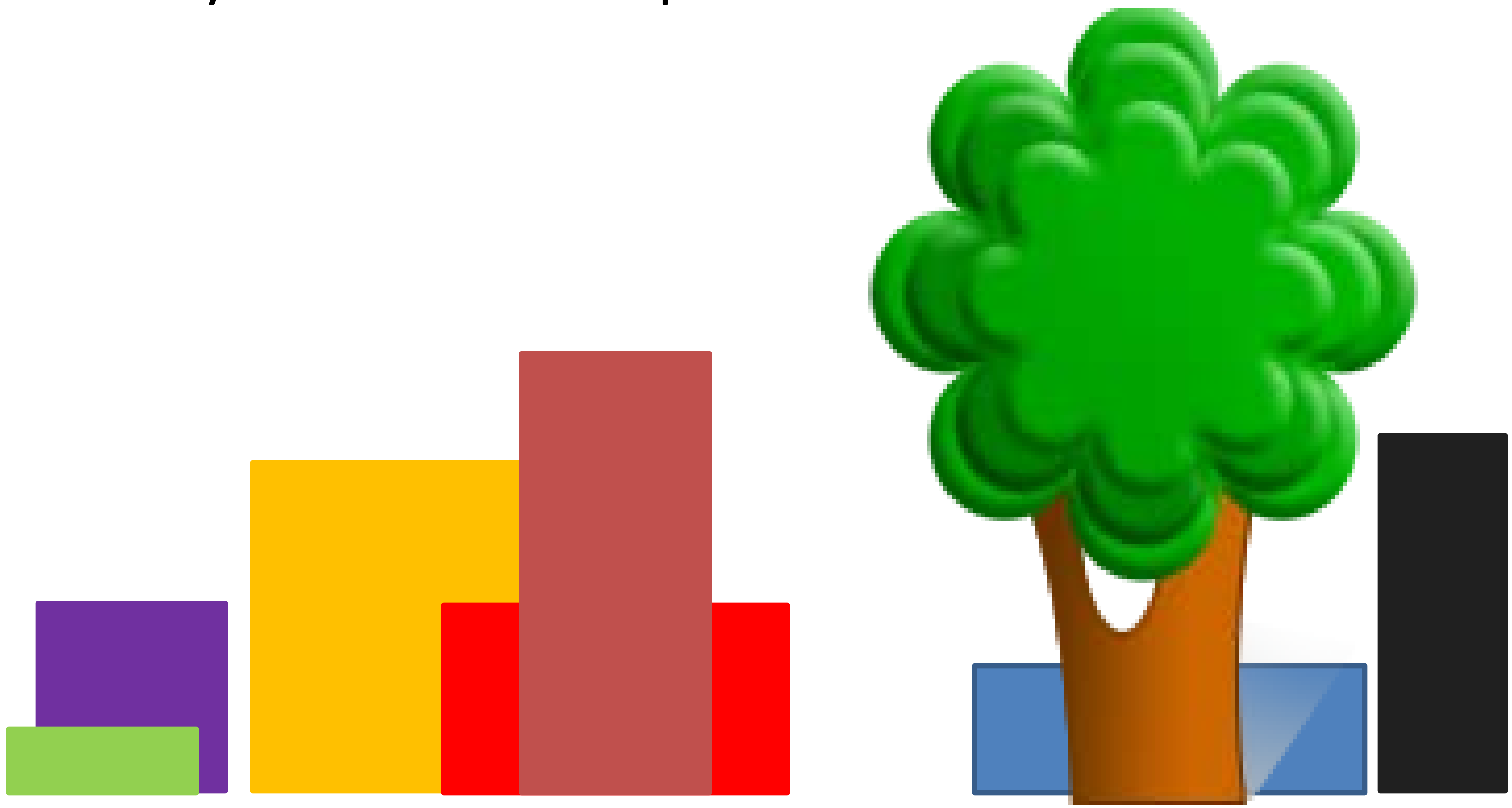
Dealing with Model Uncertainty by Model Averaging

George Kypriotakis
Ph.D. Candidate

May 13, 2011

How many boxes are in the picture?



Introduction

- Problem: Model uncertainty in health-related research
- Objective: Introduction to model averaging

Presentation Outline

- Model uncertainty
- Quick Bayesian review
- Bayesian model averaging
- Application of model averaging

Different scenarios for the price of bonds of the Greek government in 2014

Model	Price	Probability	Action
1. Default	-	.3	
2. Recover	+	.2	
3. Separated from EU	-	.09	
4. Borrow more	-	.4	
5. 1 + 3	- * 2	.002	
…			
…			

The Problem

- Suppose we want to investigate potential determinants of a dependent variable y given the (potentially large) set of regressors X

- Suppose that there are K potential regressors and the model space M is the set of all 2^K linear models.

Question: How do we proceed to determine the unknown quantities give a model space M?

The Common Practice of Model Choice

- First, choose a model
- Then, modify the model based on the data
- And finally, make inferences/predictions based on the final model as if the final model is the true model
- This is logically inconsistent!
- It leads to bad calibration (how often do I get the right answer?) and results in anticonservatism (uncertainty bands were not wide enough)

Implications of Selecting a Single Model

- What do you do about competing models? i.e. models with different specifications but with similar or different results
- Too risky to base all inferences on a single model. Especially when the stakes are high
- Model Uncertainty should be reflected in inferences/predictions

Model Uncertainty

I am not only uncertain about Δ (a quantity of interest), but also which model to choose.

Two sources of uncertainty:

1) I am uncertain about Δ

and

2) **I am uncertain about how to quantify my uncertainty about Δ**

This is model uncertainty

Three types of model uncertainty

- Theory Uncertainty: Uncertain what explanatory variables to include

- Specification Uncertainty: homogeneity, linearity

- Data uncertainty: error in measurement, outliers

Another View of Uncertainty

Donald Rumsfeld: "Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns -- the ones we don't know we don't know."

Uncertainty is a state of incomplete information about something of interest

Typically we are interested in some quantity Δ that we are at least partially uncertain

We want to quantify that uncertainty to be able to:

1) Draw inferences $p(\Delta | \text{Data, Assumptions/Judgments})$

2) Make predictions $p(\text{Data} | \Delta, \text{Assumptions/Judgments})$

3) Make decisions $U(\text{action}, \Delta, \text{Data})$

Ignoring Model Uncertainty may lead to:

1) Biased estimates

2) Overconfident standard errors

3) Misleading inference and prediction

4) Wrong decision/action

More on Model Choice

So, in problems with realistic complexity we have to consider a set of Models {$M_{1,}$ M_2 … M_k)

Is M_1 better than M_2 ?

But
Better for what???

Better in getting the right answer when try to predict.
Good models make good predictions.
Bad models make bad predictions.

All statistics is (or should be) basically about prediction

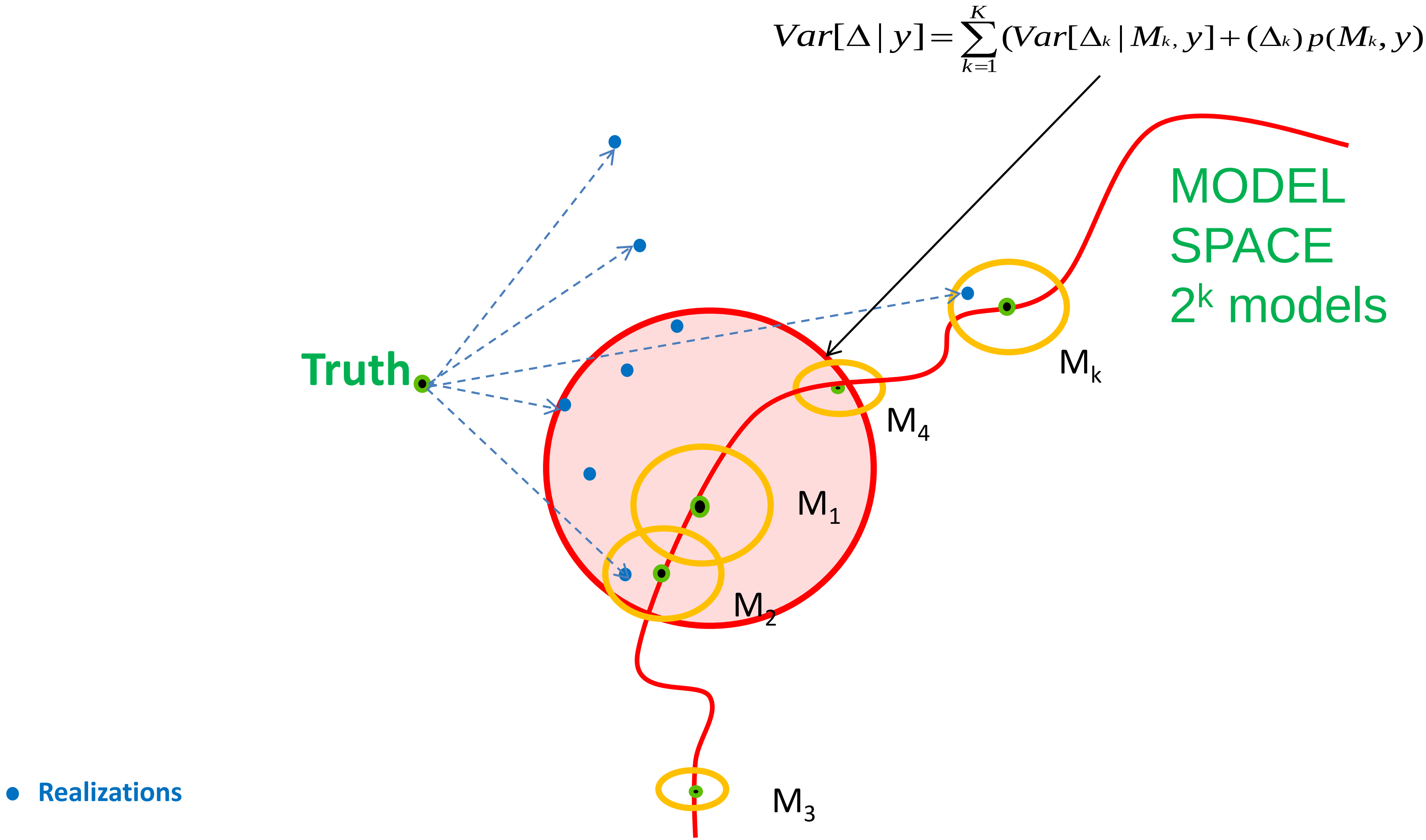
Model Averaging

Model averaging fully addresses model uncertainty (model choice) by *integrating out* uncertainty from the distribution of parameters of interest.

- Combine information from different models with relative (posterior) weights

- Both Bayesian and Frequentist approaches

Accuracy, Calibration and Model Expansion



Widening the bands or missing the "truth"?

Interpretation of Model Space

- M-closed view:

  the true model is unknown but included in the model space

- M-open view:

  there are no true models under consideration

A Bayesian Exercise on Mammography

Facts:

- 1% of scanned women have breast cancer
- 80% of women with breast cancer get positive mammography
- 9.6% of women without breast cancer also get positive mammography

Question:

A woman has a positive mammography. What is the probability that she has breast cancer?

Quick guess:

a) =< 1%
b) Between 1% and 60%
c) Between 60% and 80%
d) > 80%

from: *http://yudkowsky.net/bayes/bayes.html*

Solving the Exercise

Write down the probabilities:

- Define C+ = cancer, C- = no cancer
- Define M+ = positive mammogram, M- = negative mammogram
- The prior probability of cancer for scanned women is $p(C+) = 1\%$

- If there is cancer, the probability of a positive mammogram is $p(M+|C+) = 80\%$
- If there is no cancer, we have $p(M+|C-)=9.6\%$

Question:

What is the probability of cancer given a positive mammogram?

$$P(C+|M+) = ?$$

Solving the Exercise

Consider 10,000 screened subjects

- p(C+)=1%, 100 out of 10,000 have cancer, of which
  - p(M+|C+)=80%, 80 out of 100 get a positive mammogram
  - 20 out of 100 get a negative mammogram

- p(C+)=1%, 9,900 out of 10,000 do not have cancer, of which
  - p(M+|C-=9.6%, 950 get a positive mammogram
  - 8,950 get a negative mammogram

In a table:

	M+	M-
C+	80	20
C-	950	8,950

Solving the Exercise

- **Use Bayes' Rule**

	M+	M-
C+	80	20
C-	950	8950

- Marginal: total number of positive mammographies p(M+) = p(C+,M+) + p(C-,M+)
- $P(C+|M+) = \dfrac{p(C+,M+)}{p(M+)} = \dfrac{80}{80 + 950} = 7.8\%$

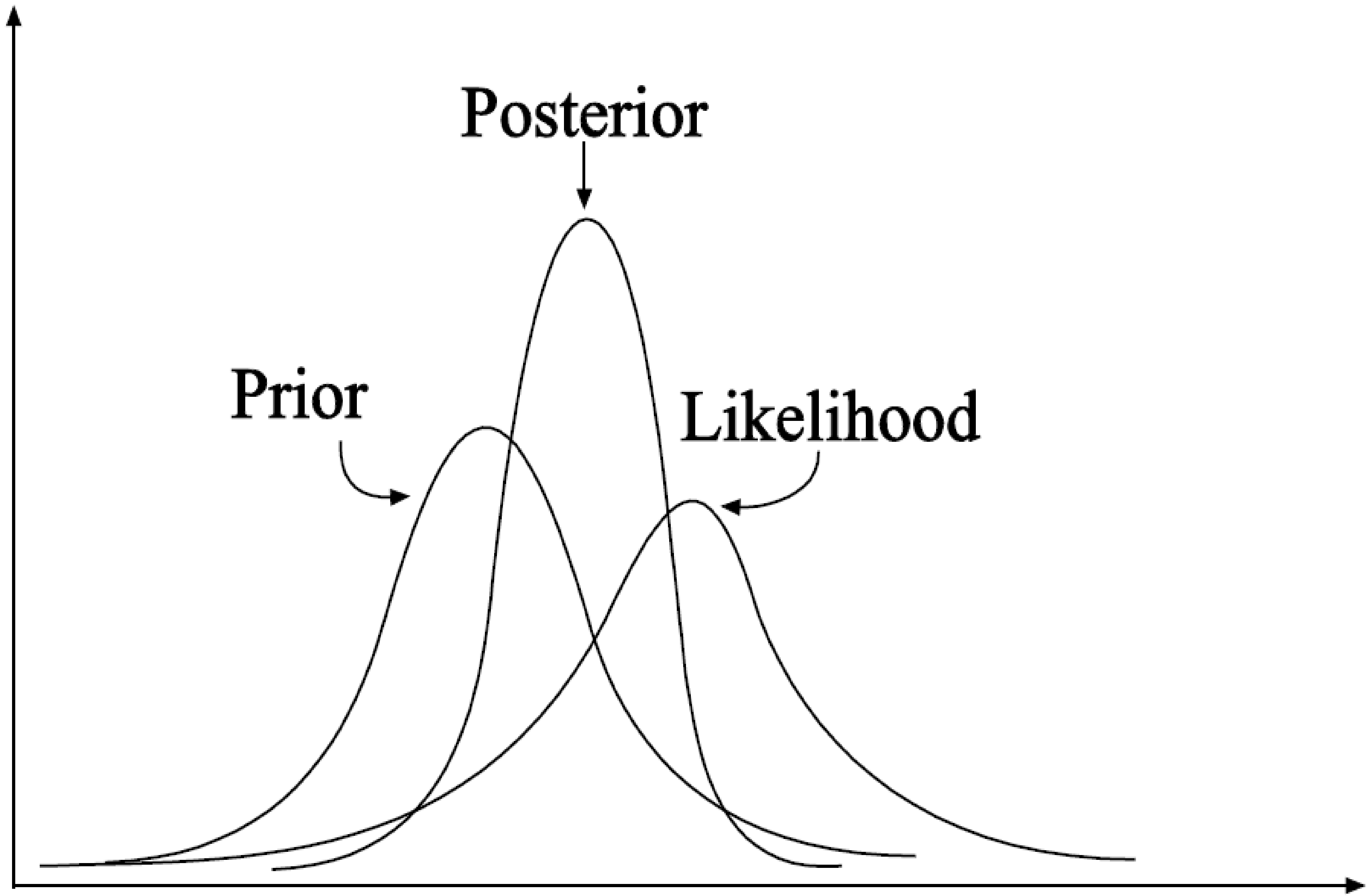
Bayes' Rule

Posterior, Likelihood, Prior

$$\underbrace{P(C+|M+)}_{\text{Posterior}} = \frac{\overbrace{p(M+|C+)}^{\text{Likelihood}}\ \overbrace{p(C+)}^{\text{Prior}}}{P(M+)}$$

http://yudkowsky.net/bayes/bayes.html

Prior, Likelihood and Posterior of a Quantity of Interest



Model Averaging

Averaging over a set of models provides better predictive ability than using a single model

Some difficulties:

- Set of Models can be enormous
- Specification of priors on models and parameters
- Integration in many cases very hard, much easier with MCMC
- Determination of a subset of modes to average over

Some Technical Stuff

The distribution of a quantity Δ given model M, is the posterior of Δ under that model

$$p(\Delta | D, M) = \frac{p(D | \Delta, M)\, p(\Delta | M)}{p(D | M)}$$

More Technical Stuff

The model averaged distribution of a quantity Δ of interest

$$p(\Delta \mid D) = \sum_{k=1}^{K} p(M_k \mid D)\, p(\Delta \mid M_k, D)$$

$$p(M_k \mid D) = \frac{p(D \mid M_k)\, p(M_k)}{\sum_{k=1}^{K} p(D \mid M_k)\, p(M_k)}$$

The posterior probability of model *k* , giver the set of models in the analysis is:

Predictive Distribution

The predictive distribution of y* (new y) given D is:

$$p(y^*|D) = \sum_{k=1}^{K} p(y^*|M_k, D)\, p(M_k, D)$$

This is bayesian model averaging: the composite predictive distribution p(y*|D) is a weighted average of the conditional predictive distributions $p(y^*|M_{i,}D)$ given the structural possibilities, weighted by the posterior probabilities $p(M_i\,|D)$ of each of the structural choices.

Steps in Bayesian Model Averaging

- Assign prior distributions to models and parameters
- Search the model space for data-supported models
- Compute posterior model probabilities (prior * likelihood)
- Obtain estimates and standard deviations for parameters
- Perform and evaluate prediction

Data Description

Data on Potential Predictors of Cancer-Related Quality of Life of 340 patients. (Copying and Communication Support Intervention Study (P.I. Julia H. Rose))

Outcome : Fact-G (a cancer related scale composed of physical well-being, emotional well-being, functional well-being and social well-being)

Measurement: average 8 wks after diagnosis of advanced cancer

Balanced data, i.e. missing values were deleted (34 regressors)

Final model space M is huge: $2^{34} = 1.7e+10$

Results Based on Best 3000 Models

VAR	PIP	Δ|data	sd|data	Sign
depress	1.00	-0.96	0.09	0.000
siminoff	1.00	-0.63	0.10	0.000
iadl	1.00	-1.66	0.23	0.000
age	0.99	0.25	0.06	1.000
religb	0.66	0.90	0.78	1.000
mon	0.42	-0.29	0.40	0.000
cronarc	0.42	-1.23	1.67	0.000
utilbin	0.39	-1.19	1.74	0.000
explain	0.25	0.13	0.28	1.000 ...
lwa	0.06	0.04	0.36	0.999
blunt	0.06	0.00	0.10	0.936
acpb	0.06	0.02	0.31	0.900
education	0.05	0.00	0.06	0.764

- PIP = Posterior Inclusion Probability
- Δ|data = regression coefficient
- sd|data = standard deviation of regression coefficient
- Sign = proportion of positive coefficient in the first 3000 models

BMA Results Based on Best 3000 Models Compared to Stepwise Method

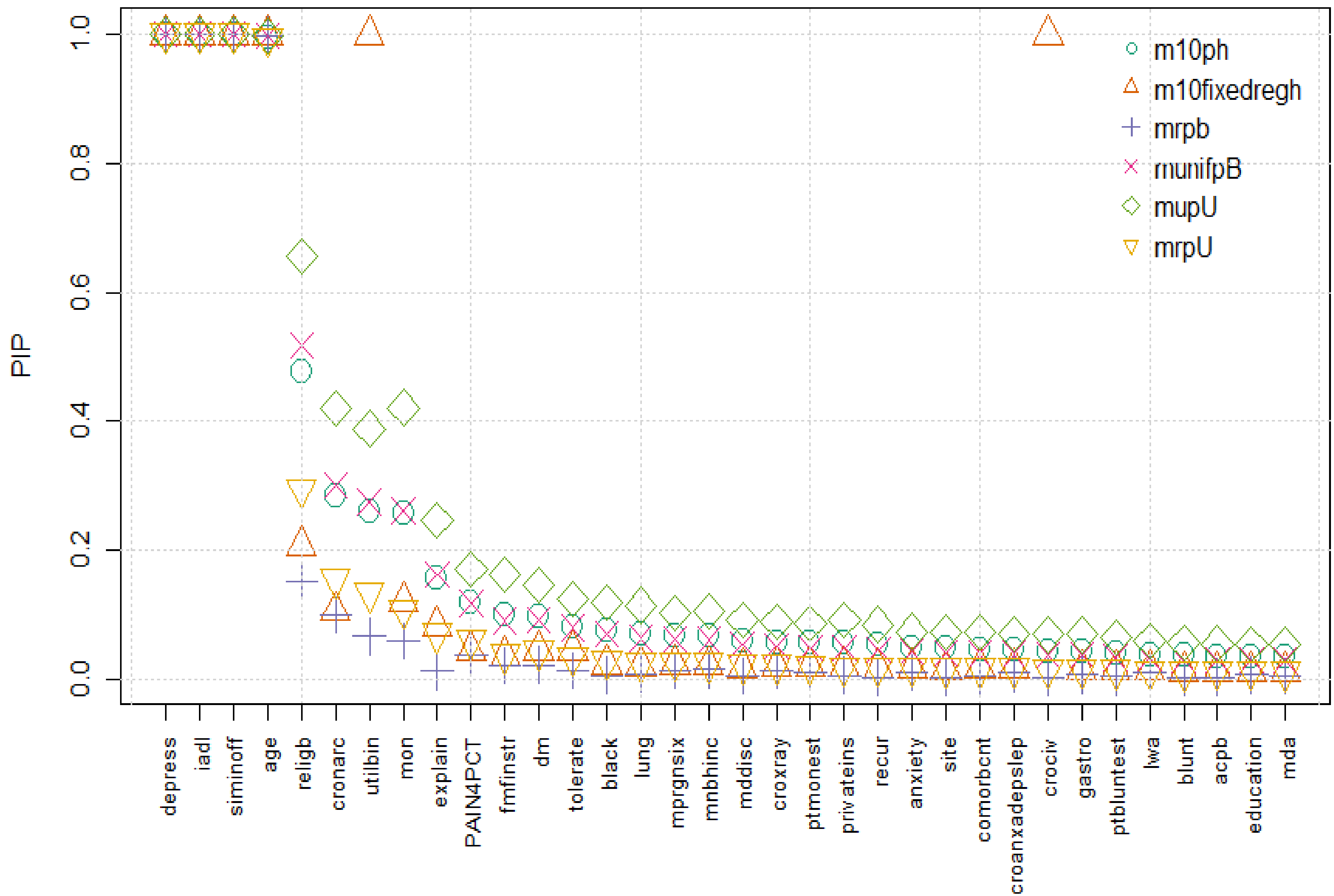
	BMA			Stepwise		
VAR	PIP	Δ	sd	Δ	se	P-Value
depress	1.00	-0.96	0.09	-0.92	0.07	<0.001
siminoff	1.00	-0.63	0.10	-0.58	0.10	<0.001
iadl	1.00	-1.66	0.23	-1.58	0.22	<0.001
age	0.99	0.25	0.06	0.24	0.06	<0.001
religb	0.66	0.90	0.78	1.48	0.51	0.004
mon	0.42	-0.29	0.40	-0.76	0.30	0.013
cronarc	0.42	-1.23	1.67	-2.93	1.27	0.022
utilbin	0.39	-1.19	1.74	-3.02	1.36	0.027
explain	0.25	0.13	0.28	0.52	0.28	0.069
site	0.07	-0.08	0.46	-2.13	1.34	0.112
income	0.11	0.01	0.02	0.05	0.03	0.161
Insur	0.09	-0.14	0.63	-2.58	1.45	0.076
finStr	0.16	-0.07	0.20	-0.46	0.27	0.089
recurren	0.08	-0.12	0.59	-2.18	1.47	0.139

P-Values overstate the evidence of an effect!

Results Based on Covariates Identified in Literature

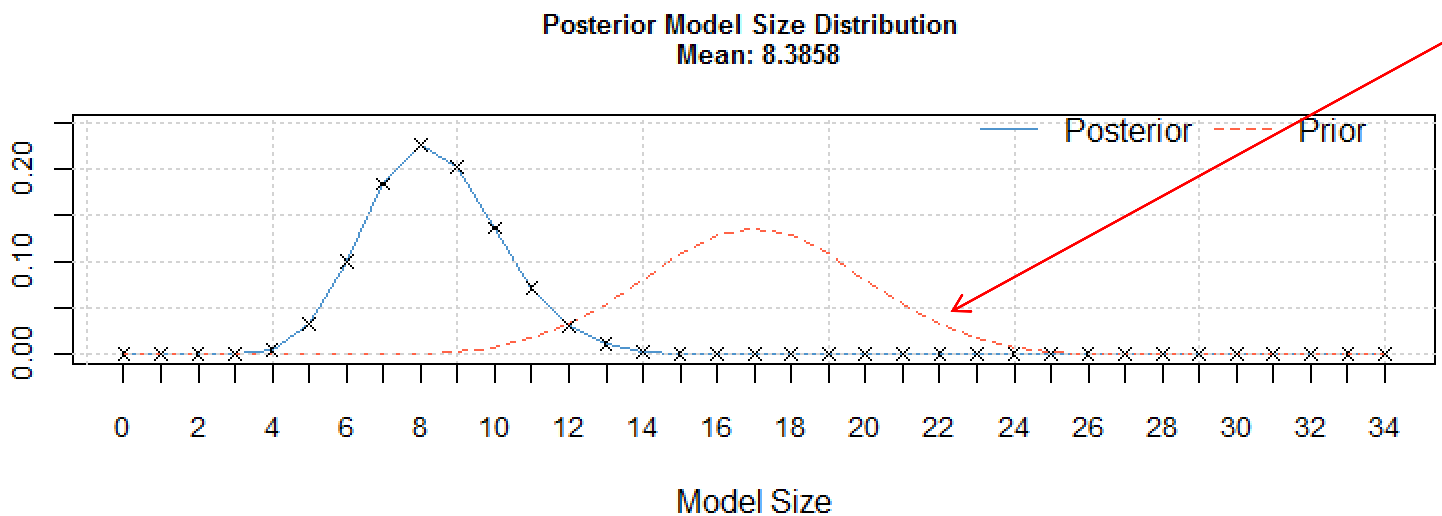
VAR	Δ	se	P-Value	BMA PIP	BMA Δ	BMA sd
age	0.26	0.07	<0.001			
education	0.06	0.27	0.824			
mprgnsix	0.05	0.02	0.055			
depress	-1.20	0.08	<0.001			
crociv	2.51	1.42	0.079			
cronarc	-3.45	1.47	0.019	0.42	-1.23	1.67
comorbcn	-0.46	0.46	0.319			
lung	2.23	1.58	0.159			
gastro	3.25	1.71	0.058			
utilbin	-3.68	1.56	0.019			

Comparison of PIPs based on 6 different model and parameter priors

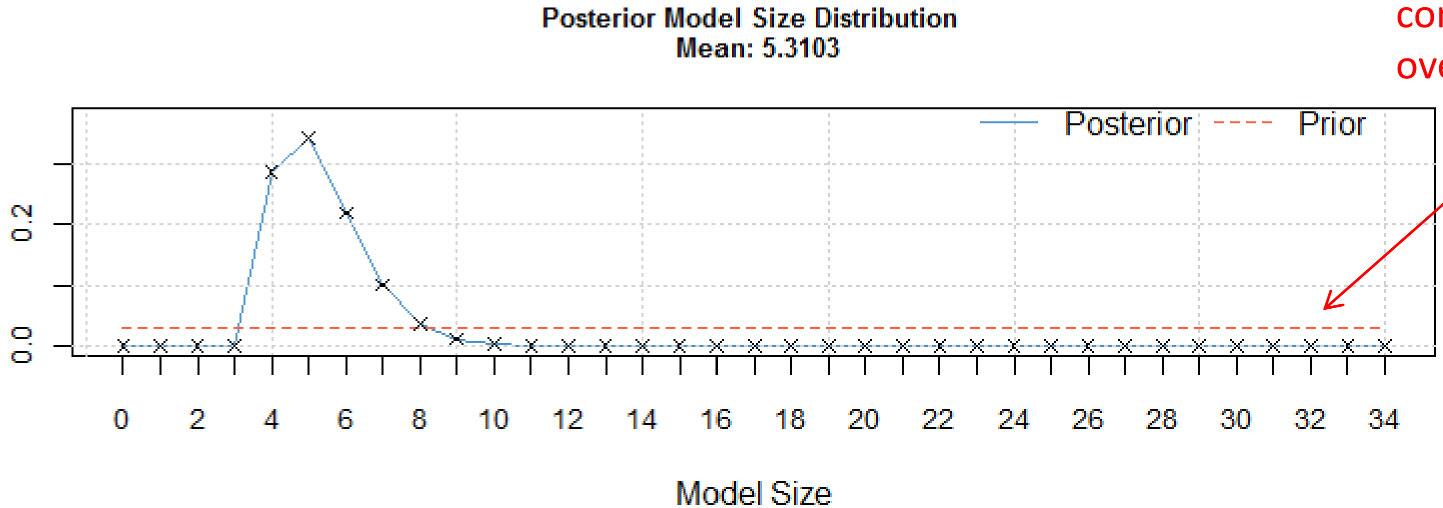


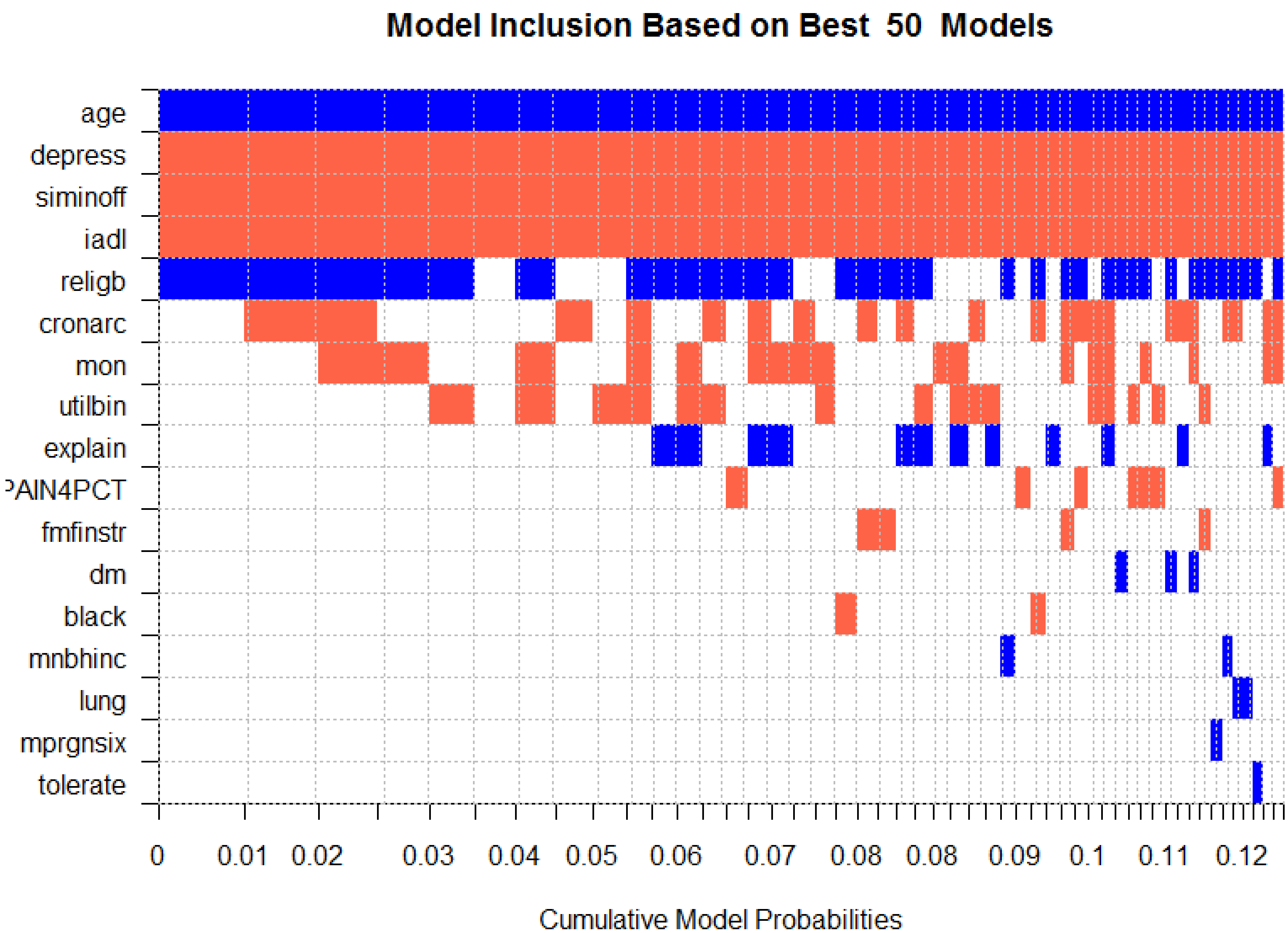
The Effect of Model Priors

The model prior is uniform



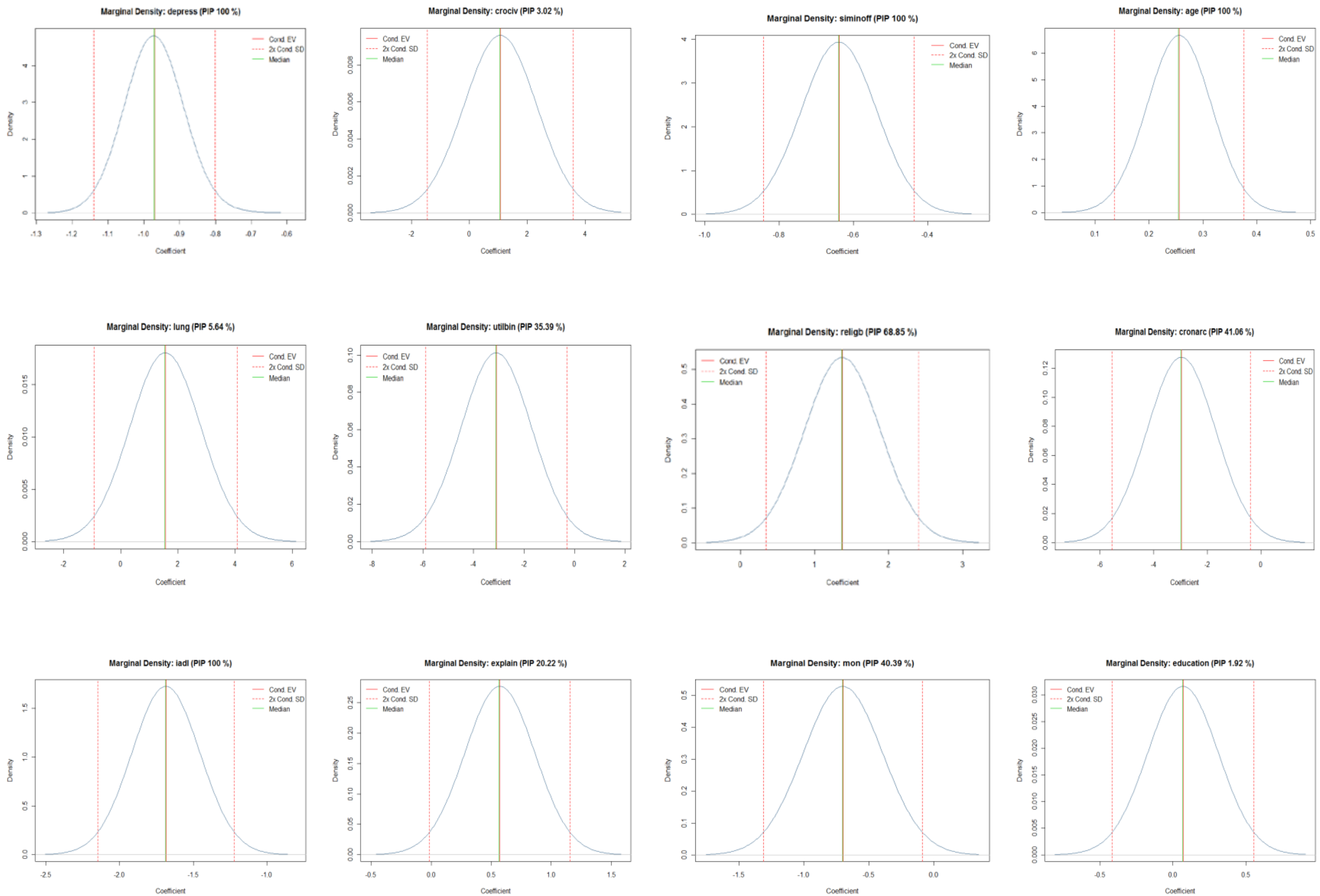
The model prior is completely flat over model sizes



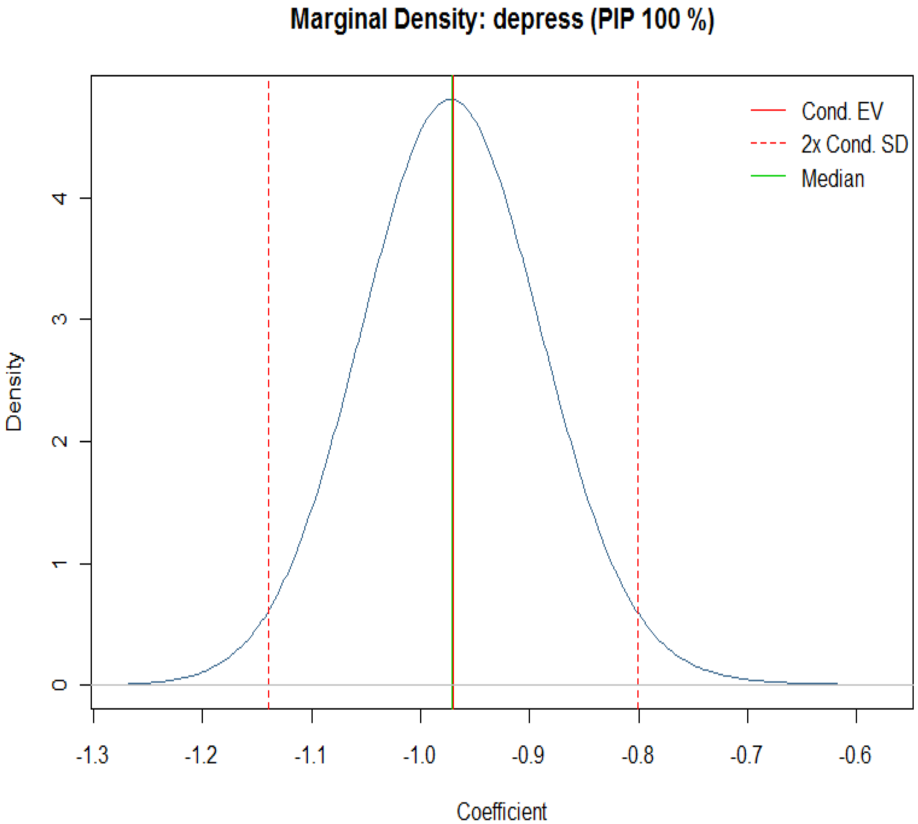


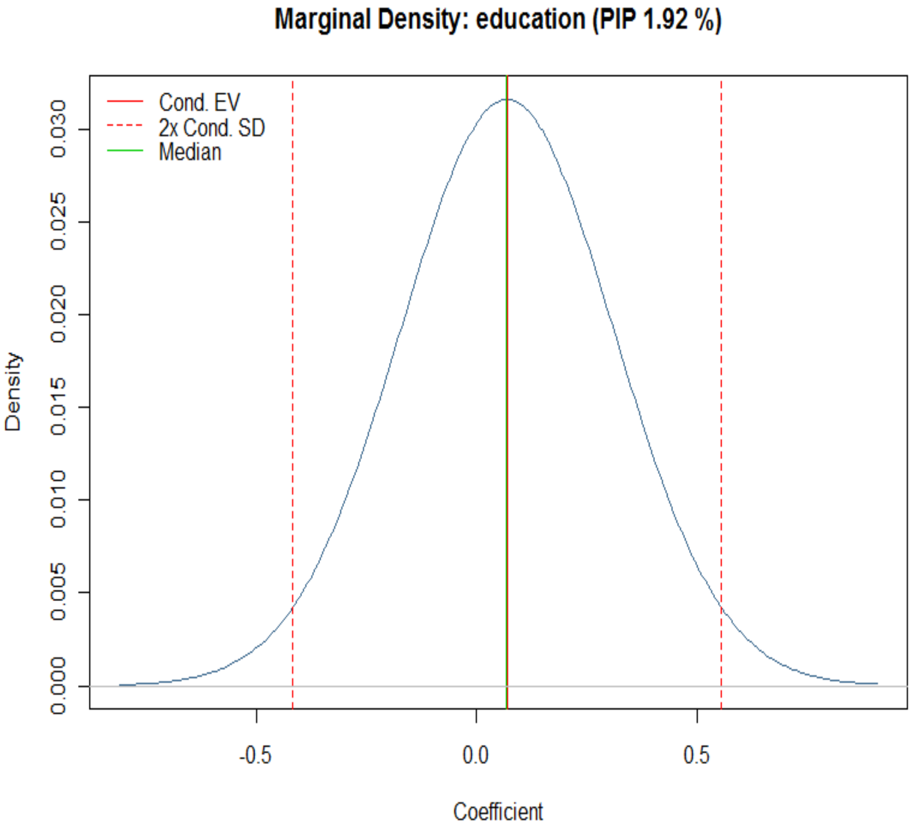
Model Inclusion Based on Best 50 Models
age
depress
siminoff
iadl
religb
cronarc
mon
utilbin
explain
PAIN4PCT
fmfinstr
dm
black
mnbhinc
lung
mprgnsix
tolerate
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.08
0.09
0.1
0.11
0.12
Cumulative Model Probabilities

Posterior Coefficient Distributions



Posterior Coefficient Distributions





So, what?

Is the averaged model any better than the other two?

Assessment of Models using Predictive Performance

- Prediction provides a neutral criterion to compare models

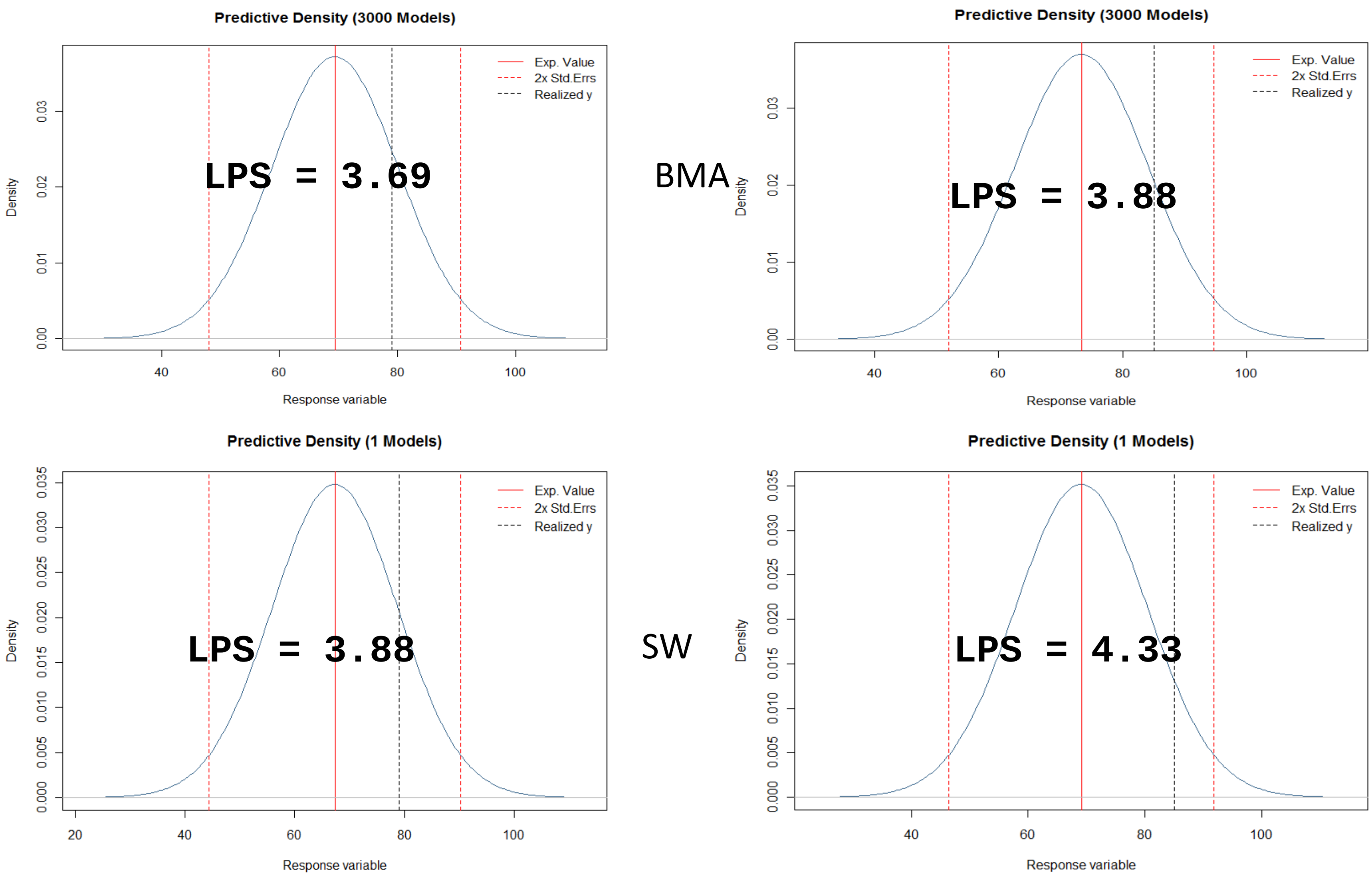
- Use models to predict values of cases not included in the analyses

- Formal evaluation of models using two formal scoring rules:
    - The Log Predictive Score (LPS)
    - The Continuous Ranked Probability Score (CRPS)

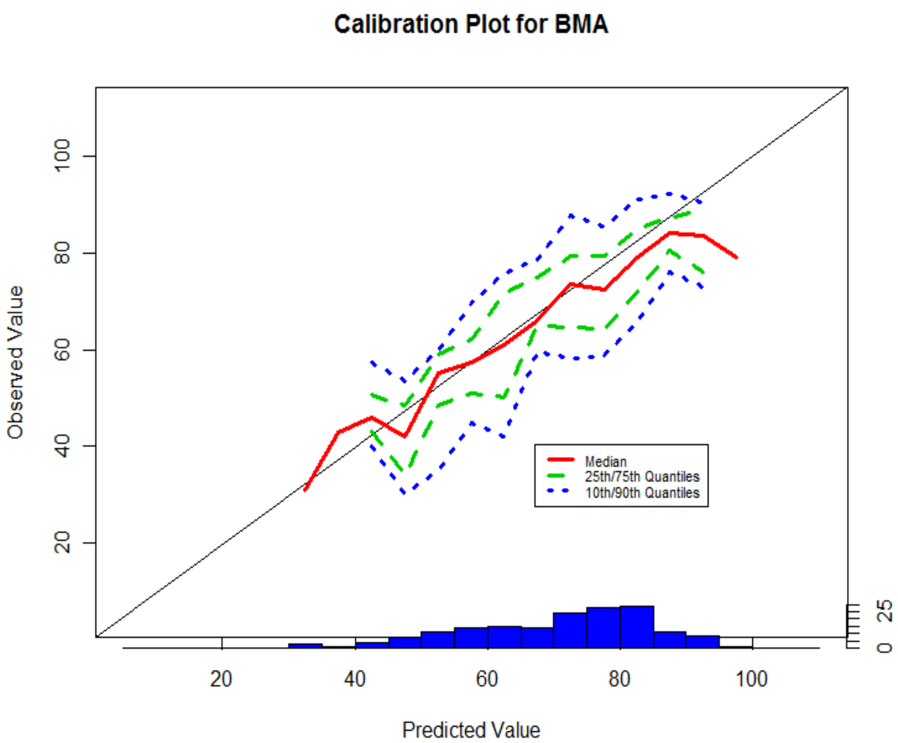
- Split the sample in a training and testing set for evaluation of predictive performance on independent data

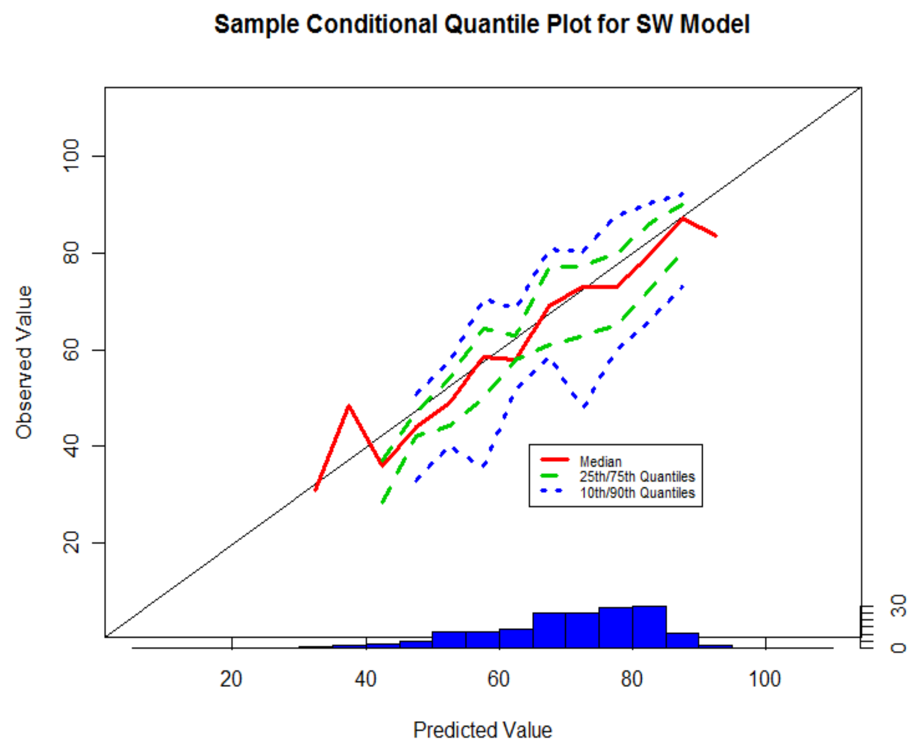
- Graphical exploration of calibration between models

Predictive Densities for two Randomly Selected Subjects



Calibration

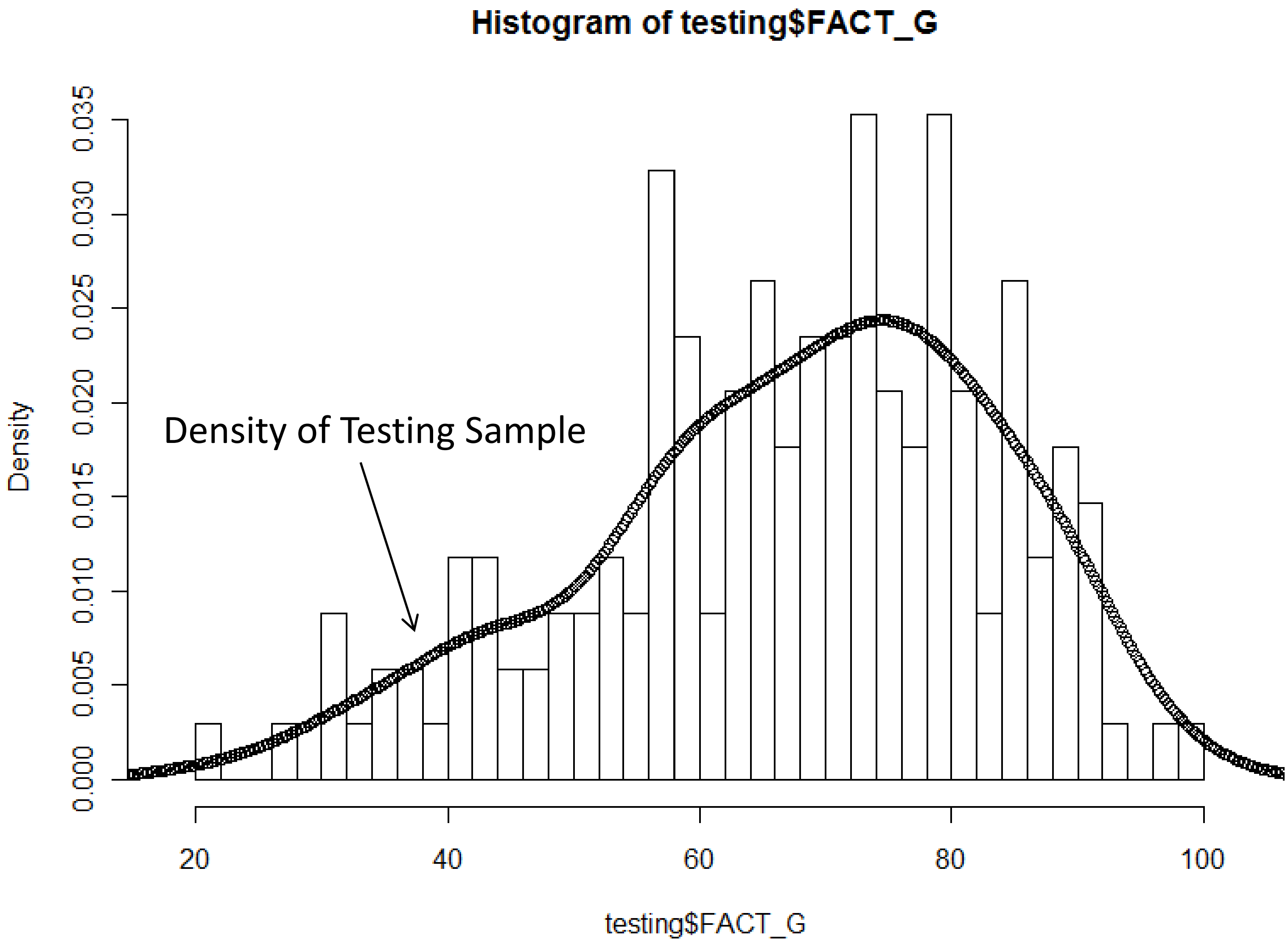




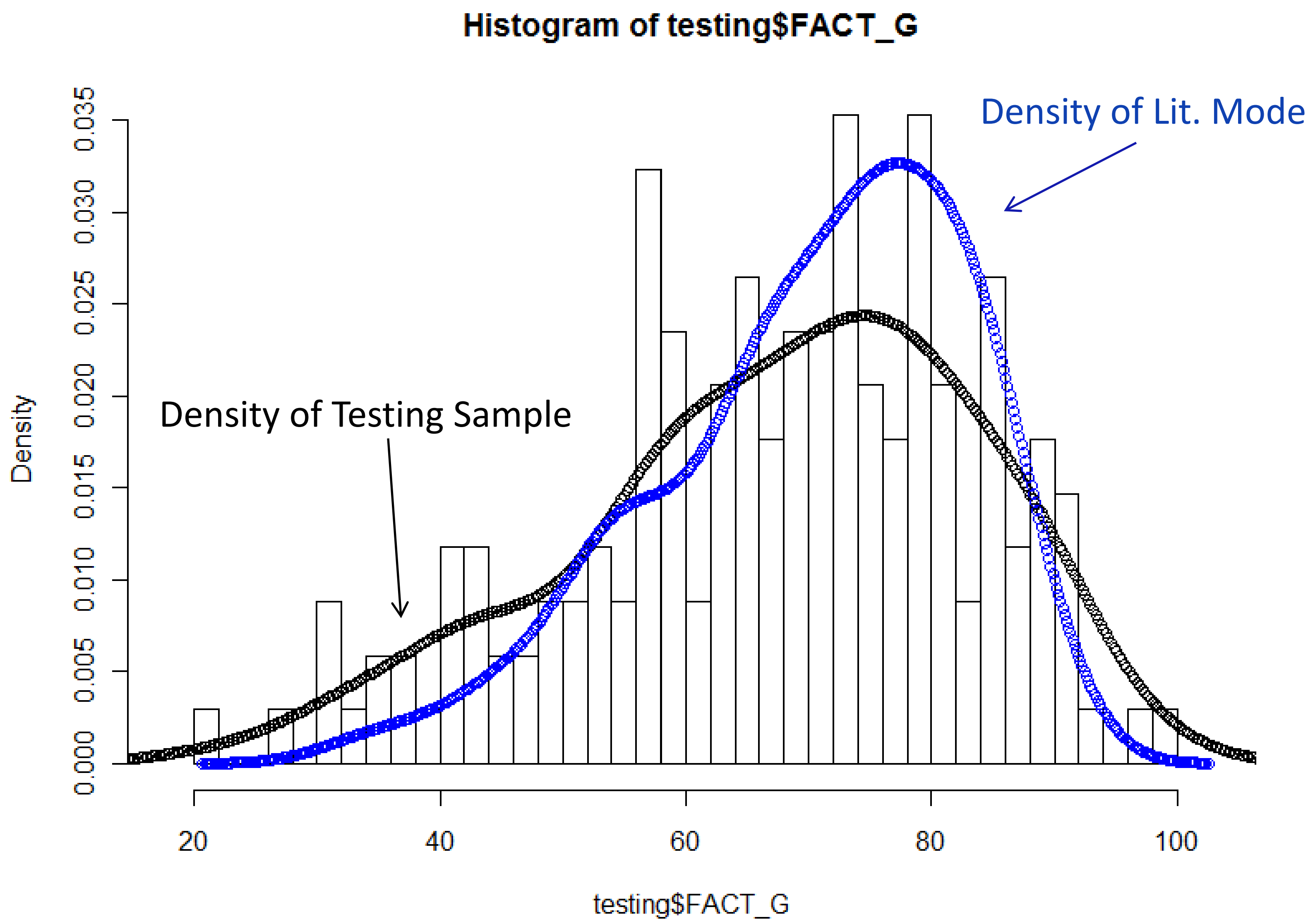
CRPS

[BMA]	**9.17**
[Best Model]	**9.23**
[Step]	**9.19**

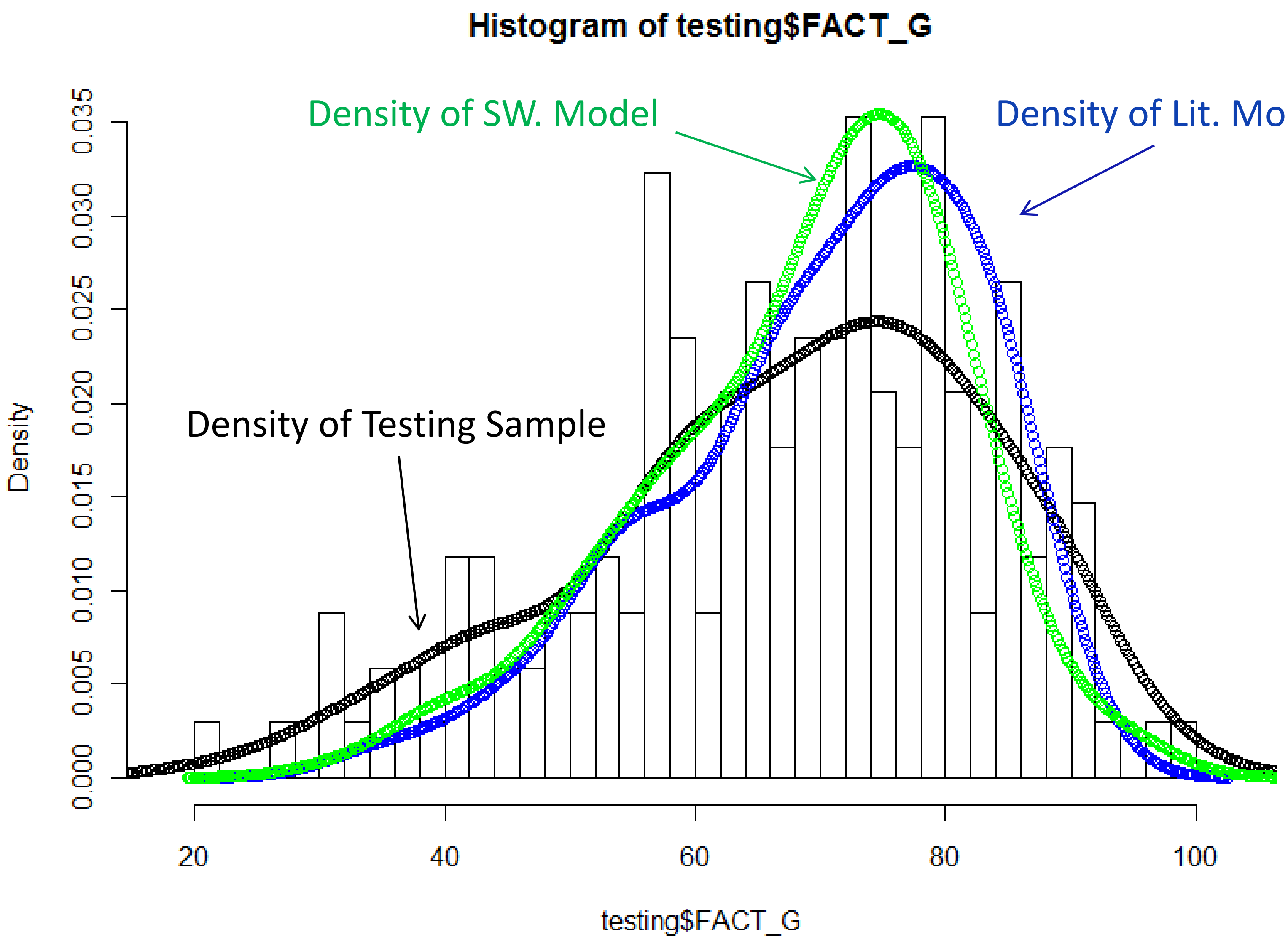
Observed Distribution of Fact-G



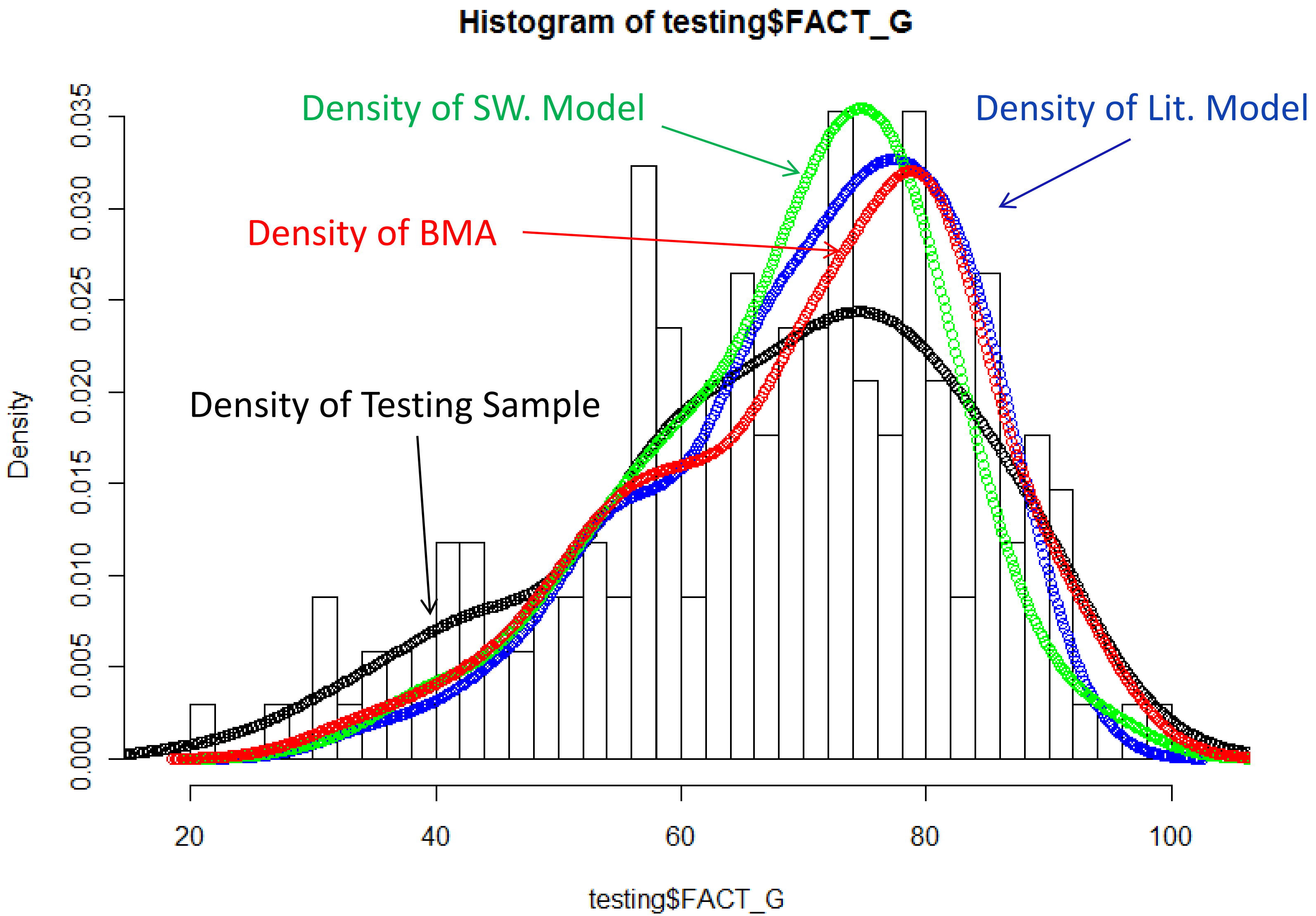
Actual Distribution of Fact-G and Predictive Distributions



Actual Distribution of Fact-G and Predictive Distributions



Actual Distribution of Fact-G and Predictive Distributions



Means SDs and Quantile for Predictive Distributions for Stepwise, Literature, and BMA based on Training Sample

	Mean	SD (var)	Quantile (0.05-0.95)
Stepwise	69.88	12.38 (153.3)	46.43 - 86.27
Literature	70.4	12.42 (154.3)	47.34 - 86.49
BMA	70.59	13.79 (190.1)	45.84 - 90.09

Mean SD and Quantiles of Actual Distribution of the Testing Sample

Fact-G Testing Sample	67.6	16.03	38.58 – 90.01

By Performing Bayesian Model Averaging :

- I have identified robust determinants of Cancer Related Quality of Life
- Have shown that model specification can alter the effects of candidate regressors
- Accounted for uncertainty of model specifications in both parameters and predictions
- And thus, improved predictive performance

"That is what we meant by science. That both question and answer are tied up with uncertainty, and that they are painful. But that there is no way around them. And that you hide nothing; instead, everything is brought out into the open". (Peter Hoeg, 1995)

Thank you