

# Bayesian Variable Selection

Hoff Chapter 9, Mixtures of g-Priors Liang et al JASA

February 28, 2017



## lm summary

```
lm(formula = log(SO2) ~ temp + log(firms) + log(popn) + wind  
    precip + rain, data = usair)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	6.7142760	1.6475086	4.075	0.000261	***
temp	-0.0649495	0.0227711	-2.852	0.007333	**
log(firms)	0.3698588	0.1934076	1.912	0.064289	.
log(popn)	-0.1771293	0.2335520	-0.758	0.453428	
wind	-0.1738606	0.0656713	-2.647	0.012204	*
precip	0.0156032	0.0132718	1.176	0.247893	
rain	0.0009153	0.0057335	0.160	0.874104	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5108 on 34 degrees of freedom

Multiple R-squared: 0.5503, Adjusted R-squared: 0.471

F-statistic: 6.936 on 6 and 34 DF, p-value: 7.12e-05

# Zellner's g-prior

Centered model:

$$\mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}^c \boldsymbol{\beta} + \epsilon$$

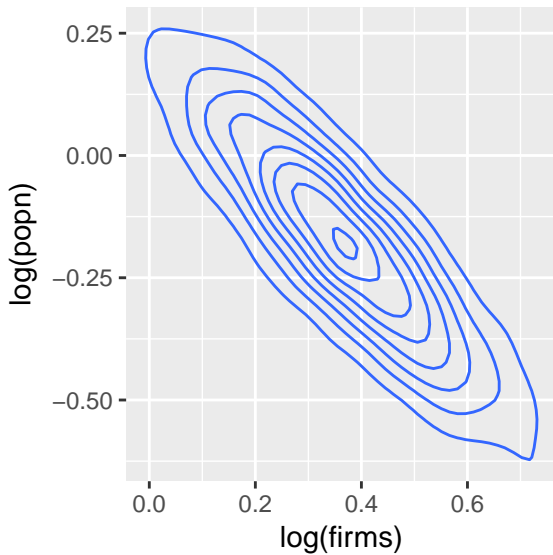
where  $\mathbf{X}^c$  is the centered design matrix where all variables have had their mean subtracted

- ▶  $p(\phi) \propto 1/\phi$
- ▶  $p(\alpha \mid \phi) \propto 1$
- ▶  $\boldsymbol{\beta} \mid \alpha, \phi, \gamma \sim N(0, g\phi^{-1}(\mathbf{X}^{cT}\mathbf{X}^c)^{-1})$

$$\boldsymbol{\beta} \mid \mathbf{Y}, \alpha, \phi \sim N\left(\frac{g}{1+g}\hat{\boldsymbol{\beta}}, \phi^{-1}\frac{g}{1+g}(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

$$\phi \mid \mathbf{Y} \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{\text{SSE} - \frac{1}{1+g}\hat{\boldsymbol{\beta}}^T(\mathbf{X}^T\mathbf{X})\hat{\boldsymbol{\beta}}}{2}\right)$$

joint posterior draws of beta's



# Bayesian Variable Selection

- ▶ Avoid the use of redundant variables (problems with interpretations)
- ▶ Inclusion of un-necessary terms yields less precise estimates, particularly if explanatory variables are highly correlated with each other
- ▶ reduced MSE: reduced variance but possibly higher bias
- ▶ it is too “expensive” to use all variables

# Bayesian Model Choice

- ▶ Models for the variable selection problem are based on a subset of the  $\mathbf{X}_1, \dots, \mathbf{X}_p$  variables
- ▶ Encode models with a vector  $\gamma = (\gamma_1, \dots, \gamma_p)$  where  $\gamma_j \in \{0, 1\}$  is an indicator for whether variable  $\mathbf{X}_j$  should be included in the model  $\mathcal{M}_\gamma$ .  $\gamma_j = 0 \Leftrightarrow \beta_j = 0$
- ▶ Each value of  $\gamma$  represents one of the  $2^p$  models.
- ▶ Under model  $\mathcal{M}_\gamma$ :

$$\mathbf{Y} \mid \alpha, \beta, \sigma^2, \gamma \sim \mathcal{N}(\mathbf{1}\alpha + \mathbf{X}_\gamma\beta_\gamma, \sigma^2\mathbf{I})$$

Where  $\mathbf{X}_\gamma$  is design matrix using the columns in  $\mathbf{X}$  where  $\gamma_j = 1$  and  $\beta_\gamma$  is the subset of  $\beta$  that are non-zero.

# Posterior Probabilities of Models

- Posterior model probabilities

$$p(\mathcal{M}_j | \mathbf{Y}) = \frac{p(\mathbf{Y} | \mathcal{M}_j)p(\mathcal{M}_j)}{\sum_j p(\mathbf{Y} | \mathcal{M}_j)p(\mathcal{M}_j)}$$

Marginal likelihood of a model is proportional to

$$p(\mathbf{Y} | \mathcal{M}_\gamma) = \iint p(\mathbf{Y} | \beta_\gamma, \sigma^2)p(\beta_\gamma | \gamma, \sigma^2)p(\sigma^2 | \gamma)d\beta d\sigma^2$$

- Probability  $\beta_j \neq 0$ :  $\sum_{\mathcal{M}_j: \beta_j \neq 0} p(\mathcal{M}_j | \mathbf{Y})$  (marginal posterior inclusion probability)



# Prior Distributions

- ▶ Bayesian Model choice requires proper prior distributions on regression coefficients (exception parameters that are included in all models)
- ▶ Vague but proper priors may lead to paradoxes!
- ▶ Conjugate Normal-Gammas lead to closed form expressions for marginal likelihoods, Zellner's g-prior is one of the most popular.

# Zellner's g-prior within Models

Centered model:

$$\mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}_\gamma^c \beta_\gamma + \epsilon$$

- ▶ Common parameters

$$p(\alpha, \phi) \propto \phi^{-1}$$

- ▶ Model Specific parameters

$$\beta_\gamma \mid \alpha, \phi, \gamma \sim \mathbf{N}(0, g\phi^{-1}(\mathbf{X}_\gamma^c{}'\mathbf{X}_\gamma^c)^{-1})$$

- ▶ Marginal likelihood of  $\mathcal{M}_\gamma$  is proportional to

$$p(\mathbf{Y} \mid \mathcal{M}_\gamma) = C(1 + g)^{\frac{n-p-1}{2}} (1 + g(1 - R_\gamma^2))^{-\frac{(n-1)}{2}}$$

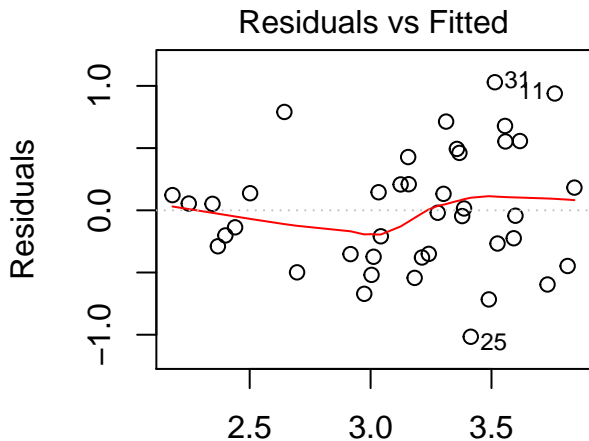
where  $R_\gamma^2$  is the usual  $R^2$  for model  $\mathcal{M}_\gamma$  and  $C$  is a constant that is  $p(\mathbf{Y} \mid \mathcal{M}_0)$  (model with intercept alone)

- ▶ uniform distribution over space of models  $p(\mathcal{M}_\gamma) = 1/(2^p)$



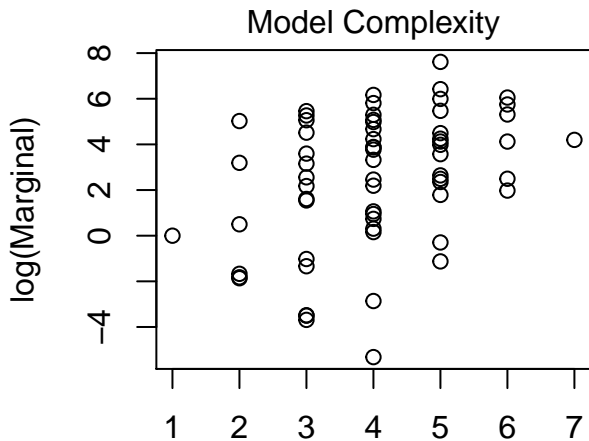
residual plot)

```
plot(poll.bma, which=1)
```



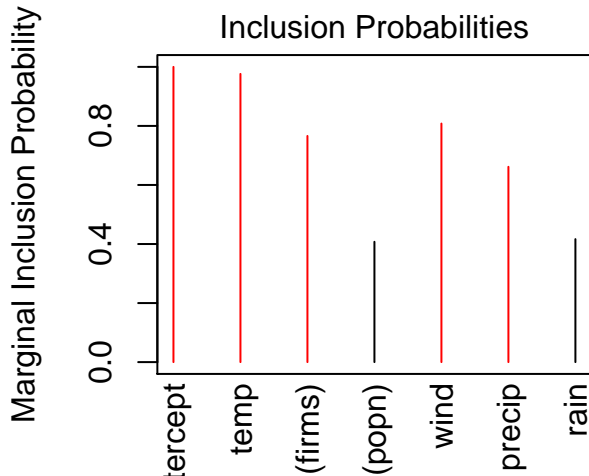
## Model Complexity)

```
plot(poll.bma, which=3)
```



## Inclusion Probabilities)

```
plot(poll.bma, which=4)
```



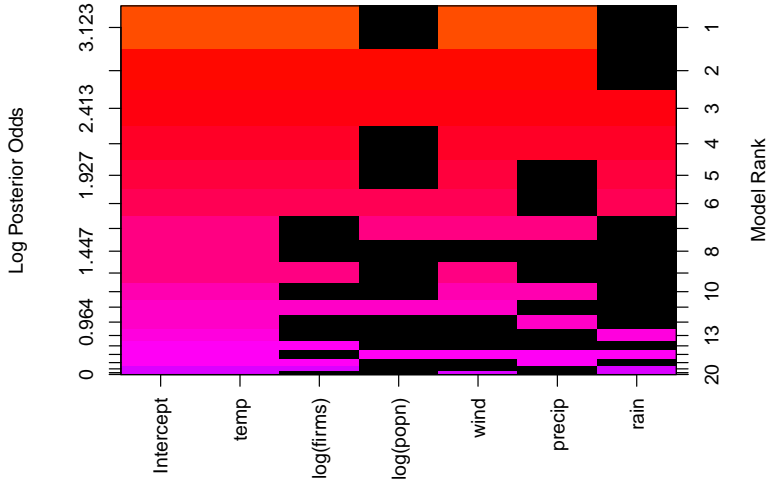
# Model Space

```
summary(poll.bma)
```

##	P(B != 0   Y)	model 1	model 2	model 3
## Intercept	1.0000000	1.000000	1.0000000	1.00000000
## temp	0.9762282	1.000000	1.0000000	1.00000000
## log(firms)	0.7659857	1.000000	1.0000000	1.00000000
## log(popn)	0.4075393	0.000000	1.0000000	1.00000000
## wind	0.8080832	1.000000	1.0000000	1.00000000
## precip	0.6615960	1.000000	1.0000000	1.00000000
## rain	0.4166394	0.000000	0.0000000	1.00000000
## BF	NA	1.000000	0.2093987	0.03277353
## PostProbs	NA	0.209600	0.1097000	0.10300000
## R2	NA	0.542700	0.5500000	0.55030000
## dim	NA	5.000000	6.0000000	7.00000000
## logmarg	NA	7.616228	6.0527128	4.19809382

## Summary

```
image(poll.bma)
```





# Coefficients

```
beta = coef(poll.bma, n.models=1)
beta
```

##

## Marginal Posterior Summaries of Coefficients:

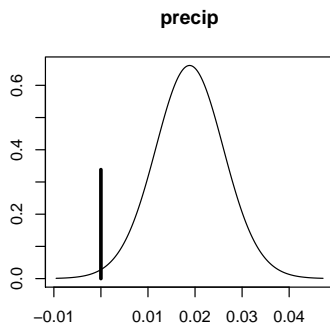
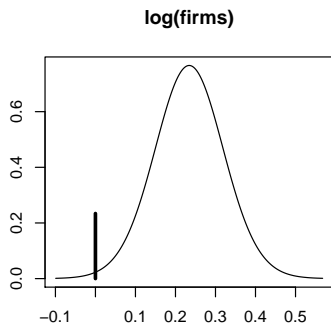
##

## Based on the top 1 models

##	post mean	post SD	post p(B != 0)
## Intercept	3.15300	0.07818	1.00000
## temp	-0.07130	0.01268	0.97623
## log(firms)	0.23428	0.08573	0.76599
## log(popn)	0.00000	0.00000	0.40754
## wind	-0.17998	0.06128	0.80808
## precip	0.01884	0.00729	0.66160
## rain	0.00000	0.00000	0.41664

# Coefficients

```
par(mfrow=c(2,2)); plot(beta, subset=c(3, 6))
```



# Bayesian Confidence Intervals

```
confint(beta)
```

```
##              2.5 %      97.5 %      beta
## Intercept    2.994993257  3.31101398  3.15300362
## temp        -0.096926645 -0.04567203 -0.07129934
## log(firms)    0.061014518  0.40753936  0.23427694
## log(popn)     0.000000000  0.00000000  0.00000000
## wind         -0.303835463 -0.05612195 -0.17997871
## precip        0.004105874  0.03357242  0.01883915
## rain          0.000000000  0.00000000  0.00000000
## attr(,"Probability")
## [1] 0.95
## attr(,"class")
## [1] "confint.bas"
```