Bayesian Econometrics (in Finance)

Richard Paap

Econometric Institute

Erasmus University Rotterdam

Master Quantitative Finance

Master Bussiness Analytics & Quantitative Marketing

Master Econometrics

September-October 2020

Lecture 1, September 3, 2020

Course Material

Exam material Bayesian Econonometrics:

- Slides provided during the course (important)
- Greenberg, 2013, Introduction to Bayesian Econometrics, Cambridge University Press, 2nd edition, ISBN 978-1-107-01531-9 (without Chapter 9, 12).
- Sample exercises

One is allowed to take the above mentioned book+slides (with extra notes on it) to the exam (open book exam). No other material is allowed, so no extra notes or sample exercises.

- **Try to use the discussion board for your questions regarding the course when possible**

Practical Issues

- Please register for the **correct** exam:
  - Master Econometrics and Master Business Analytics and Quantitative Marketing $\rightarrow$ FEM21026 Bayesian Econometrics
  - Master Quantitative Finance $\rightarrow$ FEM21032 Bayesian Econometrics in Finance
- For questions concerning the content of the course, use discussion board on CANVAS
- Study the slides in detail before the exam not during the exam.
- Refresh your knowledge of Calculus, Matrix Algebra, Probability Theory, Markov Processes and Econometrics. This really helps passing the exam.
- You do not have to study literature mentioned on the slides. These are key references which you can use during case studies or writing your thesis.

Assignment

- An individual (computer) assignment is part of the exam.
- The assignment will be made available around the 4th week.
- Deadline of handing in the assignment is the start of the regular exam
- The final grade of the course will be

$$0.1 \times \text{grade exercise} + 0.9 \times \text{grade exam}$$

- There is no resit for the assignment only for the exam.
- The result for the assignment also holds for the resit exam
- The result of the assignment expires after this academic year.
- Due to corona result of last year exercise remains valid for this year

Course Outline

1. Introduction Bayesian Econometrics (lectures $1-2$)
   - Part I of Greenberg (2013)
2. Markov Chain Monte Carlo Methods (lectures $3-4$)
   - Part II of Greenberg (2013)
3. Special Topics (lectures $5-7$)
   - QM+Ectrics: Topics from Part III relevant for Econometrics & Business Analytics
   - QF: Topics from Part III relevant for Quantitative Finance

- Lectures 1-4: joint lectures
- Lectures 5-7: different lectures for Bayesian Econometrics in Finance (by Martina Zaharieva) and Bayesian Econometrics (by Richard Paap)

Outline Lecture 1

1. Motivation Bayesian Approach
2. Who/What is Bayes?
3. Bayesian versus Frequentist Approach
4. How to Perform a Bayesian Analysis?
5. Bayesian Analysis of the Linear Regression Model
   - No Prior Information
   - Conjugate prior Information
6. Precision Specification

Bayes and Economics

- Many advanced econom(etr)ic macroeconomic models cannot be analyzed due to limited data information. Bayesian analysis allows one to include economic theory and expert's opinion in a statistical sound way.
- For many economic research projects the data set is limited and one cannot rely on asymptotic theory for inference. Bayesian inference is exact even in small samples.
- Developments of sophisticated (easy) simulation methods makes Bayesian analysis of advanced econometric models feasible and easier to implement.
- Increase in computing power makes Bayesian analysis operational.

Bayes and Big Data & Marketing

- Although big data sets contain many observations the information in the data is limited. Furthermore, when modelling big data the number of parameters may get out-of-hand. Bayesian techniques are especially useful to extract information from big data and keep the number of parameters limited.
- The hierarchical structure of many statistical models considered in marketing makes a Bayesian analysis computationally more attractive than a frequentist approach (maximum likelihood estimation).
- Bayesian analysis provides a statistical sound way to incorporate information not present in the data in the statistical inference of a model. Information of managers can be taken into account for decision problems at hand.

Bayes and Finance

- In Finance accurate estimates of uncertainty and risk are important. Bayesian analysis provides a natural way to deal with uncertainty including parameter uncertainty and even model uncertainty.
- In Finance one has to make many decisions with uncertainty involved. It can be shown that under certain conditions Bayesian decision theory (including a Bayesian analysis) is optimal for decision problems under uncertainty.
- Although it is often easy to obtain large datasets in finance, one is often interested in rare events. Frequentist analysis of rare events is difficult as asymptotic theory may not be applied in case of little observations

Why Bayes?

Arnold Zellner:

It pays to go Bayes!

Who is Bayes?

Reverend Thomas Bayes (1702-1761)

What is Bayes?

The Bayesian approach is a **different** way to look at statistical inference. The statistical approach that is common in econometrics is called the frequentist approach (classical approach). The Bayesian and classical approaches are totally different.

Let $\widehat{\theta}$ be an estimate of θ and $\widehat{s}_{\widehat{\theta}}$ its estimated standard error.

Statement: the probability that the true parameter value θ lies in the interval $(\widehat{\theta} - 1.96\widehat{s}_{\widehat{\theta}}, \widehat{\theta} + 1.96\widehat{s}_{\widehat{\theta}})$ is 95%.

$\Rightarrow$ This a **Bayesian** interpretation of a frequentist confidence interval!

The **correct** statement in a frequentist approach is:

There is a 95% probability that the interval $(\widehat{\theta} - 1.96\widehat{s}_{\widehat{\theta}}, \widehat{\theta} + 1.96\widehat{s}_{\widehat{\theta}})$ contains the true parameter value θ.

Frequentist Approach

Ingredients:

- model with parameter vector θ
- data y (one possible realization)

Classical/Frequentist approach:

- There is a true parameter θ_0 which generated the data.
- Estimators, test statistics, confidence intervals are functions of the data y. They are evaluated in terms of their properties in repeated samples.
- Given the model, hypothetical data $y(m)$, $m = 1, \ldots, M$ can be generated for which we could also compute estimators. This provides the sampling distributions of the estimators.
- Inference: Compare the realized estimator of the data with the sampling distributions of the hypothetical data.

Maximum Likelihood Estimation

The data and the model provide the likelihood function of the data y as a function of the model parameters θ

$$p(y|\theta).$$

Parameters (point) estimates are obtained by maximizing the likelihood function with respect to the parameters θ

$$\widehat{\theta}_{\mathsf{ML}} = \mathsf{argmax}_{\theta} \;\; p(y|\theta).$$

Loosely speaking: Choose that value of θ for which it is most likely that it generated the observed data.

The estimator $\widehat{\theta}_{\mathsf{ML}}$ is a function of the data y. Note this holds for any frequentist estimator (OLS, GLS, GMM, and so forth).

Frequentist Testing

The estimator $\widehat{\theta}_{\text{ML}}$ is a function of the data y. The uncertainty in the estimator results from the fact that the data is just a single realization from the model (Other sample of the data would give another $\widehat{\theta}_{\text{ML}}$).

The possible values of the estimator that data from the true model can generate provide the distribution of the estimator. This provides the uncertainty in the estimator.

If one wants to test for $\theta = \theta_0$, one compares the realized $\widehat{\theta}_{\text{ML}}$ from the observed data with the theoretical distribution of $\widehat{\theta}_{\text{ML}}$ given that a model with the parameter θ_0 generated the data.

The above arguments hold for any quantity that depends on data: frequentist estimators (ML, OLS, GLS, GMM, and so forth), test statistics (t, Wald, LM, LR, and so forth).

Bayesian Approach

Ingredients:

- model with parameter vector θ
- data y (one realization)

Bayesian approach:

- Parameters are treated as random variables. The (subjective) probability describes our state of knowledge about the true parameter θ_0.
- Prior belief about θ before observing the data (prior distribution).
- Use the data to update the prior belief about θ to obtain posterior belief (posterior distribution).
- Inference on θ is conditional on the single realization of the data. Inference is exact even in small samples (no asymptotic distribution theory).

Bayesian Analysis

The data and the model provide the likelihood function of the data y as a function of the model parameters θ

$$p(y|\theta).$$

The researcher specifies the prior distribution of parameter θ

$$p(\theta).$$

The posterior distribution of parameter θ follows from

$$\begin{aligned} p(\theta|y) &= \frac{p(\theta, y)}{p(y)} \\ &= \frac{p(\theta)p(y|\theta)}{p(y)} \quad \textit{Bayes rule} \\ &\propto p(\theta)p(y|\theta). \end{aligned}$$

Marginal likelihood (testing)

$$p(y) = \int p(y, \theta) d\theta = \int p(\theta)p(y|\theta) d\theta.$$

Kernel

A (posterior) density function $p(\theta|y)$ can always be written as

$$p(\theta|y) = ck(\theta|y),$$

where c is a constant which does not depend on θ. As $p(\theta|y)$ is a proper density, it holds that

$$\int k(\theta|y)d\theta = 1/c.$$

The function $k(\theta|y)$ is called the kernel function belonging to the density $p(\theta|y)$. It uniquely identifies the distribution of θ. We write

$$p(\theta|y) \propto k(\theta|y),$$

where $\propto$ means proportional to. As it is always possible to compute c using the integral above.

If one needs to evaluate the posterior density function in a particular value of θ one needs the value of c. For deriving the posterior distribution the value of c is not always necessary. When in the end of the derivation one can recognize the distribution from the kernel, c follows immediately from the definition of the density function.

Example: Linear Regression Model

Consider the linear regression model

$$y = X\beta + \varepsilon,$$

where $y = (y_1, \ldots, y_N)'$, $X = (x_1, \ldots, x_N)'$, x_i is a k-dimensional vector of explanatory variables, β is a k-dimensional parameter vector, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)' \sim N(0, \sigma^2 I_N)$.

The likelihood function is given by

$$p(y|\beta, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right).$$

Assume independent flat priors for β and $\ln \sigma^2$

$$p(\beta) \propto 1 \text{ and } p(\sigma^2) \propto \sigma^{-2}.$$

No prior beliefs about β and $\ln \sigma^2$ ($p(\ln \sigma^2) \propto 1$ implies $p(\sigma^2) \propto \sigma^{-2}$ (transformation of random variables)) to stay close to the frequentist approach (priors are no proper densities and called improper.)

Multivariate Normal distribution

The pdf of a multivariate normal distributed random variable Z with k-dimensional location parameter vector μ and $k \times k$ positive definite symmetric covariance matrix Σ, that is, $Z \sim N(\mu, \Sigma)$ or $Z \sim MVN(\mu, \Sigma)$ is given by

$$p(z|\mu, \Sigma) = \left(\frac{1}{\sqrt{2\pi}}\right)^k |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}(z-\mu)'\Sigma^{-1}(z-\mu)).$$

The mean and variance of Z are given by

$$\begin{aligned} \mathrm{E}[Z] &= \mu \\ \mathrm{Var}[Z] &= \Sigma. \end{aligned}$$

If $\Sigma = \sigma^2 I_k$ we can write

$$p(z|\mu, \Sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^k \exp(-\frac{1}{2\sigma^2}(z-\mu)'(z-\mu))$$

as $|\sigma^2 I_k|^{-\frac{1}{2}} = ((\sigma^2)^k |I_k|)^{-\frac{1}{2}} = \sigma^{-k}$.

Posterior Distribution

The posterior kernel is given by

$$p(\beta, \sigma^2|y) \propto p(\beta, \sigma^2)p(y|\beta, \sigma^2) \propto \left(\frac{1}{\sigma}\right)^{N+2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right).$$

This kernel represents the joint posterior density of the model parameters

To learn about the posterior distribution, one may want to compute:

- Marginal posterior distributions/densities of β and σ^2

$$p(\beta|y) \text{ and } p(\sigma^2|y)$$

- Mean/mode/median/variance of posterior distribution
  - Posterior mean of β and σ^2

$$\text{E}[\beta|y] \text{ and } \text{E}[\sigma^2|y]$$

Useful Relations

Joint posterior density

$$p(\beta, \sigma^2|y)$$

Marginal posterior densities

$$p(\beta|y) = \int_0^\infty p(\beta, \sigma^2|y) d\sigma^2 \text{ and } p(\sigma^2|y) = \int_{-\infty}^\infty p(\beta, \sigma^2|y) d\beta.$$

Decomposition

$$p(\beta, \sigma^2|y) = p(\beta|y)p(\sigma^2|\beta, y) \text{ and } p(\beta, \sigma^2|y) = p(\sigma^2|y)p(\beta|\sigma^2, y)$$

Posterior Expectations

$$\mathsf{E}[\beta|y] = \int_{-\infty}^\infty \beta p(\beta|y) d\beta = \int_{-\infty}^\infty \int_0^\infty \beta p(\beta, \sigma^2|y) d\sigma^2 d\beta$$
$$\mathsf{E}[\sigma^2|y] = \int_0^\infty \sigma^2 p(\sigma^2|y) d\sigma^2 = \int_{-\infty}^\infty \int_0^\infty \sigma^2 p(\beta, \sigma^2|y) d\sigma^2 d\beta$$

Useful Tools

To derive marginal posterior results in univariate linear regression models, two useful analytical results are applied:

Decomposition rule:

$$(y - X\beta)'(y - X\beta) = (y - X\widehat{\beta})'(y - X\widehat{\beta}) + (\beta - \widehat{\beta})'X'X(\beta - \widehat{\beta}),$$

where $\widehat{\beta} = (X'X)^{-1}X'y$ (OLS estimator).

Gamma-2 inverse step:

$$\int_0^\infty \sigma^{-(M+2)} \exp\left(-\frac{a}{2\sigma^2}\right) d\sigma^2 \propto a^{-\frac{1}{2}M}.$$

for integer value M.

Marginal Posterior Distribution of $\boldsymbol{\beta}$

The marginal posterior density of β is given by

$$
\begin{aligned}
p(\beta|y) &= \int_0^\infty p(\beta, \sigma^2|y)d\sigma^2 \\
&\propto \int_0^\infty \left(\frac{1}{\sigma}\right)^{N+2} \exp\left(\frac{-1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right) d\sigma^2 \\
&\propto [(y - X\beta)'(y - X\beta)]^{-N/2} \\
&\propto [(y - X\widehat{\beta})'(y - X\widehat{\beta}) + (\beta - \widehat{\beta})'X'X(\beta - \widehat{\beta})]^{-N/2},
\end{aligned}
$$

where we use the Gamma-2 inverse step followed by the decomposition rule. When we divide by $\widehat{\sigma}^2 = (y - X\widehat{\beta})'(y - X\widehat{\beta})/(N - k)$ we obtain

$$
p(\beta|y) \propto \left((N - k) + \frac{(\beta - \widehat{\beta})'X'X(\beta - \widehat{\beta})}{(y - X\widehat{\beta})'(y - X\widehat{\beta})/(N - k)}\right)^{-N/2}
$$

$\rightarrow$ the marginal posterior distribution of β is a multivariate t-distribution with location parameter $\widehat{\beta}$, scale parameter matrix $\widehat{\sigma}^2(X'X)^{-1}$ and $(N - k)$ degrees of freedom with $k = \dim(\beta)$.

Multivariate t-distribution

The pdf of a t-distributed random variable Z with k-dimensional location parameter vector μ, $k \times k$ positive definite symmetric scale matrix S, and $\nu > 0$ degrees of freedom, that is, $Z \sim t(\mu, S, \nu)$ is given by

$$p(z|\mu, S, \nu) = c^{-1}|S|^{-1/2}[\nu + (z-\mu)' S^{-1}(z-\mu)]^{-\frac{\nu+k}{2}},$$

where

$$c = \frac{\pi^{k/2}\Gamma(\nu/2)}{\nu^{\nu/2}\Gamma((\nu+k)/2)}.$$

The mean and variance of Z are given by

$$\begin{aligned} \mathrm{E}[Z] &= \mu && \text{for } \nu > 1 \\ \mathrm{Var}[Z] &= \tfrac{\nu}{\nu-2} S && \text{for } \nu > 2. \end{aligned}$$

Difference Frequentist/Bayes Result

Bayesian analysis

Under the diffuse prior specification $p(\beta, \sigma^2) \propto \sigma^{-2}$ the marginal posterior distribution of β is a multivariate t distribution with mean the OLS **estimate** $\widehat{\beta}$, "covariance matrix" $\widehat{\sigma}^2(X'X)^{-1}$ and $N-k$ degrees of freedom, that is,

$$\beta|y \sim t(\widehat{\beta}, \widehat{\sigma}^2(X'X)^{-1}, N-k).$$

Frequentist analysis

The OLS **estimator** $\widehat{\beta}$ is multivariate t distributed with mean β, "covariance matrix" $\widehat{\sigma}^2(X'X)^{-1}$ and $N-k$ degrees of freedom, that is,

$$\widehat{\beta} \sim t(\beta, \widehat{\sigma}^2(X'X)^{-1}, N-k).$$

Marginal Posterior Distribution of σ^2

The marginal posterior density of σ^2 is given by

$$p(\sigma^2|y) = \int_{-\infty}^{\infty} p(\beta, \sigma^2|y) d\beta \propto \int_{-\infty}^{\infty} \left(\frac{1}{\sigma}\right)^{N+2} \exp\left(\frac{-(y - X\beta)'(y - X\beta)}{2\sigma^2}\right) d\beta$$

$$\propto \sigma^{-(N+2)} \exp\left(\frac{-(y - X\widehat{\beta})'(y - X\widehat{\beta})}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left(\frac{-(\beta - \widehat{\beta})'X'X(\beta - \widehat{\beta})}{2\sigma^2}\right) d\beta$$

$$\propto \sigma^{-(N+2)} \exp\left(\frac{-(y - X\widehat{\beta})'(y - X\widehat{\beta})}{2\sigma^2}\right) |\sigma^2(X'X)^{-1}|^{1/2}$$

$$\int_{-\infty}^{\infty} |\sigma^2(X'X)^{-1}|^{-1/2} \exp\left(\frac{-(\beta - \widehat{\beta})'X'X(\beta - \widehat{\beta})}{2\sigma^2}\right) d\beta$$

$$\propto \sigma^{-(N+2-k)} \exp\left(-\frac{1}{2\sigma^2}(y - X\widehat{\beta})'(y - X\widehat{\beta})\right),$$

using the decomposition rule, the fact that the integral is proportional to a constant and that $|\sigma^2(X'X)^{-1}| \propto \sigma^{2k}$. This is a kernel of an inverted Gamma-2 distribution with parameter $(y - X\widehat{\beta})'(y - X\widehat{\beta})$ and $N - k$ degrees of freedom which implies that $\frac{(y-X\widehat{\beta})'(y-X\widehat{\beta})}{\sigma^2}$ is $\chi^2(N - k)$-distributed.

Inverted Gamma-2 Distribution

The pdf of an inverted Gamma-2 distributed random variable Z with parameter $\mu > 0$ and degrees of freedom $\nu > 0$, that is, $Z \sim IG2(\mu, \nu)$ is given by

$$p(z|\mu,\nu) = \frac{1}{\Gamma(\nu/2)} \left(\frac{\mu}{2}\right)^{\nu/2} z^{-\frac{\nu+2}{2}} \exp\left(-\frac{\mu}{2z}\right),$$

The mean and variance of Z are given by

$$\mathrm{E}[Z] = \tfrac{\mu}{\nu-2} \qquad \text{for } \nu > 2$$

$$\mathrm{Var}[Z] = \tfrac{2\mu^2}{(\nu-4)(\nu-2)^2} \qquad \text{for } \nu > 4.$$

It holds that $\mu/Z \sim \chi^2(\nu)$ & $Z^{-1} \sim G2(\mu, \nu)$ (Gamma 2 distribution).

The inverted Gamma-2 distribution is just another parametrization of the inverted Gamma distribution. It holds that $Z \sim IG2(\mu, \nu)$ is equal to $Z \sim IG(\nu/2, \mu/2)$, see for the specification of the inverted Gamma distribution Greenberg (2013).

Including Prior Information

Instead of using diffuse priors as in the previous example, one may incorporate informative prior information in a Bayesian analysis.

It is possible to take any distribution as prior for the model parameters. In practice it is often easier to specify a **conjugate** prior, that is, a prior distribution that leads to a posterior distribution of the same type or the same family.

The advantage of a conjugate prior is that there is an analytical solution for the posterior distribution. Furthermore, it is also easier to see how the data update the prior information into posterior information.

In linear regression models the conjugate prior is a normal distribution for β given σ^2 and an inverted Gamma-2 distribution for σ^2.

Note: **Calibration** can be seen as a prior specification where all probability mass is fixed at one point (no uncertainty).

Example: Conjugate Prior for β given σ^2

Consider again the linear regression model

$$y = X\beta + \varepsilon,$$

where β is a k-dimensional parameter vector and $\varepsilon \sim N(0, \sigma^2 I_N)$. The prior specification for β and σ^2 is usually split up in two parts

$$p(\beta, \sigma^2) = p(\beta|\sigma^2)p(\sigma^2).$$

We take the conditional conjugate prior for β given σ^2 which is a normal distribution $\beta|\sigma^2 \sim N(b, \sigma^2 B)$ or

$$p(\beta|\sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^k |\sigma^2 B|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(\beta - b)' B^{-1} (\beta - b)\right),$$

where b and B are prior parameters. For σ^2 we assume a diffuse prior

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

but if one has prior information on σ^2, one can also take the conjugate prior which is an inverted Gamma-2 prior.

Posterior Distribution

The posterior density follows from

$$p(\beta,\sigma^2|y) \propto p(\beta|\sigma^2)p(\sigma^2)p(y|\beta,\sigma^2)$$
$$\propto \sigma^{-k} \exp\left(\frac{-(\beta-b)'B^{-1}(\beta-b)}{2\sigma^2}\right)\left(\frac{1}{\sigma}\right)^{N+2}\exp\left(\frac{-(y-X\beta)'(y-X\beta)}{2\sigma^2}\right).$$

To derive marginal results we write $B^{-1} = B^{-\frac{1}{2}'}B^{-\frac{1}{2}}$ (Choleski decomposition) and hence

$$(\beta-b)'B^{-1}(\beta-b) = (B^{-\frac{1}{2}}b - B^{-\frac{1}{2}}\beta)'(B^{-\frac{1}{2}}b - B^{-\frac{1}{2}}\beta).$$

We can write

$$(y-X\beta)'(y-X\beta)+(B^{-\frac{1}{2}}b-B^{-\frac{1}{2}}\beta)'(B^{-\frac{1}{2}}b-B^{-\frac{1}{2}}\beta) = (w-V\beta)'(w-V\beta),$$

where

$$w = \begin{pmatrix} y \\ B^{-\frac{1}{2}}b \end{pmatrix} \text{ and } V = \begin{pmatrix} X \\ B^{-\frac{1}{2}} \end{pmatrix}.$$

Marginal Posterior Distribution of $\boldsymbol{\beta}$

The marginal posterior density of β is given by

$$
\begin{aligned}
p(\beta|y) &= \int_0^\infty p(\beta,\sigma^2|y)d\sigma^2 \\
&\propto \int_0^\infty \left(\frac{1}{\sigma}\right)^{N+k+2} \exp\left(\frac{-1}{2\sigma^2}(w-V\beta)'(w-V\beta)\right) d\sigma^2 \\
&\propto [(w-V\beta)'(w-V\beta)]^{-(N+k)/2} \\
&\propto [(w-V\tilde{\beta})'(w-V\tilde{\beta}) + (\beta-\tilde{\beta})'V'V(\beta-\tilde{\beta})]^{-(N+k)/2},
\end{aligned}
$$

where we use the Gamma-2 inverse step followed by the decomposition rule and where $\tilde{\beta} = (V'V)^{-1}V'w = (X'X+B^{-1})^{-1}(X'y+B^{-1}b)$. Dividing by $\tilde{\sigma}^2 = (w-V\tilde{\beta})'(w-V\tilde{\beta})/N$ results in

$$
p(\beta|y) \propto \left(N + \frac{(\beta-\tilde{\beta})'(X'X+B^{-1})(\beta-\tilde{\beta})}{(w-V\tilde{\beta})'(w-V\tilde{\beta})/N}\right)^{-(N+k)/2}
$$

$\rightarrow$ this a multivariate t-distribution with location parameter $\tilde{\beta}$, scale parameter $\tilde{\sigma}^2(V'V)^{-1} = \tilde{\sigma}^2(X'X+B^{-1})^{-1}$ and N degrees of freedom with $k = \dim(\beta)$.

Prior Influence

The posterior mean of β is $\tilde{\beta} = (X'X + B^{-1})^{-1}(X'y + B^{-1}b)$ and the posterior scale parameter is $\tilde{\sigma}^2(X'X + B^{-1})^{-1}$ with

$$\tilde{\sigma}^2 = \frac{1}{N}\Big((y - X\tilde{\beta})'(y - X\tilde{\beta}) + (b - \tilde{\beta})'B^{-1}(b - \tilde{\beta})\Big).$$

- If the prior variance B is small, the influence of the prior on the posterior mean and posterior variance is high.
- If the prior variance B is large, the influence of the prior on the posterior mean and posterior variance is small.
- If the number of observations N is large, the influence of the prior becomes less as $X'X = \sum_{i=1}^{N} x_i x_i'$ and $X'y = \sum_{i=1}^{N} x_i y_i'$ become large $\Rightarrow$ Prior does not matter for very large sample sizes.

Marginal Posterior Distribution of σ^2

The marginal posterior density of σ^2 is given by

$$
\begin{aligned}
p(\sigma^2|y) = \int_{-\infty}^{\infty} p(\beta,\sigma^2|y)d\beta &\propto \int_{-\infty}^{\infty}\left(\frac{1}{\sigma}\right)^{N+k+2}\exp\left(\frac{-(w-V\beta)'(w-V\beta)}{2\sigma^2}\right)d\beta \\
&\propto \sigma^{-(N+k+2)}\exp\left(\frac{-(w-V\tilde{\beta})'(w-V\tilde{\beta})}{2\sigma^2}\right)\int\exp\left(\frac{-(\beta-\tilde{\beta})'V'V(\beta-\tilde{\beta})}{2\sigma^2}\right)d\beta, \\
&\propto \sigma^{-(N+k+2)}\exp\left(\frac{-(w-V\tilde{\beta})'(w-V\tilde{\beta})}{2\sigma^2}\right)|\sigma^2(V'V)^{-1}|^{1/2} \\
&\qquad \int_{-\infty}^{\infty}|\sigma^2(V'V)^{-1}|^{-1/2}\exp\left(\frac{-(\beta-\tilde{\beta})'V'V(\beta-\tilde{\beta})}{2\sigma^2}\right)d\beta \\
&\propto \sigma^{-(N+2)}\exp\left(-\frac{1}{2\sigma^2}(w-V\tilde{\beta})'(w-V\tilde{\beta})\right),
\end{aligned}
$$

using the decomposition rule and that the integral is proportional to a constant. This is a kernel of an inverted Gamma-2 distribution with parameter $(w-V\tilde{\beta})'(w-V\tilde{\beta})$ and N degrees of freedom.

$\rightarrow \frac{(w-V\tilde{\beta})'(w-V\tilde{\beta})}{\sigma^2}$ is $\chi^2(N)$ distributed.

Precision instead of variance

In some recent papers, books, and computer programs, one defines the variance of the error term of the linear regression model in terms of $h = 1/\sigma^2$ (precision). This implies that $\sigma^2 = 1/h$.

The Jacobian of this transformation is

$$|J| = \left|\frac{\partial \sigma^2}{\partial h}\right| = \left|-\frac{1}{h^2}\right| = \frac{1}{h^2}.$$

It is easy to transform our results into the precision specification. The improper prior $p(\beta, \sigma^2) \propto \sigma^{-2}$ becomes

$$p(\beta, h) = p(\beta, \sigma^2)|_{\sigma^2 = 1/h} \times |J| = h \times |J| \propto h/h^2 = 1/h.$$

The marginal posterior of β is the same as it does not contain σ^2. The marginal posterior of h is given by

$$p(h|y) = p(\sigma^2|y)|_{\sigma^2 = 1/h} \times |J|.$$

Transformation of Random Variables

Suppose that θ_1 is a continuous random variable with density function $p_1(\theta_1)$ and assume that $\theta_2 = g(\theta_1)$, where g defines a one-to-one transformation.

The inverse relation is given by $\theta_1 = g^{-1}(\theta_2)$.

The density of θ_2 denoted by $p_2(\theta_2)$ is given by

$$p_2(\theta_2) = p_1(\theta_1))\big|_{\theta_1 = g^{-1}(\theta_2)} \times \left|\frac{\partial g^{-1}(\theta_2)}{\partial \theta_2}\right|.$$

The latter term is called the Jacobian of the transformation.

Marginal Posterior in terms of Precision

Consider again the linear regression model

$$y = X\beta + \varepsilon \text{ with } \varepsilon \sim N(0, 1/hI_N)$$

Under the prior specification $p(\beta, h) \propto h^{-1}$ the marginal posterior kernel of h is given by

$$p(h|y) = p(\sigma^2|y)|_{\sigma^2=1/h}\frac{1}{h^2} \propto h^{(N-k-2)/2}\exp\left(-\frac{1}{2}h(y - X\widehat{\beta})'(y - X\widehat{\beta})\right),$$

which is a Gamma-2 distribution with $N - k$ degrees of freedom and scale parameter $(y - X\widehat{\beta})'(y - X\widehat{\beta})$.

Under the prior specification $\beta|h \sim N(b, 1/hB)$ and $p(h) \propto h^{-1}$ the marginal posterior kernel of h is given by

$$p(h|y) = p(\sigma^2|y)|_{\sigma^2=1/h}\frac{1}{h^2} \propto h^{(N-2)/2}\exp\left(-\frac{1}{2}h(w - V\tilde{\beta})'(w - V\tilde{\beta})\right),$$

This is a kernel of a Gamma-2 distribution with N degrees of freedom and parameter $(w - V\tilde{\beta})'(w - V\tilde{\beta})$.

Gamma-2 Distribution

The pdf of a Gamma-2 distributed random variable Z with parameter $\mu > 0$ and degrees of freedom $\nu > 0$, that is, $Z \sim G2(\mu, \nu)$ is given by

$$p(z|\mu,\nu) = \frac{1}{\Gamma(\nu/2)} \left(\frac{\mu}{2}\right)^{\nu/2} z^{\frac{\nu-2}{2}} \exp\left(-\frac{z\mu}{2}\right),$$

The mean and variance of Z are given by

$$\mathsf{E}[Z] = \frac{\nu}{\mu}$$
$$\mathsf{Var}[Z] = \frac{2\nu}{\mu^2}.$$

It holds that $\mu Z \sim \chi^2(\nu)$ and $Z \sim G(\nu/2, \mu/2)$ (Gamma distribution as defined in Greenberg (2013)).

Note: There are also other parametrizations of the Gamma distribution. For example, Koop (2003) replaces μ by ν/μ in his definition of the Gamma distribution.

Some Other Bayesian Econometric Textbooks

- Bauwens, Lubrano & Richard, 1999, *Bayesian Inference in Dynamic Econometric Models*, Oxford: Oxford University Press (time series).
- Geweke, 2005, *Contemporary Bayesian Econometrics and Statistics*, New York: Wiley
- Koop, 2003, *Bayesian Econometrics*, Chichester: Wiley.
- Lancaster, 2004, *An Introduction to Modern Bayesian Econometrics*, Malden: Blackwell.
- Rossi, Allenby & McCulloch, 2005, Bayesian Statistics and Marketing, New York: Wiley.
- Zellner, 1971, *An Introduction to Bayesian Inference in Econometrics*, New York: Wiley (decent but not up-to-date).

Exercises

- Exam April 2006
  - 2
- Exam July 2012
  - 2
- Greenberg (2013)
  - 2.3, 2.4, 2.5
- Extra Exercises
  - 18
- Answers available on CANVAS