Dealing with Separation in Logistic Regression Models

Carlisle Rainey
Assistant Professor
Texas A&M University
crainey@tamu.edu

paper, data, and code at
crain.co/research

The prior matters a lot,
so choose a good one.

The prior matters a lot,

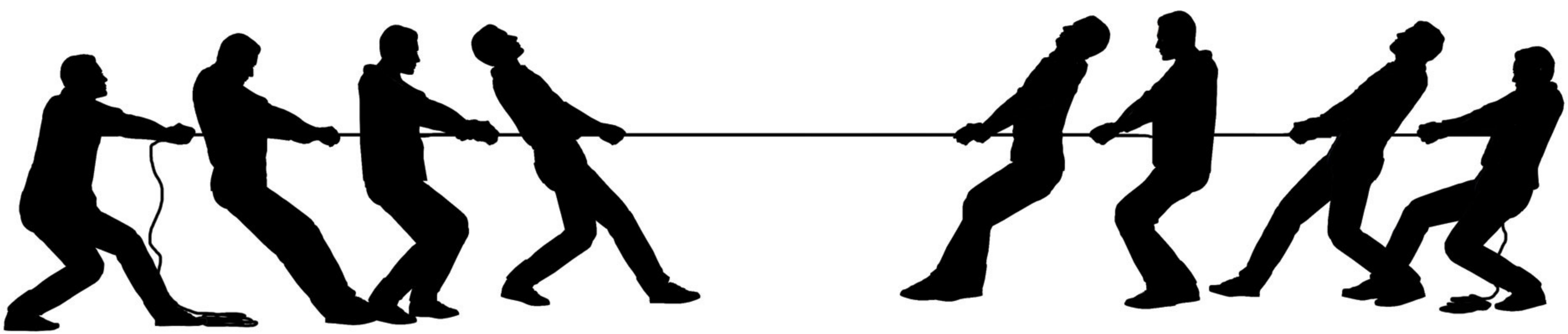
1. in practice
2. in theory

so choose a good one.

3. concepts
4. software

The Prior Matters
in Practice

politics **need**

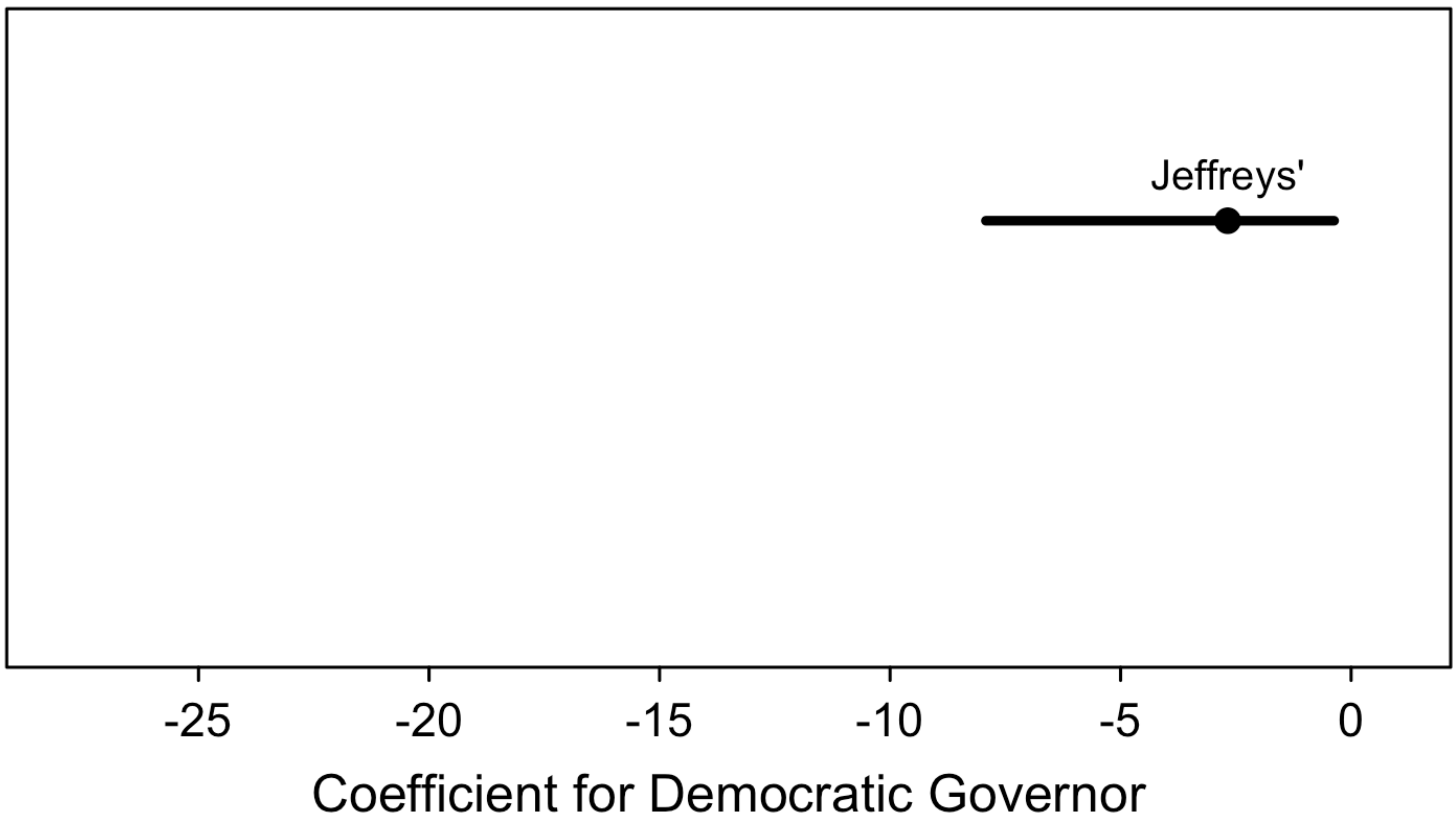


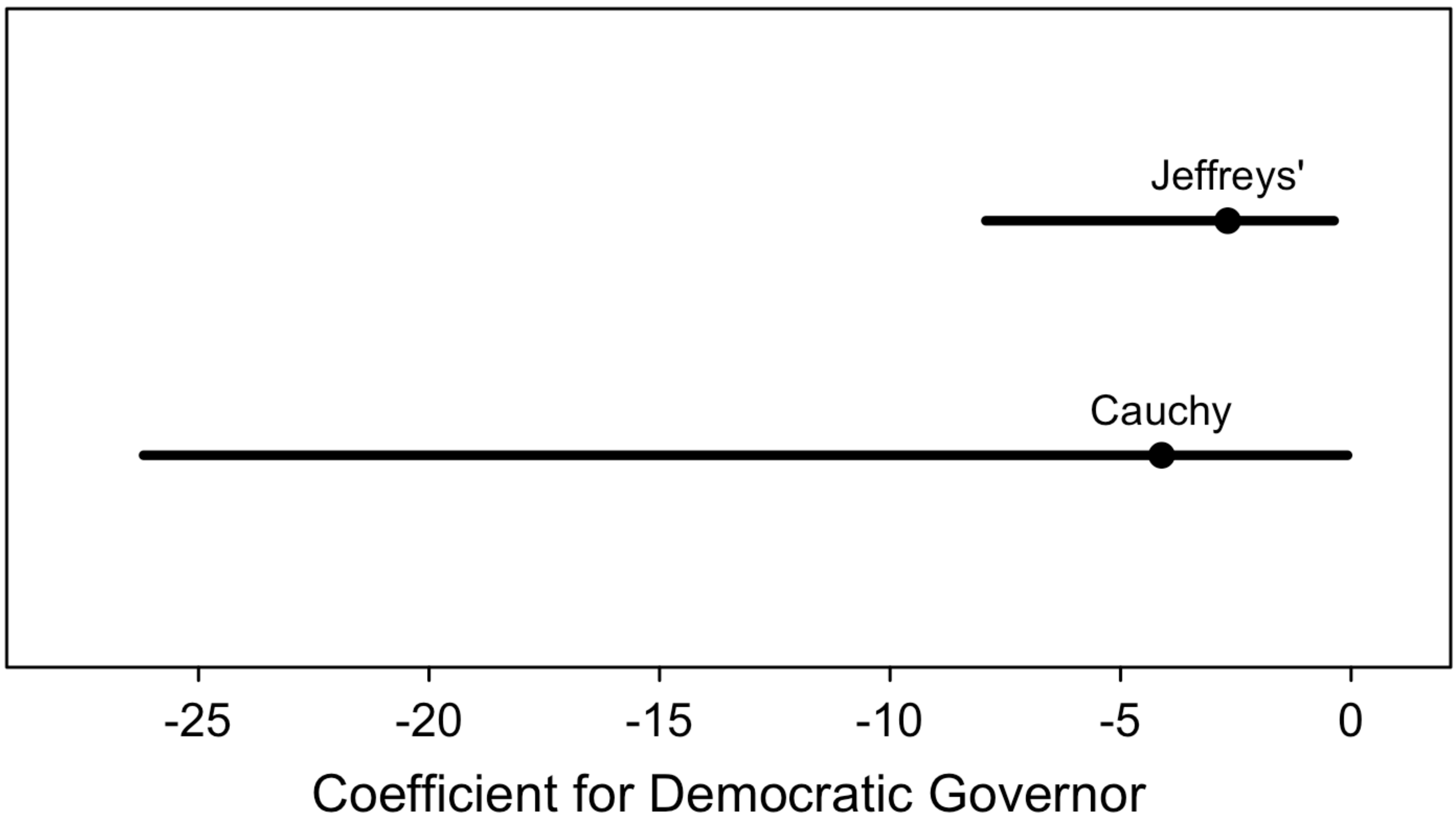
Variable	*Coefficient*	*Confidence Interval*
Democratic Governor	**-26.35**	**[-126,979.03; 126,926.33]**
% Uninsured (Std.)	**0.92**	**[-3.46; 5.30]**
% Favorable to ACA	0.01	[-0.17; 0.18]
GOP Legislature	2.43	[-0.47; 5.33]
Fiscal Health	0.00	[-0.02; 0.02]
Medicaid Multiplier	-0.32	[-2.45; 1.80]
% Non-white	0.05	[-0.12; 0.21]
% Metropolitan	-0.08	[-0.17; 0.02]
Constant	2.58	[-7.02; 12.18]

Variable	*Coefficient*	*Confidence Interval*
Democratic Governor	**-26.35**	**[-126,979.03; 126,926.33]**

Variable	*Coefficient*	*Confidence Interval*
Democratic Governor	**-26.35**	**[-126,979.03; 126,926.33]**

This is a failure of maximum likelihood.


Jeffreys'
-25
-20
-15
-10
-5
0
Coefficient for Democratic Governor


Jeffreys'
Cauchy
-25
-20
-15
-10
-5
0
Coefficient for Democratic Governor

Donald J. Trump @realDonaldTrump · Now

Those inferences from supposedly "default" priors aren't even close to similar. Very sad!

RETWEETS	FAVORITES
823	**4835**

6:23 AM - 21 May 2016 · Details

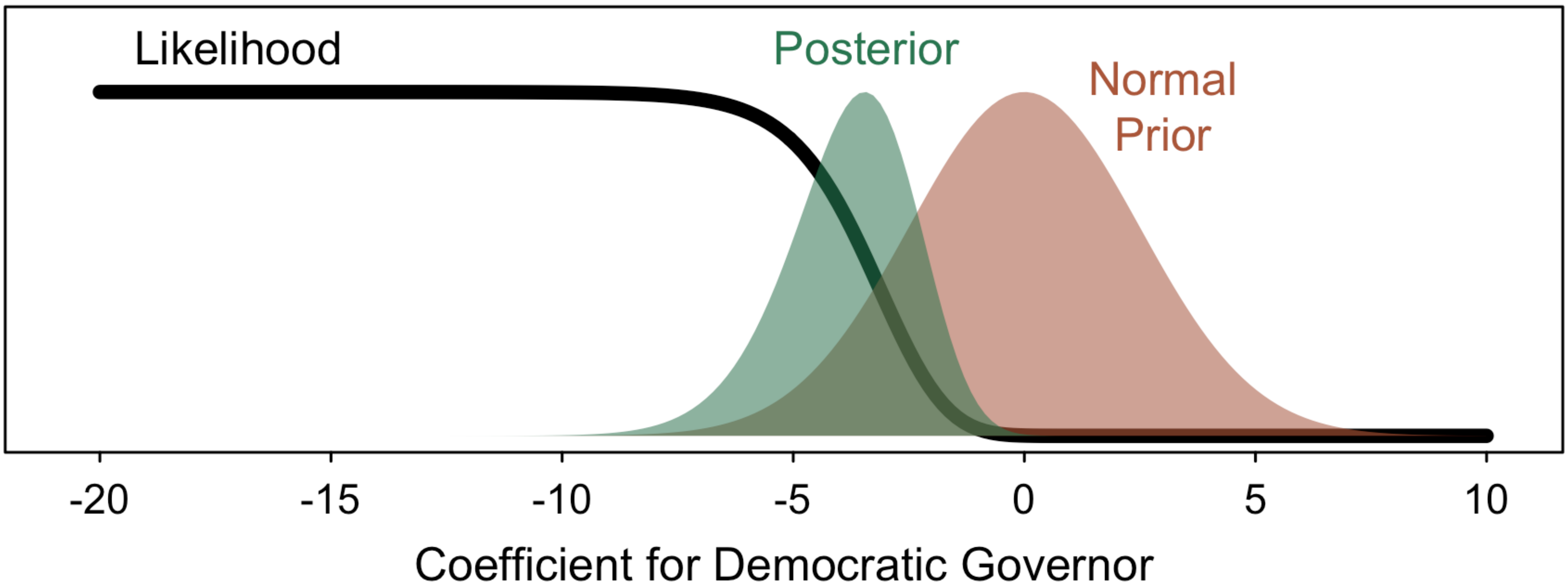
Different ***default*** priors produce different results.

The Prior Matters
in Theory

For

1. a monotonic likelihood $p(y|\beta)$ decreasing in β_s,
2. a proper prior distribution $p(\beta|\sigma)$, and
3. a large, negative β_s,

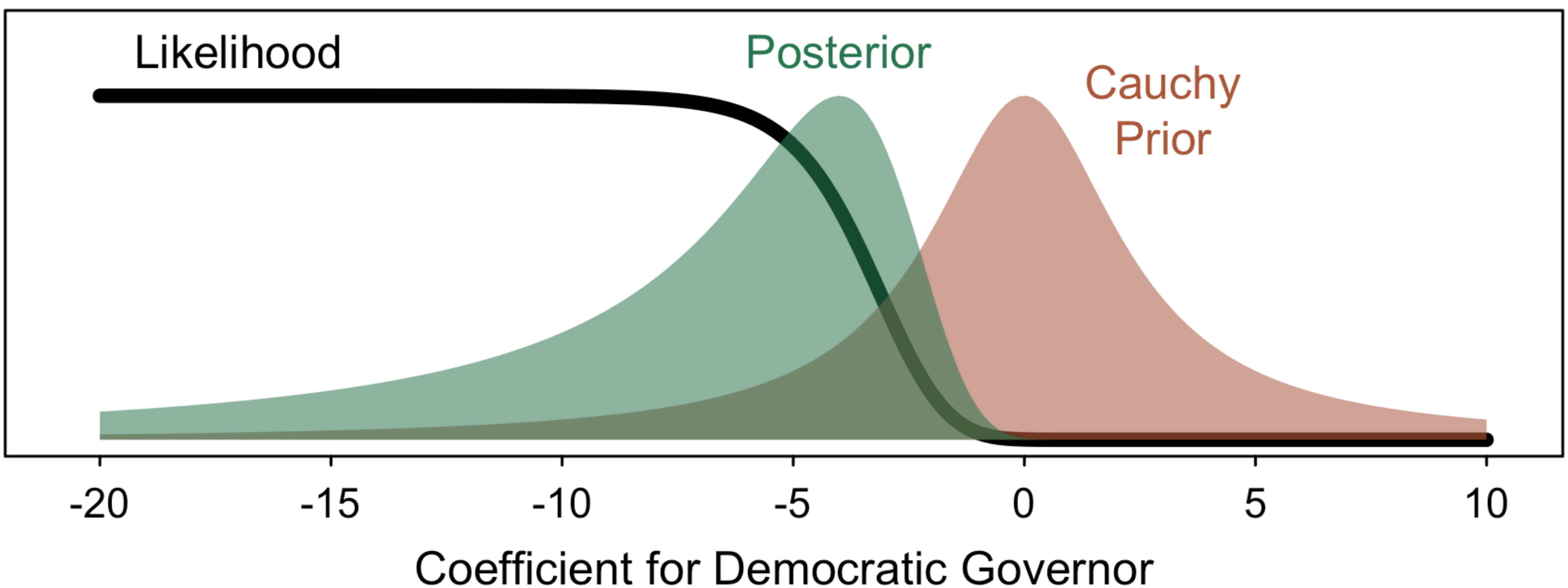
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



For

1. a monotonic likelihood $p(y|\beta)$ decreasing in β_s,
2. a proper prior distribution $p(\beta|\sigma)$, and
3. a large, negative β_s,

the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



The prior ***determines***
crucial parts of the posterior.

Key Concepts

for Choosing a Good Prior

$$Pr(y_i) = \Lambda(\beta_c + \beta_s s_i + \beta_1 x_{i1} + ... + \beta_k x_{ik})$$

Transforming the Prior Distribution

$$\tilde{\beta} \sim p(\beta)$$

$$\tilde{\pi}_{new} = p(y_{new} | \tilde{\beta})$$

$$\tilde{q}_{new} = q(\tilde{\pi}_{new})$$

We Already Know Few Things

$$\beta_1 \approx \hat{\beta}_1^{mle}$$

$$\beta_2 \approx \hat{\beta}_2^{mle}$$

$$\vdots$$

$$\beta_k \approx \hat{\beta}_k^{mle}$$

$$\beta_{\boldsymbol{s}} < 0$$

Partial Prior Distribution

$$p^*(\beta|\beta_s < 0, \beta_{-s} = \hat{\beta}^{mle}_{-s}),$$
$$\text{where } \hat{\beta}^{mle}_{s} = -\infty$$

Software

for Choosing a Good Prior

separation

(on GitHub)

Stan Project

`rstanarm`

`StataStan`

Conclusion

The prior matters a lot,
so choose a good one.

What should *you* do?

1. Notice the problem and do something.
2. Recognize the the prior affects the inferences and choose a good one.
3. Assess the robustness of your conclusions to a range of prior distributions.