Bayesian Econometrics Lab 2

Ping Wu

University of Strathclyde

ping.wu@strath.ac.uk

Before beginning

- If you're still feeling a little stuck or lost, as with all code, I would suggest you really get to grips with the theory (and the economic intuition underlying it) in Lecture 2 first.
- This will make the code much easier to understand.

Load Data and Get It Into Correct Format

```
%load in the data set. Here use house price data from hprice.txt
load hprice.txt;

% Load dataset

n=size(hprice,1);

% Give me the size of dimension 1 of the houseprice dataset, which is the
% number of rows (also, the number of observations)

y=hprice(:,1);

% Give me column 1 of the hprice dataset, this is the dependent variable

x=hprice(:,2:5);

% Give me column 2-5 of the hprice dataset, this is the explanatory
% variables

x=[ones(n,1) x];

% Add an intercept as one explanatory variables

k=size(x,2);

% Give me the dimension 2 of x, which is the number of explanatory variables
```

Figure: Load Data and Get It Into Correct Format.

Spell out the regression model

$$salesprice_i = \beta_1 + \beta_2 lotsize_i + \beta_3 bedrooms_i + \beta_4 bathrooms_i + \beta_5 storeys_i + \epsilon_i$$

1. n = 546 houses sold in Windsor, Canada in 1987.
2. k = 5 explanatory variables (includes intercept term).

Specifying Hyperparameters

Now we specify our prior beliefs = a sensible guess about parameter values based on: past research, specialist knowledge, economic theory or economic intuition.

- E.g. Here, "the researcher could ask a local real estate agent to help provide prior information" (Koop, 2003, p.48).
- E.g. In fat data problems where we have many predictors, our prior belief is that many predictors may be irrelevant so we shrink coefficients towards zero.

We also specify our confidence in these beliefs/strength of information in the prior.

- Informative prior = prior certainty = confidence in beliefs= our guess is good.
- Noninformative prior = prior uncertainty = lack of confidence in beliefs = our guess is crude.

Recall natural conjugate Normal-Gamma prior

$$\beta | h \sim \mathcal{N}(\underline{\beta}, h^{-1}\underline{V})$$

- This means β conditional on h.

$$h \sim \mathcal{G}(\underline{s}^{-2}, \underline{v})$$

- Recall that $\epsilon \sim \mathcal{N}(0, h^{-1}I_N)$,
- then $h = \frac{1}{\sigma^2}$,
- then $h^{-1} = \sigma^2 = var(\epsilon) =$ error variance $=$ dispersion of the error term, reflecting house mispricing
- h gives the error precision

Specifying Hyperparameters: In our code

1. v0=5
   - $\underline{v} = 5$, this represents our confidence in prior belief about error precision h.
2. b0=0*ones(k,1); b0(2,1)=10; b0(3,1)=5000; b0(4,1)=10000; b0(5,1)=10000;

$$\underline{\beta} = \begin{pmatrix} 0 \\ 10 \\ 5000 \\ 10000 \\ 10000 \end{pmatrix}.$$

3. s02=1/4.0e-8;
   - $\underline{s}^{-2}$, this is prior belief about error precision h.

Specifying Hyperparameters: In our code

4. capv0=2.4*eye(k); capv0(2,2)=6e-7; capv0(3,3)=.15; capv0(4,4)=.6; capv0(5,5)=.6;

$$\underline{V} = \begin{pmatrix} 2.4 & 0 & 0 & 0 & 0 \\ 0 & 6 \times 10^{-7} & 0 & 0 & 0 \\ 0 & 0 & .15 & 0 & 0 \\ 0 & 0 & 0 & .6 & 0 \\ 0 & 0 & 0 & 0 & .6 \end{pmatrix}.$$

- Confidence in prior belief about regression coefficients.

How do we Interpret our Hyperparameters?

$$\underline{\beta} = \begin{pmatrix} 0 \\ 10 \\ 5000 \\ 10000 \\ 10000 \end{pmatrix}$$

- Our prior belief is that our regression equation looks like this: $\text{salesprice}_i = 10\times \text{ lotsize}_i + 5000\times \text{ bedrooms}_i + 10000\times$ $\text{bathroom } + 10000\times \text{ storeys}_i + \epsilon_i$
- All things held constant, if the number of bedrooms increases by one, the sales price increases by 10,000.
- All things held constant, if the number of storeys increases by one, the sales price increases by 10,000.

How do we Interpret our Hyperparameters?

$$\underline{V} = \begin{pmatrix} 2.4 & 0 & 0 & 0 & 0 \\ 0 & 6 \times 10^{-7} & 0 & 0 & 0 \\ 0 & 0 & .15 & 0 & 0 \\ 0 & 0 & 0 & .6 & 0 \\ 0 & 0 & 0 & 0 & .6 \end{pmatrix}.$$

- Recall that $var(\beta|h) = h^{-1}\underline{V}$, the variance of β conditional on h
- we can show that the prior covariance matrix for β alone is: $var(\beta) = \frac{\underline{vs}^2}{\underline{v}-2}\underline{V} = 41666666\frac{2}{3} \times \underline{V}$
- if we want $var(\beta_1) = 10000^2$, we need to set $\underline{V}(1,1) = \frac{10000^2}{41666666\frac{2}{3}}$, which is 2.4

How do we Interpret our Hyperparameters?

$$\underline{s}^{-2} = 4 \times 10^{-8}$$

- Most houses sold for prices in the 50 000 - 150 000 region.
- So in a model which fits well errors (ϵ_i) might be just a few thousand dollars and 10 000 at most.
- If $\sigma = 5000$ then 95% of errors will be less than $1.96 \times 5000 = 9800$, about 10 000.
- So a good guess for the prior mean of error precision, $h = \frac{1}{\sigma^2} = \frac{1}{5000^2} = 4.0 \times 10^{-8}$

Posterior Simulation

- This takes place in the script EXPOST1
- We combine OLS quantities with priors to obtain posterior.
- See page 7 of lecture notes for corresponding formulae.