Discussion of

George & Foster (2000)
"Calibration and Empirical Bayes Variable Selection"



Seminar Series in Quantitative Studies in Consumer Behavior

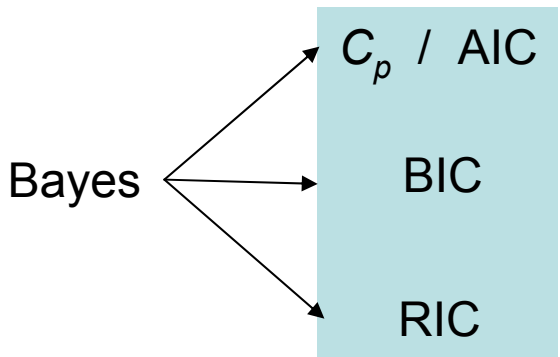
11 April 2006

Overview

$$\underset{n \times p}{Y} = \underset{p \times 1}{X\beta} + \varepsilon \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2 I_n)$$

γ indexes a model

Calibration



Empirical Bayes

Data based methods for prior specification

Common Approaches

Choose model that maximizes:

$$\mathrm{SS}_\gamma/\hat{\sigma}^2 - Fq_\gamma$$

dimension of model γ

"dimension penalty"

$$\begin{aligned}\mathrm{SS}_\gamma &= \hat{\beta}'_\gamma X'_\gamma X_\gamma \hat{\beta}_\gamma \\ &= Y' X_\gamma (X'_\gamma X_\gamma)^{-1} X'_\gamma Y\end{aligned}$$

Common Approaches

$$\mathrm{SS}_\gamma / \hat{\sigma}^2 - F q_\gamma$$

F	Criterion	
2	C_p (AIC)	Mallows (1973) Akaike (1973)
log n	BIC	Schwartz (1978)
2 log p	RIC	Foster & George (1994)

Bayes Approach

Model Space Posterior

$$p(\gamma|y) \propto p(y|\gamma)p(\gamma)$$

Marginal Likelihood

$$p(y|\gamma) = \int p(y|\beta_\gamma, \gamma)p(\beta_\gamma|\gamma)d\beta_\gamma$$

Prior Distributions

$$p(\beta_\gamma, \gamma) = p(\beta_\gamma|\gamma)p(\gamma)$$

Calibrate a particular class of priors to *F*

Priors

Bernoulli / g-prior

$$p(\beta_\gamma|\gamma, c) = \mathrm{N}\left(0, c\sigma^2(X'_\gamma X_\gamma)^{-1}\right)$$

$$p(\gamma|w) = w^{q_\gamma}(1-w)^{p-q_\gamma}$$

Hyperparameters

$$c > 0$$

$$0 < w < 1$$

Note:

c controls the expected size of nonzero coefficients

w controls the number of nonzero coefficients

g-priors are…

…conditionally compatible

$$p(\beta|c) = \mathrm{N}\left(0, c\sigma^2(X'X)^{-1}\right)$$

$$\downarrow$$

$$p(\beta_\gamma|\gamma, c) = \mathrm{N}\left(0, c\sigma^2(X_\gamma'X_\gamma)^{-1}\right)$$

…a limiting case of a latent factor regression model (West, 2001)

…computationally tractable $p(y|\gamma) \propto (1+c)^{-q_\gamma/2}\left(y'y - \frac{c}{1+c}\hat{y}_\gamma'\hat{y}_\gamma\right)^{-n/2}$

...can be problematic for model selection (Jeffreys, 1961; Berger & Pericchi, 2001):

$$B(\gamma:0) = (1+c)^{(n-q_\gamma)/2}\left(1+c(1-R^2_\gamma)\right)^{-n/2}$$

$$\text{Fixed } n,\ R^2_\gamma \to 1:\ B(\gamma:0) \to (1+c)^{(n-q_\gamma)/2} < \infty$$

Calibration

$$
\begin{aligned}
p(\gamma|y,c,w) &\propto p(\gamma|w)p(y|\gamma,c) \\
&\propto \exp\left[\frac{c}{2(1+c)}\left\{\mathrm{ss}_\gamma/\sigma^2 - F(c,w)q_\gamma\right\}\right]
\end{aligned}
$$

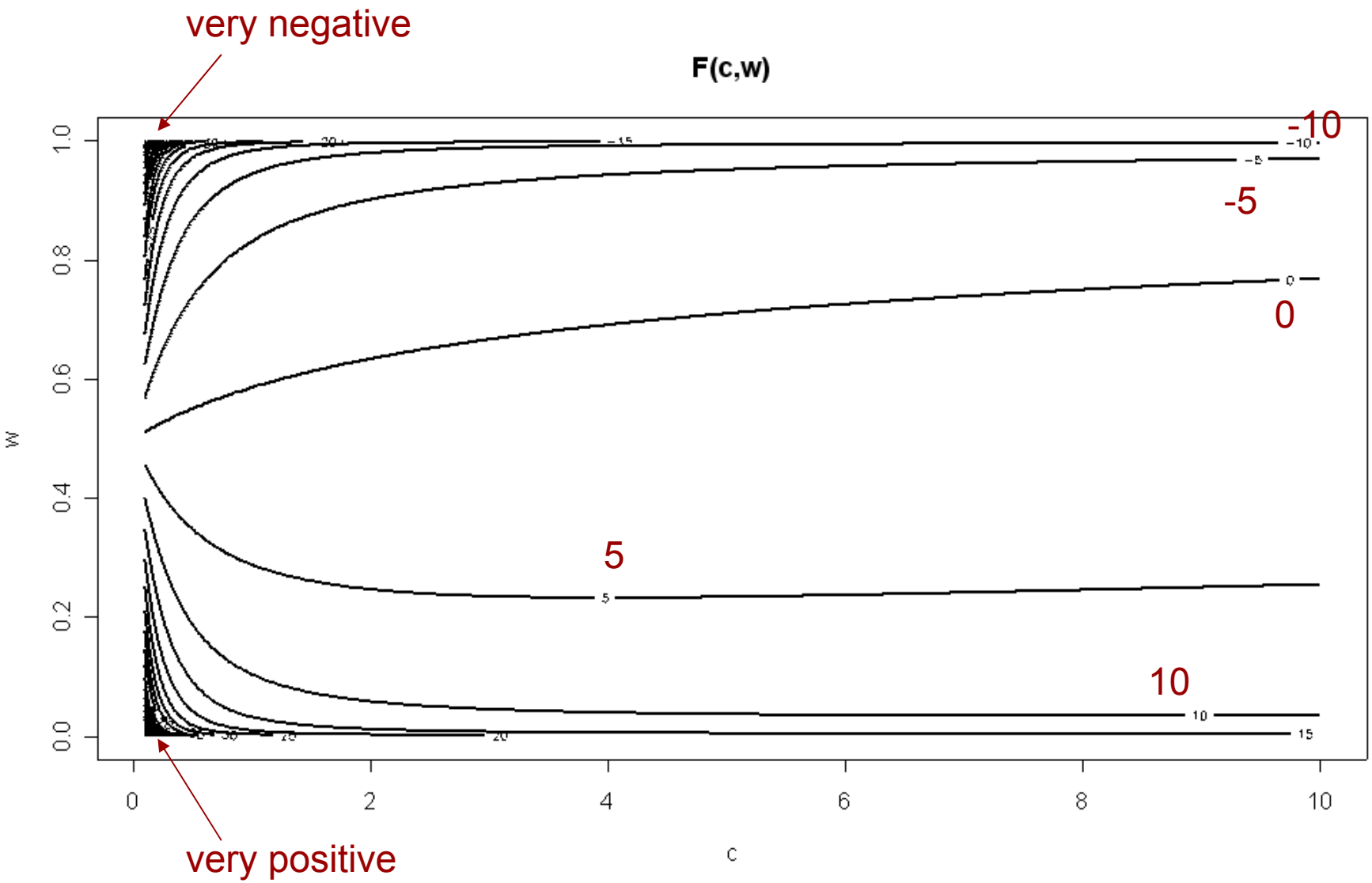
$$
p(\gamma_1|y,c,w) > p(\gamma_2|y,c,w)
$$

if and only if

$$
\mathrm{ss}_{\gamma_1}/\sigma^2 - F(c,w)q_{\gamma_1} > \mathrm{ss}_{\gamma_2}/\sigma^2 - F(c,w)q_{\gamma_2}
$$

$$
F(c,w) = \frac{1+c}{c}\left\{2\log\frac{1-w}{w} + \log(1+c)\right\}
$$

F(c,w)


very negative
F(c,w)
-10
-5
0
5
10
very positive
w
c
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10

Calibration

$$\mathrm{SS}_\gamma/\hat{\sigma}^2 - F(c,w)q_\gamma$$

$F(c,w)$	Criterion	c	w
2	C_p (AIC)	3.92	0.5
$\log n$	BIC	n	0.5
$2 \log p$	RIC	p^2	0.5

Empirical Bayes Estimation of *c* and *w*

- "Default" choices can bias posterior away from "true" underlying model
- Fully Bayes approach: prior $p(c,w)$

$$p(\gamma|c,w,Y)$$
$$p(c|w,\gamma,Y)$$
$$p(w|c,\gamma,Y)$$

- Stochastic search methods only visit a small area of the model space

Empirical Bayes Estimation of *c* and *w*

THE OHIO STATE UNIVERSITY

Marginal Likelihood

$$
\begin{aligned}
p(y|c,w) &= \sum_{\gamma} p(y,\gamma|c,w) \\
&= \sum_{\gamma} p(y|\gamma,c)p(\gamma|w) \\
&\propto \sum_{\gamma} w^{q_\gamma}(1-w)^{p-q_\gamma}(1+c)^{-q_\gamma/2}\exp\left\{\frac{c\ \mathrm{SS}_\gamma}{2\sigma^2(1+c)}\right\}
\end{aligned}
$$

Note:

Maximize (numerically) to get $\hat{c}, \hat{w}$

Not computationally feasible for large *p*

Empirical Bayes Estimation of *c* and *w*

THE OHIO STATE UNIVERSITY

Orthogonal Design

$$p(y|c,w) \propto \prod_{j=1}^{p} \left\{ (1-w)e^{-t_j^2/2} + w(1+c)^{-1/2} e^{-t_j^2/2\,(1+c)} \right\}$$

$$t_j = \frac{x_j' y}{\sqrt{\sigma^2 x_j' x_j}}$$

Note:

Only *p* distinct terms - maximization is (computationally) feasible

MAP model maximizes $C_{\mathrm{MML}} = \mathrm{SS}_\gamma/\hat{\sigma}^2 - F(\hat{c},\hat{w})q_\gamma$

Conditional EBayes Approach

$$p(y, \gamma | c, w) \propto w^{q_\gamma}(1-w)^{p-q_\gamma}(1+c)^{-q_\gamma/2} \exp\left\{\frac{c\ \mathrm{SS}_\gamma}{2\sigma^2(1+c)}\right\}$$

conditional on a model γ

maximize rather than marginalize:

$$\begin{aligned} \hat{w}_\gamma &= q_\gamma/p \\ 1+\hat{c}_\gamma &= \mathrm{SS}_\gamma/(\sigma^2 q_\gamma) \end{aligned}$$

Note:

Under the "usual" estimate $\hat{\sigma}^2 = \mathrm{SS}_{\mathrm{resid}}/(n-p)$,

$$1+\hat{c}_\gamma = F_{\mathrm{stat}_\gamma} \quad \Rightarrow \quad \hat{c}_\gamma = \max\{F_{\mathrm{stat}_\gamma} - 1, 0\}$$

Conditional EBayes Approach

$$\log p(\gamma|y,\hat{c}_\gamma,\hat{w}_\gamma) = \text{const.} + \text{ss}_\gamma/\sigma^2 - q_\gamma\left(1+\log \text{ss}_\gamma/(\sigma^2 q_\gamma)\right) - \left(-2\{(p-q_\gamma)\log(p-q_\gamma)+q_\gamma\log q_\gamma\}\right)$$

$$C_{\text{CML}} = \text{ss}_\gamma/\sigma^2 - B\left(\text{ss}_\gamma/\sigma^2\right) - R(q_\gamma)$$

Notes:

$p(\gamma|y,\hat{c}_\gamma,\hat{w}_\gamma)$ is maximized by the γ that maximizes C_{CML}

Computation easy, but still have to search model space

When $p(y, \gamma \mid c, w)$ dominated by single model, C_{MML} and C_{CML} are similar