Bayesian analysis of VARs, State-space models and DSGEs Part I: The Bayesics

Mattias Villani

Division of Statistics and Machine Learning
Department of Computer and Information Science
Linköping University

Lecture overview

- The Bayesics: **Prior**, **Likelihood** and **Posterior**
- **Bernoulli** model - **Beta** prior
- **Normal** model - Normal prior
- Bayesian analysis of **linear regression**
- Bayesian analysis of **autoregressive processes**

The likelihood function - Bernoulli trials

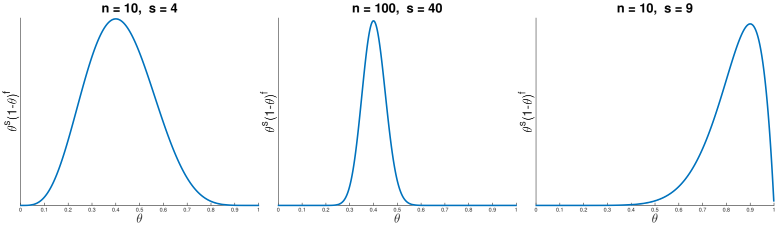
- **Bernoulli trials** (coin flips):

$$x_1, \ldots, x_n | \theta \overset{iid}{\sim} Bern(\theta).$$

- **Likelihood** from $s = \sum_{i=1}^{n} x_i$ successes and $f = n - s$ failures.

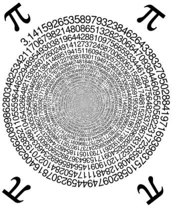
$$p(x_1, \ldots, x_n | \theta) \quad = \quad p(x_1|\theta) \cdots p(x_n|\theta) = \theta^s (1-\theta)^f$$

- **Maximum likelihood estimator** $\hat{\theta} = s/n$ maximizes $p(x_1, \ldots, x_n|\theta)$.
- Given the data $x_1, \ldots, x_n$, we can plot $p(x_1, \ldots, x_n|\theta)$ as a function of θ.



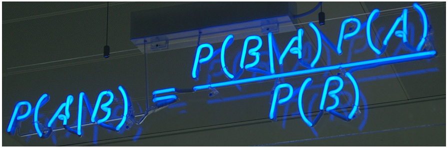
Uncertainty and subjective probability

- Statements like $\Pr(\theta < 0.6|\text{data})$ only make sense if θ is random.
- But what if θ is a fixed constant? Marginal propensity to consume.
- **Bayesian: doesn't matter if θ is fixed or random**.
- Do You know the value of θ or not?
- $p(\theta)$ reflects Your knowledge/**uncertainty** about θ.
- **Subjective probability**.
- $\Pr(\text{last quarter's GDP growth} < 0) = 0.34$.
- $\Pr(\text{Inflation in 2016} > 3\%) = 0.05$.



Bayesian learning

- **Bayesian learning** about a model parameter θ from data:
  - state your **prior** knowledge as a **prior** probability distribution $p(\theta)$.
  - **collect data** $\mathbf{x}$ and form the **likelihood** function $p(\mathbf{x}|\theta)$.
  - **combine** your prior knowledge $p(\theta)$ with the data information $p(\mathbf{x}|\theta)$ to a **posterior distribution** $p(\theta|\mathbf{x})$.
- Prior comes from: previous data, other data, experience etc. **Subjective**.
- **How to combine** the two sources of information? **Bayes' theorem**. **Objective**!



LEARNING FROM DATA - BAYES' THEOREM

- How do we **update** from the **prior** $p(\theta)$ to the **posterior** $p(\theta|Data)$?
- **Bayes' theorem** for events A and B

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}.$$

- Bayes' Theorem for a model parameter θ

$$p(\theta|Data) = \frac{p(Data|\theta)p(\theta)}{p(Data)}.$$

- It is the prior that turns the likelihood function $p(Data|\theta)$ into a posterior **probability density** $p(\theta|Data)$.
- A probability distribution for θ is extremely useful. **Decision making**.

THE NORMALIZING CONSTANT IS NOT IMPORTANT

- Bayes theorem

$$p(\theta|Data) = \frac{p(Data|\theta)p(\theta)}{p(Data)} = \frac{p(Data|\theta)p(\theta)}{\int_\theta p(Data|\theta)p(\theta)d\theta}.$$

- The integral $p(Data) = \int_\theta p(Data|\theta)p(\theta)d\theta$ can make you cry.
- $p(Data)$ is just a constant that makes $p(\theta|Data)$ integrate to one.
- Example: $x \sim N(\mu, \sigma^2)$

$$p(x) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right].$$

- We may write

$$p(x) \propto \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right].$$

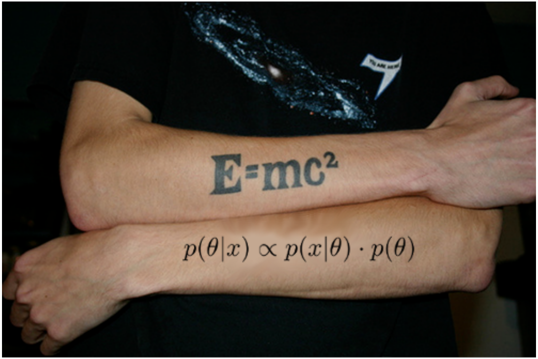
Great theorems make great tattoos

- All you need to know:

$$p(\theta|Data) \propto p(Data|\theta)p(\theta)$$

or

$$\text{Posterior} \propto \text{ Likelihood } \cdot \text{ Prior}$$



Bernoulli trials - Beta prior

- **Model**

$$x_1, ..., x_n | \theta \overset{iid}{\sim} Bern(\theta)$$

- **Prior**

$$\theta \sim Beta(\alpha, \beta)$$

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \text{ for } 0 \leq \theta \leq 1.$$

- **Posterior**

$$\begin{aligned} p(\theta | x_1, ..., x_n) &\propto p(x_1, ..., x_n | \theta) p(\theta) \\ &\propto \theta^s (1-\theta)^f \theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &= \theta^{s+\alpha-1}(1-\theta)^{f+\beta-1}. \end{aligned}$$

- This is proportional to the $Beta(\alpha+s, \beta+f)$ density.
- The **prior-to-posterior** mapping reads

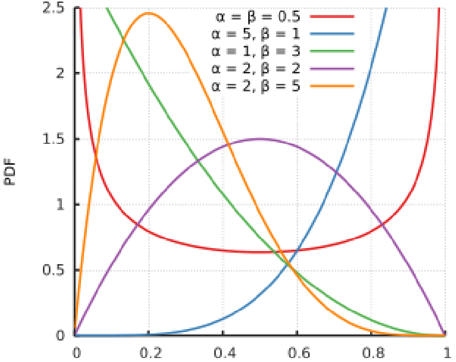
$$\theta \sim Beta(\alpha, \beta) \overset{x_1, ..., x_n}{\Longrightarrow} \theta | x_1, ..., x_n \sim Beta(\alpha+s, \beta+f).$$

Beta distribution

- $X \sim Beta(\alpha, \beta)$

$$E(X) = \frac{\alpha}{\beta}, \quad \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- Variance increases as $\alpha, \beta \to 0$.



Normal data, known variance - normal prior

- Model:

$$x_1, \ldots, x_n | \theta \sim N\left(\theta, \sigma^2\right), \quad \sigma^2 \;\; \texttt{known}$$

- Prior

$$\theta \sim N(\mu_0, \tau_0^2)$$

- Posterior

$$\begin{aligned} p(\theta | x_1, \ldots, x_n) &\propto p(x_1, \ldots, x_n | \theta, \sigma^2) p(\theta) \\ &\propto N(\theta | \mu_n, \tau_n^2), \end{aligned}$$

where

$$\frac{1}{\tau_n^2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2},$$

$$\mu_n = w\bar{x} + (1-w)\mu_0,$$

and

$$w = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}.$$

Normal data, known variance - normal prior

$$\theta \sim N(\mu_0, \tau_0^2) \overset{x_1,\ldots,x_n}{\Longrightarrow} \theta|\mathbf{x} \sim N(\mu_n, \tau_n^2).$$

$$\text{Posterior precision} = \text{Data precision} + \text{Prior precision}$$

$$\text{Posterior mean} =$$

$$\tfrac{\text{Data precision}}{\text{Posterior precision}}(\text{Data mean}) + \tfrac{\text{Prior precision}}{\text{Posterior precision}}(\text{Prior mean})$$

Normal data, known mean - Inv χ^2 prior

- **Model**:

$$x_1, ..., x_n|\theta \sim N\left(\theta, \sigma^2\right), \quad \theta \texttt{ known}$$

- **Prior:** **Scaled Inverse χ^2** prior $\sigma^2 \sim Inv - \chi^2\left(\nu_0, \tau_0^2\right)$

$$p(x) \propto \frac{\exp\left(\frac{-\nu\tau^2}{2x}\right)}{x^{\nu/2+1}}.$$

- Note that

$$Inv - \chi^2\left(\nu_0, \tau_0^2\right) = InvGamma\left(\frac{\nu_0}{2}, \frac{\nu_0\tau_0^2}{2}\right).$$

- **Posterior** is also scaled inverse χ^2:

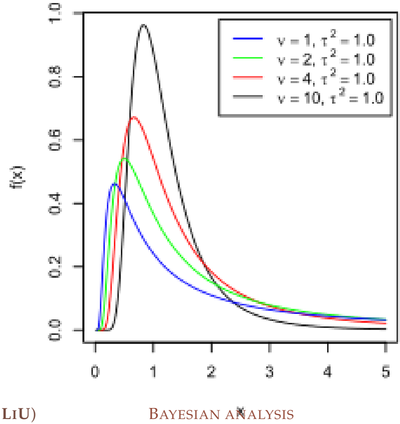
$$\sigma^2 \sim Inv - \chi^2\left(\nu_0, \tau_0^2\right) \overset{x_1,\ldots,x_n}{\Longrightarrow} \sigma^2|\mathbf{x} \sim Inv - \chi^2\left(\nu_0 + n, \frac{\nu_0\tau_0^2 + ns^2}{\nu_0 + n}\right).$$

- $\nu_0 \to 0$ makes the prior less informative.

Scaled Inv χ^2 distribution

- **Mean** (for $\nu > 2$) and **mode** for $X \sim Inv - \chi^2\left(\nu, \tau^2\right)$

$$E(X) = \frac{\nu}{\nu - 2}\tau^2, \quad \mathrm{Mode}(X) = \frac{\nu}{\nu + 2}\tau^2$$



Linear regression

- The **linear regression** model in matrix form

$$\underset{(n\times 1)}{\mathbf{y}} = \underset{(n\times k)(k\times 1)}{\mathbf{X}\beta} + \underset{(n\times 1)}{\varepsilon}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}$$

- Usually $x_{i1} = 1$, for all i. β_1 is the intercept.
- **Likelihood** for the full sample

$$\mathbf{y}|\beta, \sigma^2, \mathbf{X} \sim N(\mathbf{X}\beta, \sigma^2 I_n)$$

LINEAR REGRESSION - UNIFORM PRIOR

- Standard **non-informative prior**: uniform on $(\beta, \log \sigma^2)$

$$p(\beta, \sigma^2) \propto \sigma^{-2}$$

- **Joint posterior** of β and σ^2 [recall $p(X, Y) = p(X|Y)p(Y)$]:

$$\begin{aligned} \beta|\sigma^2, \mathbf{y} &\sim N\left[\hat{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right] \\ \sigma^2|\mathbf{y} &\sim Inv\text{-}\chi^2(n-k, s^2) \end{aligned}$$

  where $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $s^2 = \frac{1}{n-k}(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})$.
- Simulate from the joint posterior by iteratively simulating from
  - $p(\sigma^2|y)$
  - $p(\beta|\sigma^2, y)$
- **Marginal posterior** of β :

$$\beta|y \sim t_{n-k}\left[\hat{\beta}, s^2(X'X)^{-1}\right]$$

Linear regression - conjugate prior

- Joint **prior** for β and σ^2

$$\beta|\sigma^2 \sim N\left(\mu_0, \sigma^2\Omega_0^{-1}\right)$$
$$\sigma^2 \sim Inv - \chi^2\left(\nu_0, \sigma_0^2\right)$$

- **Posterior**

$$\beta|\sigma^2, \mathbf{y}, \mathbf{X} \sim N\left[\mu_n, \sigma^2\Omega_n^{-1}\right]$$
$$\sigma^2|\mathbf{y}, \mathbf{X} \sim Inv-\chi^2\left(\nu_n, \sigma_n^2\right)$$

$$\mu_n = \left(\mathbf{X}'\mathbf{X} + \Omega_0\right)^{-1}\mathbf{X}'\mathbf{X}\hat{\beta} + \left(\mathbf{X}'\mathbf{X} + \Omega_0\right)^{-1}\Omega_0\mu_0$$
$$\Omega_n = \mathbf{X}'\mathbf{X} + \Omega_0$$

AR PROCESS

- **AR process**

$$y_t = c + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t,\ \ \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

- Can be estimated as a regression with
  - $\mathbf{X} = (1_n, \mathbf{y}_{-1}, ..., \mathbf{y}_{-p})$
  - $\beta = (c, \phi_1, ..., \phi_k)'$.
- **Multivariate normal prior**

$$\beta|\sigma^2 \sim N\left(\mu_0, \sigma^2\Omega_0^{-1}\right).$$

- The **prior mean** is usually set to $\mu_0 = (0, r, 0, ..., 0)'$.
- Most probable model a priori: : $y_t = r \cdot y_{t-1} + \varepsilon_t$.
- The **prior covariance** is often set to

$$\Omega_{\mathbf{0}}^{-\mathbf{1}} = \begin{pmatrix} \tau_c^2 & 0 & 0 & \cdots & 0 \\ 0 & \tau_\phi^2 & 0 & \cdots & 0 \\ 0 & 0 & \frac{\tau_\phi^2}{2^\gamma} & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \frac{\tau_\phi^2}{k^\gamma} \end{pmatrix}$$

AR PROCESS

- **AR process**

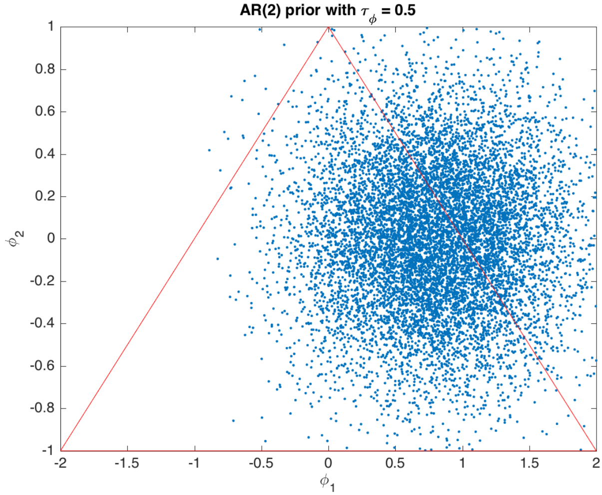
$$y_t = c + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t,\ \ \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

- The prior covariance

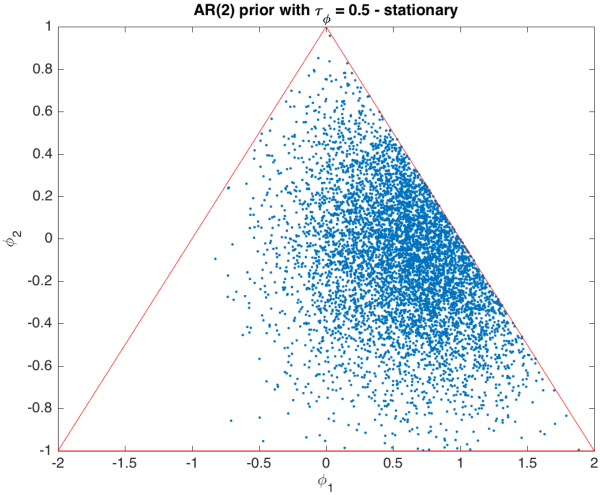
$$\Omega_0^{-1} = \begin{pmatrix} \tau_c^2 & 0 & 0 & \cdots & 0 \\ 0 & \tau_\phi^2 & 0 & \cdots & 0 \\ 0 & 0 & \frac{\tau_\phi^2}{2^\gamma} & 0 & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & & 0 \\ 0 & 0 & 0 & 0 & \frac{\tau_\phi^2}{k^\gamma} \end{pmatrix}$$

- User needs to set:
  - r, prior mean of ϕ_1 ($r = 0$ for GDP growth, $r = 0.8$ for interest rate)
  - τ_c, the prior standard deviation of the intercept (e.g. $\tau_c = 100$)
  - τ_ϕ, the prior standard deviation of ϕ_1 (e.g. $\tau_\phi = 1$)
  - γ, the lag decay (e.g. $\gamma = 1$). How fast the prior variance shrinks to zero for longer lags.
- **Stationarity** can be imposed by truncating the prior to the stationarity region (and use simulation).

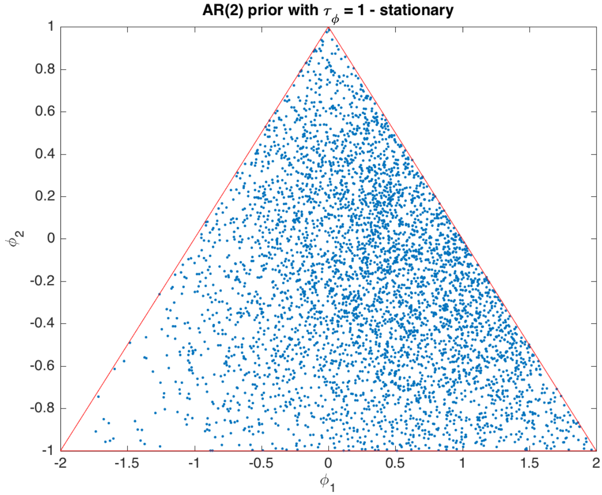
AR(2) JOINT PRIOR



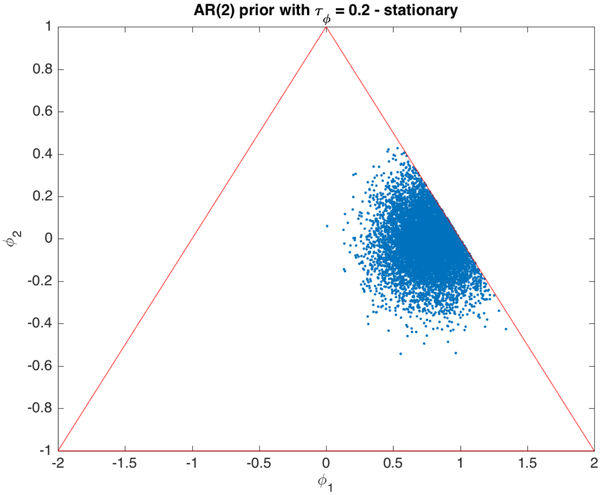
AR(2) JOINT PRIOR WITH $\tau_\phi = 0.5$



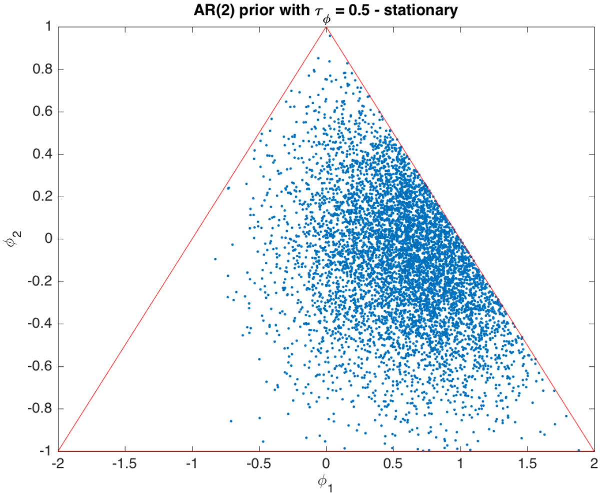
AR(2) JOINT PRIOR WITH $\tau_\phi = 1$



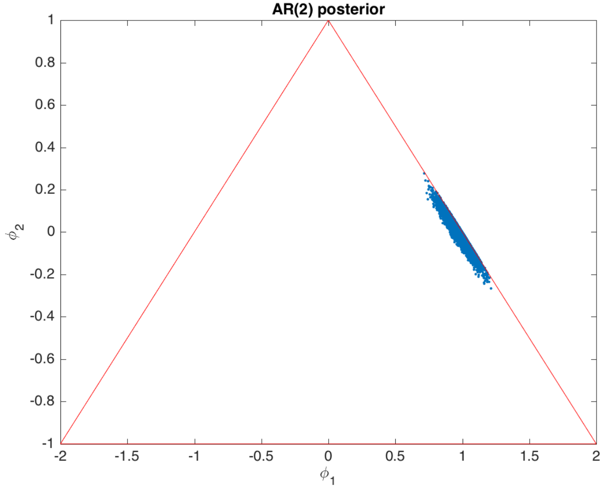
AR(2) JOINT PRIOR WITH $\tau_\phi = 0.2$



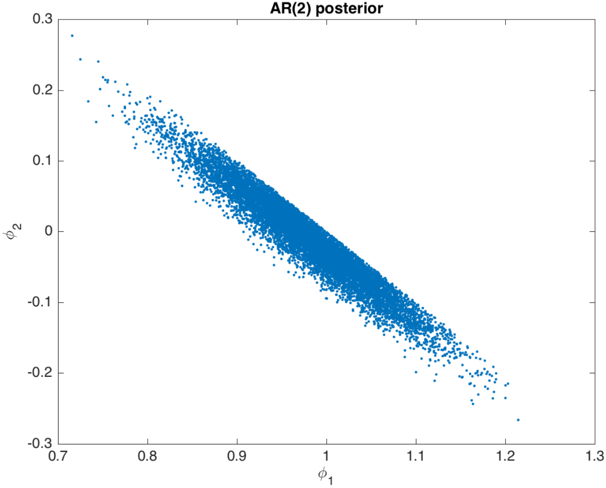
AR(2) JOINT PRIOR WITH $\tau_\phi = 0.5$



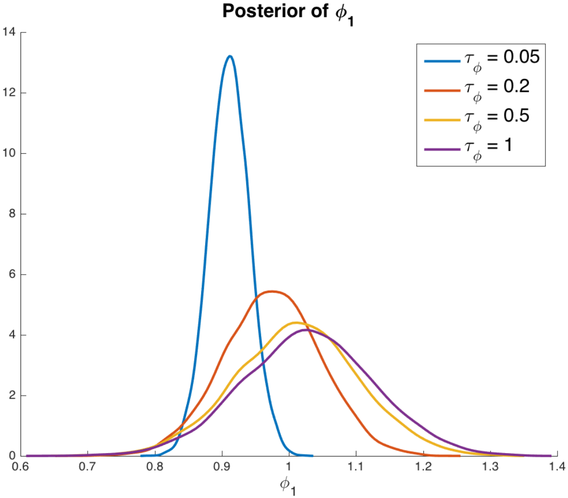
AR(2) POSTERIOR - STATIONARITY PRIOR ($\tau_\phi = 0.5$)



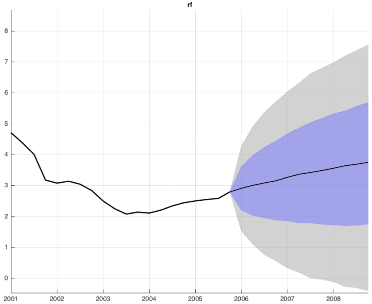
AR(2) JOINT POSTERIOR - ZOOMED

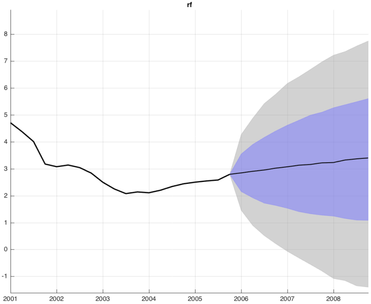


Univariate AR(4) posterior
Foreign interest rate 1980Q2-2005Q4



UNIVARIATE AR PREDICTIONS $\tau_\phi = 0.05$ AND $\tau_\phi = 0.2$





UNIVARIATE AR PREDICTIONS $\tau_\phi = 0.5$ AND $\tau_\phi = 1$