**Домашняя работа по курсу «Байесовский подход в эконометрике»**

Для рассмотрения были выбраны страны Латинской Америки (Аргентина, Бразилия, Колумбия, Мексика, Чили, Венесуэла). Исследовалась зависимость уровня социального согласия в обществе от уровня бюрократии и личной безопасности в обществе. Предполагалось, что менее бюрократизированных и более безопасных обществах люди должны жить в большем согласии.

Была проведена оценка модели методом MCMC (Markov Chain Monte Carlo). Было использовано 10 000 итераций, в качестве начальных были взяты байесовские оценки (использование других начальных значений, вплоть до набора случайных чисел, не приводит к заметным изменениям результата – цепь сходится; в пользу этого же говорят и проведенные тесты, о них ниже).

Гистограммы коэффициентов и дисперсии:

Наименьшая дисперсия у распределения параметра дисперсии, оценки коэффициентов, чисто визуально, имеют примерно одинаковый разброс. При этом средние значения коэффициентов очень близки к полученным при помощи МНК (и далеки от байесовских оценок):

Константа Bureaucracy Security

MCMC 2.865458 -0.843103 1.246776

МНК 2.8662887 -0.8307734 1.2406413

Байес 1.803631 1.822229 -0.3007939

Возможные причины таких расхождений в оценках МНК и байесовских уже обсуждались в предыдущей части работы – возможны изменения в структуре или в способе расчёта показателей, которые при использовании байесовского подхода «размываются» использованием априорной информации. С этой точки зрения, MCMC, не использующий априорную информацию (в данном примере в качестве начальных брались байесовские значения, но результаты остаются стабильными даже при использовании в этом качестве случайных чисел), кажется более разумным. И прогнозы:

Оказываются ближе к истине (4,03 и 4,96 соответственно): для Аргентины попадание очень точное, для Бразилии – менее, но всё же результат несколько лучше, чем для Байеса (напомним, он давал прогноз для Бразилии 3,43 – отклонение в полтора раза больше, чем для MCMC).

Совмещение доверительной области байесовского прогноза и результатов MCMC даёт следующий результат:

По большей части прогнозные значения MCMC оказались внутри байесовской доверительной области, но оказались смещены относительно её центра.

Далее была проведена диагностика сходимости цепи для MCMC при помощи тестов Geweke, Gelman and Rubin, Heidelberger and Welch и Raftery and Lewis. Результаты (приведены значения статистики, распределение – стандартное нормальное, нулевая гипотеза – вышли на стационар):

> geweke.diag(m)

Fraction in 1st window = 0.1

Fraction in 2nd window = 0.5

 var1 var2 var3 var4

 0.8596 1.3221 -0.8500 0.3577

Нулевая гипотеза не отвергается ни для одной переменной, значит, цепь вышла на стационар.

Тест Gelman and Rubin (значение показателя заметно больше 1 обозначает отсутствие сходимости; для получения оценок требуется несколько цепей – было сгенерировано 10 с разными начальными значениями коэффициентов):

> gelman.diag(mlist)

Potential scale reduction factors:

 Point est. Upper C.I.

[1,] 1 1.01

[2,] 1 1.00

[3,] 1 1.00

[4,] 1 1.00

Multivariate psrf

1

По всем показателям по отдельности и по совокупности получаем значение 1 (с верхней границей доверительного интервала от 1,00 до 1,01) – значит, есть сходимость.

Тест Heidelberger and Welch (нулевая гипотеза: длина цепи достаточна):

> heidel.diag(m)

 Stationarity start p-value

 test iteration

var1 passed 1 0.0644

var2 passed 1 0.5398

var3 passed 1 0.2075

var4 passed 1 0.2197

 Halfwidth Mean Halfwidth

 test

var1 passed 2.013 0.0222

var2 passed 2.849 0.0223

var3 passed -0.818 0.0256

var4 passed 1.235 0.0160

Для всех переменных получаем стационарность.

Тест Raftery and Lewis (значения I больше 5 говорят о проблемах со сходимостью):

> raftery.diag(m)

Quantile (q) = 0.025

Accuracy (r) = +/- 0.005

Probability (s) = 0.95

 Burn-in Total Lower bound Dependence

 (M) (N) (Nmin) factor (I)

 3 4028 3746 1.08

 2 3929 3746 1.05

 2 3802 3746 1.01

 2 3997 3746 1.07

Как видно, проблем со сходимостью нет. Как было обнаружено в ходе дополнительных исследований, даже при 1000 итераций тесты в основном говорят о сходимости: Geweke и Raftery and Lewis – по всем показателям, а Heidelberger and Welch кроме переменной при бюрократии. Сходимость достаточно хорошая.

Приложение: код на R (может представлять хоть какой-то интерес в части, посвященной диагностике):

#######MCMC#######

#Based on some variables from the 1st part of the homework

#Requires MCMC code provided by Boris Demeshev

init<-rbind(summary(model2005)$sigma, as.matrix(summary(model2005)$coefficients[,1]))

res<-mcmc\_regression(10000,model2005$y,model2005$x,init,c(alpha\_post,beta\_post))

par(mfrow=c(2,2))

hist(res[,1],main="distribution of sigma2")

hist(res[,2],main="distribution of intercept")

hist(res[,3],main="distribution of Bureaucracy")

hist(res[,4],main="distribution of Security")

par(mfrow=c(1,2))

#forecast

mcmcf<-res[,2:4]%\*%t(Xf)

hist(mcmcf[,1],main="distribution of forecast for Argentina")

hist(mcmcf[,2],main="distribution of forecast for Brazil")

par(mfrow=c(1,1))

#OLS Forecast

t(as.matrix(summary(model2005)$coefficients[,1]))%\*%t(Xf)

###Ellipse and MCMC points

plot(xr,yr,asp=1, xlab = "Argentina", ylab = "Brasil")

points(mcmcf, col = "blue")

###Diagnostics

library(coda)

m<-mcmc(res)

#Geweke test

geweke.diag(m)

geweke.diag(m,frac1 = 0.2)

geweke.diag(m,frac1 = 0.3)

geweke.diag(m,frac1 = 0.4)

#Create mcmc list and try Gelman and Rubin's test

res1<-mcmc\_regression(10000,model2005$y,model2005$x,init,c(alpha\_post,beta\_post))

res2<-mcmc\_regression(10000,model2005$y,model2005$x,(init+rnorm(4,1)),c(alpha\_post,beta\_post))

res3<-mcmc\_regression(10000,model2005$y,model2005$x,(init+rnorm(4,1)),c(alpha\_post,beta\_post))

res4<-mcmc\_regression(10000,model2005$y,model2005$x,(init+rnorm(4,1)),c(alpha\_post,beta\_post))

res5<-mcmc\_regression(10000,model2005$y,model2005$x,(init+rnorm(4,1)),c(alpha\_post,beta\_post))

res6<-mcmc\_regression(10000,model2005$y,model2005$x,(init+rnorm(4,1)),c(alpha\_post,beta\_post))

res7<-mcmc\_regression(10000,model2005$y,model2005$x,(init+rnorm(4,1)),c(alpha\_post,beta\_post))

res8<-mcmc\_regression(10000,model2005$y,model2005$x,(init+rnorm(4,1)),c(alpha\_post,beta\_post))

res9<-mcmc\_regression(10000,model2005$y,model2005$x,(init+rnorm(4,1)),c(alpha\_post,beta\_post))

res10<-mcmc\_regression(10000,model2005$y,model2005$x,(init+rnorm(4,1)),c(alpha\_post,beta\_post))

mlist<-mcmc.list(mcmc(res1),mcmc(res2),mcmc(res3),mcmc(res4),mcmc(res5),mcmc(res6),mcmc(res7),mcmc(res8),mcmc(res9),mcmc(res10))

gelman.diag(mlist)

#Heidelberger and Welch's test

heidel.diag(m)

#Raftery and Lewis's test

raftery.diag(m)