VU Minor Applied Econometrics

Bayesian Econometrics for Business & Economics (Bayesian statistics & simulation methods)

Period 2, 2025-2026 (E_MFAE_BEBE)

Lennart Hoogerheide

Vrije Universiteit Amsterdam

E-mail: l.f.hoogerheide@vu.nl

Assignments (for $(30 - \epsilon)\%$ of the grade)

There are two alternatives:

(I) Four assignments that involve programming the simulation methods (e.g. in Python/Matlab/R).

(II) Write an essay to summarize (and give feedback on – what could be done better/more?) four Bayesian articles.

About the assignments:

- The assignments should be made with 'groups' of 1/2/3 students.
- A .pdf file with answers must be submitted.
  For alternative (I) also the code must be submitted.
  (For example, the Python/Matlab/R files).
- Deadline for assignment 3: **Wednesday November 26, 23:59**.
- Deadline for assignment 4: **Wednesday December 3, 23:59**.

Alternative (II):

Assignment 1: summarize (and give feedback on – what could be done better/more?) the following article:

George Casella, Edward I. George (1992).
Explaining the Gibbs Sampler.
The American Statistician 46 (3), 167-174.

Assignment 2: summarize (and give feedback on – what could be done better/more?) the following article:

Lennart Hoogerheide, David Ardia, Nienke Corré (2012).
Stock Index Returns' Density Prediction using GARCH Models: Frequentist or Bayesian Estimation?
Economics Letters 116 (3), 322-325.

Assignment 3: summarize (and give feedback on – what could be done better/more?) the following article:

Lennart Hoogerheide, Joern Block, Roy Thurik (2012).
Family background variables as instruments for education in income regressions: A Bayesian analysis.
Economics of Education Review 31 (5), 515-523.

Assignment 4: summarize (and give feedback on – what could be done better/more?) the following article:

Robert E. Kass, Adrian E. Raftery (1995).
Bayes Factors.
Journal of the American Statistical Association 90 (430), 773-795.

Alternative (I): Assignment 1: Model with normal distribution and Gibbs sampling

A company that has multiple shops in different cities wants to analyze whether shop A and shop B have a different expected profit.
The company has computed the difference y_j (= profit of shop A - profit of shop B, in 10000 euros) for 10 periods:

1.6088
-1.4579
0.4502
1.0701
-0.3803
1.1201
0.4199
0.3998
-0.1904
0.3932

In this exercise we assume that $y_j \sim N(\mu, \frac{1}{h})$, i.i.d. where mean μ and precision h are unknown parameters.

(a) First we assume that we specify the (improper) non-informative prior

$$p(\mu, h) \propto \begin{cases} \frac{1}{h} & \text{if } h > 0 \\ 0 & \text{else} \end{cases}$$

Simulate from the posterior distribution of μ and h by applying the **Gibbs sampling method (from slide 52 of lecture 2)** with 11000 draws (with a *burn-in* of 1000 draws) and initial value $\mu = \bar{y}$.

Please first read the hints/warnings on slides 6-7 of this assignment!

Inspect the *trace plot* of the draws of μ (i.e., the graph with μ_i on the vertical axis and $i = 1, 2, \ldots, 11000$ on the horizontal axis): does the Gibbs sampling method keep moving through the parameter space or does it 'get stuck' in a certain point?

Use the resulting 10000 draws to estimate the posterior probabilities $\Pr(\mu < 0|y)$ and $\Pr(\mu > 0|y)$. Interpret the results.

Hints/warnings:

(1) There exist two versions of the Gamma distribution with parameters:

- *shape* a and *scale* b (or *shape* k and *scale* θ), used during this course.
- *shape* a and *rate* $1/b$ (or *shape* α and *rate* β)

For example, see also
`https://en.wikipedia.org/wiki/Gamma_distribution`.
Check which version of the Gamma distribution your computer package uses. If necessary, use arguments a and $\frac{1}{b}$.

(2) On the lecture slides the second parameter of the normal distribution $N(\mu, \sigma^2)$ is the **variance** σ^2, whereas computer packages often require the **standard deviation** $\sigma = \sqrt{\sigma^2}$.

Hint: So, the pseudocode is given by:

- Compute $\mu_0 = \bar{y}$, the sample mean of the $n = 10$ observations.
- Set $n_{draws} = 11000$ (and please notice the difference between $n_{draws} = 11000$, which is the number of iterations of the for-loop that we can choose ourselves, and $n = 10$, the number of available observations, where $n = 10$ appears in the conditional posterior distributions)
- Do for draw $i = 1, \ldots, n_{draws}$:
  - Simulate h_i from Gamma($a = n/2, b = (\frac{1}{2}\sum_{j=1}^{n}(y_j - \mu_{i-1})^2)^{-1}$) distribution.
  - Simulate μ_i from normal distribution with mean $\bar{y}$ and standard deviation $\frac{1}{\sqrt{h_i n}}$.
- Estimate $\Pr(\mu < 0|y)$ as the fraction of the 10000 draws μ_i $(i = 1001, 1002, \ldots, 11000)$ that are negative, and $\Pr(\mu > 0|y)$ as the fraction of the 10000 draws μ_i $(i = 1001, 1002, \ldots, 11000)$ that are positive.

(b) Now suppose we have prior

$$p(\mu, h) = p(\mu) \times p(h)$$

with a normal prior density for μ:

$$\mu \sim N(m_{prior}, v_{prior})$$

with $m_{prior} = 0, v_{prior} = 100^2$. That is, a prior with mean 0 and large variance.

And the same prior $p(h)$ for h:

$$p(h) \propto \begin{cases} \frac{1}{h} & \text{if } h > 0 \\ 0 & \text{else} \end{cases}$$

Simulate from the posterior distribution of μ and h by applying the **Gibbs sampling method (from slide 54 of lecture 2)** with 11000 draws (with a *burn-in* of 1000 draws) and initial value $\mu = \bar{y}$. See the hint on slide 10 of this assignment.

Use the resulting 10000 draws to estimate $\Pr(\mu < 0|y)$ and $\Pr(\mu > 0|y)$. Compare the results with part (a).

Hint: So, the pseudocode is given by:

- Compute $\mu_0 = \bar{y}$, the sample mean of the $n = 10$ observations.
- Set $n_{draws} = 11000$, $m_{prior} = 0$, $v_{prior} = 100^2 = 10000$
- Do for draw $i = 1, \ldots, n_{draws}$:
  - Simulate h_i from Gamma($a = n/2, b = (\frac{1}{2}\sum_{j=1}^{n}(y_j - \mu_{i-1})^2)^{-1}$) distribution.
  - Simulate μ_i from normal distribution with mean $\frac{\frac{m_{prior}}{v_{prior}} + h_i n\bar{y}}{\frac{1}{v_{prior}} + h_i n}$ and standard deviation $\frac{1}{\sqrt{\frac{1}{v_{prior}} + h_i n}}$.
- Estimate $\Pr(\mu < 0|y)$ as the fraction of the 10000 draws μ_i ($i = 1001, 1002, \ldots, 11000$) that are negative, and $\Pr(\mu > 0|y)$ as the fraction of the 10000 draws μ_i ($i = 1001, 1002, \ldots, 11000$) that are positive.

(c) Now suppose we have prior

$$p(\mu, h) = p(\mu) \times p(h)$$

with a normal prior density for μ:

$$\mu \sim N(m_{prior}, v_{prior})$$

with $m_{prior} = 0.5, v_{prior} = 0.25^2$. That is, a prior with mean 0.5 and small variance.
And the same prior $p(h)$ for h:

$$p(h) \propto \begin{cases} \frac{1}{h} & \text{if } h > 0 \\ 0 & \text{else} \end{cases}$$

Simulate from the posterior distribution of μ and h by applying the **Gibbs sampling method (from slide 54 of lecture 2)** with 11000 draws (with a *burn-in* of 1000 draws) and initial value $\mu = \bar{y}$.

Use the resulting 10000 draws to estimate $\Pr(\mu < 0|y)$ and $\Pr(\mu > 0|y)$. Compare the results with part (a) and (b).

(d) Perform a classical/frequentist two-sided test of

$H_0 : \mu = 0$ versus
$H_1 : \mu \neq 0$

at 5% significance, comparing the t-statistic $\frac{\bar{y}}{\sqrt{s^2/n}}$ with the 2.5% and 97.5% percentiles of the t_9 distribution (for which you can simply use the values -2.2622 and 2.2622). What is the conclusion?

Compare the conclusion with parts (a), (b) and (c).

Assignment 2: random walk Metropolis(-Hastings) method

(a) The file `Assignment2Dataset.csv` contains the daily returns of the S&P 500 in the six years 2018-2023. Show a graph of the daily returns y_t.

(b) We consider the ARCH(1) model (with variance targeting, so that α is the only unknown parameter). For α we specify the uniform prior on [0,1). In other words, we consider the ARCH(1) model and prior of lecture 3.

Use the **random walk Metropolis(-Hastings) method** (given on the next two slides) to estimate the posterior mean and posterior standard deviation of the parameter α. Use 1100 draws and a burn-in of 100 draws. Also show a histogram of the 1000 draws after the burn-in. Use the normal distribution $\tilde{\alpha} \sim N(\alpha_{i-1}, 0.0275^2)$ as the candidate distribution.

Also inspect the trace plot of the draws α and compute the acceptance percentage and the first order serial correlation in the sequence of accepted (and repeated) draws α. Interpret the results.

Hint: The following pseudocode can be used:

- Choose feasible initial value $\alpha_0 = 0.4745$ and set $n_{draws} = 1100$.
- Do for draw $i = 1, \dots, n_{draws}$:
  - Simulate candidate draw $\tilde{\alpha}$ from normal distribution with mean α_{i-1} and standard deviation 0.0275.
  - If $\tilde{\alpha} < 0$ or $\tilde{\alpha} \geq 1$, then set acceptance probability $a = 0$ (since the prior density $p(\tilde{\alpha})$ and posterior density $p(\tilde{\alpha}|y)$ are equal to 0 for $\tilde{\alpha} < 0$ and for $\tilde{\alpha} \geq 1$). Else compute acceptance probability

    $$a = \min \left\{ \exp \left[\ln p(y|\tilde{\alpha}) - \ln p(y|\alpha_{i-1}) \right], 1 \right\}$$

    with loglikelihood $\ln p(y|\alpha)$ given on the next slide.
  - Simulate U from uniform distribution on $[0, 1]$.
  - If $U \leq a$, then accept: $\alpha_i = \tilde{\alpha}$ (accept candidate draw).
    If $U > a$, then reject: $\alpha_i = \alpha_{i-1}$ (repeat previous draw).
- Estimate the posterior mean and posterior standard deviation of α as the sample mean and sample standard deviation of the 1000 draws α_i ($i = 101, 102, \dots, 1100$).

In our ARCH(1) model: loglikelihood[1]

$$\ln p(y|\alpha) = \sum_{t=2}^{n} \left\{ -\frac{1}{2} \ln(\boxed{2\pi} [s^2(1-\alpha) + \alpha y_{t-1}^2]) - \frac{y_t^2}{2[s^2(1-\alpha) + \alpha y_{t-1}^2]} \right\}$$

[1] The factor $\boxed{2\pi}$ is part of the constant scaling factor of the posterior density kernel and can be left out in this assignment.

(c) Repeat part (b) where the normal candidate distribution has the **very small** standard deviation 0.002 instead of 0.0275, and compare the results with part (b).

(d) Repeat part (b) where the normal candidate distribution has the **very large** standard deviation 6.0 instead of 0.0275, and compare the results with part (b).

(e) Estimate the GARCH(1,1) model of lecture 3 using the random walk Metropolis(-Hastings) method (again using 1100 draws with a burn-in of 100 draws) using a bivariate normal candidate distribution with mean $\theta_{i-1} = (\alpha_{i-1}, \beta_{i-1})'$ and variance-covariance matrix

$$\begin{pmatrix} 0.000430 & -0.000510 \\ -0.000510 & 0.000626 \end{pmatrix}$$

Compute the posterior mean and standard deviation of α and β.

Also inspect the trace plots of the (accepted and repeated) draws of α and β and compute the acceptance percentage and the two first order serial correlations in the sequences of (accepted and repeated) draws of α and β. Interpret the results.

Also make a scatter plot of (accepted and repeated) draws of α and β and a scatter plot of the *candidate* draws of α and β. Compare these two scatter plots.

(f) Repeat part (e) where the candidate distribution has variance-covariance matrix

$$\begin{pmatrix} 0.000430 & \mathbf{0} \\ \mathbf{0} & 0.000626 \end{pmatrix}$$

Compare the results with part (e).

Assignment 3: posterior model probabilities & importance sampling

This exercise involves the (approximate) replication and extension of some of the results of lecture 4 (for the computation of posterior model probabilities where marginal likelihoods are computed using **importance sampling**).

(a) Consider the dataset (of $n = 100$ observations) in the Excel file `Assignment3Dataset.xlsx`. These data have been simulated from a Student-t distribution with $DoF = 7$ degrees of freedom. Make a histogram of this dataset y_j $(j = 1, \ldots, 100)$.

Hint: this histogram should roughly look like the histogram on slide 13 of lecture 4.

If you work with Python, then please check that all 100 observations are loaded from the data file. It may be necessary to use the option `header=None` in the command

```
pd.read_excel("Assignment3Dataset.xlsx", header=None)
```

Otherwise Python could consider the first row (containing number 0.12219562) the header of the file (with a column label), so that Python would only consider the other 99 observations as data.

(b) Use importance sampling with $n_{draws} = 10000$ draws to compute the marginal likelihood in the model with the Student-t distribution (with mean = sample mean, variance = sample variance) with a uniform prior for DoF on [4.1, 50]. Use the prior as the importance density.

Hint: Here we first simulate $n_{draws} = 10000$ draws DoF_i $(i = 1, 2, \dots, n_{draws})$ from the uniform distribution on [4.1, 50], by making use of a built-in function or by simulating $n_{draws} = 10000$ draws U_i $(i = 1, 2, \dots, n_{draws})$ from a uniform distribution on [0,1] and computing $DoF_i = 4.1 + 45.9U_i$.
After that we evaluate the likelihood on slide 15 of lecture 4

$$p(y|DoF_i) = \left(\frac{\Gamma(\frac{DoF_i+1}{2})}{\Gamma(\frac{DoF_i}{2})\sqrt{(DoF_i - 2)\pi}} \right)^n \times \frac{1}{(s^2)^{n/2}} \times \prod_{j=1}^{n} \left(1 + \frac{(y_j - \bar{y})^2}{(DoF_i - 2)s^2} \right)^{-\frac{DoF_i+1}{2}}$$

for each draw DoF_i, where we have sample mean $\bar{y} = \frac{1}{100} \sum_{i=1}^{100} y_j$ and sample variance $s^2 = \frac{1}{99} \sum_{i=1}^{100} (y_j - \bar{y})^2$. Finally, we compute the marginal likelihood as the sample mean of the $n_{draws} = 10000$ likelihood values $p(y|DoF_i)$.

Also use importance sampling with $n_{draws} = 10000$ draws to compute the marginal likelihood in the model with the GED (with mean = sample mean, variance = sample variance) with a uniform prior for β on [0.38, 1.88]. Use the prior as the importance density.

Hint: Here we first need to simulate $n_{draws} = 10000$ draws β_i $(i = 1, 2, \ldots, n_{draws})$ from the uniform distribution on [0.38,1.88], by making use of a built-in function or by simulating $n_{draws} = 10000$ draws U_i $(i = 1, 2, \ldots, n_{draws})$ from a uniform distribution on [0,1] and computing $\beta_i = 0.38 + 1.5U_i$.

After that we evaluate the likelihood on slide 30 of lecture 4

$$p(y|\beta_i) = \left(\frac{\beta_i(\Gamma(3/\beta_i))^{1/2}}{2s(\Gamma(1/\beta_i))^{3/2}}\right)^n \times \exp\left(-\sum_{j=1}^{n}\left(\frac{|y_j - \bar{y}|}{s\sqrt{\frac{\Gamma(1/\beta_i)}{\Gamma(3/\beta_i)}}}\right)^{\beta_i}\right)$$

for each draw β_i. Finally, we compute the marginal likelihood as the sample mean of the $n_{draws} = 10000$ likelihood values $p(y|\beta_i)$.

Use the two marginal likelihoods to compute the Bayes factor, the posterior odds ratio and the posterior model probabilities of these two models with Student-t distribution and GED (if we only consider these two models for which we assume equal prior model probabilities).

(c) Now use importance sampling with $n_{draws} = 10000$ draws to compute the marginal likelihood in the model with the Student-t distribution (with mean = sample mean, variance = sample variance) with a uniform prior for DoF on [4.1, 1000]. Use the prior as the importance density.[2]

Use the two marginal likelihoods to compute the Bayes factor, the posterior odds ratio and the posterior model probabilities of these two models with Student-t distribution and GED (if we only consider these two models for which we assume equal prior model probabilities).

Explain the difference with part **(b)**.

[2] **Part (c) may yield numerical problems! See the hints on the next slides.**

Hint: In this part the likelihood

$$p(y|DoF_i) = \left(\frac{\Gamma(\frac{DoF_i+1}{2})}{\Gamma(\frac{DoF_i}{2})\sqrt{(DoF_i-2)\pi}} \right)^n \times \frac{1}{(s^2)^{n/2}} \times \prod_{j=1}^{n} \left(1 + \frac{(y_j - \bar{y})^2}{(DoF_i - 2)s^2} \right)^{-\frac{DoF_i+1}{2}}$$

needs to be evaluated for huge values of the parameter DoF_i (in the interval $[4.1, 1000]$). This yields numerical problems with the computation of $\Gamma(\frac{DoF_i+1}{2})$ and $\Gamma(\frac{DoF_i}{2})$. For example, $\Gamma(x)$ may be evaluated as $+$Infinity for $x > 171.63$, so that $\Gamma(\frac{DoF_i}{2})$ is evaluated as $+$Infinity for $DoF_i > 343.26$. This yields density values Infinity/Infinity $=$ NaN.

The solution is to evaluate

$$\frac{\Gamma(\frac{DoF_i+1}{2})}{\Gamma(\frac{DoF_i}{2})} \qquad \text{as} \qquad \exp\left(\log\Gamma\left(\frac{DoF_i+1}{2}\right) - \log\Gamma\left(\frac{DoF_i}{2}\right)\right)$$

where the $\log\Gamma()$ function is typically a built-in function. For example, `gammaln` in Matlab or `scipy.special.gammaln` after `import scipy` in Python.

Do **not** use `scipy.special.loggamma` in Python! That function returns undesired complex values.

Hint: So here we first simulate $n_{draws} = 10000$ draws DoF_i $(i = 1, 2, \ldots, n_{draws})$ from the uniform distribution on [4.1, 1000], by making use of a built-in function or by simulating $n_{draws} = 10000$ draws U_i $(i = 1, 2, \ldots, n_{draws})$ from a uniform distribution on [0,1] and computing $DoF_i = 4.1 + 995.9U_i$.

After that we evaluate the likelihood

$$p(y|DoF_i) = \left(\exp\left(\log\Gamma\left(\frac{DoF_i+1}{2}\right) - \log\Gamma\left(\frac{DoF_i}{2}\right)\right)\frac{1}{\sqrt{(DoF_i-2)\pi}}\right)^n \times$$
$$\times \frac{1}{(s^2)^{n/2}} \times \prod_{j=1}^{n}\left(1 + \frac{(y_j-\bar{y})^2}{(DoF_i-2)s^2}\right)^{-\frac{DoF_i+1}{2}}$$

for each draw DoF_i. We compute the marginal likelihood in the model with the Student-t distribution as the sample mean of the $n_{draws} = 10000$ likelihood values $p(y|DoF_i)$.

At part **(c)** the marginal likelihood in the model with the GED remains the same as at part **(b)**.

(d) How do your results at part **(b)** change if we assume prior model probabilities $\Pr(\text{Student-t}) = 0.1$ and $\Pr(\text{GED}) = 0.9$ with prior odds ratio $\frac{\Pr(\text{Student-t})}{\Pr(\text{GED})} = \frac{1}{9}$ (instead of prior model probabilities $\Pr(\text{Student-t}) = 0.5$ and $\Pr(\text{GED}) = 0.5$ with prior odds ratio $\frac{\Pr(\text{Student-t})}{\Pr(\text{GED})} = 1$)?

Explain the difference with part **(b)**.

Assignment 4: Forecasting US real GDP growth using Autoregressive (AR) models

Assignment 4 involves the (approximate) replication and extension of some of the results of lecture 5, involving the prediction of US real GDP growth using Autoregressive (AR) models.

(a) Consider the dataset of $n = 106$ observations[3] y_t of quarterly US real GDP growth (1990Q1 - 2016Q2) in the file `Assignment5Dataset.xlsx`. Note: this file contains columns with time, y_t, y_{t-1} and y_{t-2} in the first, second, third and fourth columns, respectively. That is, the 'pre-sample' values of y_0 (1989Q4) and y_{-1} (1989Q3) are also available, so that the AR(2) model can be estimated with 106 observations (without 'losing' the first two observations). Make a graph of y_t over time.

Hint: this graph should roughly look like the graph on slide 35 of lecture 5.

[3] or $T = 106$, using the notation T for the number of observations from part of the time series literature.

(b) Consider the AR(2) model. Estimate the model using Ordinary Least Squares (OLS). Compute the frequentist/classical prediction for y_{T+1} and 95% prediction interval for y_{T+1}.

Hint: Here we have the regression model $y = X\beta + \varepsilon$ with 106×1 vector y and 106×3 matrix X:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{106} \end{pmatrix} \quad X = \begin{pmatrix} 1 & y_0 & y_{-1} \\ 1 & y_1 & y_0 \\ \vdots & \vdots & \vdots \\ 1 & y_{105} & y_{104} \end{pmatrix}$$

We compute the OLS estimator $\hat{\beta}$ and s^2 (the estimator of the variance of the error terms ε_t) as:

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$s^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{T-k} = \frac{1}{T-k}\sum_{t=1}^{T}(y_t - \hat{\beta}_0 - \hat{\beta}_1 y_{t-1} - \hat{\beta}_2 y_{t-2})^2$$

with $T = 106$ and number of regressors $k = 3$ (including the constant term).

The prediction for y_{T+1} in the estimated AR(2) model is computed as $\hat{\beta}_0 + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1}$ and the 95% prediction interval for y_{T+1} in the estimated AR(2) model is computed as

$$[\hat{\beta}_0 + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1} - 1.96s,\ \ \hat{\beta}_0 + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1} + 1.96s].$$

The results should be as on slide 39 of lecture 5, with prediction 0.4361 and 95% prediction interval [-0.6484, 1.5205].

(c) Consider the AR(2) model under the non-informative prior $p(\beta, h) \propto \frac{1}{h}$ (for $h > 0$). Compute the Bayesian prediction for y_{T+1} and the 95% prediction interval for y_{T+1}, using 10000 draws.
Compare the results between parts **(b)** and **(c)**.

Hint: Here we use the approach on slide 41 of lecture 5, where we first simulate a set of 10000 draws $y_{T+1,i}$ $(i = 1, 2, \ldots, 10000)$. See the next two slides.

Do for $i = 1, 2, \ldots, 10000$:

- simulate 3×1 vector $\beta_i = (\beta_{0,i}, \beta_{1,i}, \beta_{2,i})'$ from marginal posterior distribution, multivariate Student-t($\hat{\beta}, s^2(X'X)^{-1}, T-k$), either using a built-in function or by simulating a scalar random variable V_i from a chi-squared($T-k$) distribution, the chi-squared distribution with $T - k = 106 - 3 = 103$ degrees of freedom, and a 3×1 random vector W_i from $N(0, s^2(X'X)^{-1})$, the multivariate normal distribution with mean vector 0 and variance-covariance matrix $s^2(X'X)^{-1}$, and computing

$$\beta_i = \hat{\beta} + \frac{1}{\sqrt{V_i/(T-k)}} \times W_i.$$

- simulate h_i from conditional posterior distribution upon β_i: Gamma($\frac{T}{2}, [\frac{1}{2}(y - X\beta_i)'(y - X\beta_i)]^{-1}$)

- compute $y_{T+1,i} = \beta_{0,i} + \beta_{1,i} y_T + \beta_{2,i} y_{T-1} + \varepsilon_{T+1,i}$ with $\varepsilon_{T+1,i}$ simulated from $N(0, \frac{1}{h_i})$, the normal distribution with mean 0 and standard deviation $\frac{1}{\sqrt{h_i}}$.

The prediction for y_{T+1} is the sample mean of the 10000 draws $y_{T+1,i}$.

The 95% prediction interval for y_{T+1} is obtained by computing the 2.5% and 97.5% quantiles of the 10000 draws $y_{T+1,i}$.

The prediction for y_{T+1} should be rather close to 0.4361. The 95% prediction interval for y_{T+1} should be rather close to [-0.6628, 1.5594], the interval on slide 43 of lecture 5.

(d) Compute the Bayesian prediction for y_{T+1} and 95% prediction interval for y_{T+1}, using **Bayesian Model Averaging (BMA)** with the AR(1) and AR(2) models. Use equal prior model probabilities. First, compute the Bayes factor as the **Savage-Dickey density ratio (SDDR)**, using the posterior after $T_{prior} = 10$ observations as "the prior".

Note: in the AR(1) model we obviously have a different number of regressors $k = 2$ (instead of $k = 3$ at the AR(2) model) and a different matrix X that has $k = 2$ columns (instead of $k = 3$ columns at the AR(2) model):

$$X = \begin{pmatrix} 1 & y_0 \\ 1 & y_1 \\ \vdots & \vdots \\ 1 & y_{T-1} \end{pmatrix}$$

Also see the hints on the next three slides!

Compare the results between parts **(c)** and **(d)**.

Hint: The "prior" density in the denominator of the Savage-Dickey density ratio (SDDR) is the Student-t posterior density of β_2 based on the first $T_{prior} = 10$ observations. For this you need the formula at the bottom line of slide 47 of lecture 5, where:

- $x = 0$;
- $m =$ OLS estimator of β_2 in AR(2) model estimated using the first $T_{prior} = 10$ observations (where X is a 10×3 matrix and y is a 10×1 vector);
- $c^2 =$ square of OLS standard error of the OLS estimator of β_2 in AR(2) model estimated using the first $T_{prior} = 10$ observations (that is, c^2 is the bottom right element of the 3×3 matrix $s^2(X'X)^{-1}$ where X is a 10×3 matrix);
- $DoF = T_{prior} - k = 10 - 3 = 7$;

where this "prior" density of β_2 at $\beta_2 = 0$ takes the value 0.8263 here.

The posterior density in the numerator of the Savage-Dickey density ratio (SDDR) is the Student-t posterior density of β_2 based on **all** 106 observations. For this you need the formula at the bottom line of slide 47 of lecture 5, where:

- $x = 0$;
- $m =$ OLS estimator of β_2 in AR(2) model estimated using all 106 observations (where X is a 106×3 matrix and y is a 106×1 vector);
- $c^2 =$ square of OLS standard error of the OLS estimator of β_2 in AR(2) model estimated using all 106 observations (that is, c^2 is the bottom right element of the 3×3 matrix $s^2(X'X)^{-1}$ where X is a 106×3 matrix);
- $DoF = T - k = 106 - 3 = 103$;

where this posterior density of β_2 at $\beta_2 = 0$ takes the value 0.5458 here.

Note: the posterior density kernel (of β and h in the AR(2) model) after 106 observations remains the same in parts (c) and (d)! Only instead of

"non-informative prior $\times$ likelihood for 106 observations"

we consider this as

"posterior density kernel after 10 observations under a non-informative prior $\times$ likelihood for last 96 observations",

which is **only** done for the computation of the **posterior model probabilities** of the AR(1) and AR(2) models.