AAEC 6564 Bayesian Econometric Analysis

Instructor: Klaus Moeltner

Problem Set 3

General Instructions:

Please type everything in LaTeX (including all Math) and hand in a pdf file. For problems involving Matlab, answer questions in LaTeX, and attach your script, log file, and any graphs to your main pdf file.

If you prefer, you can also load your Matlab figures and/ or output tables into your LaTeX file (see PS1 instructions).

Q1: Probit with door-to-door data

The data for this exercise flow from a door-to-door fundraising campaign conducted in Pitt County, North Carolina, during the fall of 2005. The details of this field experiment are described in Landry et al. (2006). Forty-three solicitors interacted with an average of 39.3 households for a total sample size of 1690 observations. All observations are based on actual interactions, i.e. the "door didn't open" cases are not considered in this data set. The focus of this research was on the effect of lottery designs and solicitor attributes on donation outcomes.

The data are posted as door2door.txt. The variables are as follows:

```
% Variable
%
% 1     solid        solicitor ID
% 2     donate       "1"=donation received
% 3     amount       donation amount
% 4     sol_f        "1"=sol=female
% 5     sol_nw       "1"=sol=non-white
% 6     height       sol height (inches)
% 7     bmi          sol body mass index (>28=overweight)
% 8     beauty       beauty index (z-score)
% 9     spunk        spunk index (low=-40, high=40)
% 10    res_f        "1"=resident = female
% 11    res_nw       "1"=resident = non-white
% 12    res_old      "1"=resident=60 or older
% 13    hh_size      avg. HH size in census block
% 14    perc_own     % of homes owned in census block
% 15    inc000       median census tract income  ($1000)
% 16    lottery      "1" solicitation included a lottery incentive
```

part (a)

Run a probit model of "donate" on **X**, with **X** given as:

```
% Contents of X1
% 1      constant
% 2      sol_f       "1"=sol=female
% 3      sol_nw      "1"=sol=non-white
% 4      height      sol height (inches)
% 5      bmi         sol body mass index (>28=overweight)
% 6      beauty      beauty index (z-score)
% 7      spunk       spunk index (low=-40, high=40)
```

```
% 8      hh_size        avg. HH size in census block
% 9      perc_own     % of homes owned in census block
% 10     inc000       median census tract income  ($1000)
% 11     lottery        "1" solicitation included a lottery incentive
```

Use 15,000 burn-ins and 5000 keepers. Also capture in your log file the fraction of "zeros" in your dependent variable, i.e. unsuccessful solicitation attempts. Call everything "`ps3_q1a`". Save your "`betamat`" and data matrix **X**. Use the prior settings from class.

Comment on model convergence and efficiency.
Which variables seem to matter most based on the p(>0) statistics (i.e. have p(>0)>0.95 or < 0.05)?

part (b)

Generate and plot posterior predictive distributions (PPDs) for the probability of receiving a donation for each of the following three types (you can use the following code directly):

```
h_mw=71.7; %height, male, white
h_mm=70.8; %height, male, minority
h_fw=65.9;
h_fm=61.5;
bmwb=23.1; %bmi, male, white, baseline
bmwe=25.4; %bmi, male, white, elevated
bmmb=25.9;
bmme=33.6;
bfwb=23.5;
bfwe=28.8;
bfmb=25.2;
bfme=28.8;
lmwb=0; %looks, male, white, baseline
lmwe=0.9; %looks, male, white, elevated
lmmb=0;
lmme=0.9;
lfwb=0;
lfwe=0.9;
lfmb=0;
lfme=0.9;
sp=27; %spunk
hh=2.54; %HH size
own=0.8; %own home
inc=42; %median tract income, 000
lot=1; %lottery dummy

%Type 1: average white male
x1= [1 0 0 h_mw bmwb lmwb sp hh own inc lot]';
%Type 2: heavy-set minority male
x2= [1 0 1 h_mm bmme lmmb sp hh own inc lot]';
%Type 3: Attractive white female
x3= [1 1 0 h_fw bfwb lfwe sp hh own inc lot]';
```

Comment on the relative shape and location of the three distributions. Which type would you hire if your goal was to maximize donation incidents?

What are the bounds of a 95% Highest Posterior Density Interval for the third type? Are you confident that this type has at least a 40% chance of receiving a donation? Explain.

part (c)

Returning to your original results, generate and plot PPDs for the marginal effect of "*beauty*" and "*lottery*" on the probability of receiving a donation. For beauty, set all regressors to their sample mean. For "lottery", do the same, except with settings for "lottery" at 0 and 1 respectively.

Comment on any differences in the two distributions. Compute the 95% HPDI for each case. Are you confident that the marginal effect will be positive for either case? Explain.

<u>Q2: Tobit with door-to-door data</u>

part (a)

Run a basic Tobit model for the door-to-door fundraising application. Your dependent variable will be “amount/10”. Your explanatory variables should be as follows:

```
% 1      constant
% 2      sol_f        "1"=sol=female
% 3      sol_nw       "1"=sol=non-white
% 4      height       sol height (inches)
% 5      bmi          sol body mass index (>28=overweight)
% 6      beauty       beauty index (z-score)
% 7      spunk        spunk index (low=-40, high=40)
% 8      res_f       "1"=resident = female
% 9      res_nw       "1"=resident = non-white
% 10     res_old      "1"=resident=60 or older
% 11    lottery      "1" solicitation included a lottery incentive
```

Use 15,000 burn-ins and 5000 keepers. Use the same prior settings as in script `mod5_tobit`. Call everything “`ps3_q3a`”. Save your “`betamat`” and “`sig2mat`”.

Comment on model convergence and efficiency.
Which variables seem to matter most based on the p(>0) statistics (i.e. have p(>0)>0.95 or < 0.05)?

part (b)

Generate and plot posterior predictive distributions (PPDs) for the expected amount *conditional on receiving a donation* for the three solicitor types from above (generate PPDs in units of $10, then multiply the entire series of draws by 10 to get results in dollars). You can use this code directly: (the implicit respondent is a white male under 60 years of age):

```
h_mw=71.7; %height, male, white
h_mm=70.8; %height, male, minority
h_fw=65.9;
h_fm=61.5;
bmwb=23.1; %bmi, male, white, baseline
bmwe=25.4; %bmi, male, white, elevated
bmmb=25.9;
bmme=33.6;
bfwb=23.5;
bfwe=28.8;
bfmb=25.2;
bfme=28.8;
lmwb=0; %looks, male, white, baseline
lmwe=0.9; %looks, male, white, elevated
lmmb=0;
lmme=0.9;
lfwb=0;
lfwe=0.9;
lfmb=0;
lfme=0.9;
sp=27; %spunk
hh=2.54; %HH size
own=0.8; %own home
inc=42; %median tract income, 000
```

```
lot=1; %lottery dummy

%Type 1: average white male
x1= [1 0 0 h_mw bmwb lmwb sp 0  0   0   lot]';
%Type 2: heavy-set minority male
x2= [1 0 1 h_mm bmme lmmb sp 0  0   0   lot]';
%Type 3: Attractive white female
x3= [1 1 0 h_fw bfwb lfwe sp 0  0   0   lot]';
```

Comment on the relative shape and location of the three distributions. Which type would you hire if your goal was to maximize the expected amount per donation? How does this plot compare to the one from Q1 that compared the tree types based on donation success rates?

What are the bounds of a 95% Highest Posterior Density Interval for the FIRST type? Are you confident that this type will collect at least $5.50 / donation (conditional on receiving a donation)? Explain.

part (c)

Returning to your original results, generate and plot PPDs for the marginal effect of "*beauty*" and "*lottery*" on donation amount, conditional on receiving a donation. For beauty, set all regressors to their sample mean. For "lottery", do the same, except with settings for "lottery" at 0 and 1 respectively.

Comment on any differences in the two distributions. Compute the 95% HPDI for each case. Are you confident that the marginal effect will be positive for either case? Explain.

Q3: HNRM with outage data

Consider the HNRM discussed in class for the 50-firm outage data set.

part (a)

Estimate a model that has only a single random effect for "outage duration". All other coefficients are fixed.

Use `gs_HNRM_v2` for your Gibbs Sampler, with the following priors and general settings:

```
% general elements
r1=10000; % burn-in
r2=5000; % keepers
R=r1+r2;

% generic OLS
bols=inv(X'*X)*X'*y;
res=y-X*bols;
s2=(res'*res)/(obs-k);

% elements for betaf (fixed coefficients)
muf0=zeros(kf,1); %diffuse prior for mean of fixed coefficients
Vf0=eye(kf)*100; % diffuse prior for varcov of fixed coefficients
% since we're drawing beta first, won't need starting draw

% elements for betar (hierarchical means of random coefficients
mur0=zeros(kr,1);
Vr0=eye(kr)*100;
betardraw=bols(kf+1:end);

% elements for E -hierarchical VCOV
v0= kr+1;
S0= eye(kr);
Edraw=iwishrnd(S0,v0);


% elements for sig2
sig2draw=s2; % use OLS variance as starting draw for Gibbs Sampler
eta0=1/2;
tau0=1/2; % diffuse prior shape and scale
```

Note that even though you just have a single random coefficient, you can keep working with the *IW* in this case – it will simply collapse to the *ig* for one-dimensional draws.

part (b)

Use the Chib (1995) method to derive the marginal likelihood for this model. Compute a (logged) Bayes Factor that compares this model to the unrestricted HNRM that we used in class (with fully correlated random coefficients for "weekday," "daytime," and "duration"). You can use `gs_HNRM_chib` for the simulation part.

Comment on the result of this comparison. Which model receives stronger support from the data? Also comment on the comparative magnitude of individual components that feed into the marginal likelihood. Which component is the primary driver of the difference in the marginal likelihoods? Is it the log-likelihood or the difference between prior and posterior?