Bayesian Time Series Methods: Introductory

Computer Tutorial 3: Fat Data Regression Methods

➢ Exercise 1: BMA in Cross Country Growth Regressions
Rule of Probability

- Bayesian model averaging (BMA) is motivated from the laws of probability. From the rules of probability, we obtain

$$p(\beta \mid y) = \sum_{r=1}^{R} p(\beta_r \mid y, M_r)\, p(M_r \mid y),$$

- which implies that the posterior for β is the average of the posterior in each individual model with weights proportional to $p(M_r \mid y)$.

- In order to have a solution, we need to derive the posterior distribution for $p(\beta_r \mid y, M_r)$ and the posterior distribution for the model $p(M_r \mid y)$.

Exercise 1: BMA in Cross Country Growth Regressions Model Setup

- How BMA is used in regression models with fat data? Let's first write the model space as

$$y = \alpha \iota_N + X_r \beta_r + \varepsilon.$$

- where ι_N is a $N \times 1$ vector of ones, X_r is a $N \times k_r$ matrix, $r = 1, \ldots, R$ denotes different models, and ε is assumed to be $N(0_N, h^{-1} I_T)$.

- The data includes N = 72 countries and k - 1 = 41 explanatory variables and average per capita GDP growth being the dependent variable.

- We have $2^{k-1} = 2^{41}$ possible choices for X_r and for total number of models.

Exercise 1: BMA in Cross Country Growth Regressions
The Marginal Likelihood

```
        %Making the matrix for data
y=rawdat(:,1);
xraw=rawdat(:,2:42);
bigk=size(xraw,2);
        %bigk is the number of potential explanatory variables

        %subtract mean from all regressors as in FLS
mxraw=mean(xraw);
sdxraw=std(xraw);
for i=1:bigk
    xraw(:,i)=(xraw(:,i) - mxraw(1,i))/sdxraw(1,i);
end
```

Exercise 1: BMA in Cross Country Growth Regressions Priors

- The g-prior is a special case of the natural conjugate prior, as emphasized by Koop et al. (2007). A common prior set by Fernandez et al. (2001) is

$$p(h) \propto 1/h,$$

- for the intercept

$$p(\alpha) \propto 1.$$

- and for the regression coefficients

$$\beta_r \mid h \sim N(\underline{\beta}_r, h^{-1}\underline{V}_r).$$

- The prior shrinks coefficients towards zero

$$\underline{\beta}_r = 0_{k_r},$$

- and

$$\underline{V}_r = \left[g_r X_r' X_r\right]^{-1}.$$

Exercise 1: BMA in Cross Country Growth Regressions Posteriors

- We obtain the posterior for $\overline{\beta}_r$ as follows

  $$\overline{\beta}_r = \overline{V}_r X_r' y,$$

  for the covariance matrix

  $$\operatorname{var}(\beta_r \mid y, M_r) = \frac{\overline{v}\,\overline{s}_r^2}{\overline{v} - 2} \overline{V}_r,$$

  where

  $$\overline{v} = N,$$

  $$\overline{V}_r = \left[(1 + g_r) X_r' X_r\right]^{-1},$$

  $$\overline{s}_r^2 = \frac{\frac{1}{g_r + 1} y' P_{X_r} y + \frac{g_r}{g_r + 1} (y - \overline{y}\iota_N)'(y - \overline{y}\iota_N)}{\overline{v}},$$

  and

  $$P_{X_r} = I_N - X_r (X_r' X_r)^{-1} X_r'.$$

Exercise 1: BMA in Cross Country Growth Regressions
The Marginal Likelihood

- The marginal likelihood is

$$p(y|M_r) \propto \left(\frac{g_r}{g_r+1}\right)^{\frac{k_r}{2}} \left[\frac{1}{g_r+1} y' P_{X_r} y + \frac{g_r}{g_r+1}(y-\bar{y}\iota_N)'(y-\bar{y}\iota_N)\right]^{-\frac{N-1}{2}},$$

and $p(M_r|y) = cp(y|M_r)p(M_r)$.

Exercise 1: BMA in Cross Country Growth Regressions
Priors for g

- Use $g = 1/N$ if $N > K^2$ or $g = 1/K^2$ if $N \leq K^2$

```
        % Specifying g-prior
if n<=(bigk^2)
   g0=1/(bigk^2);
else
   g0=1/n;
end
```

Exercise 1: BMA in Cross Country Growth Regressions Posteriors

```
%calculating posterior properties of coefficients means
Q1inv = (1+g0)*xold'*xold;
Q0inv=g0*xold'*xold;
    Q1=inv(Q1inv);
    b1= Q1*xold'*y;
    vs2 = (y-xold*b1)'*(y-xold*b1) + b1'*Q0inv*b1;
    bcov = (vs2/(n-2))*Q1;
```

➢ Exercise 1: BMA in Cross Country Growth Regressions Computations MC^3

- Computations can be carried out using Markov chain Monte Carlo model compositions (MC^3) algorithm.

- Use draws and burn-in replications 110,000 and 10,000, respectively.

- Are results sensitive to the number of draws?

- How to select values of g?

➢ Exercise 2: Stochastic Search Variable Selection (SSVS)

- Let's write the linear regression model as

$$y_i = \beta_0 + \sum_{j=1}^{K} x_{ji}\beta_j + \varepsilon_i,\ \ \varepsilon_i \sim_{i.i.d.} N(0,\sigma^2).$$

- where y is $N \times 1$ vector and X is a $N \times K$ matrix of explanatory variables, and ε is a $N \times 1$ vector of errors.

- To extract information relevant to variable selection, above equation can be considered as part of a larger hierarchical model.

Exercise 2: SSVS

- Each component of β is modelled as having come from a mixture of two normal distributions with different variances.

- For each coefficient β_j with $j = 1, \ldots, K$, the prior is

$$\beta_j \mid \gamma_j \sim (1-\gamma_j) N(0, \tau_{0j}^2) + \gamma_j N(0, c_j^2 \tau_{1j}^2),$$

- c_j and τ_j are known hyperparameters.
- $\tau_{0j}^2 = \text{small}.$
- $c_j^2 \tau_{1j}^2 = \text{large}.$

- $P(\gamma_j = 1) = p_j$ or $P(\gamma_j = 0) = 1 - p_j$.
- If $\gamma_j = 0$, the variable x_j can be excluded from the model.
- If $\gamma_j = 1$, the variable x_j can be retained in the model.

➢ Exercise 2: SSVS
Semi-automatic Choice of Small and Large Prior Variances

- Use OLS

$$b = (X'X)^{-1} X'y,$$

$$s^2 = e'e/(n-k),$$

$$\text{var}(b \mid X) = \sigma^2 (X'X)^{-1}.$$

```
%Make the semi-automatic choices of "small" and "large" prior variances
%This uses OLS so will not work if bigk>nobs
xtxinv = inv(xmat'*xmat);
b_ols = xtxinv*xmat'*y;
s2_ols = (y-xmat*b_ols)'*(y-xmat*b_ols)/(nobs - bigk-1);
b_cov = s2_ols*xtxinv;
b_sd = sqrt(diag(b_cov));
```

Exercise 2: SSVS

Prior Hyperparameters

- With $\beta_0 \sim N(0, \underline{V})$, where $\underline{V} = DD$, and $D = diag(h_1, \ldots, h_K)$.

$$h_j = \begin{cases} \tau_{0j}, \text{ if } \gamma_j = 0, \\ \tau_{1j}, \text{ if } \gamma_j = 1. \end{cases}$$

- Set $\tau_{0j} = c_1 \hat{\sigma}_\beta$ and $\tau_{1j} = c_2 \hat{\sigma}_\beta$, where $c_1 < c_2$, such as

```
        %prior hyperparameters
V0= 10^2;
c1 = 0.1;
c2=10;
tau1 = c1*b_sd(2:bigk+1,1);
tau2 = c2*b_sd(2:bigk+1,1);
```

- And for the error precision $h \sim G(\underline{s}^{-2}, \underline{v})$.

```
        %prior hyperparameters for error precision
s_bar_2 = 0.01;
prior_dof = 0;
post_dof = prior_dof + nobs;
```

➢ Exercise 3: The Least Absolute Shrinkage and Selection Operator (LASSO)

- The purpose of LASSO is to minimize

$$(y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^{k} \left|\beta_j\right|,$$

- where λ is a shrinkage parameter.

- Laplace prior for β that is a mixture of Normal distributions with different variances can be written as

$$\beta_j \sim N(0, h^{-1}\tau_j^2),$$

$$\tau_j^2 \sim Exp\left(\frac{\lambda^2}{2}\right).$$

- We can obtain posterior conditionals for $p(\beta \mid y, h, \tau)$ and $p(h \mid y, \beta, \tau)$ for $\tau = (\tau_1, \ldots, \tau_K)'$ using standard results for Normal linear regression.

Exercise 3: MCMC Algorithm

- Obtain the LASSO prior covariance as

$$\underline{V} = h^{-1}DD,$$

- where $D = diag(\tau_1, \ldots, \tau_K)'$. We get

- With $\beta \mid y, h, \tau \sim N(\overline{\beta}, \overline{V})$.

$$\overline{\beta} = \left(X'X + (DD)^{-1}\right)^{-1} X'y,$$

$$\overline{V} = h^{-1}\left(X'X + (DD)^{-1}\right)^{-1}.$$

- And $h \mid y, \beta, \tau \sim G(\overline{s}^{-2}, \overline{v})$.

$$\overline{v} = N + K \text{ and } \overline{s}^2 = \frac{(y - X\beta)'(y - X\beta) + \beta'(DD)^{-1}\beta}{\overline{v}}.$$

Exercise 3: MCMC Algorithm
Drawing β

```
% Update beta from Normal
A = inv(x'*x + inv(V_L));
post_mean_beta = A*x'*y;
post_var_beta = sigma2*A;
beta = Draw_Normal(post_mean_beta,post_var_beta);
```

Exercise 3: MCMC Algorithm
Priors for τ, λ and h

$$1/\tau_j^2 \mid y, \beta, h, \lambda \sim IG(\bar{c}_j, \bar{d}_j)$$

- with $\bar{d} = \lambda^2$,

$$\bar{c}_j = \sqrt{\frac{\lambda^2}{h\beta_j^2}}.$$

- And

$$\lambda^2 \sim G(\underline{\mu}_\lambda, \underline{v}_\lambda).$$

$$\bar{v}_\lambda = \underline{v}_\lambda + 2K,$$

$$\bar{\lambda} = \frac{\underline{v}_\lambda + 2K}{2\sum_{j=1}^{K} \tau_j^2 + \frac{\underline{v}_\lambda}{\underline{\mu}_\lambda}}.$$

- We may assume a noninformative prior for h as

$$p(h) \propto 1/h.$$

Exercise 3: MCMC Algorithm
Drawing τ and λ

```
% Update tau2_j from Inverse Gaussian
  for j = 1:p
     a1 = (lambda2*sigma2)./(beta(j,1)^2);
     a2 = lambda2;
     tau2_inverse = Draw_IG(sqrt(a1),a2);
 %note: often need to add a very small constant to avoid matrix singularity
     tau2(j,1) = 1/tau2_inverse + 1e-15;
  end

 % Update lambda2 from Gamma
  b1 = p + r;
  b2 = 0.5*sum(tau2) + delta;
  lambda2=gamrnd(b1,1/b2);
```

Exercise 3: MCMC Algorithm
Drawing σ^2

```
%Update sigma2 from Inverse Gamma
    c1 = (T-1+p)/2;
    PSI = (y-x*beta)'*(y-x*beta);
    c2 = 0.5*PSI + 0.5*(beta'/V_L)*beta;
    sigma2 = Draw_iGamma(c1,c2);
```

References

- Fernandez, C. Ley, E. and Steel, M. F. J. (2001), Model uncertainty in cross country growth regressions, Journal of Applied Econometrics, 16, 563–576.
- George, E. and McCulloch, R. (1993), Variable selection via Gibbs sampling, Journal of the American Statistical Association, 88, 881–889.
- George, E., Sun, D. and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions, Journal of Econometrics, 142, 553–580.
- Koop, G. (2003), Bayesian econometrics, Wiley.
- Koop, G., Poirier, D. and Tobias, J. (2007), Bayesian econometric methods. Cambridge: Cambridge University Press.