David Giles
Bayesian Econometrics

10. Model Selection - Applications

Baseball Example

(Hoogerheide *et al.*, 2007)

- 2004 World Series - Boston Red Socks *vs*. St; Louis Cardinals

$$y_t = 1 \;\; ; Red\ Socks\ win\ Game\ t$$

$$= 0 \;\; ; \;\; Cardinals\ win\ Game\ t \;\; ; \quad t = 1, 2, \ldots., T$$

Bernoulli distribution:

$$p(y_t \mid \theta) = \theta^{y_t}(1 - \theta)^{1 - y_t} \qquad ; \qquad 0 \leq \theta \leq 1$$

- Likelihood function if Red Socks win T_1 games and Cardinals win T_2 games:

$$L(\theta \mid \boldsymbol{y}) = p(\boldsymbol{y} \mid \theta) = \prod_{t=1}^{T} p(y_t \mid \theta) = \theta^{T_1}(1-\theta)^{T_2}$$

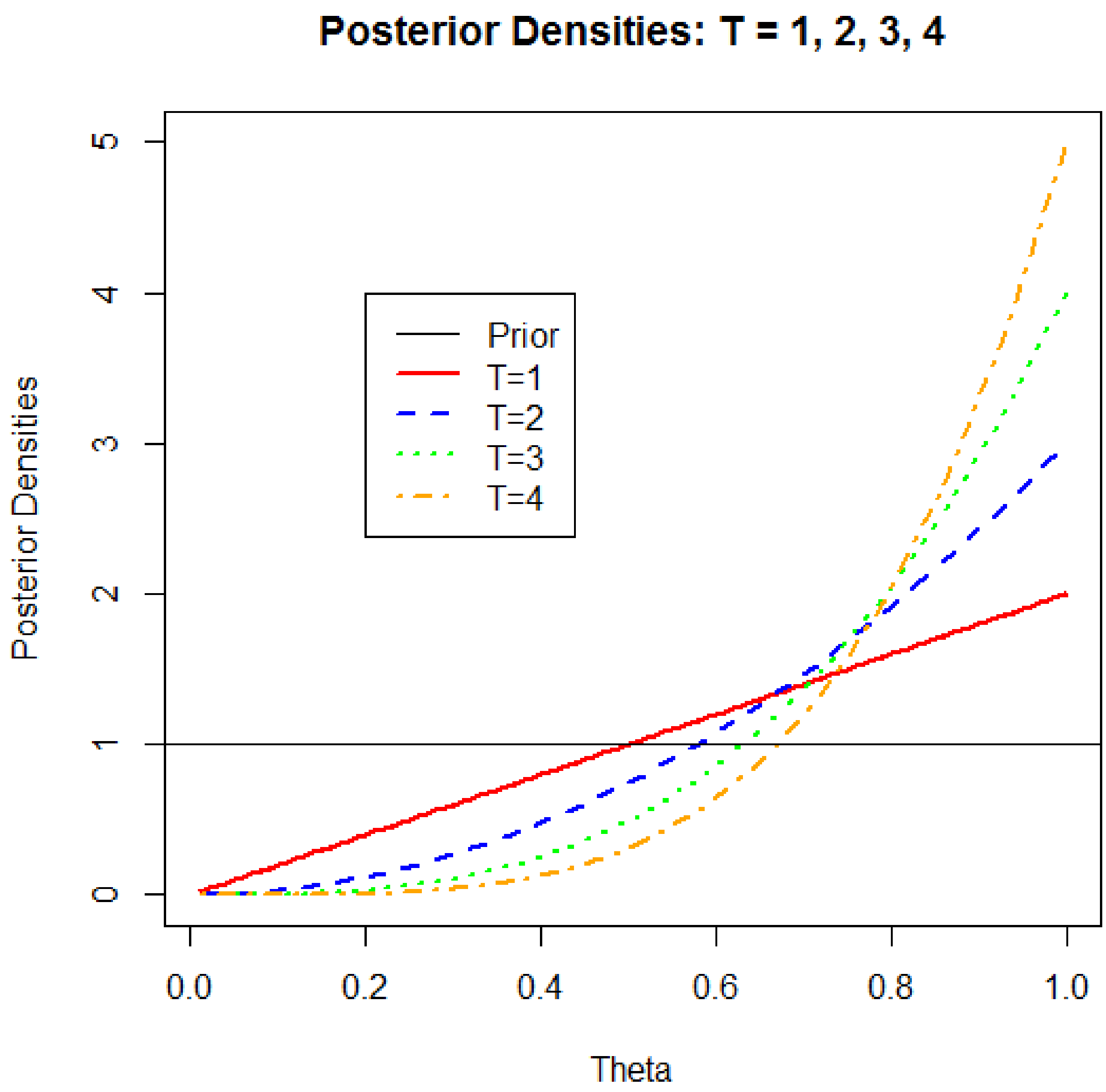
Prior density:

$$p(\theta) = 1 \quad ; \quad \theta \in [0\,,1]$$

- Actually, Boston won the World Series in 4 straight games.
- Apply Bayes' Theorem. After T games,

$$p(\theta \mid \boldsymbol{y}) \propto p(\theta)L(\theta \mid \boldsymbol{y}) \propto \theta^{T} \quad ; \quad T = 1, 2, 3, 4.$$

- Normalizing constant is $1/\left(\int_0^1 \theta^T d\theta\right) = (T+1)$
- So, $p(\theta \mid \boldsymbol{y}) = (T+1)\,\theta^{T} \quad ; \quad T = 1, 2, 3, 4.$


Posterior Densities: T = 1, 2, 3, 4
Posterior Densities
Theta
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Prior
T=1
T=2
T=3
T=4

- Now consider BPO analysis. Two "non-nested" models.
- $M_1: \theta \leq \frac{1}{2}$ ("Cardinals are at least as good as the Red Socks")

  $M_2: \theta > \frac{1}{2}$ ("Red Socks are better than the Cardinals")

- Let $p(M_1) = p(M_2) = \frac{1}{2}$

$$p(\theta \mid M_1) = 2 \quad ; \quad 0 \leq \theta \leq \frac{1}{2} \quad \text{and} \quad p(\theta \mid M_2) = 2 \quad ; \quad \frac{1}{2} < \theta \leq 1$$

- Recall that Red Socks won all matches, so

$$p(\boldsymbol{y} \mid M_1) = \int p(\boldsymbol{y} \mid \theta, M_1) p(\theta \mid M_1) d\theta = \int_0^{0.5} \theta^T 2 d\theta = \frac{2}{(T+1)} \left(\frac{1}{2}\right)^{T+1}$$

$$p(\boldsymbol{y} \mid M_2) = \int p(\boldsymbol{y} \mid \theta, M_2) p(\theta \mid M_2) d\theta = \int_{0.5}^{1} \theta^T 2 d\theta = \frac{2}{(T+1)} \left[1 - \left(\frac{1}{2}\right)^{T+1}\right]$$

- $BPO_{12} = \frac{p(M_1|\boldsymbol{y})}{p(M_2|\boldsymbol{y})} = \frac{\left(\frac{1}{2}\right)^{T+1}}{1-\left(\frac{1}{2}\right)^{T+1}}$

$$p(M_1|\boldsymbol{y}) = \left(\frac{1}{2}\right)^{T+1} \quad ; \quad p(M_2|\boldsymbol{y}) = 1 - \left(\frac{1}{2}\right)^{T+1}$$

- So, "the probability that the Cardinals are at least as good as the Red Socks", given T = 1, 2, 3, 4 matches won by the Red Socks, is $\left(\frac{1}{2}\right)^{T+1} =$ 0.25, 0.125, 0.06, 0.03.
- Check the frequentist outcome when $H_0 = M_1$, $H_A = M_2$, and the test statistic is the number of games won by the Red Socks. The p-value = $(0.5)^T$. So, even when $T = 4$, $p = 0.0625$. We would not reject H_0 (M_1) at the 5% significance level!

Econometric Example - Distributed Lag Models

- Example from Giles (1975) – competing "Distributed Lag" regression models with AR(1) error terms.
- Explain payments for imports into N.Z..
- 12 different models: 3 lag "shapes" (S); 4 maximum lag lengths (L)
- 4 parameters in each model. Scale parameter for errors eliminated by analytic integration.
- Rest of analysis involved 3-dimensional numerical integration. (Prior to MCMC!)
- Emphasis on:

  (i) Posterior probabilities for each model.

  (ii) Parameter estimates & predictions based on Bayesian Model Averaging (BMA), using model posterior probabilities as weights.

Bayesian Model Averaging Example - The BMS Package

- Basic idea – estimate many competing models and then "weight" the results using the model Posterior Probabilities.
- Applies to estimates of coefficients, predictions, *etc*.
- **BMS package for R** deals with regression models where there are K *potential regressors*.
- Each regressor can be included or excluded from the model, so there are 2^K models in the full Model Space.
- *e.g.*, If K = 40, there are 1.1×10^{12} possible models!
- A Metropolis-Hasting type of MCMC is used to search the model space and obtain a manageable random selection of models that have high posterior probabilities.

- These models are then combined using the (re-weighted) posterior probabilities.

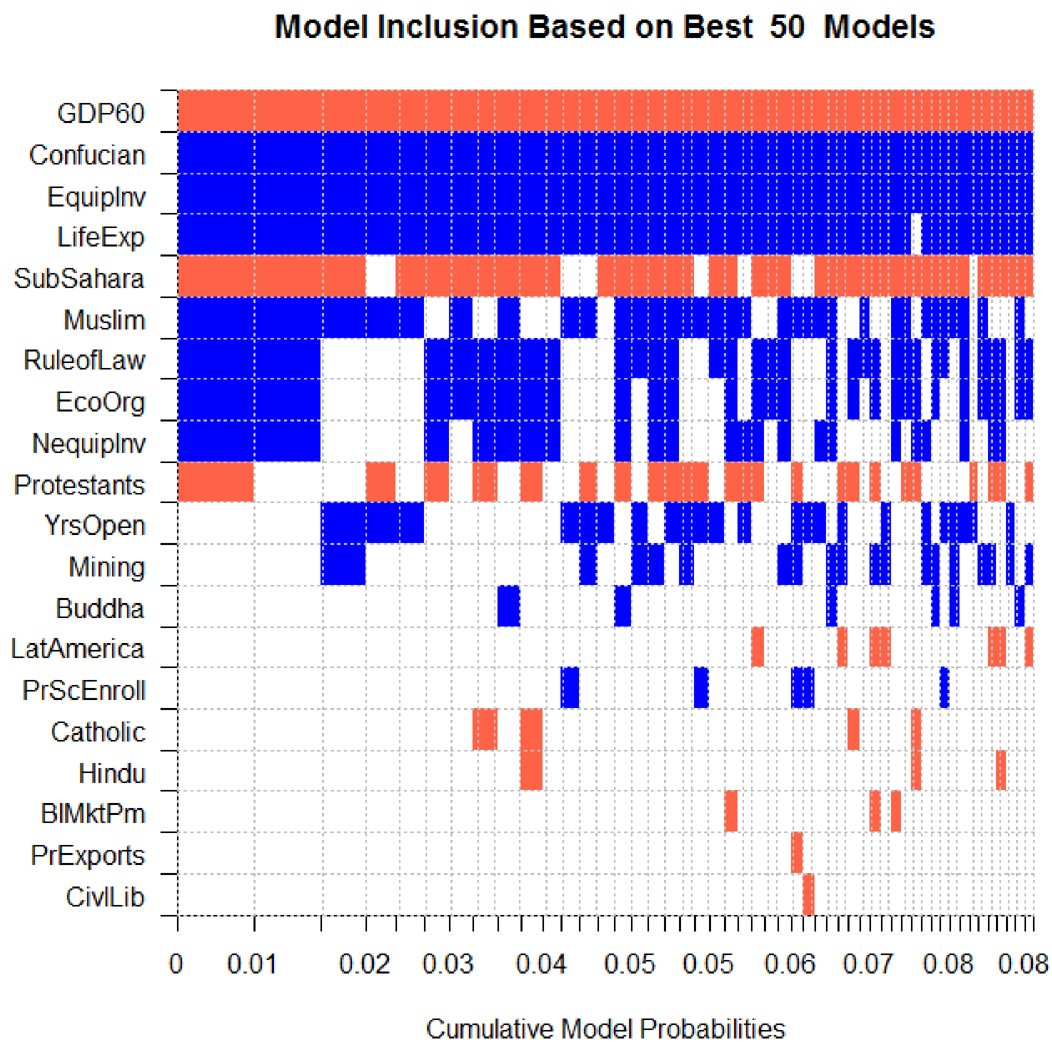
Some R code

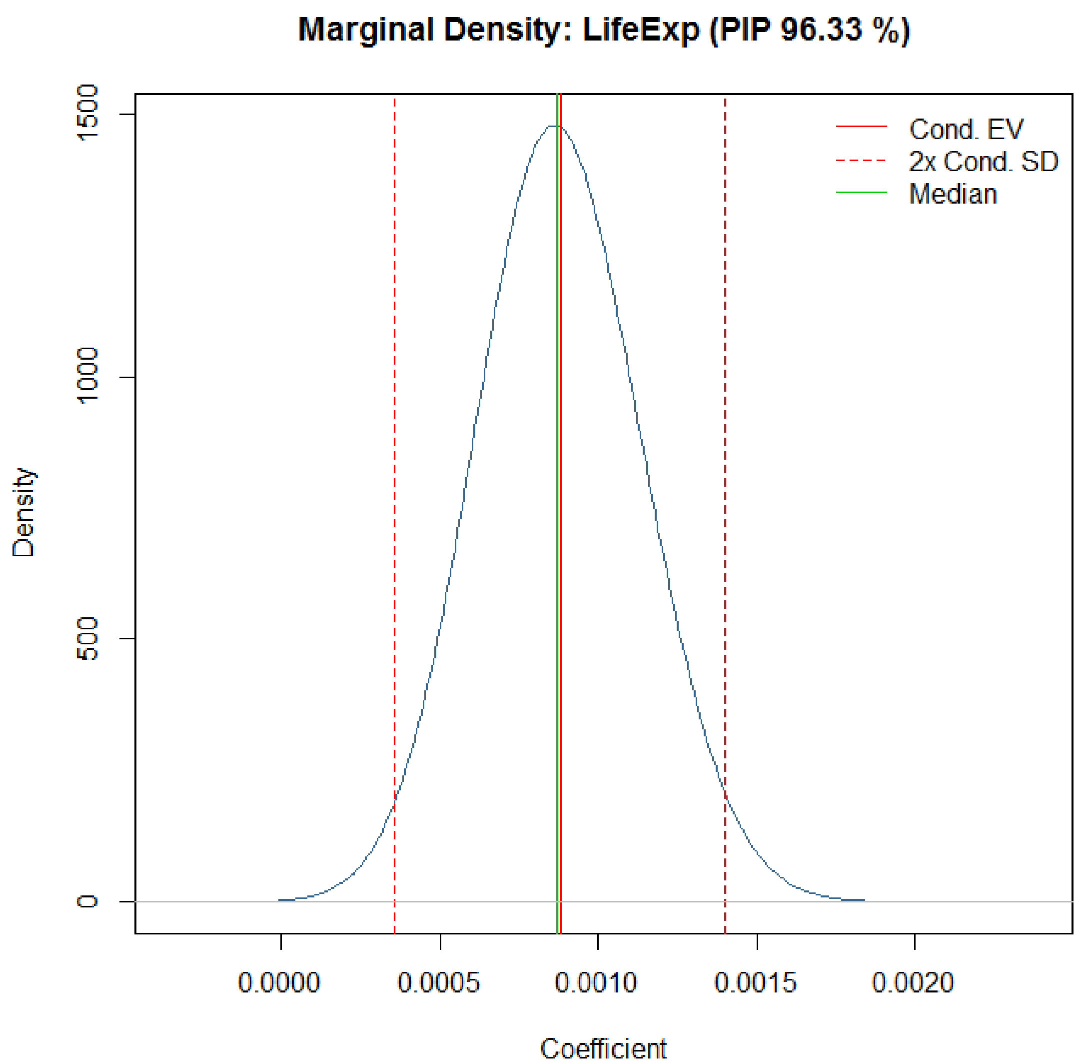
```
library(BMS)

data(datafls)

set.seed(12345)

# Total number of models = 2^41 = 2.2*10^12

#  "PIP" denotes "Prior Inclusion Probability"

# NOTE: With 1,000,000 drawings, the next line will take approximately 1.5
minutes to execute

#######################################################################

growth <- bms(datafls, burn = 50000, iter = 1e+06, g = "BRIC", mprior =
"uniform", nmodel = 2000, mcmc = "bd", user.int = F)

growth
```

	PIP	Post Mean	Post SD	Cond.Pos.Sign	Idx
GDP60	0.999312	-1.615045e-02	3.124842e-03	0.00000000	12
Confucian	0.989107	5.647854e-02	1.457507e-02	1.00000000	19
EquipInv	0.930196	1.611110e-01	6.741018e-02	1.00000000	38
LifeExp	0.928943	8.345725e-04	3.462344e-04	1.00000000	11
SubSahara	0.728207	-1.140227e-02	8.508603e-03	0.00000000	7
Muslim	0.655289	9.023585e-03	7.775526e-03	0.99894093	23
YrsOpen	0.512421	7.303333e-03	7.983056e-03	0.99988486	15
RuleofLaw	0.498742	7.400353e-03	8.320895e-03	1.00000000	26
EcoOrg	0.458943	1.191827e-03	1.439255e-03	0.99994771	14
Mining	0.453810	1.865991e-02	2.320524e-02	1.00000000	13
Protestants	0.443457	-5.531704e-03	7.006628e-03	0.00000000	25
NequipInv	0.434420	2.479481e-02	3.173819e-02	1.00000000	39
PrScEnroll	0.220389	4.594347e-03	9.956695e-03	0.99091606	10
LatAmerica	0.205092	-1.739452e-03	4.123220e-03	0.05577497	6
Buddha	0.194909	2.560094e-03	5.936916e-03	0.99990252	17
BlMktPm	0.180562	-1.372529e-03	3.329068e-03	0.00036553	41
CivlLib	0.133006	-3.042498e-04	9.129109e-04	0.00580425	34
Catholic	0.125844	-2.247763e-04	3.044196e-03	0.40133022	18
Hindu	0.125245	-3.436461e-03	1.196617e-02	0.05214579	21
PrExports	0.099785	-9.835269e-04	3.550193e-03	0.00412888	24
PolRights	0.090776	-1.440864e-04	5.603911e-04	0.01880453	33
Age	0.082319	-3.795793e-06	1.537440e-05	0.00041303	16
RFEXDist	0.081073	-4.136241e-06	1.724494e-05	0.02954128	37
LabForce	0.074980	7.426586e-09	4.015423e-08	0.84526540	29
WarDummy	0.071575	-2.706038e-04	1.205524e-03	0.00292001	5
Foreign	0.068527	2.827676e-04	1.411898e-03	0.92402994	36
English	0.068469	-4.336655e-04	1.985764e-03	0.00074486	35
EthnoL	0.060665	3.511486e-04	1.924021e-03	0.94166323	20
Spanish	0.059576	2.525208e-04	1.619391e-03	0.85871828	2
stdBMP	0.048867	-6.083865e-07	3.851397e-06	0.04107066	40
French	0.048807	1.937706e-04	1.178806e-03	0.97615096	3
HighEnroll	0.047371	-1.682002e-03	1.138897e-02	0.03696354	30
WorkPop	0.044402	-3.043240e-04	2.366836e-03	0.15125445	28
Abslat	0.043675	1.285909e-06	3.302464e-05	0.55532914	1
OutwarOr	0.037489	-6.814124e-05	5.830504e-04	0.08946624	8
Popg	0.036089	5.098816e-03	4.696447e-02	0.87323007	27
Jewish	0.034901	-2.435773e-04	2.817309e-03	0.19838973	22
Brit	0.034349	-6.146858e-05	6.223336e-04	0.13595738	4
RevnCoup	0.032384	-5.458105e-06	1.012380e-03	0.49814723	32
PublEdupct	0.031817	6.523282e-04	2.540805e-02	0.52943395	31
Area	0.029769	-4.484178e-09	1.006330e-07	0.29335886	9

Mean no. regressors	Draws	Burnins	Time	No. models visited
"10.4456"	"1e+06"	"50000"	"1.518669 mins"	"182877"
Modelspace 2^K	% visited	% Topmodels	Corr PMP	No. Obs.
"2.2e+12"	"8.3e-06"	"38"	"0.9883"	"72"
Model Prior	g-Prior	Shrinkage-Stats		
"uniform / 20.5"	"BRIC"	"Av=0.9994"		

```
> # In the next plot, BLUE corresponds to a +ve coefficient,
> # and RED corresponds to a -ve coefficient.
> # WHITE implies non-inclusion in the model:
> image(growth[1:50])
```



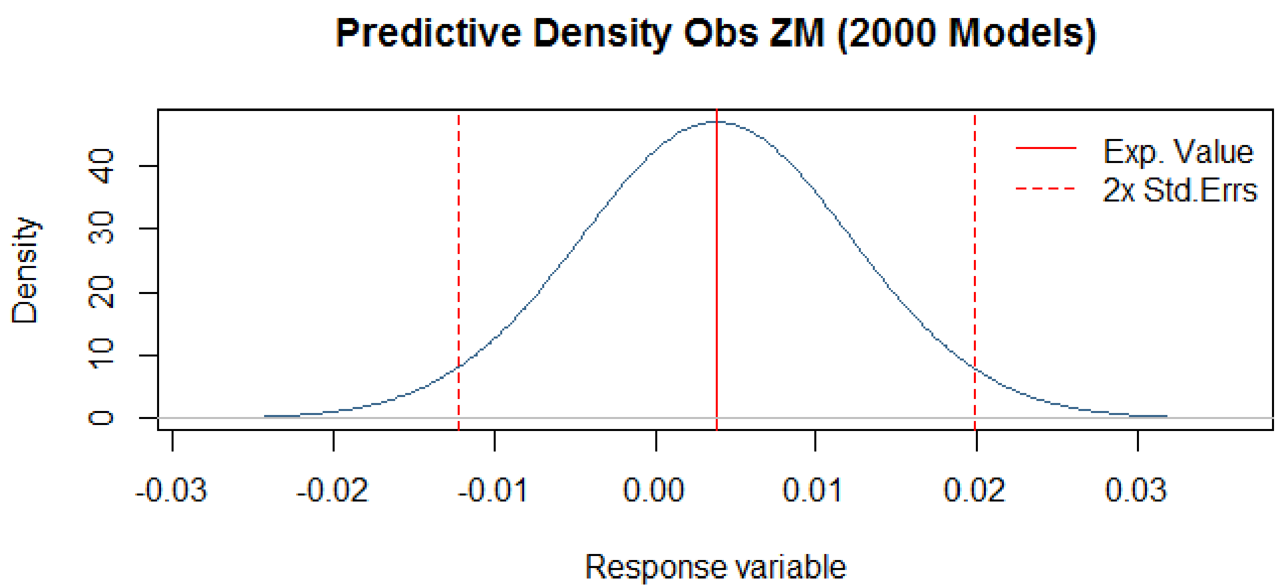


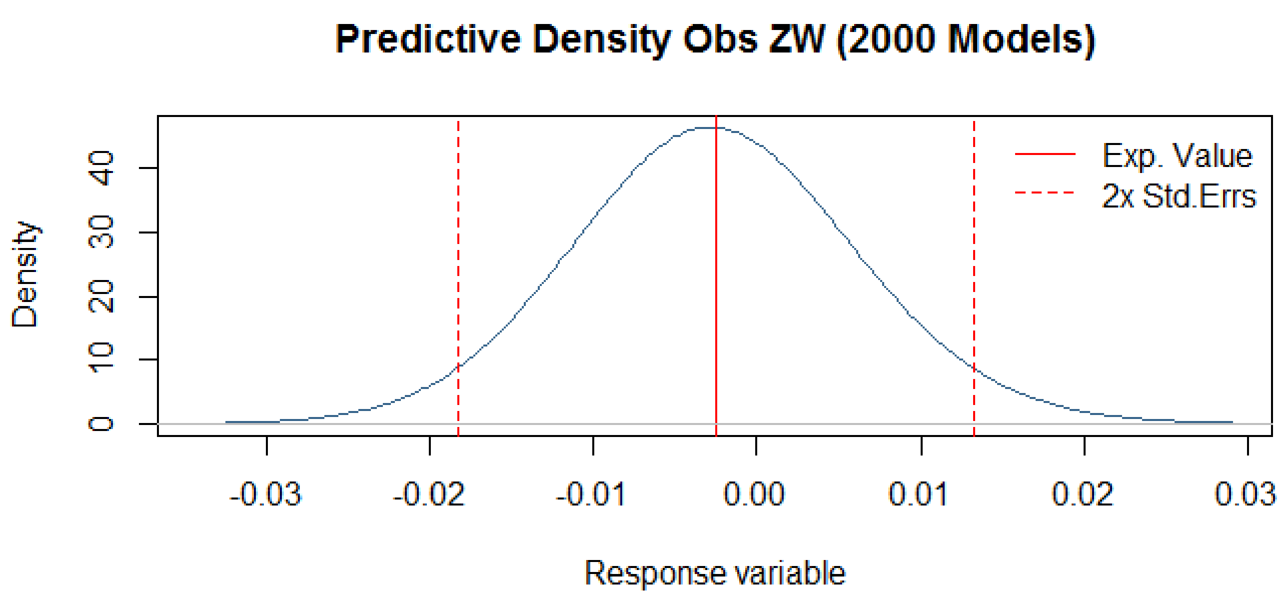

```
> # (Weighted) posterior density for coefficient of just one important regressor
> density(growth, reg="LifeExp")
```

```
> # Predictive densities for the U.K. and the U.S.
> pdens = pred.density(growth, newdata = datafls[66:67, ])
> pdens
Call:
pred.density(growth, newdata = datafls[66:67, ])

Densities for conditional forecast(s)
300 data points, based on 2000 models;
     Exp.Val.    Std.Err.
UK 0.01929568 0.007780345
US 0.01698227 0.007820930
> quantile(pdens, c(0.05,0.50, 0.95))
            5%        50%        95%
UK 0.005250805 0.01924536 0.03355155
US 0.003771534 0.01691727 0.03044644
> par(mfrow=c(2,1))
> plot(pdens,1)
> plot(pdens,2)
> par(mfrow=c(1,1))
```

Posterior Means and Std. Deviations

Medians


Predictive Density Obs ZM (2000 Models)
Exp. Value
2x Std.Errs
Density
0
10
20
30
40
-0.03
-0.02
-0.01
0.00
0.01
0.02
0.03
Response variable


Predictive Density Obs ZW (2000 Models)
Exp. Value
2x Std.Errs
Density
0
10
20
30
40
-0.03
-0.02
-0.01
0.00
0.01
0.02
0.03
Response variable