Bayesian Tools for Econometric Analysis

European Central Bank, DG-Research Summer 2009

Part 3: Bayesian treatment of the univariate linear regression model (ULRM)

Gianni Amisano (ECB DG-Research and University of Brescia)

gianni.amisano@ecb.europa.int, amisano@eco.unibs.it

www.eco.unibs.it/~amisano/BT4EA_2009/ECB_Bayesian_course_Amisano.htm

Frankfurt, 2009/06/19

Contents

1 The conjugate prior analysis

1.1 The known variance case

Suppose we have T observations on the k covariate regression model with NID Gaussian errors

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

$$\varepsilon \backsim N\left(\mathbf{0}, \sigma^2 \mathbf{I}_T\right) \tag{2}$$

with $h = \sigma^{-2}$,error precision, known and $\mathbf{X}$ exogenous.

Let us focus on the Bayesian estimation of β.

Non-Bayesians:

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3}$$

$$V(\widehat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{4}$$

Likelihood:

$$p(\mathbf{y}|\mathbf{X}, \beta, \sigma^2) = \left(\sqrt{2\pi}\right)^{-T}\left(\sigma^2\right)^{-T/2}\exp\left\{-\frac{1}{2\sigma^2}\varepsilon'\varepsilon\right\} \tag{5}$$

$$\varepsilon = \mathbf{y} - \mathbf{X}\beta \tag{6}$$

Prior for β :

$$p(\beta) = N\left(\underset{(k\times 1)}{\underline{\beta}}, \underset{(k\times k)}{\underline{\mathbf{H}}^{-1}}\right) = (2\pi)^{-k/2}|\underline{\mathbf{H}}|^{1/2}\times$$
$$\times \exp\left\{-\frac{1}{2}\left(\beta-\underline{\beta}\right)'\underline{\mathbf{H}}_{\beta}\left(\beta-\underline{\beta}\right)\right\}$$

- absence of a priori info: $\underline{\underline{\mathbf{H}}} = [\mathbf{0}]$

- dogmatic prior: $\underline{\underline{\mathbf{H}}}^{-1} = [\mathbf{0}]$

⇒Bayes Theorem:

$$
\begin{aligned}
p(\beta|\mathbf{y},\mathbf{X},\sigma^2) &\propto \exp\left\{-\frac{1}{2}\left(\beta-\underline{\beta}\right)'\underline{\mathbf{H}}_\beta\left(\beta-\underline{\beta}\right)\right\} \\
&\times \exp\left\{-\frac{1}{2\sigma^2}\varepsilon'\varepsilon\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\beta-\underline{\beta}\right)'\underline{\mathbf{H}}_\beta\left(\beta-\underline{\beta}\right)+\frac{1}{\sigma^2}\varepsilon'\varepsilon\right]\right\} \qquad (7)
\end{aligned}
$$

Remember OLS decomposition

$$
\begin{aligned}
\varepsilon' \varepsilon = & (\beta - \widehat{\beta})' \mathbf{X}' \mathbf{X} (\beta - \widehat{\beta}) + \widehat{\varepsilon}' \widehat{\varepsilon} && (8) \\
\Longrightarrow \exp\left(-\frac{1}{2\sigma^2} \varepsilon' \varepsilon\right) = & \exp\left(-\frac{1}{2\sigma^2} \widehat{\varepsilon}' \widehat{\varepsilon}\right) \times \\
& \exp\left(-\frac{1}{2\sigma^2} (\beta - \widehat{\beta})' \mathbf{X}' \mathbf{X} (\beta - \widehat{\beta})\right) && (9)
\end{aligned}
$$

Clearly Gaussian for β, with moments which can be obtained by writing exponential part as a sum of squares:

$$\begin{aligned} \text{sample info:} \quad \underset{(T\times 1)}{\mathbf{y}} &= \mathbf{X}\beta + \varepsilon, \varepsilon \backsim N(\mathbf{0},\sigma^2\mathbf{I}_T) \\ \text{a priori info:} \quad \underset{(k\times 1)}{\underline{\beta}} &= \beta + \varepsilon_0, \varepsilon_0 \backsim N(\mathbf{0}, \underline{\mathbf{H}}^{-1}) \end{aligned} \tag{10}$$

Standardising:

$$\text{sample: } \frac{1}{\sigma} \underset{(T\times 1)}{\mathbf{y}} = \frac{1}{\sigma}\mathbf{X}\beta + \mathbf{e}, \mathbf{e} = \frac{1}{\sigma}\varepsilon \backsim \mathbf{N}(\mathbf{0}, \mathbf{I}_T)$$

$$\text{a priori : } \mathbf{P}^{-1} \underset{(k\times 1)}{\underline{\beta}} = \mathbf{P}^{-1}\beta + \mathbf{e}_0, \mathbf{e}_0 = \mathbf{P}^{-1}\varepsilon_0, \tag{11}$$

$$\mathbf{e}_0 \backsim N(\mathbf{0}, \mathbf{I}_k) \tag{12}$$

$$\underline{\underline{\mathbf{H}}}^{-1} = \mathbf{P}\mathbf{P}' \tag{13}$$

$\Rightarrow$

$$\underset{(T+k)\times 1}{\mathbf{y}^*} = \begin{bmatrix} \frac{1}{\sigma} \underset{(T\times 1)}{\mathbf{y}} \\ \mathbf{P}^{-1} \underset{(k\times 1)}{\underline{\mu}_\beta} \end{bmatrix}, \underset{(T+k)\times k}{\mathbf{X}^*} = \begin{bmatrix} \frac{1}{\sigma}\mathbf{X} \\ \mathbf{P}^{-1} \end{bmatrix}, \mathbf{e}^* = \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_0 \end{bmatrix} \tag{14}$$

$$\mathbf{y}^* = \mathbf{X}^*\beta + \mathbf{e}^*, \mathbf{e}^* \backsim N(\mathbf{0}, \mathbf{I}_{T+k}) \tag{15}$$

$$\begin{aligned}
\overline{\beta} &= (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{y}^* = \left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} + \mathbf{P}^{-1\prime}\mathbf{P}^{-1}\right)^{-1} \\
&\times \left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{y} + \mathbf{P}^{-1\prime}\mathbf{P}^{-1}\underline{\beta}\right) \\
&= \left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}+\underline{\mathbf{H}}\right)^{-1}\left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{y}+\underline{\mathbf{H}}\underline{\beta}\right) = \\
&= \left(h\mathbf{X}'\mathbf{X}+\underline{\mathbf{H}}\right)^{-1}\left(h\mathbf{X}'\mathbf{y}+\underline{\mathbf{H}}\underline{\beta}\right)
\end{aligned} \tag{16}$$

$$\overline{\mathbf{H}} = \left(\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}+\underline{\mathbf{H}}\right) = \left(h\mathbf{X}'\mathbf{X}+\underline{\mathbf{H}}\right) \tag{17}$$

Note: inputs for moments of the posterior distribution of β:

1. $\underline{\mathbf{H}}, \underline{\mathbf{H}} \times \underline{\beta}$
2. $\mathbf{X}'\mathbf{X}, \mathbf{X}'\mathbf{y}$
3. h

$\Rightarrow$output:

1. $\overline{\beta}$
2. $\overline{\mathbf{H}}$

1.2 Unknown variance case

1.2.1 Digression:Gamma distribution

If $s \cdot x$ is Chisquare(ν)

$$p(s \cdot x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\nu/2)}(s \cdot x)^{\frac{\nu}{2}-1}\exp(-s \cdot x/2)$$

$$p(x) = \frac{\left(\frac{s}{2}\right)^{\frac{\nu}{2}}}{\Gamma(\nu/2)}(x)^{\frac{\nu}{2}-1}\exp(-s \cdot x/2)$$

$$E(x) = \frac{\nu}{s}$$
$$V(x) = \frac{2\nu}{s^2}$$
$$Mode(x) = \frac{\nu - 2}{s}$$

we say x is Gamma distributed.

In the ULRM:

- MRL: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- $\mathbf{X}$ $(T \times k)$ exogenous regressors matrix
- $\varepsilon \backsim N(\mathbf{0}, h^{-1}\mathbf{I}_T)$
- $h = \frac{1}{\sigma^2}$ precision

Likelihood:

$$p(\mathbf{y}|\mathbf{X}, \beta, \sigma^2) = \left(\sqrt{2\pi}\right)^{-T} h^{T/2} \exp\left\{-\frac{h}{2}\varepsilon'\varepsilon\right\}$$
$$\varepsilon = \mathbf{y} - \mathbf{X}\beta$$

Conjugate prior:

$$p(\theta) = p(\beta, h) = p(\beta|h) \times p(h)$$
$$p(\beta|h) = N(\underline{\beta}, h^{-1}\underline{\mathbf{H}}),$$
$$p(h) : \underline{s} \times h \backsim \chi^2_{\underline{\nu}} \text{ (} h \text{ Gamma distributed)}$$

posterior distribution

$$
\begin{aligned}
p(\theta|\mathbf{X},\mathbf{y}) &\propto p(\theta) \times p(\mathbf{y}|\mathbf{X},\theta) \\
&= h^{T/2} \exp\left\{-\frac{h}{2}\varepsilon'\varepsilon\right\} \times \\
&\quad h^{k/2} \exp\left\{-\frac{h}{2}\left(\beta - \underline{\beta}\right)' \underline{\mathbf{H}} \left(\beta - \underline{\beta}\right)\right\} \times \\
&\quad h^{\frac{\underline{\nu}-2}{2}} \exp(-\frac{1}{2}\underline{s}h) \\
&= h^{\frac{T+k+\underline{\nu}-2}{2}} \exp\left\{-\frac{h}{2}\varepsilon'\varepsilon - \frac{h}{2}\left(\beta - \underline{\beta}\right)' \underline{\mathbf{H}} \left(\beta - \underline{\beta}\right) - \frac{1}{2}\underline{s}h\right\}
\end{aligned}
$$

Easy to describe conditional posterior distributions

$$p(\beta|h, \mathbf{X}, \mathbf{y}) = N(\overline{\beta}, \overline{\mathbf{H}}^{-1}) \tag{18}$$

$$p(h|\beta, \mathbf{X}, \mathbf{y}) : \overline{s} \times h \backsim \chi^2_{\overline{\nu}} \tag{19}$$

(similar to known variance case)

$$\overline{\beta} = \left(\mathbf{X}'\mathbf{X} + \underline{\mathbf{H}}\right)^{-1} \left(\mathbf{X}'\mathbf{y} + \underline{\mathbf{H}}\underline{\beta}\right) \tag{20}$$

$$\overline{\mathbf{H}}_{\beta}^{-1} = h^{-1} \left(\mathbf{X}'\mathbf{X} + \underline{\mathbf{H}}\right)^{-1} \tag{21}$$

$$\overline{s} = \left[\underline{s} + \varepsilon'\varepsilon\right] \tag{22}$$

$$\overline{\nu} = \underline{\nu} + T + k \tag{23}$$

and the marginal posterior distribution $p(\beta|\mathbf{X}, \mathbf{y})$?

$$p(\beta|\mathbf{X}, \mathbf{y}) = \int p(\beta, h|\mathbf{X}, \mathbf{y}) dh \tag{24}$$

is Student t

Moreover, marginal likelihood is known analytically

$$p(\mathbf{y}|\mathbf{X}) = \int p(\beta, h) p(\mathbf{y}|\mathbf{X}, \beta, h) dh d\beta =$$

$$= \pi^{-T/2} \Gamma\left(\frac{T+\underline{\nu}}{2}\right) \Gamma\left(\frac{\underline{\nu}}{2}\right) \left[\frac{|\underline{\mathbf{H}}|}{|\overline{\mathbf{H}}|}\right]^{1/2} \underline{s}^{\frac{\underline{\nu}}{2}} \times$$
$$\left[\underline{s} + \widehat{\varepsilon}' \widehat{\varepsilon} + \left(\widehat{\beta} - \overline{\beta}\right)' \mathbf{X}'\mathbf{X} \left(\widehat{\beta} - \overline{\beta}\right) + \left(\underline{\beta} - \overline{\beta}\right)' \underline{\mathbf{H}} \left(\underline{\beta} - \overline{\beta}\right)\right]^{-\frac{T+\underline{\nu}}{2}} \tag{25}$$

complicated but known analytically (see Geweke, 2005).

2 Other prior distributions

Independent priors

$$
\begin{aligned}
p(\theta) &= p(\beta,h) = p(\beta) \times p(h) \text{ (ind a priori)} \\
p(\beta|h) &= N(\underline{\beta}, \underline{\mathbf{H}}^{-1}), \\
p(h) &: \underline{s} \times h \backsim \chi^2_{\underline{\nu}} \text{ (} h \text{ Gamma distributed)}
\end{aligned}
$$

Posterior distribution is

$$\begin{aligned}
p(\theta|\mathbf{X},\mathbf{y}) &\propto p(\theta)\times p(\mathbf{y}|\mathbf{X},\theta) \\
&= h^{T/2}\exp\left\{-\frac{h}{2}\varepsilon'\varepsilon\right\}\times \\
&\exp\left\{-\frac{1}{2}\left(\beta-\underline{\beta}\right)'\underline{\mathbf{H}}\left(\beta-\underline{\beta}\right)\right\}\times \\
&h^{\frac{\underline{\nu}-2}{2}}\exp(-\frac{1}{2}\underline{s}h) \\
&= h^{\frac{T+\underline{\nu}-2}{2}}\exp\left\{-\frac{h}{2}\varepsilon'\varepsilon-\frac{1}{2}\left(\beta-\underline{\beta}\right)'\underline{\mathbf{H}}\left(\beta-\underline{\beta}\right)-\frac{1}{2}\underline{s}h\right\}
\end{aligned}$$

Full set of conditionals

$$p(\beta|h, \mathbf{X}, \mathbf{y}) = N(\overline{\beta}, \overline{\mathbf{H}}^{-1}) \tag{26}$$

$$p(h|\beta, \mathbf{X}, \mathbf{y}) : \overline{s} \times h \backsim \chi^2_{\overline{\nu}} \tag{27}$$

with

$$\overline{\beta} = \left(h\mathbf{X}'\mathbf{X}+\underline{\mathbf{H}}\right)^{-1}\left(h\mathbf{X}'\mathbf{y}+\underline{\mathbf{H}}\underline{\beta}\right) \tag{28}$$

$$\overline{\mathbf{H}}^{-1} = \left(h\mathbf{X}'\mathbf{X}+\underline{\mathbf{H}}\right)^{-1} \tag{29}$$

$$\overline{s} = \left[\underline{s} + \varepsilon'\varepsilon\right] \tag{30}$$

$$\overline{\nu} = \underline{\nu} + T \tag{31}$$

What is marginal distribution of $p(\beta|\mathbf{X}, \mathbf{y})$?

$$p(\beta|\mathbf{X}, \mathbf{y}) = \int p(\beta, h|\mathbf{X}, \mathbf{y})dh \tag{32}$$

not known analytically.

Need to use simulation based methods

Also to obtain marginal likelihood

3 MCMC in the ULRM

Simplest approach: use a Gibbs sampling algorithm, with two blocks:

$$\theta = \begin{bmatrix} \beta \\ h \end{bmatrix}$$

Sequentially sample from:

$$p(\beta^{(i)}|h^{(i-1)}, \mathbf{y}, \mathbf{X}) \tag{33}$$

$$p(h^{(i)}|\beta^{(i)}, \mathbf{y}, \mathbf{X}) \tag{34}$$

from arbitrary initialisation.

This is a Markov chain that converges to posterior distribution .

3.1 Example: a Phillips curve for the US

Amisano and Giacomini (2007)

$$\pi_t = \beta_0 + \sum_{i=1}^{4} \beta_i \pi_{t-i} + \gamma_1 u_{t-1} + h^{-1/2} e_t \tag{35}$$

see codes for conjugate and non-conjugate analysis.

The codes are called

`main_linear_regression_PC_01.m` and

`main_linear_regression_PC_02.m`,

respectively.

4 Priors as stochastic constraints

Priors as stochastic constraints

$$\underset{(q\times k)}{\mathbf{R}}\beta = \delta + \varepsilon_0 \tag{36}$$

$$\varepsilon_0 \backsim N(0, \underline{\underline{\mathbf{H}}}^{-1}) \tag{37}$$

if $\underline{\underline{\mathbf{H}}}$ goes to zero, priors become less and less informative.

Write again in dummy observation form

$$\mathbf{y}^* = \mathbf{X}^*\beta + \mathbf{e}^* \tag{38}$$

$$\mathbf{y}^* = \begin{bmatrix} h^{1/2}\mathbf{y} \\ \underline{\underline{\mathbf{H}}}^{1/2}\delta \end{bmatrix}, \mathbf{X}^* = \begin{bmatrix} h^{1/2}\mathbf{X} \\ \underline{\underline{\mathbf{H}}}^{1/2}\mathbf{R} \end{bmatrix}, \mathbf{e}^* = \begin{bmatrix} h^{1/2}\varepsilon \\ \underline{\underline{\mathbf{H}}}^{1/2}\varepsilon_0 \end{bmatrix} \tag{39}$$

with Gamma prior on h.

$\longrightarrow$

$$p(\beta|h,\mathbf{y},\mathbf{X}) = N(\overline{\beta},\overline{\mathbf{H}})$$

$$\overline{\mathbf{H}} = \left[h\mathbf{X}'\mathbf{X} + \mathbf{R}'\underline{\mathbf{H}}\mathbf{R}\right]^{-1}$$

$$\overline{\beta} = \overline{\mathbf{H}}\left[h\mathbf{X}'\mathbf{y} + \mathbf{R}'\underline{\mathbf{H}}\delta\right]$$

Theil and Goldberger (1964) mixed estimator.

Examples of this dummy prior

- sum of coefficient priors : VARs, integration and cointegration
- shrinkage priors (Bayesian VARs);
- Del Negro and Schorfheide (IER, 2004): DSGE based prior

4.1 Sum of coefficients priors

Example with sum of coefficient priors: vertical Phillips curve

$$\sum_{i=1}^{4} \beta_i = 1 + noise \tag{40}$$

4.2 Shrinkage priors

When large parameter space, it is convenient to impose priors that shrink parameter space in a stochastic manner

$$\begin{aligned}\sqrt{V(a_{ij,k})} &= \lambda_0 \times k^{\lambda_1}, \lambda_1 < 0; \\ \lambda_0 &= overall \text{ shrinkage} \\ \lambda_1 &= lag\ decay \\ \sqrt{V(intercept)} &= \lambda_2\end{aligned}$$

Example with "Taylor Rule"

$$r_t = \beta_0 + \sum_{i=1}^{12} (\beta_{1i} r_{t-i} + \beta_{2i} \pi_{t-i} + \beta_{3i} u_{t-i}) + h^{-1/2} e_t \tag{41}$$

see code `main_linear_regression_TR_01.m`

5 Forecasting

Statement about the future (out of sample). What is unknown about the future?

- validity of the model
- parameters
- future shocks

In a frequentist approach

$$p(y_{T+1}|\mathbf{D}_T, \widehat{\theta}_T, M_j) \tag{42}$$

In a Bayesian framework

$$p(y_{T+1}|\mathbf{D}_T, M_j) = \int p(y_{T+1}|\mathbf{D}_T, \theta, M_j)p(\theta|\mathbf{D}_T, M_j)d\theta \tag{43}$$

(marginalise parameters out)

$$p(y_{T+1}|\mathbf{D}_T) = \sum_{j=1}^{J} p(y_{T+1}|\mathbf{D}_T, M_j)p(M_j|\mathbf{D}_T) \tag{44}$$

(marginalise model out)

$$p(M_j|\mathbf{D}_T) = \frac{p(\mathbf{D}_T|M_j)p(M_j)}{\sum_{i=1}^{J} p(\mathbf{D}_T|M_i)p(M_i)}$$

Classical forecasts

1. Point forecasts

2. interval forecasts

3. density forecasts