro models as data generating processes
2. State space models and likelihood based estimation
3. Bayesian estimation of DSGE models
4. Bayesian analysis of DSGE models
5. Structural empirical models of news, noise and imperfect information

Bayesian Methods for DSGE models

Two good books:

Bayesian Estimation of DSGE Models by Herbst and Schorfheide, Princeton University Press 2015

Bayesian Data Analysis (BDA3) by Gelman, Carlin, Stern, Dunson,Vehtari and Rubin, CRC Press 2013



Bayesian Methods for DSGE models

What's in Box?

- Lecture notes
- Slides
  - Uploaded in real time...
- Additional reading
- Matlab code

Macro models as data generating processes

Today:

- Introduction to Bayesian estimation
- Macro models as likelihood functions
- Solving linear rational expectations models

Bayesian statistics

Bayesian statistics

Bayesians used to dominate statistics, then became fringe types, and now Bayesian methods dominate applied macro economics

- This is largely due to increased computing power

Bayesian methods have several advantages:

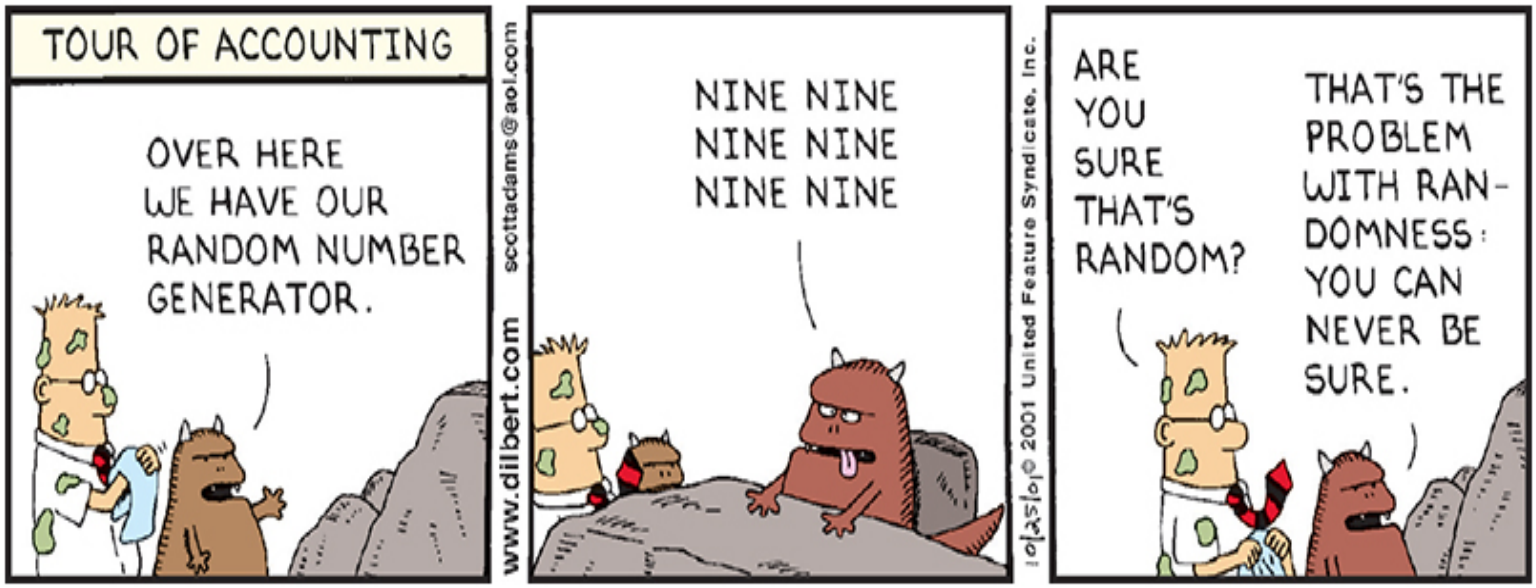
- Facilitates incorporating information from outside of sample
- Easy to compute confidence/probabilty intervals of functions of parameters
- Good (i.e. known) small sample properties

Why probabilistic models?

Is the world characterized by randomness?

- Is the weather random?
- Is a coin flip random?
- ECB interest rates?
- The outcome of the Brexit vote?

It is difficult to say with certainty whether something is "truly" random.


TOUR OF ACCOUNTING
OVER HERE WE HAVE OUR RANDOM NUMBER GENERATOR.
www.dilbert.com scottadams@aol.com
NINE NINE NINE NINE NINE NINE
ARE YOU SURE THAT'S RANDOM?
THAT'S THE PROBLEM WITH RANDOMNESS: YOU CAN NEVER BE SURE.
© 2001 United Feature Syndicate, Inc.

Probability and statistics: What's the difference?

Probability is a branch of mathematics

- There is little disagreement about whether the theorems follow from the axioms

Statistics is an inversion problem: What is a good probabilistic description of the world, given the observed outcomes?

- There is some disagreement about how to interpret probabilistic statements and about how we should interpret data when we make inference about unobservable parameters

Two schools of statistics

Two schools of thought

What is the meaning of probability, randomness and uncertainty?

- The classical (or frequentist) view is that probability corresponds to the frequency of occurrence in repeated experiments
- The Bayesian view is that probabilities are statements about our state of knowledge, i.e. a subjective view.

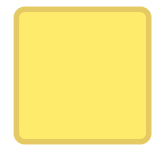
The difference has implications for how we interpret estimated statistical models and there is no general agreement about which approach is "better".

The subjective view of probability

Important: Probabilities can be viewed as statements about our knowledge of a parameter, even if the parameter is a constant

- E.g. treating risk aversion as a random variable does not imply that we think of it as varying across time.

Subjective does *not* mean that anything goes!



Frequentist vs Bayesian statistics

Distribution of estimator vs distribution of parameter

- Bayesians think of parameters as having distributions while frequentists conduct inference by thinking about the distribution of an *estimator* under a particular null hypothesis (H_0)

Frequentist confidence intervals:

- If H_0 is true, and with repeated draws from the population of equal sample length, what is the interval that $\widehat{\theta}$ lies in 95% of the time?

Bayesian probability intervals:

- Conditional on the observed data, a prior distribution of θ and a functional form (i.e. model), what is the shortest interval that contains θ with probability 95%?

Decisions under uncertainty

Coin flip example

After flipping a coin 10 times and counting the number of heads, what is the probability that the next flip will come up heads?

- A Bayesian with a uniform prior would say that the probability of the next flip coming up heads is $x/10$ and where x is the number of times the flip came up heads in the observed sample
- A frequentist cannot answer this question: A frequentist makes probabilistic statements about the likelihood of observing a particular realized sample, given a null hypothesis about the parameter that governs the probability of heads vs tails

The main components in Bayesian inference

Bayes' Rule

Deriving Bayes' Rule

$$p(A, B) = p(A \mid B)p(B)$$

Symmetrically

$$p(A, B) = p(B \mid A)p(A)$$

Implying

$$p(A \mid B)p(B) = p(B \mid A)p(A)$$

or

$$p(A \mid B) = \frac{p(B \mid A)p(A)}{p(B)}$$

which is known as Bayes' Rule.

Data and parameters

The purpose of Bayesian analysis is to use the data y to learn about the "parameters" θ

- Parameters of a statistical model
- Or anything not directly observed

Replace A and B in Bayes rule with θ and y to get

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}$$

The probability density $p(\theta \mid y)$ then describes what we know about θ, given the data y.

Parameters as random variables

$p(\theta \mid y)$ is a density function

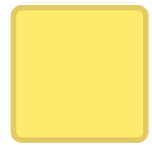
- We treat θ as a random variable with a distribution

This is the subjective view of probability

- Important: the randomness of a parameter is subjective in the sense of depending on what information is available to the person making the statement.

The density $p(\theta \mid y)$ is proportional to the prior times the likelihood function

$$\underbrace{p(\theta \mid y)}_{\text{posterior}} \propto \underbrace{p(y \mid \theta)}_{\text{likelihood function}} \underbrace{p(\theta)}_{\text{prior}}$$



But what exactly is a prior and a likelihood function?

The prior

The prior summarizes the investigator's knowledge about parameters θ *before* observing the data y.

In practice, we need to choose functional forms and select "hyper-parameters" that reflect this our prior knowledge.

Example:
Let's say that our prior about θ an be described by a uniform distribution

- $\theta \sim U[-1, 1]$

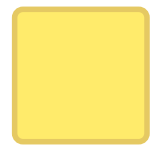
-1,1 would then be the **hyper-parameters** of our prior

The likelihood function

The likelihood function is the probability density of observing the data y conditional on a statistical model and the parameters θ.

For instance if the data y conditional on θ is normally distributed we have that

$$p(y \mid \theta) = (\mathbf{2}\pi)^{-T/2} \left|\Omega^{-1}\right|^{1/2} \exp\left[-\frac{1}{2}(\mathbf{y}-\mu)'\Omega^{-1}(\mathbf{y}-\mu)\right]$$

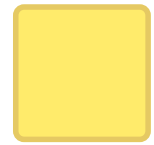


where Ω and μ are either elements or functions of θ.

The prior and the likelihood function

Remember how we derived Bayes' rule:

$$p(\theta, y) = p(y \mid \theta)p(\theta)$$



The model determines both the likelihood function and what the the prior is "about".

The posterior density $p(\theta \mid y)$

The posterior density is often the object of fundamental interest in Bayesian estimation.

- The posterior is the "result".
- We are interested in the entire conditional distribution of the parameters θ

The linear regression model

The linear regression model

Let's start with something very basic:

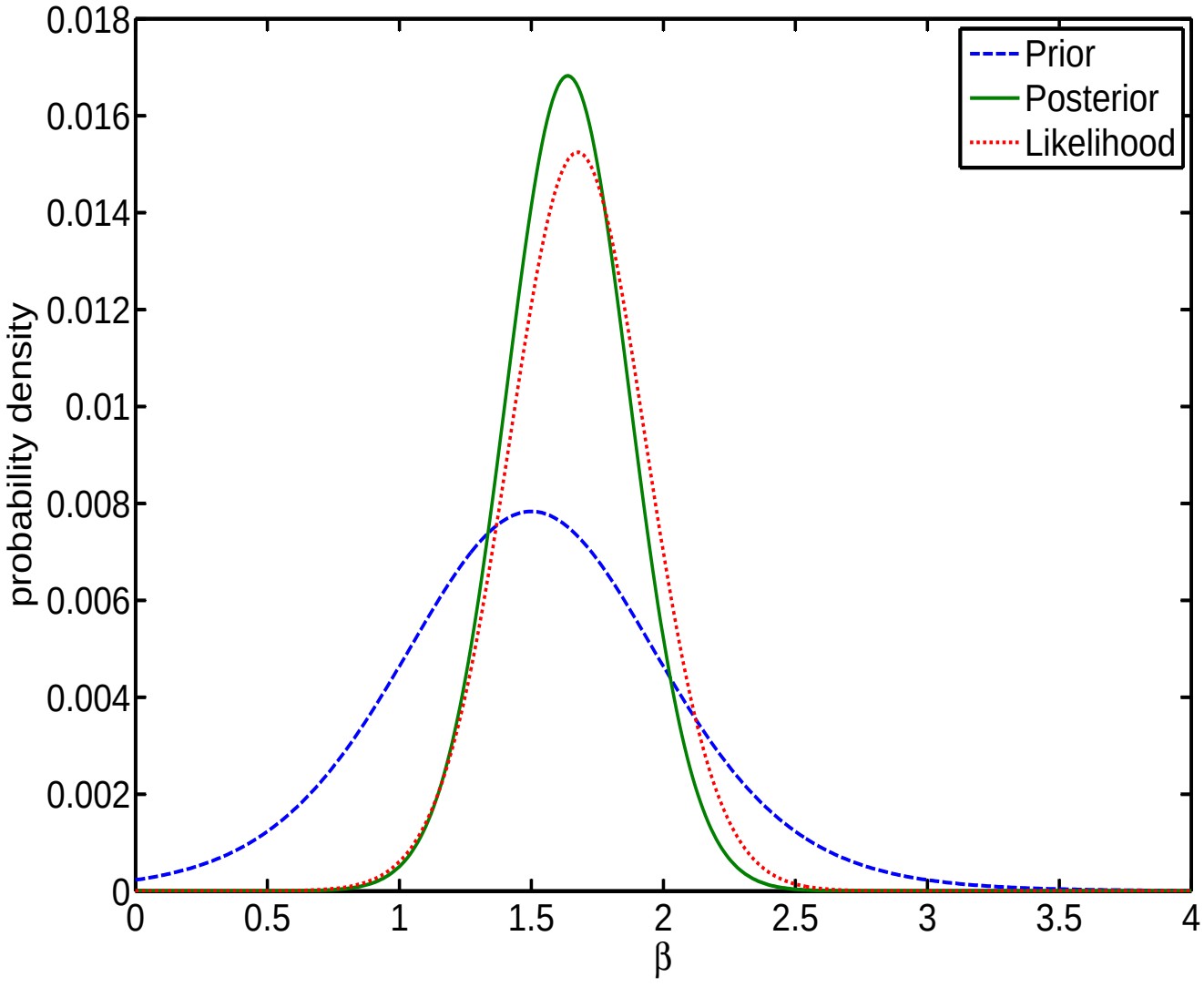
- Walk before we run etc...

The linear regression model

$$\begin{aligned} y_i &= \beta x_i + \varepsilon_i \sim N(0, \sigma^2) \\ i &= 1, 2, ...N \end{aligned}$$

The variable x_i is fixed (i.e. not random) or independent of ε_i and with a probability density function $p(x_i \mid \lambda)$ where the parameter(s) λ does not include β or σ^2

Figure 2.1: Marginal Prior and Posteriors for β



The likelihood function

Defining the likelihood function

The assumptions made about the linear regression model implies that $p\left(y_i \mid \beta, \sigma^2\right)$ is Normal.

$$p\left(y_i \mid \beta, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(y_i - \beta x_i\right)^2}{2\sigma^2}\right]$$

or that

$$p\left(y \mid \beta, \sigma^2\right) = \frac{1}{(2\pi)^{\frac{N}{2}}\,\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - \beta x_i\right)^2\right]$$

where

$$y \equiv \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix}'$$

Reformulating the likelihood function

It will be convenient to use that

$$\sum_{i=1}^{N} (y_i - \beta x_i)^2 = \nu s^2 + \left(\beta - \widehat{\beta}\right)^2 \sum_{i=1}^{N} x_i^2$$

where

$$\begin{aligned} \nu &= N - 1 \\ \widehat{\beta} &= \frac{\sum x_i y_i}{\sum x_i^2} \\ s^2 &= \frac{\sum (y_i - \beta x_i)^2}{\nu} \end{aligned}$$

Reformulating the likelihood function

The likelihood function can then be written as

$$p\left(y \mid \beta, \sigma^2\right) = \frac{1}{(2\pi)^{\frac{N}{2}}} \left\{ h^{\frac{1}{2}} \exp\left[-\frac{h}{2} \left(\beta - \widehat{\beta}\right)^2 \sum_{i=1}^{N} x_i^2 \right] \right\} \times \left\{ h^{\frac{v}{2}} \exp\left[-\frac{hv}{2s^{-2}} \right] \right\}$$

where $h \equiv \sigma^{-2}$.

Formulating a prior

The Prior density

The prior density should reflect information about θ that we have prior to observing y.

Priors should preferably

- be easy to interpret
- be of a form that facilitates computing the posterior

Natural conjugate priors have both of these advantages.

Conjugate priors

Conjugate priors, when combined with the likelihood function, result in posteriors that are of the same family of distributions

- Natural conjugate priors has the same functional form as the likelihood function

The posterior can $p(\theta \mid y)$ then be thought of as being the result of two "sub-samples"

$$p(\theta \mid y) \propto p(y \mid \theta)p(\theta)$$

- The sufficient statistic about θ of the first "sub-sample" is described by the hyper-parameters of the prior
- The information about θ in second "sub-sample" can be described by the sufficient statistic for the actual observed sample (i.e. y)

This feature makes natural conjugate priors easy to interpret

Specifying a prior density: It's your choice

For the Normal regression model we need to "elicit", i.e. formulate, a prior $p(\theta)$ for β and h which we can denote $p(\beta, h)$.

- It will be convenient to use that

$$p(\beta, h) = p(\beta \mid h)p(h)$$

  and let $\beta \mid h \sim N(\underline{\beta}, h^{-1}\underline{V})$ and let $h \sim G\left(\underline{s}^{-2}, \underline{\nu}\right)$.
- The natural conjugate prior for β and h is then denoted

$$\beta, h \sim NG\left(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}\right)$$

Notation: Bars below parameters denote hyper-parameters of a prior, and bars above denote hyper-parameters of a posterior.

What makes a prior conjugate?

Short answer: Convenient functional forms
Long answer: Remember the likelihood function

$$p\left(y \mid \beta, \sigma^2\right) = \frac{1}{(2\pi)^{\frac{N}{2}}} \left\{ h^{\frac{1}{2}} \exp\left[-\frac{h}{2} \left(\beta - \widehat{\beta}\right)^2 \sum_{i=1}^{N} x_i^2 \right] \right\} \times \left\{ h^{\frac{v}{2}} \exp\left[-\frac{hv}{2s^{-2}} \right] \right\}$$

and consider the normal-gamma prior distribution $p\left(\beta \mid h\right) p(h)$ where $\beta \mid h \sim N(\underline{\beta}, h^{-1}\underline{V})$ and $h \sim G\left(\underline{s}^{-2}, \underline{v}\right)$. We then have

$$p\left(\beta \mid h\right) p(h) = \left\{ \frac{1}{\sqrt{2\pi h^{-1}\underline{V}}} \exp\left[-\frac{\left(\beta - \underline{\beta}\right)^2}{2h^{-1}\underline{V}} \right] \right\} \times \left\{ c_G^{-1} h^{\frac{\underline{v}-2}{2}} \exp\left(-\frac{h\underline{v}}{2\underline{s}^{-2}} \right) \right\}$$

Note the similar functional forms.

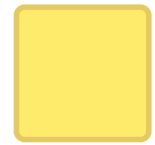
The posterior

The Normal-Gamma posterior

It is possible to use Bayes rule

$$p(\beta, h \mid y) = \frac{p(y \mid \beta, h) p(\beta \mid h) p(h)}{p(y)}$$

to find an analytical expression for the posterior that is Normal-Gamma



$$p(\beta, h \mid y) = \left\{ \frac{1}{\sqrt{2\pi \overline{V}}} \exp\left[-\frac{\left(\beta - \overline{\beta}\right)^2}{2\overline{V}} \right] \right\} \left\{ c_G^{-1} h^{\frac{\overline{\nu}-2}{2}} \exp\left(-\frac{h\overline{\nu}}{2\overline{s}^{-2}} \right) \right\}$$

i.e. the same form as the prior

- $\beta, h \mid y \sim NG\left(\overline{\beta}, \overline{V}, \overline{s}^{-2}, \overline{\nu}\right)$
- $\beta, h \sim NG\left(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}\right)$

The posterior

The (hyper) parameters of the posterior

$$\beta, h \mid y \sim NG\left(\overline{\beta}, \overline{V}, \overline{s}^{-2}, \overline{\nu}\right)$$

are a combination of the (hyper) parameters of the prior and sample information

$$\begin{aligned}
\overline{V} &= \frac{1}{\underline{V}^{-1} + \sum x_i^2} \\
\overline{\beta} &= \overline{V}\left(\underline{V}^{-1}\underline{\beta} + \widehat{\beta}\sum x_i^2\right) \\
\overline{\nu} &= \underline{\nu} + N \\
\overline{\nu s}^2 &= \underline{\nu s}^2 + \nu s^2 + \frac{\left(\widehat{\beta} - \underline{\beta}\right)^2}{\underline{V} + \sum x_i^2}
\end{aligned}$$

Combining prior and sample information

The posterior mean of β is given by

$$\begin{aligned} \overline{\beta} &= \overline{V}\left(\underline{V}^{-1}\underline{\beta} + \widehat{\beta}\sum x_i^2\right) \\ \overline{V} &= \frac{1}{\underline{V}^{-1} + \sum x_i^2} \end{aligned}$$

Do these expressions make sense?

- $\underline{V}^{-1}$ is the precision of the prior
- $\sum x_i^2$ is the "precision" of the data

What happens when $\underline{V}^{-1} = 0$? when $\underline{V} \to 0$?

Example

Simple "empirical" example

Consider a sample of 50 observations generated from the model

$$y_i = \beta x_i + \varepsilon_i \sim N(0, \sigma^2)$$

with $\beta = 2$, $\sigma^2 = 1$ and $x_i \sim N(0, 1)$

Set the prior hyper parameters in $\beta, h \sim NG\left(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}\right)$ to

$$\begin{aligned} \underline{\beta} &= 1.5 \\ \underline{V} &= 0.25 \\ \underline{s}^{-2} &= 1 \\ \underline{\nu} &= 10 \end{aligned}$$

The data

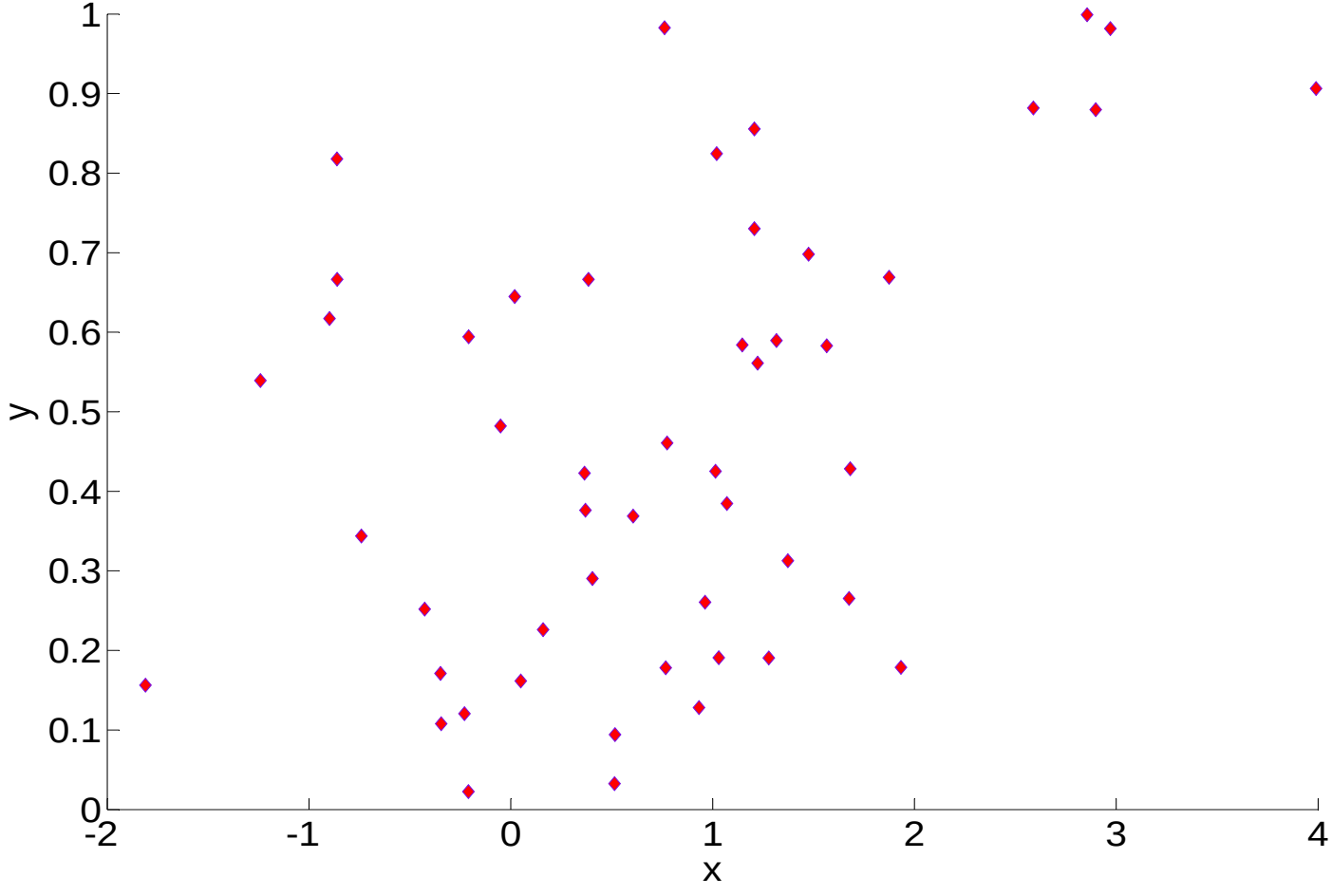
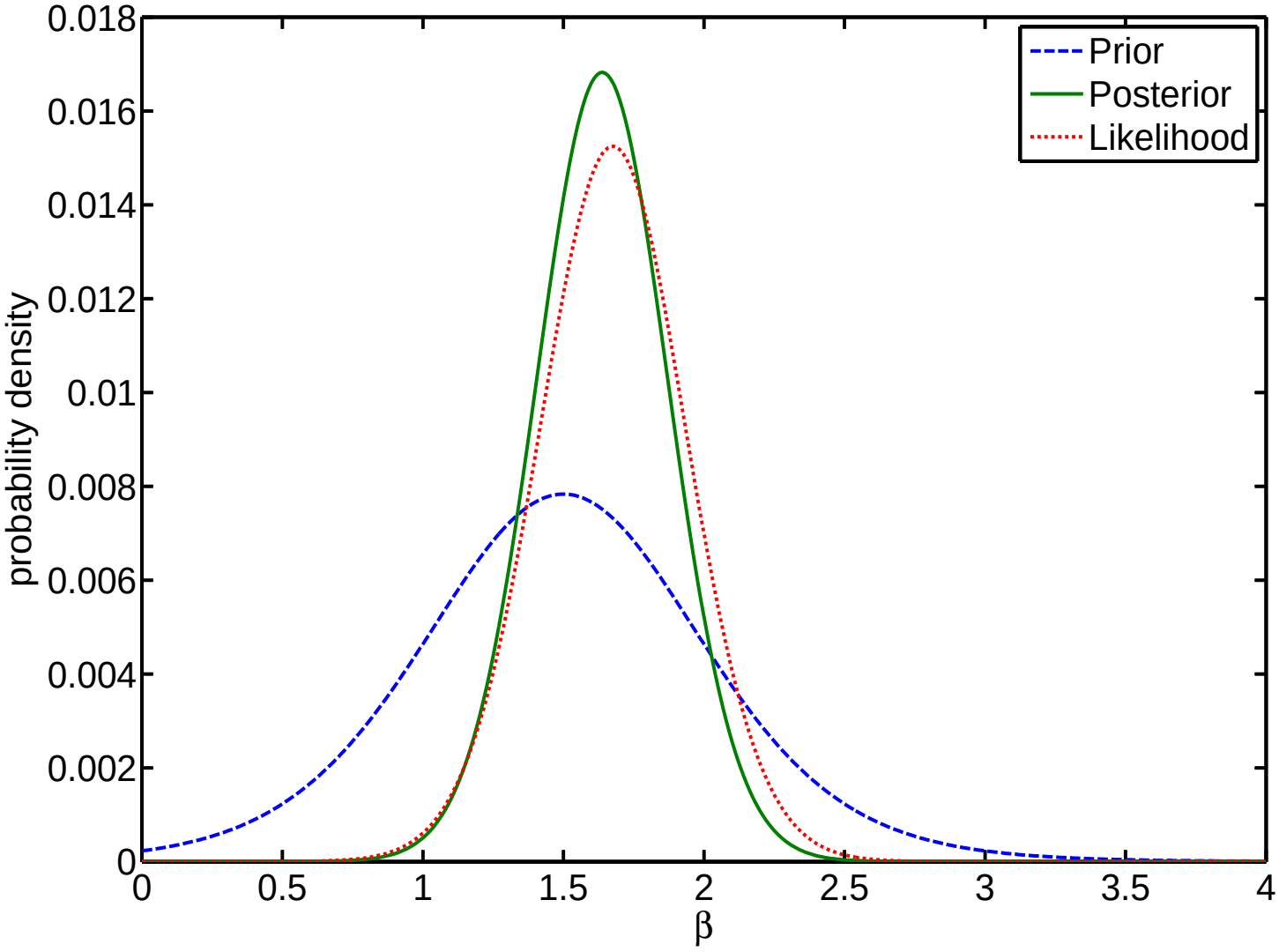


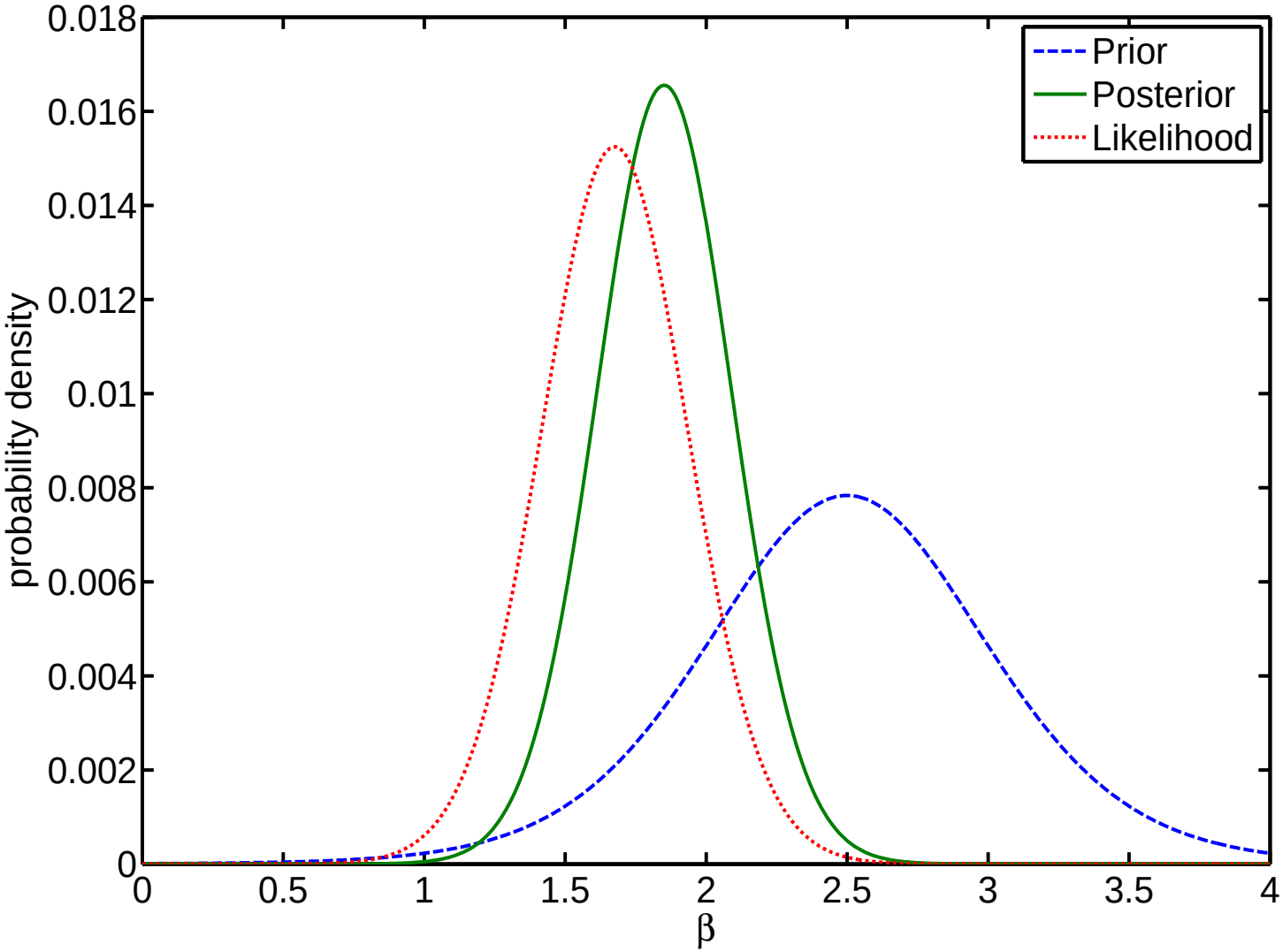
Figure 2.1: Marginal Prior and Posteriors for β



Change the prior mean

$$
\begin{aligned}
\underline{\beta} &= 1.5 \Rightarrow 2.5 \\
\underline{V} &= 0.25 \\
\underline{s}^{-2} &= 1 \\
\underline{\nu} &= 10
\end{aligned}
$$

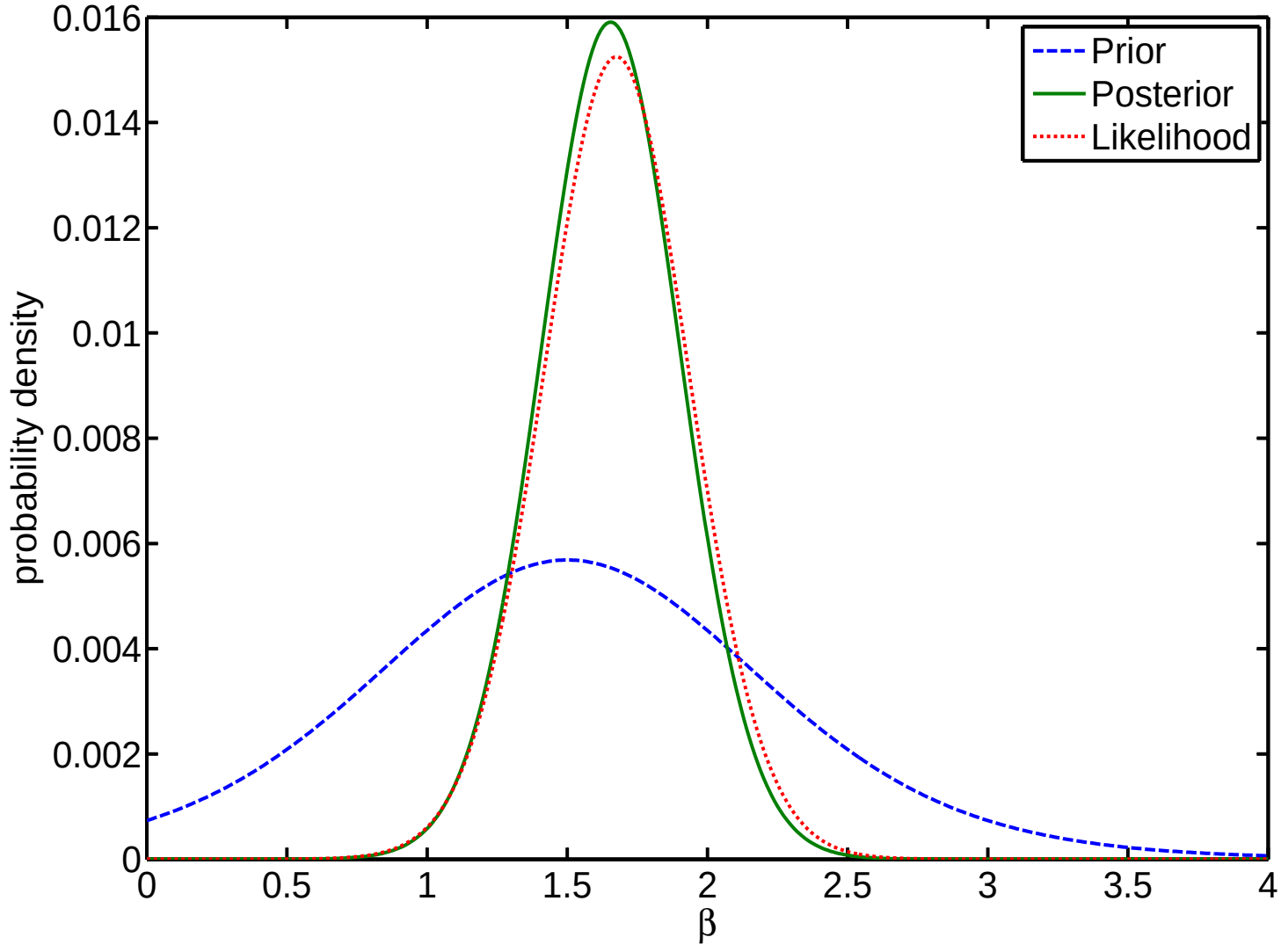
Figure 2.1: Marginal Prior and Posteriors for β



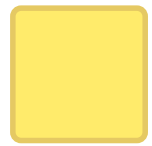
Reduce the prior precision

$$
\begin{aligned}
\underline{\beta} &= 1.5 \\
\underline{V} &= 0.25 \Rightarrow 0.5 \\
\underline{s}^{-2} &= 1 \\
\underline{\nu} &= 10
\end{aligned}
$$

Figure 2.1: Marginal Prior and Posteriors for β



More data: N=100



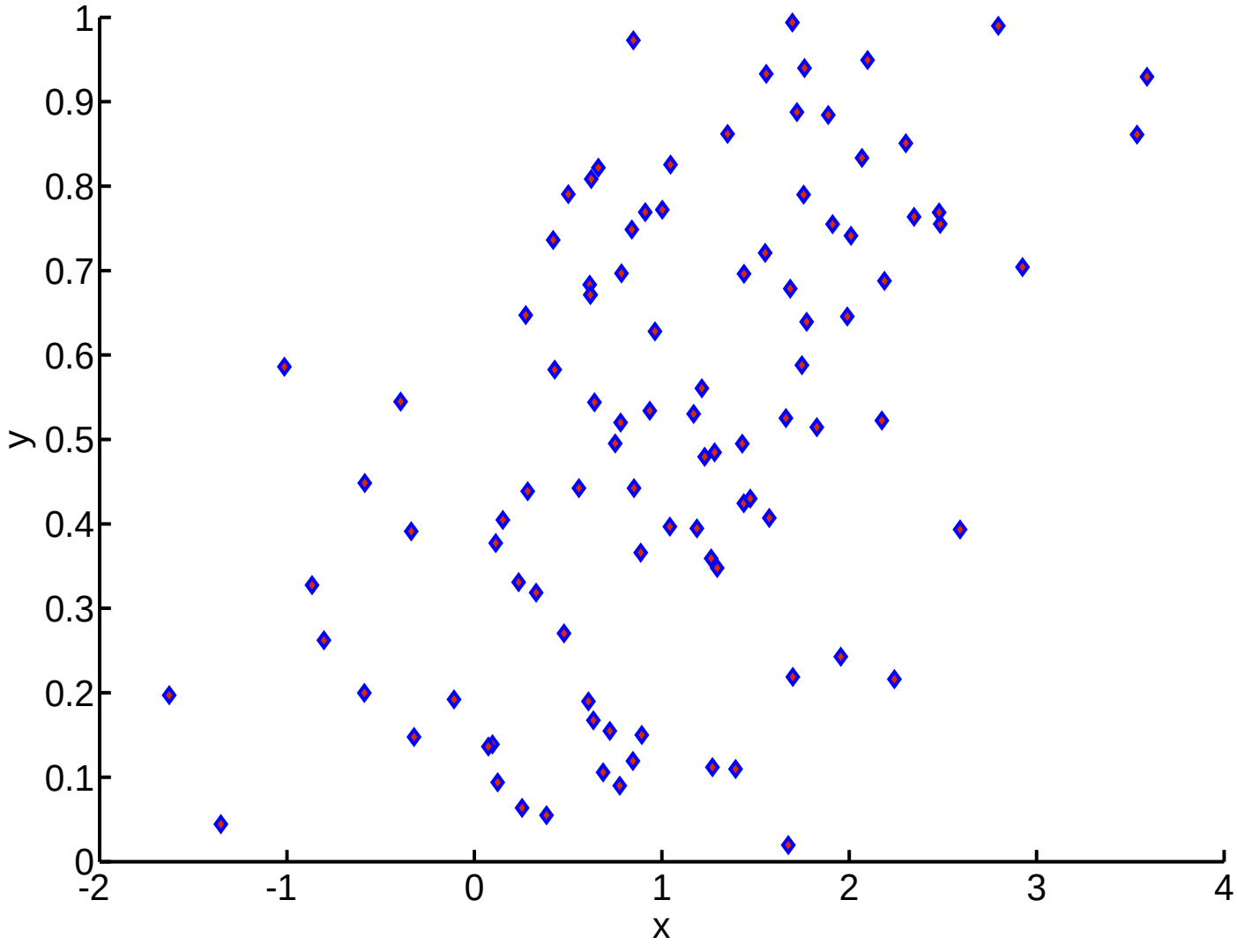
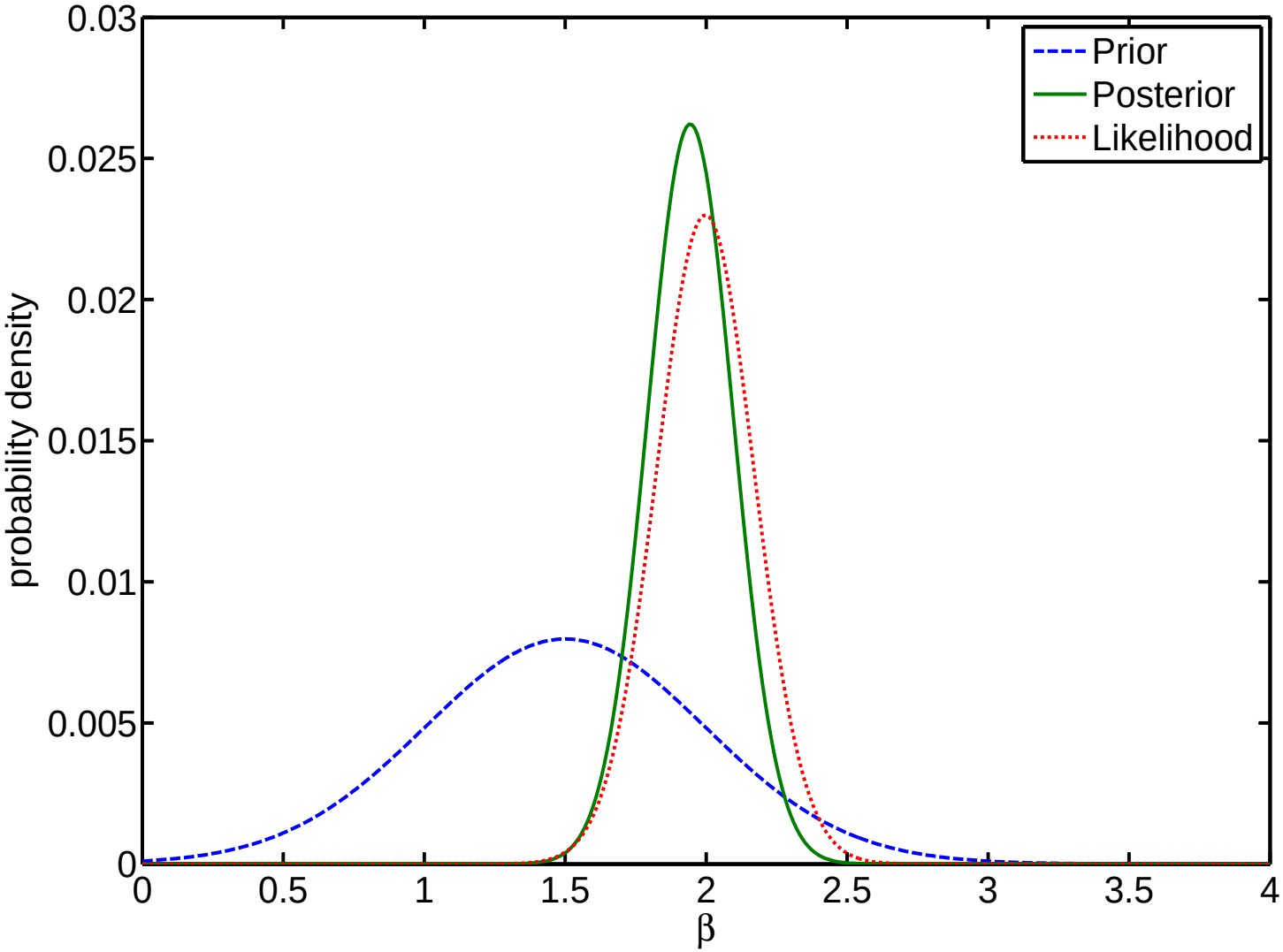
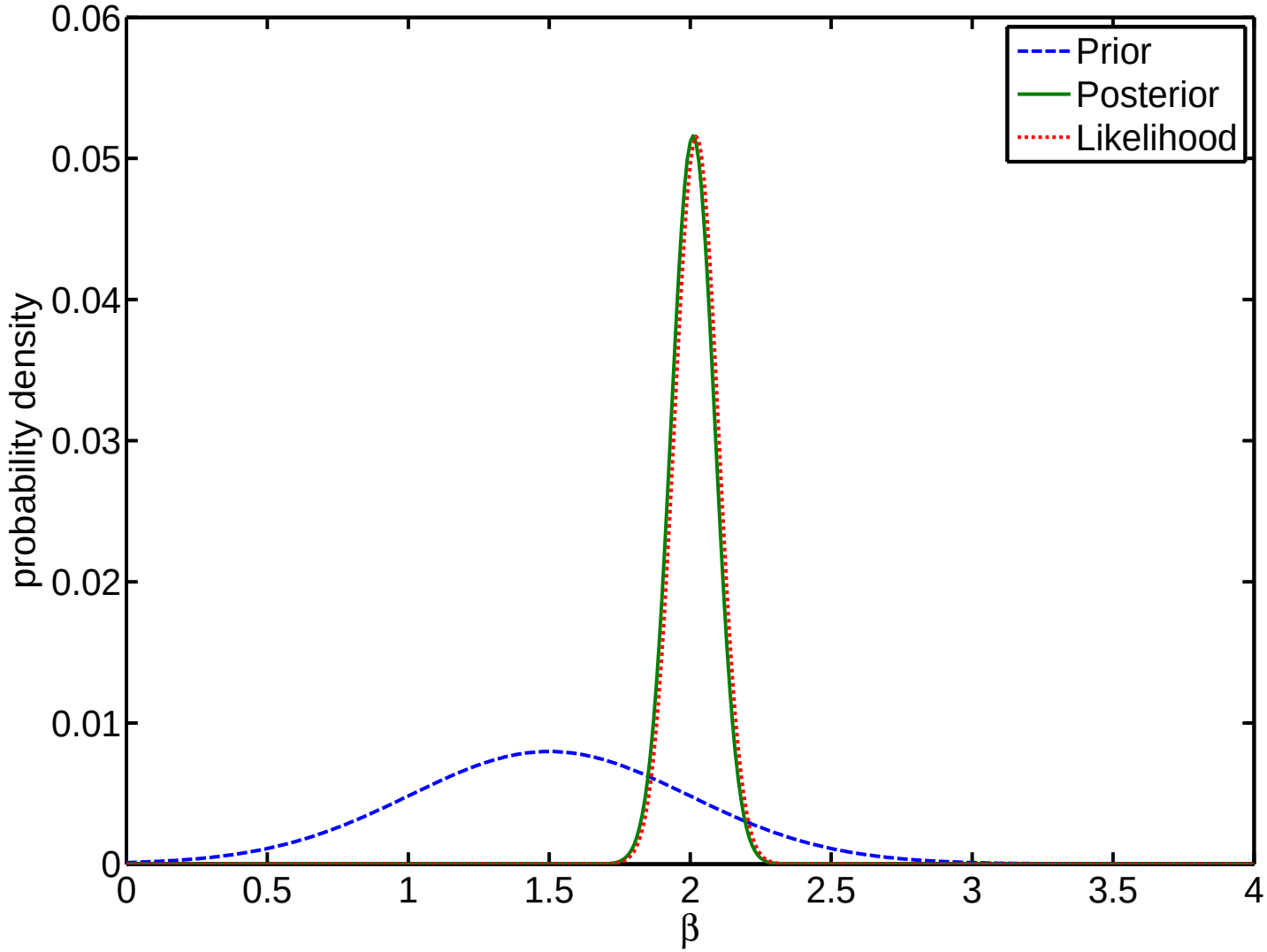

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
y
-2
-1
0
1
2
3
4
x

Figure 2.1: Marginal Prior and Posteriors for β



Even more data:N=500

Figure 2.1: Marginal Prior and Posteriors for β



Model comparison

Bayesian statistics generally do not "reject" models

Instead: assign probabilities to different models

- We can index different models by $i = 1, 2, ...m$

$$p(\theta \mid y, M_i) = \frac{p(y \mid \theta, M_i)p(\theta \mid M_i)}{p(y \mid M_i)}$$

The posterior odds ratio

The *posterior odds ratio* is the relative probabilities of two models conditional on the data

$$\frac{p(M_i \mid y)}{p(M_j \mid y)} = \frac{p(y \mid M_i)p(M_i)}{p(y \mid M_j)p(M_j)}$$

It is made up of

- The *prior odds* ratio $\frac{p(M_i)}{p(M_j)}$
- The *Bayes factor* $\frac{p(y|M_i)}{p(y|M_j)}$

Concepts of Bayesian analysis

Prior densities, likelihood functions, posterior densities, posterior odds ratios, Bayes factors etc are part of almost all Bayesian analysis.

- Data, choice of prior densities and the likelihood function are the inputs into the analysis
- Posterior densities etc are the outputs

The rest of the course is about how to choose these inputs and how to compute the outputs

Bayesian computation

Bayesian computation

More specifics about the outputs

- Posterior mean $E(\theta \mid y) = \int \theta p(\theta \mid y) d\theta$
- Posterior variance $var(\theta \mid y) = E\left[[\theta - E(\theta \mid y)] \mid y\right]^2$
- Posterior $prob(\theta_i > 0)$

These objects can all be written in the form

$$E(g(\theta) \mid y) = \int g(\theta) p(\theta \mid y) d\theta$$

where $g(y)$ is the function of interest.

Posterior simulation

There are only a few cases when the expected value of functions of interest can be derived analytically.

- Instead, we rely on *posterior simulation* and *Monte Carlo integration.*
- Posterior simulation consists of constructing a sample from the posterior distribution $p(\theta \mid y)$

Monte carlo integration then uses that

$$\widehat{g}_S = \frac{1}{S} \sum_{s=1}^{S} g\left(\theta^{(s)}\right)$$

and that $\lim_{S \to \infty} \widehat{g}_S = E(g(\theta) \mid y)$ where $\theta^{(s)}$ is a draw from the posterior distribution.

The end-product of Bayesian statistics

Most of Bayesian econometrics consists of simulating distributions of parameters using numerical methods.

- A simulated posterior is a numerical approximation to $p(\theta \mid y)$
- We rely on ergodicity, i.e. that the moments of the constructed sample correspond to the moments of the distribution $p(\theta \mid y)$

The most popular (and general) procedure to simulate the posterior is called the Random Walk Metropolis Algorithm

Bayesian DSGE models

"Modern macro models" (Kocherlakota)

1. Specifies constraints for households, technologies for firms and resource constraints for the whole economy
2. Specifies household preferences and firm objectives
3. Assumes forward looking behavior
4. Includes stochastic processes
5. Are models of the entire economy: No prices are exogenous

Bayesian DSGE models

A strategy for estimating structural macro models:

- Specify problems of representative agent and firm and the technologies at their disposal
- Log-linearize model to find approximate first order conditions and budget constraints
- Estimate parameters of model using likelihood based methods
  - "Likelihood based" means that we specify a complete statistical model that is assumed to be the data generating process

Bayesian estimation

- Combine prior and sample information in a transparent manner

A minimalistic DSGE model

The building blocks of a minimalistic DSGE model

A simple New Keynesian model

- A representative agent that consumes and supply labor
- A representative firm that sets prices to maximize profit
- A central bank that sets nominal interest rates
- Exogenous shocks (or "wedges")

Preferences and technology

The representative agent's objective

In period t the representative agent maximizes expected future utility

$$E_t \left\{ \sum_{s=0}^{\infty} \beta^s U\left(C_{t+s}(i), N_{t+s}(i)\right) \right\}$$

where C_t is consumption, N_t is supplied hours of labor.

Agents are assumed to be forward looking, but also to value the present more than the future ($0 < \beta < 1$).

The representative agent's objective

Period t utility

$$U\left(C_t(i), N_t(i)\right) = \frac{C_t(i)^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}} - \frac{N_t(i)^{1+\varphi}}{1+\varphi}$$

Main assumptions:
Households like to consume but they do not like to work.

- Decreasing marginal utility of consumption
- Increasing marginal disutility of working

But we also impose specific functional forms.

The decisions made by the representative agent

- How much to consume and how much to save (Euler-equation)

$$U_c(C_t) = \beta E_t R_t \frac{P_t U_c(C_{t+1})}{P_{t+1}}$$

- How much to work:

$$\frac{W_t}{P_t} U_c(C_t) = N_t^{\varphi}$$

Production of goods

Output in firm $j \in (0,1)$ is produced using labor as the sole input

$$Y_t(j) = \exp(a_t) N_t(j)$$

where a_t is an exogenous labor augmenting technology.

Price setting

Prices are reset only infrequently (with $prob = 1 - \theta$).

This results in a (linearized) Phillips curve

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(y_t - x_t)$$

where x is potential output.

Monopolistic competition: Each firm produces a unique good and has some market power.

Monetary policy

Monetary policy is often represented by a simple Taylor-type rule

$$r_t = \phi_y y_t + \phi_\pi \pi_t + \phi_r r_{t-1} + u_t^r$$

Can be viewed as a reduced form representation of an optimizing central bank and fits the data quite well.

We will simplify the model further by assuming that $\phi_y = \phi_r = 0$.

The linearized structural system

After linearizing, the main model equations are given by

$$\begin{aligned}
\pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - \overline{y}_t) \\
y_t &= E_t y_{t+1} - \gamma\left(i_t - E_t \pi_{t+1}\right) \\
i_t &= \phi \pi_t + u_t^r : u_t^r \sim N(0, \sigma_r^2) \\
x_t &= \rho x_{t-1} + u_t^x : u_t^x \sim N(0, \sigma_x^2)
\end{aligned}$$

typo: should be x_t

where π_t, y_t, y_t, i_t are inflation, output, potential output and nominal interest rate respectively.

We want to estimate the parameters $\theta = \{\rho, \gamma, \kappa, \phi, \sigma_x, \sigma_r\}$

Solving the model

3 ways to solve a linear model

Solving a model using full information rational expectations as the equilibrium concept involves integrating out expectations terms from the structural equations of the model by replacing agents' expectations with the mathematical expectation, conditional on the state of the model.

Three different ways of doing this.

1. Method of undetermined coefficients, can be very quick when feasible and illustrates the fixed point nature of the rational expectations solution.
2. Replacing expectations with linear projections onto observable variables
3. Decouple the stable and unstable dynamics of the model and set the unstable part to zero.

The 3 equation NK model

As a vehicle to demonstrate the different solution methods, we will use a simple New-Keynesian model

$$\begin{aligned}
\pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - \overline{y}_t) \\
y_t &= E_t y_{t+1} - \gamma\left(i_t - E_t \pi_{t+1}\right) \\
i_t &= \phi \pi_t + u_t^r : u_t^r \sim N(0, \sigma_r^2) \\
x_t &= \rho x_{t-1} + u_t^x : u_t^x \sim N(0, \sigma_x^2)
\end{aligned}$$

where π_t, y_t, y_t, i_t are inflation, output, potential output and nominal interest rate respectively.

- Single variable, potential output x_t, as the state.

Method I:
Method of undetermined coefficients

Pros

- Method is quick when feasible
- Illustrates well the fixed point nature of rational expectations equilibria.

Cons

- Difficult to implement in larger models

Method of undetermined coefficients

$$
\begin{aligned}
\pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - \overline{y}_t) \\
y_t &= E_t y_{t+1} - \gamma\left(i_t - E_t \pi_{t+1}\right) \\
i_t &= \phi \pi_t + u_t^r : u_t^r \sim N(0, \sigma_r^2) \\
x_t &= \rho x_{t-1} + u_t^x : u_t^x \sim N(0, \sigma_x^2)
\end{aligned}
$$

Start by substituting in the interest rate in the Euler equation

$$
\begin{aligned}
x_t &= \rho x_{t-1} + u_t^x \\
y_t &= E_t(y_{t+1}) - \gamma\left[\phi_\pi \pi_t + u_t^r - E_t\left(\pi_{t+1}\right)\right] \\
\pi_t &= E_t\left(\pi_{t+1}\right) + \kappa\left[y_t - x_t\right]
\end{aligned}
$$

Solving model using method of undetermined coefficients

Conjecture that model can be put in the form

$$\begin{aligned} x_t &= \rho x_{t-1} + u_t^x \\ y_t &= a x_t \quad (+ \ c u_t^r) \\ \pi_t &= b x_t \end{aligned}$$

Why is this a good guess?

Solving model using method of undetermined coefficients

Substitute in conjectured form of solution into structural equation

$$\begin{aligned} ax_t &= a\rho x_t - \gamma\left[\phi_\pi b x_t - b\rho x_t\right] \\ bx_t &= b\rho x_t + \kappa\left[a x_t - x_t\right] \end{aligned}$$

where we used that $x_t = \rho x_{t-1} + u_t^x$ implies that $\mathsf{E}[x_{t+1} \mid x_t] = \rho x_t$

Solving model using method of undetermined coefficients

Equate coefficients on right and left hand side

$$\begin{aligned} a &= a\rho - \gamma\phi_\pi b + \gamma b\rho \\ b &= b\rho + \kappa\left[a - 1\right] \end{aligned}$$

or

$$\begin{bmatrix} (1-\rho) & \gamma\left(\phi_\pi - \rho\right) \\ -\kappa & (1-\rho) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ -\kappa \end{bmatrix}$$

Solving model using method of undetermined coefficients

Solve for a and b

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} (1-\rho) & \gamma(\phi_\pi - \rho) \\ -\kappa & (1-\rho) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -\kappa \end{bmatrix}$$

or

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -\kappa \frac{\phi-\rho}{-c} \\ \kappa\gamma \frac{1-\rho}{-c} \end{bmatrix}$$

where $c = \gamma - \kappa\rho - 2\gamma\rho + \kappa\phi + \gamma\rho^2 < 0$

The solved model

The solved model is of the form

$$\begin{aligned} x_t &= \rho x_{t-1} + u_t^x \\ y_t &= -\kappa \frac{\rho - \phi_\pi}{c} x_t \\ \pi_t &= \kappa\gamma \frac{\rho - 1}{c} x_t \end{aligned}$$

where $c = \gamma - \kappa\rho - 2\gamma\rho + \kappa\phi + \gamma\rho^2 < 0$

~~Method II: Replacing expectations with linear projections~~

skipped here

The second method uses that projections of the future values of variables on observables gives optimal expectations (in the sense of minimum error variance) if the observables span the space of the state.

How does it work?

- Replace $E_t\pi_{t+1}$ and $E_t y_{t+1}$ with linear projections of these variables on current inflation.
- There is nothing special about inflation. Projecting onto current output would also work.

But first we need to know (a little) more about projections.

Least squares estimation via the projection theorem

To find the estimate $\widehat{x}$ as a linear function of y simply use that

$$\begin{aligned} \langle x - \beta y, y \rangle &= E\left[(x - \beta y)\, y'\right] \\ &= 0 \end{aligned}$$

and solve for β

$$\beta = E\left(xy'\right)\left[E\left(yy'\right)\right]^{-1}$$

The projection theorem ensures that an estimate with orthogonal errors is the (linear) minimum variance estimate.

Two useful properties of linear projections

1. If two random variables X and Y are Gaussian, then the projection of Y onto X coincides with the conditional expectation $E(Y \mid X)$.
2. If X and Y are not Gaussian, the linear projection of Y onto X is the minimum variance linear prediction of Y given X.

Replacing expectations with linear projections

We will use that in equilibrium

$$
\begin{aligned}
E\left(\pi_{t+1} \mid \pi_t\right) &= \frac{cov(\pi_t, \pi_{t+1})}{var\left(\pi_t\right)}\pi_t \\
E\left(y_{t+1} \mid \pi_t\right) &= \frac{cov(\pi_t, y_{t+1})}{var\left(\pi_t\right)}\pi_t
\end{aligned}
$$

if the innovations u_t to x_t are Gaussian.

Replacing expectations with linear projections

Let

$$\begin{aligned} c_0\pi_t &= E^*\left(\pi_{t+1} \mid \pi_t\right) \\ d_0\pi_t &= E^*\left(y_{t+1} \mid \pi_t\right) \end{aligned}$$

denote initial candidate projections of expected inflation and output on current inflation. We can then write the structural equations as

$$\begin{aligned} \pi_t &= \beta c_0\pi_t + \kappa(y_t - x_t) \\ y_t &= d_0\pi_t - \gamma\left(\phi\pi_t - c_0\pi_t\right) \end{aligned}$$

Replacing expectations with linear projections

Put the whole system in matrix form

$$\begin{bmatrix} x_t \\ \pi_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \kappa & 1-\beta c_0 & -\kappa \\ 0 & -d_0+\gamma\phi-\gamma c_0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \pi_{t-1} \\ y_{t-1} \end{bmatrix}$$
$$+ \begin{bmatrix} 1 & 0 & 0 \\ \kappa & 1-\beta c_o & -\kappa \\ 0 & -d_0+\gamma\phi-\gamma c_0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u_t$$

or

$$X_t = AX_{t-1} + Cu_t$$

Solution algorithm

1. Make an initial guess of c_0 and d_0 in (1)
2. Compute the implied covariances of current inflation and future inflation and output using

$$\begin{aligned} E\left[X_t X_t'\right] &= \Sigma_{XX} \\ \Sigma_{XX} &= A\Sigma_{XX}A' + CC' \end{aligned}$$

   and

$$E\left[X_{t+1} X_t'\right] = A\Sigma_{XX}$$

3. Replace the c_s and d_s with the c_{s+1} and d_{s+1} in ((1))

$$\begin{aligned} c_{s+1} &= \frac{cov(\pi_t, \pi_{t+1})}{var(\pi_t)} \\ d_{s+1} &= \frac{cov(\pi_t, y_{t+1})}{var(\pi_t)} \end{aligned}$$

   using the covariances from Step 2.

4. Repeat Step 2-3 until c_s and d_s converges.

Method III: Stable/unstable decoupling

Originally due to Blanchard and Kahn (1980)

- Computational aspects of the method has been further developed by others, for instance Klein (2000).
- The most accessible reference is probably Soderlind (1999), who also has code posted on his web site.

The method has several advantages:

- Fast
- Provides conditions for when a solution exists
- Provides conditions for when the solution is unique.

Method III: Stable/unstable decoupling

Start by putting the model into matrix form

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \beta & 0 \\ 0 & \gamma & 1 \end{bmatrix} \begin{bmatrix} x_{t+1} \\ E_t\pi_{t+1} \\ E_t y_{t+1} \end{bmatrix}$$

$$= \begin{bmatrix} \rho & 0 & 0 \\ \kappa & 1 & -\kappa \\ 0 & \gamma\phi & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \pi_t \\ y_t \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u_{t+1}$$

or

$$A_0 \begin{bmatrix} x^1_{t+1} \\ E_t x^2_{t+1} \end{bmatrix} = A_1 \begin{bmatrix} x^1_t \\ x^2_t \end{bmatrix} + C_1 u_{t+1}$$

- x^1_t is vector containing the pre-determined and/or exogenous variables (i.e. x_t)
- x^2_t a vector containing the forward looking ("jump") variables (i.e. $E_t y_{t+1}$ and $E_t\pi_{t+1}$).

Method III: Stable/unstable decoupling

Pre-multiply both sides of

$$A_0 \begin{bmatrix} x_{t+1}^1 \\ E_t x_{t+1}^2 \end{bmatrix} = A_1 \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + C_1 u_{t+1}$$

by A_0^{-1} to get

$$\begin{bmatrix} x_{t+1}^1 \\ E_t x_{t+1}^2 \end{bmatrix} = A \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + C u_{t+1}$$

where $A = A_0^{-1} A_1$ and $C = A_0^{-1} C_1$.

For the model to have unique stable solution the number of stable eigenvalues of A must be equal to the number of exogenous/pre-determined variables.

Method III: Stable/unstable decoupling

Use a Schur decomposition to get

$$A = ZTZ^H$$

where T is upper block triangular

$$T = \begin{bmatrix} T_{11} & T_{12} \\ \mathbf{0} & T_{22} \end{bmatrix}$$

and Z is a unitary matrix so that $Z^H Z = ZZ^H = I$
($\implies Z^H = Z^{-1}$).

- For any square matrix W, $W^{-1}AW$ is a so called similarity transformation of A.
- Similarity transformations do not change the eigenvalues of a matrix
- We can always choose Z and T so that the unstable eigenvalues of A are shared with T_{22}

Method III: Stable/unstable decoupling

Define the auxiliary variables

$$\begin{bmatrix} \theta_t \\ \delta_t \end{bmatrix} = Z^H \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix}$$

We can then rewrite the system as

$$Z^H \begin{bmatrix} x_{t+1}^1 \\ E_t x_{t+1}^2 \end{bmatrix} = Z^H Z T Z^H \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix}$$

or equivalently

$$E \begin{bmatrix} \theta_{t+1} \\ \delta_{t+1} \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ \mathbf{0} & T_{22} \end{bmatrix} \begin{bmatrix} \theta_t \\ \delta_t \end{bmatrix}$$

since $Z^H Z = I$.

Method III: Stable/unstable decoupling

For this system to be stable, the auxiliary variables associated with the unstable roots in T_{22} must be zero for all t. (WHY?)

Imposing $\delta_t = 0 \forall t$ reduces the relevant state dynamics to

$$\theta_t = T_{11}\theta_{t-1}$$

To get back the original variables we simply use that

$$\begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} Z_{11} \\ Z_{21} \end{bmatrix} \theta_t$$

or

$$\begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} Z_{11} \\ Z_{21} \end{bmatrix} Z_{11}^{-1} x_t^1$$

which is the solution to the model. It is in the form

$$\begin{aligned} x_t^1 &= M x_{t-1}^1 + \varepsilon_t \\ x_t^2 &= G x_t^1 \end{aligned}$$

where $M = Z_{11} T_{11} Z_{11}^{-1}$ ($= \rho$ in our example) and $G = Z_{21} Z_{11}^{-1}$.

The model as a state space system

A model can be viewed as a mapping from a set of parameters to a state space system

$$\begin{aligned} X_t &= AX_{t-1} + Cu_t \\ Z_t &= DX_t \end{aligned}$$

where for our model the matrices A, C and D are given by

$$\begin{aligned} X_t &= x_t, A = \rho, Cu_t = u_t^x \\ Z_t &= \begin{bmatrix} r_t \\ \pi_t \\ y_t \end{bmatrix}, D = \begin{bmatrix} \phi_\pi \\ \kappa\gamma\frac{1-\rho}{-c} \\ -\kappa\frac{\phi_\pi-\rho}{-c} \end{bmatrix} \end{aligned}$$

Today's Box Folder

- Lecture slides
- Lecture notes on
  - Solving linearized rational expectations models
  - Projection theorem
- Kocherlakota on *Modern Macro Models*
- Soderlind (1999)
- Lucas (1976)

Additional references

Bayesian econometrics:

- *Bayesian Econometrics*, Gary Koop, Wiley 2003.
- *Contemporary Bayesian Econometrics and Statistics*, John Geweke, Wiley 2005.

New Keynesian macro models:

- *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Jordi Gali, Princeton 2008.