Dealing with Separation in Logistic Regression Models

Carlisle Rainey
Assistant Professor
University at Buffalo, SUNY
rcrainey@buffalo.edu

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

Dealing with Separation in Logistic Regression Models

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

43 million times larger

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

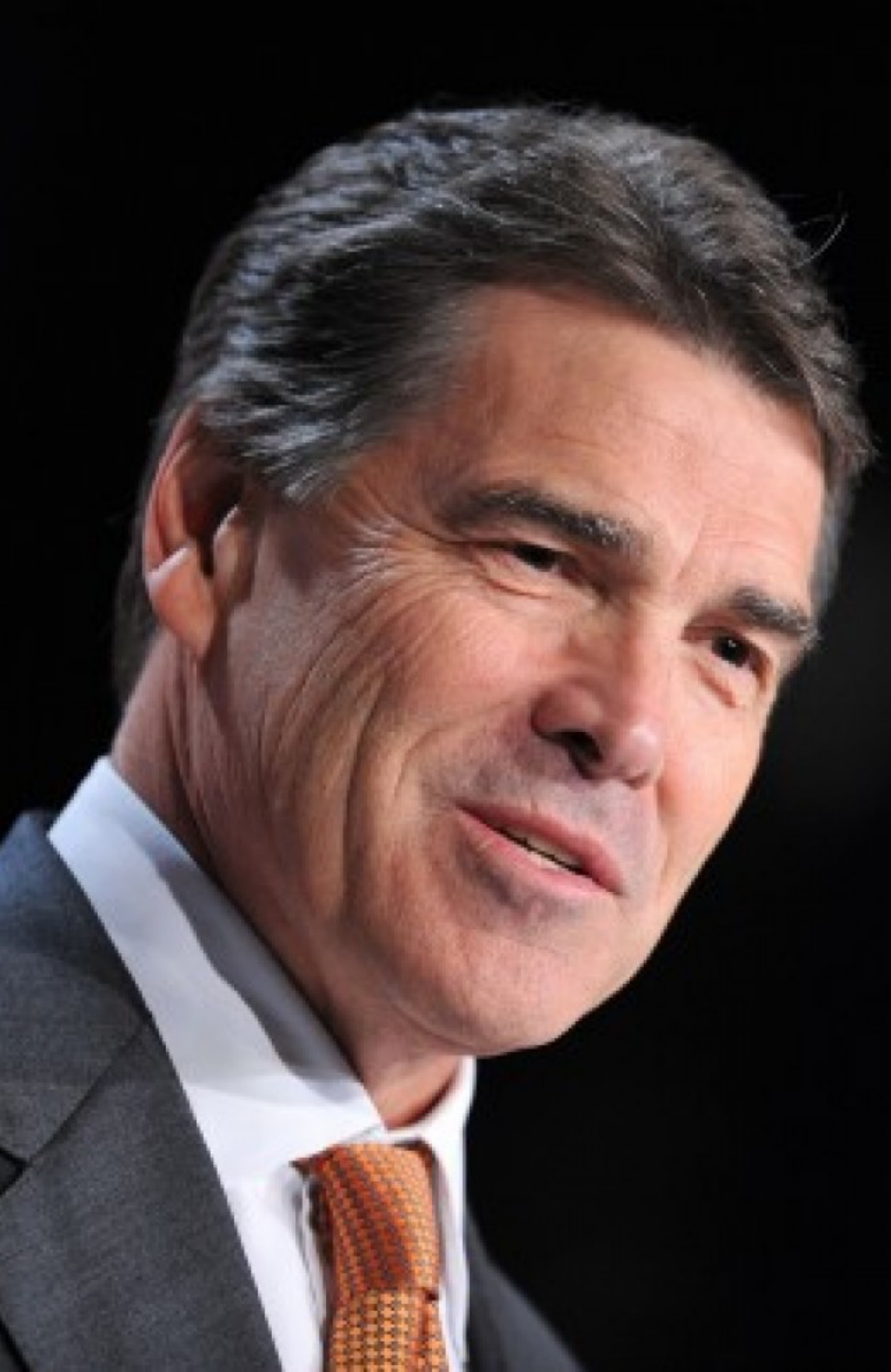

SEAL OF THE PRESIDENT OF THE UNITED STATES

2 million

3,000

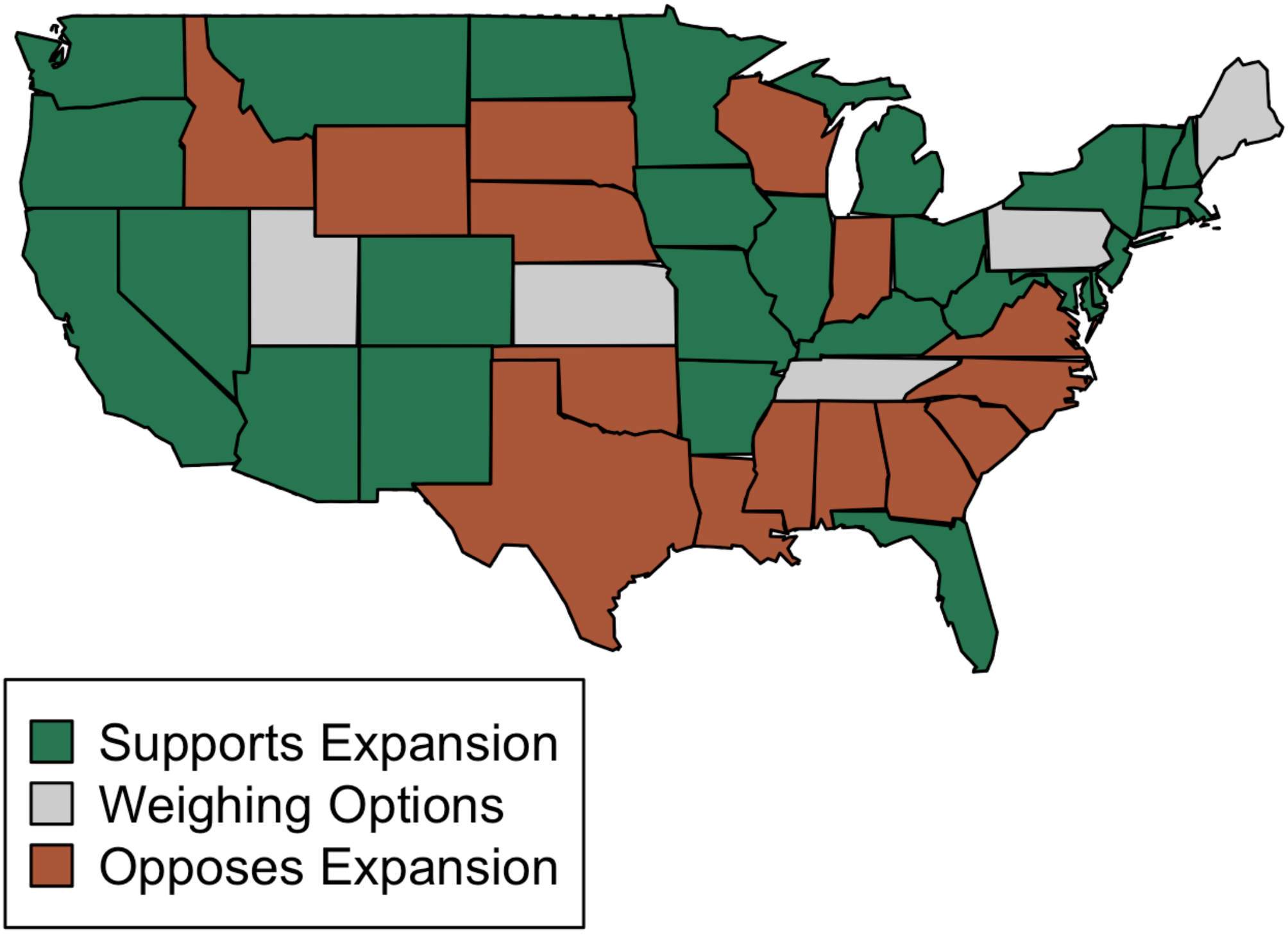
100%

90%

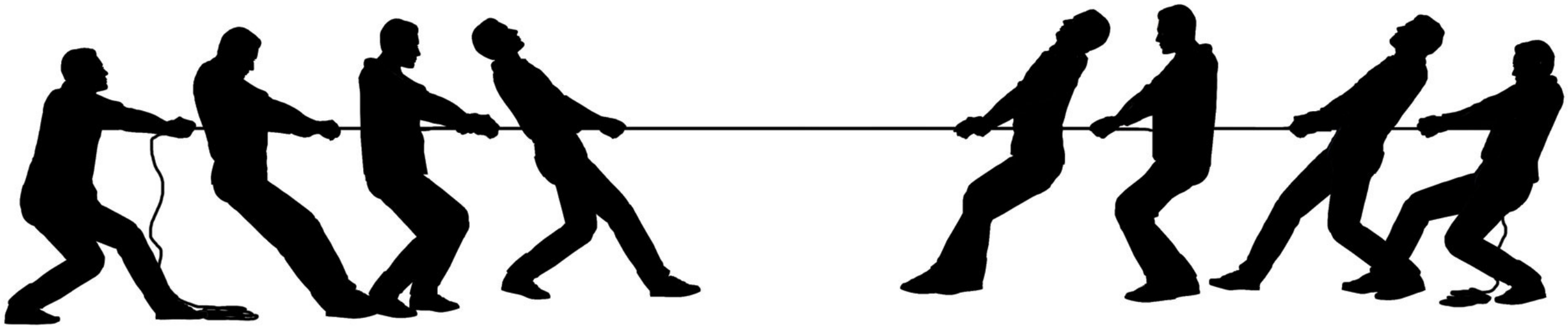


“To expand this program is not unlike adding a thousand people to the Titanic.”

— July 2012

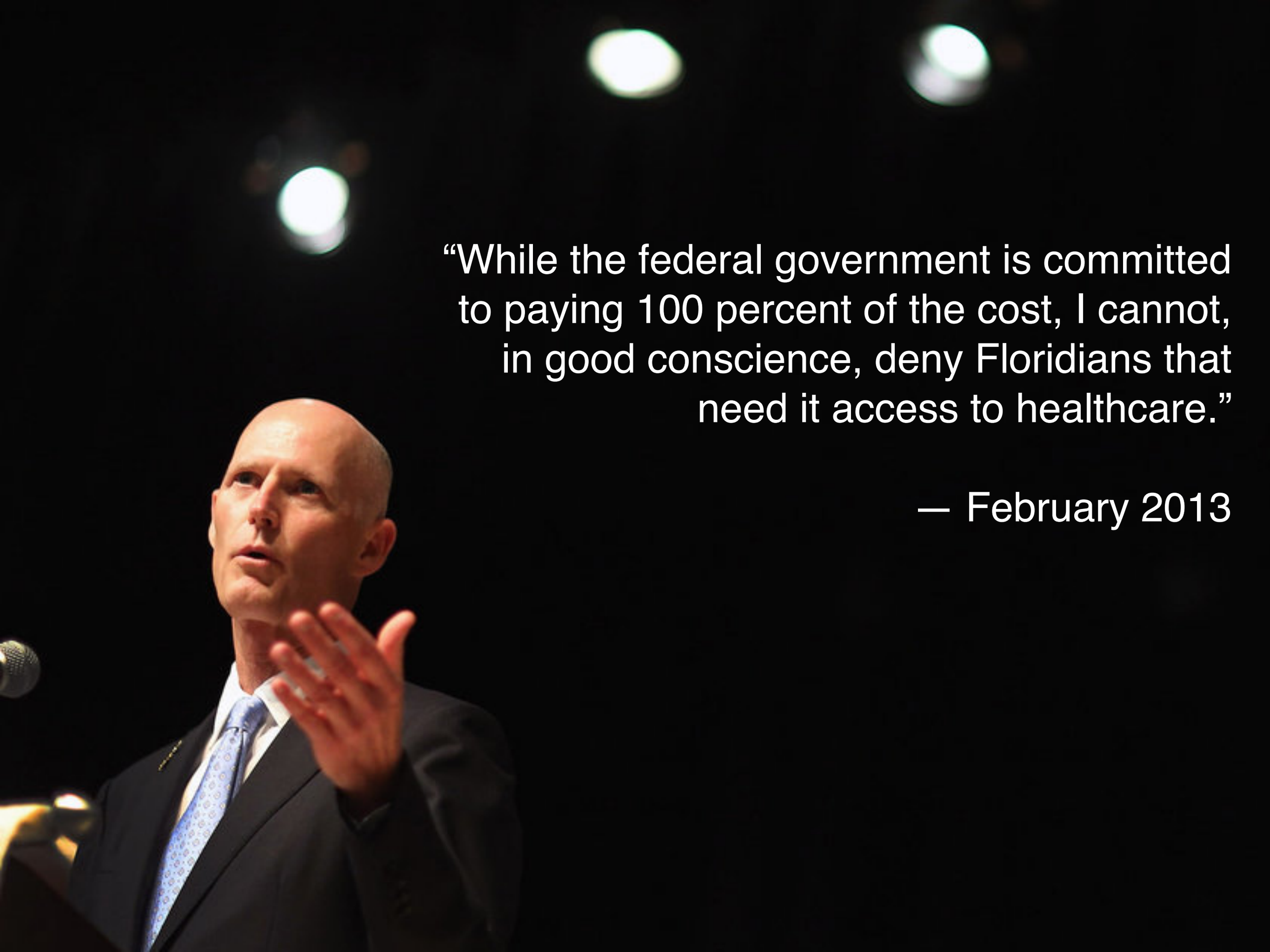

Supports Expansion
Weighing Options
Opposes Expansion

politics need



“Obamacare is going to be horrible for patients. It’s going to be horrible for taxpayers. It’s probably the biggest job killer ever.”

— October 2010



"While the federal government is committed to paying 100 percent of the cost, I cannot, in good conscience, deny Floridians that need it access to healthcare."

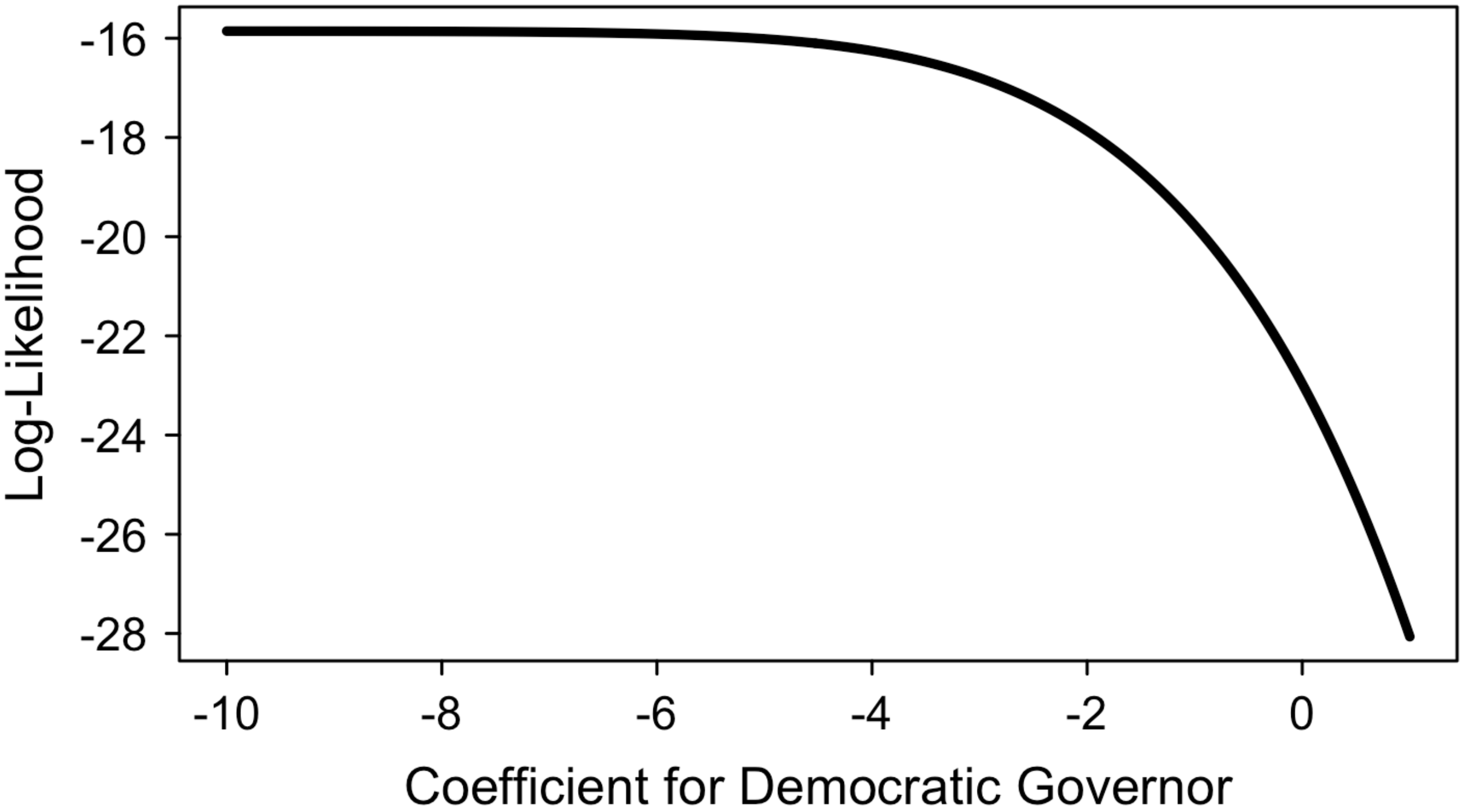
— February 2013

In the tug-of-war between politics and need,

which one wins?

Variable	*Coefficient*	*Confidence Interval*
Democratic Governor	**-20.35**	**[-6,340.06; 6,299.36]**
% 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]

	Doesn't Oppose	Opposes
Republican	14	16
Democrat	20	0


Log-Likelihood
-16
-18
-20
-22
-24
-26
-28
-10
-8
-6
-4
-2
0
Coefficient for Democratic Governor

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]**

unreasonable

useless

This is a failure of maximum likelihood.

Jeffreys' Prior

Zorn (2005)

Jeffreys'

-25

-20

-15

-10

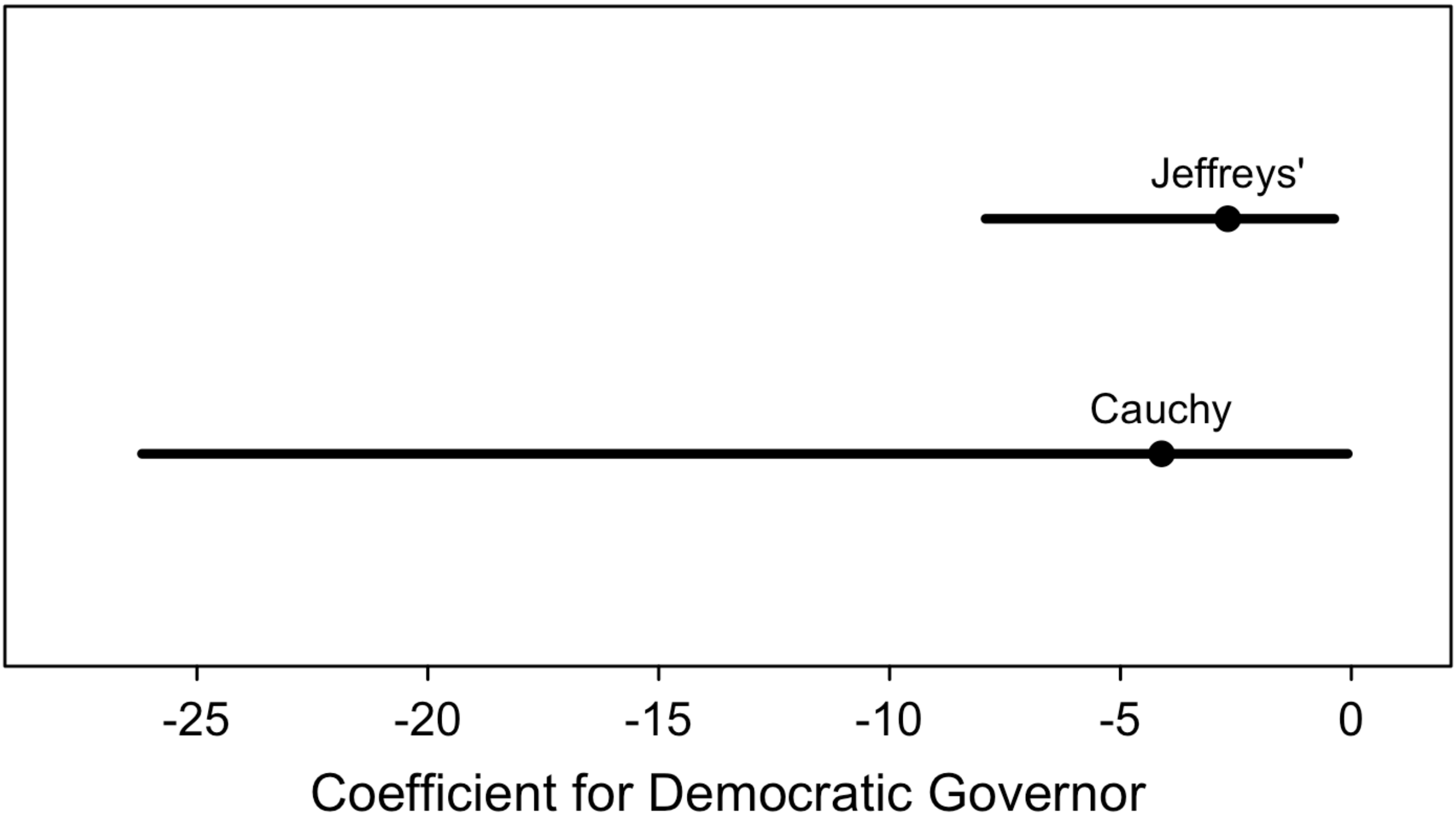
-5

0

Coefficient for Democratic Governor

Cauchy Prior

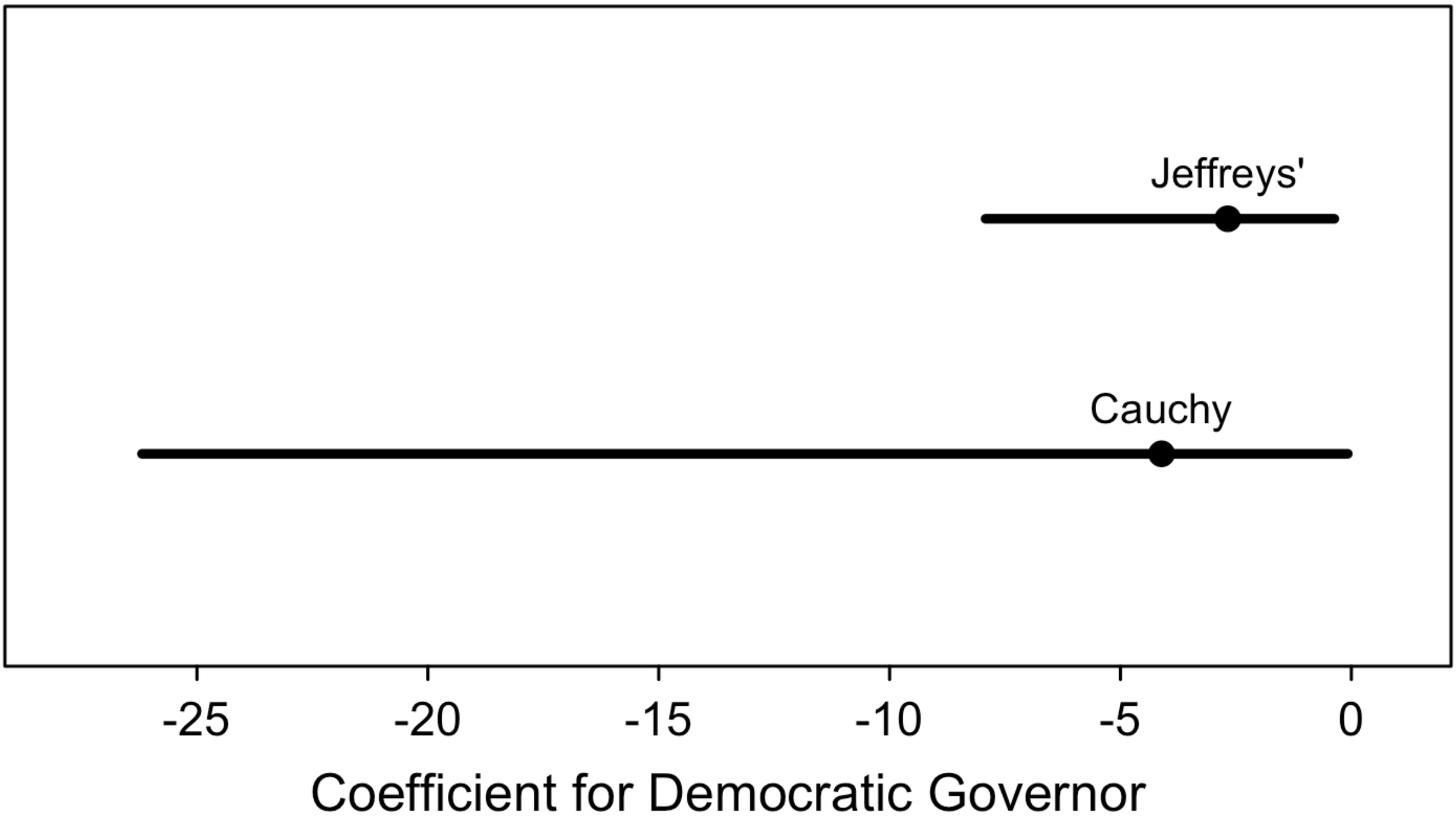
Gelman et al. (2008)


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

The Cauchy prior produces…

a confidence interval that is

250% wider


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

The Cauchy prior produces…

a coefficient estimate that is

50% larger

The Cauchy prior produces…

a risk-ratio estimate that is

43 million times larger

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

The Prior Matters

in Theory

For

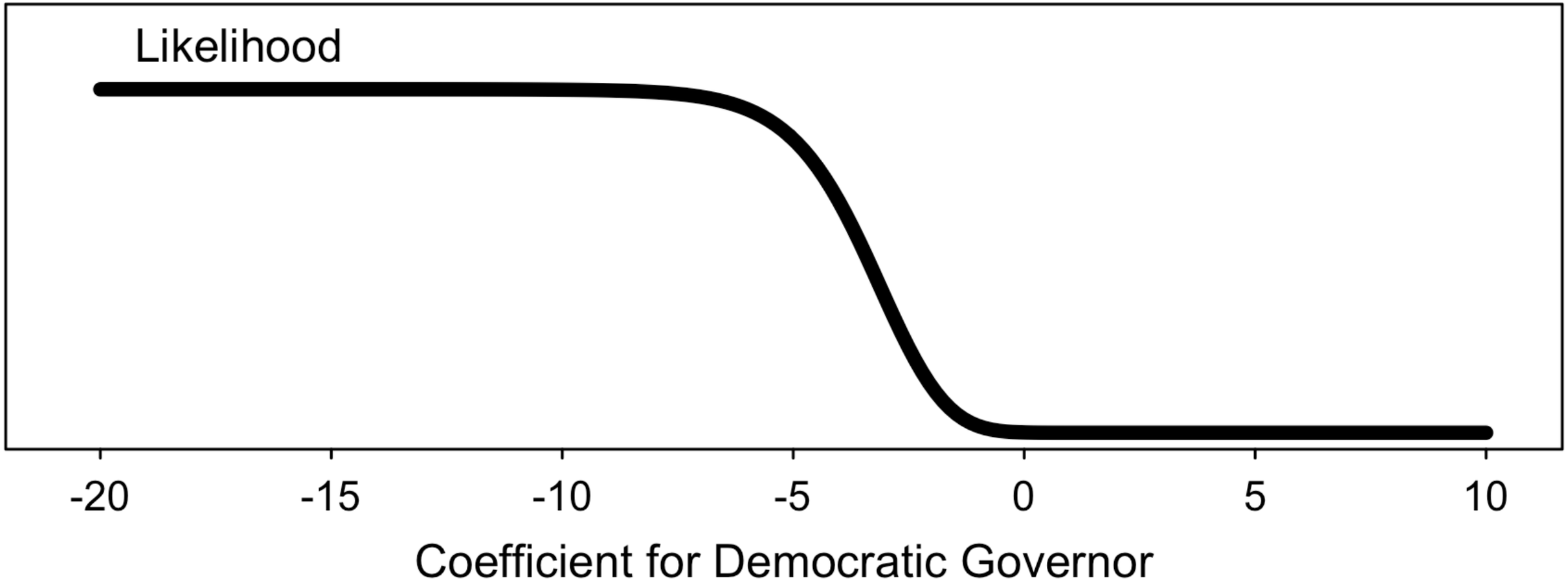
1. a monotonic likelihood $p(y|\beta)$ decreasing in β_s,
2. a proper prior distribution $p(\beta|\sigma)$, and
3. a large, negative β_s,

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,

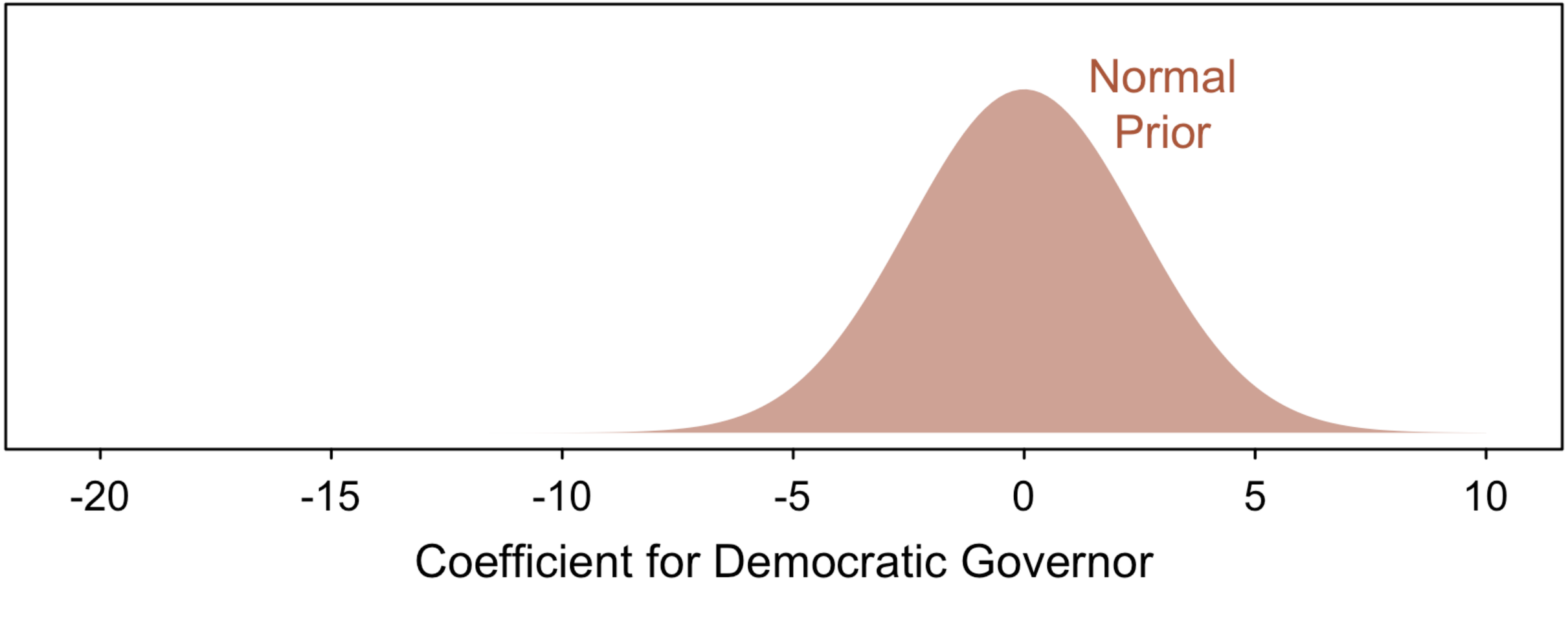
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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,

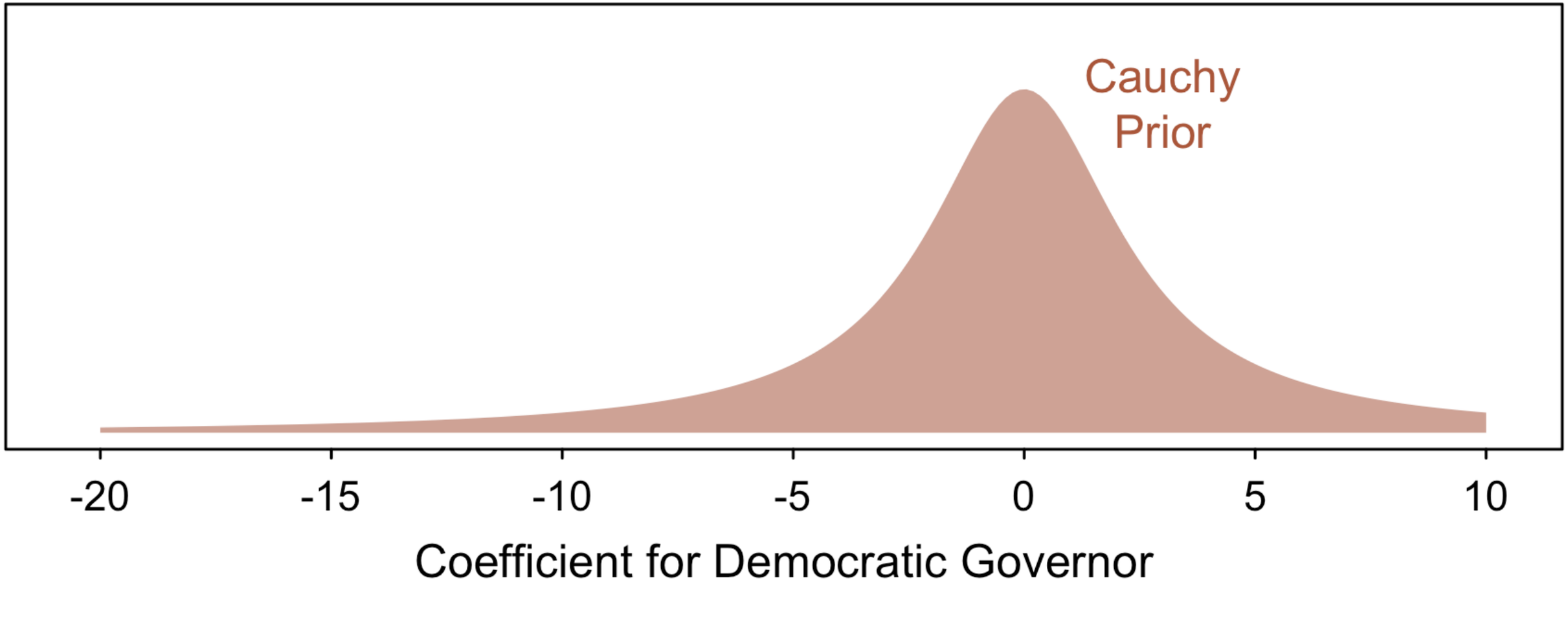
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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,

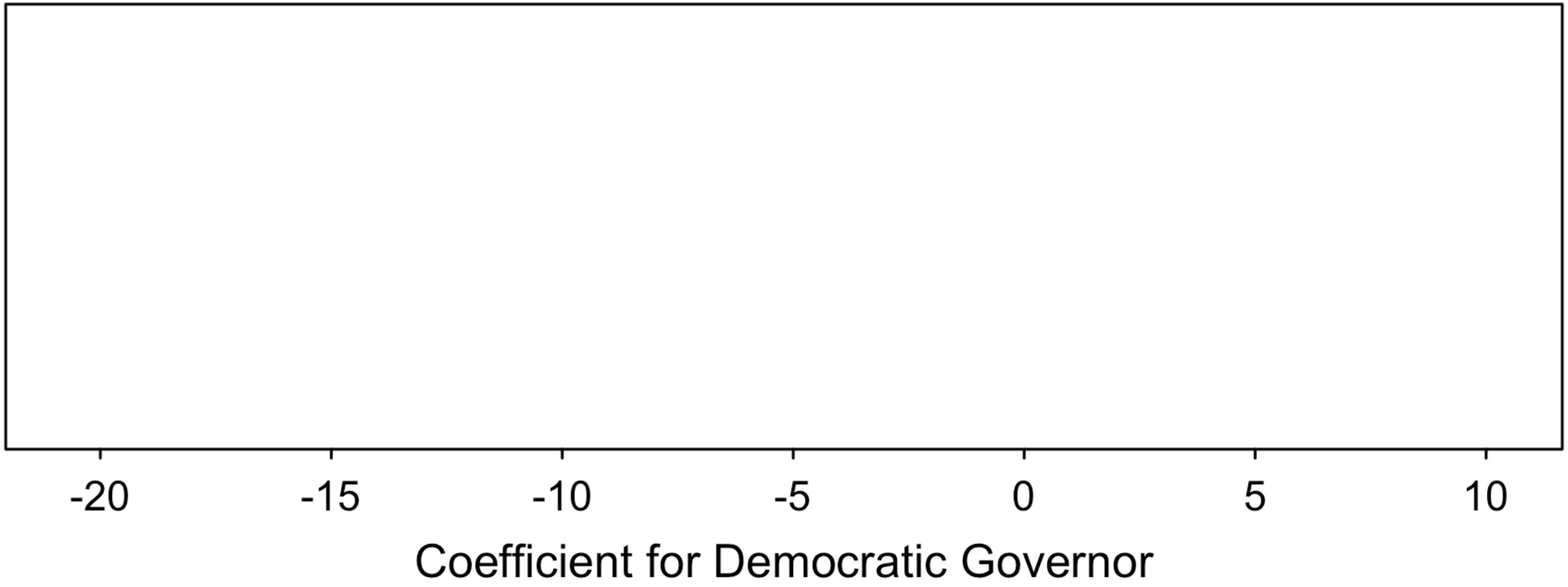
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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 $p(\beta_s|y) \propto p(\beta_s|\sigma)$.

For

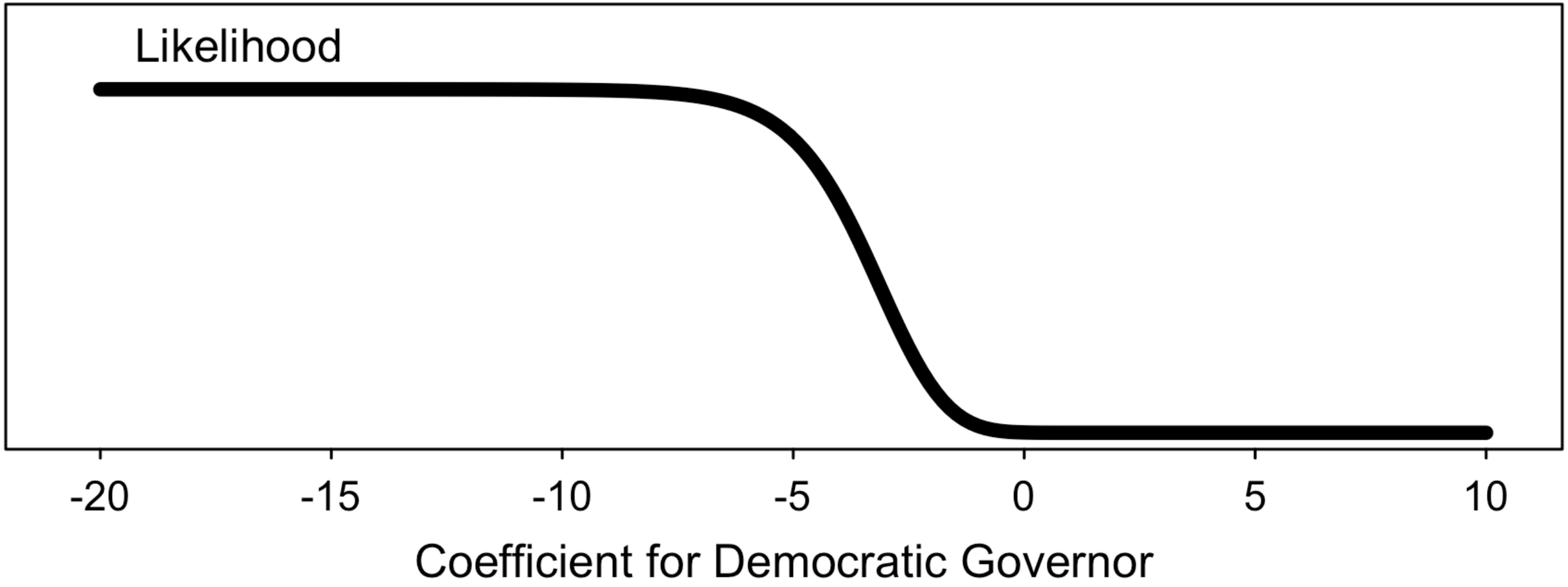
1. a monotonic likelihood $p(y|\beta)$ decreasing in β_s,
2. a proper prior distribution $p(\beta|\sigma)$, and
3. a large, negative β_s,

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,

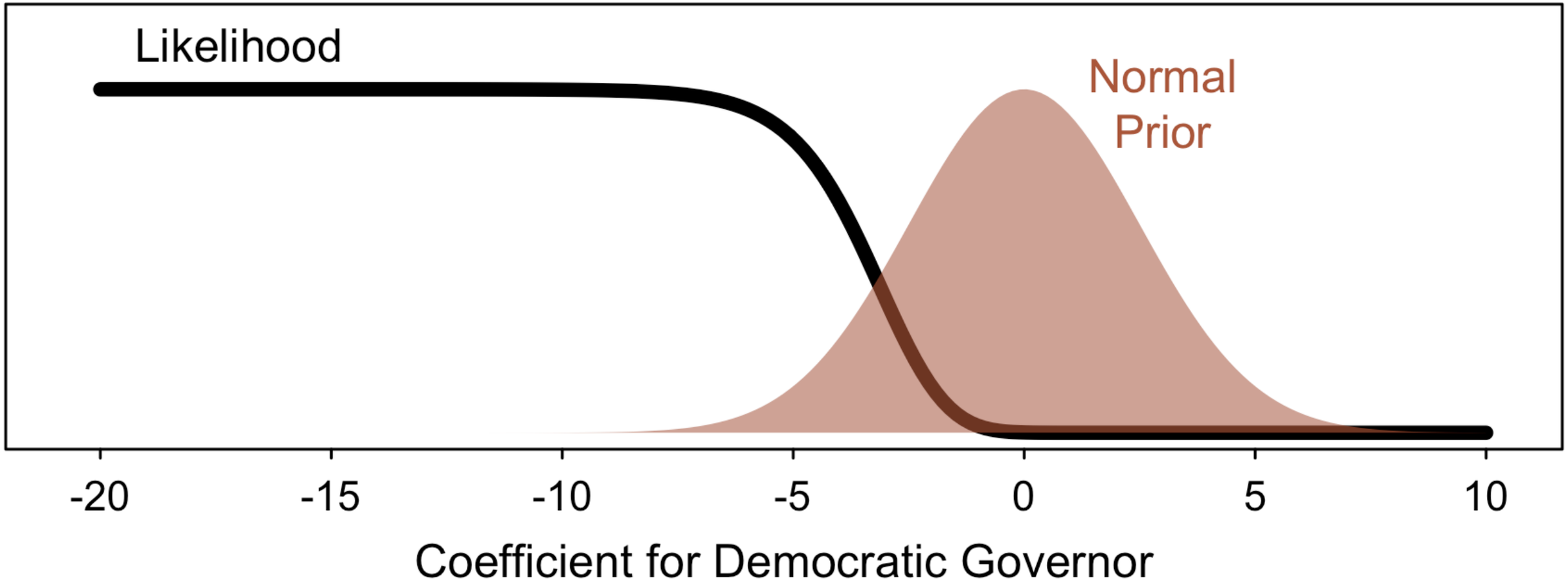
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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,

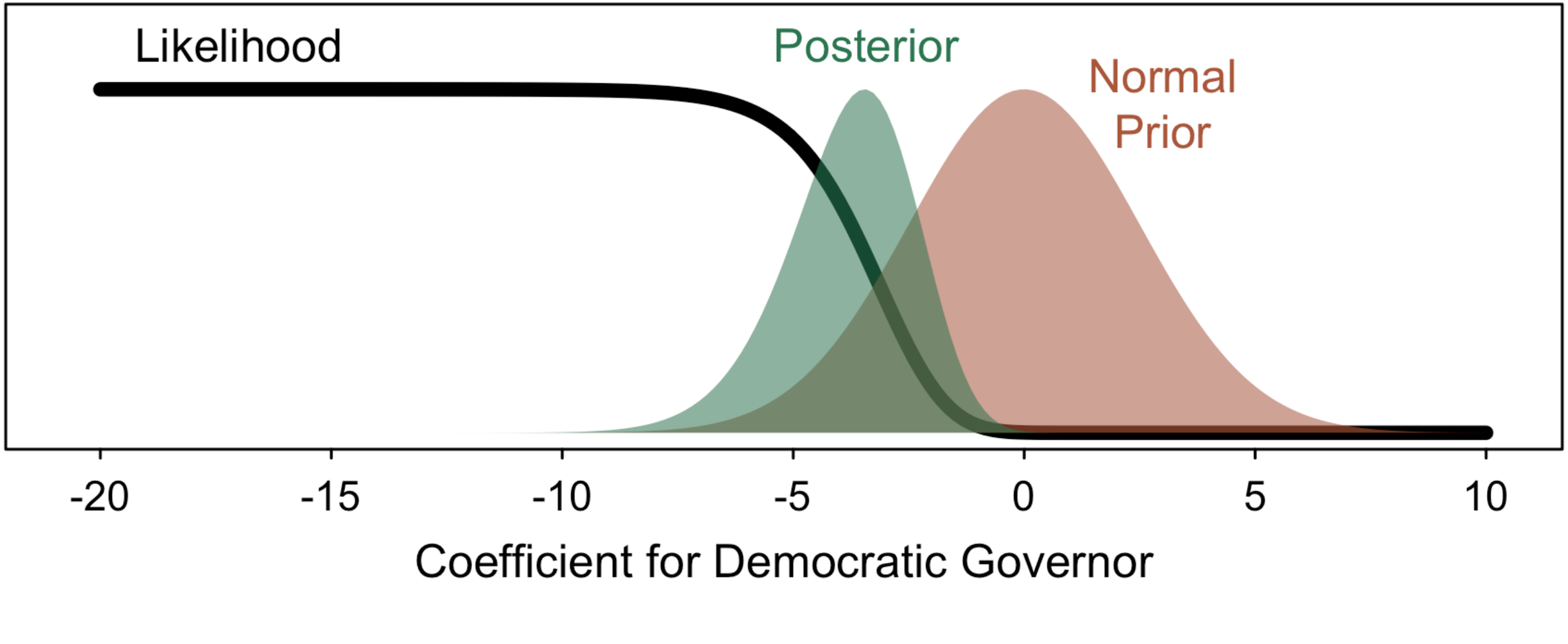
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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,

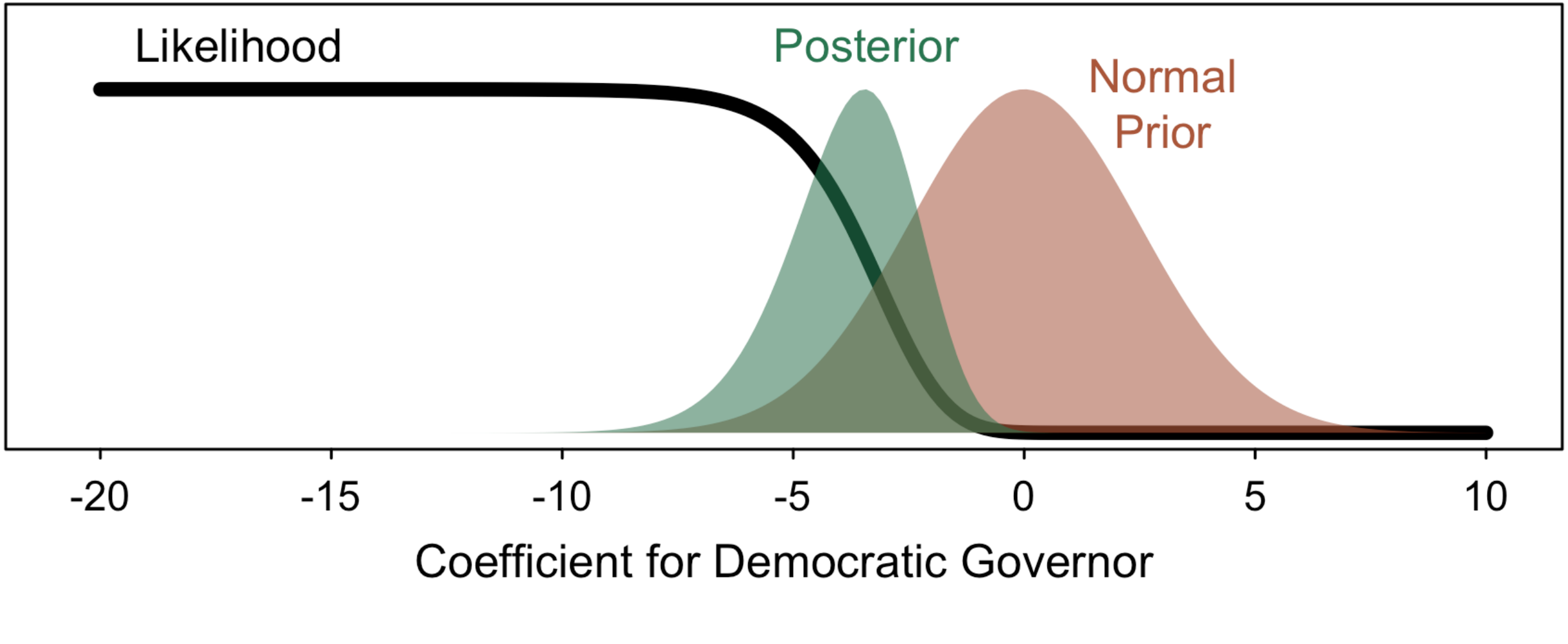
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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,

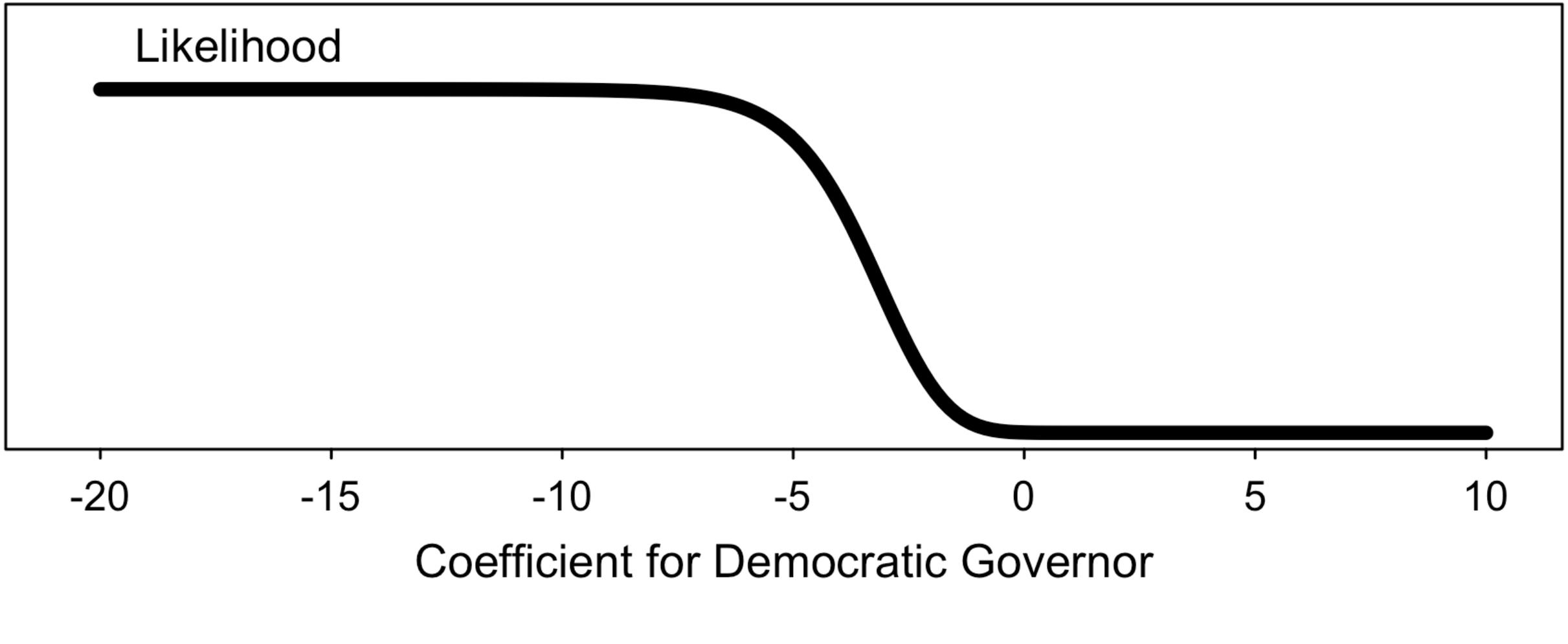
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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,

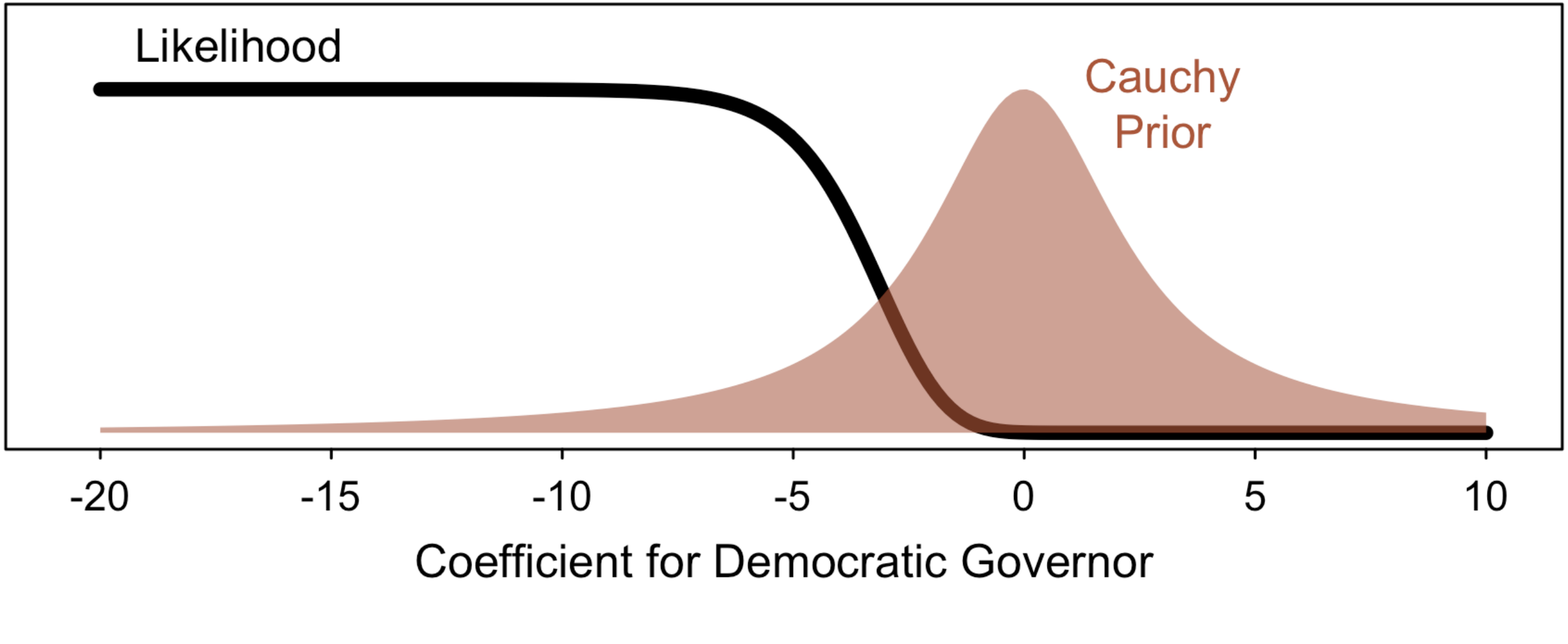
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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,

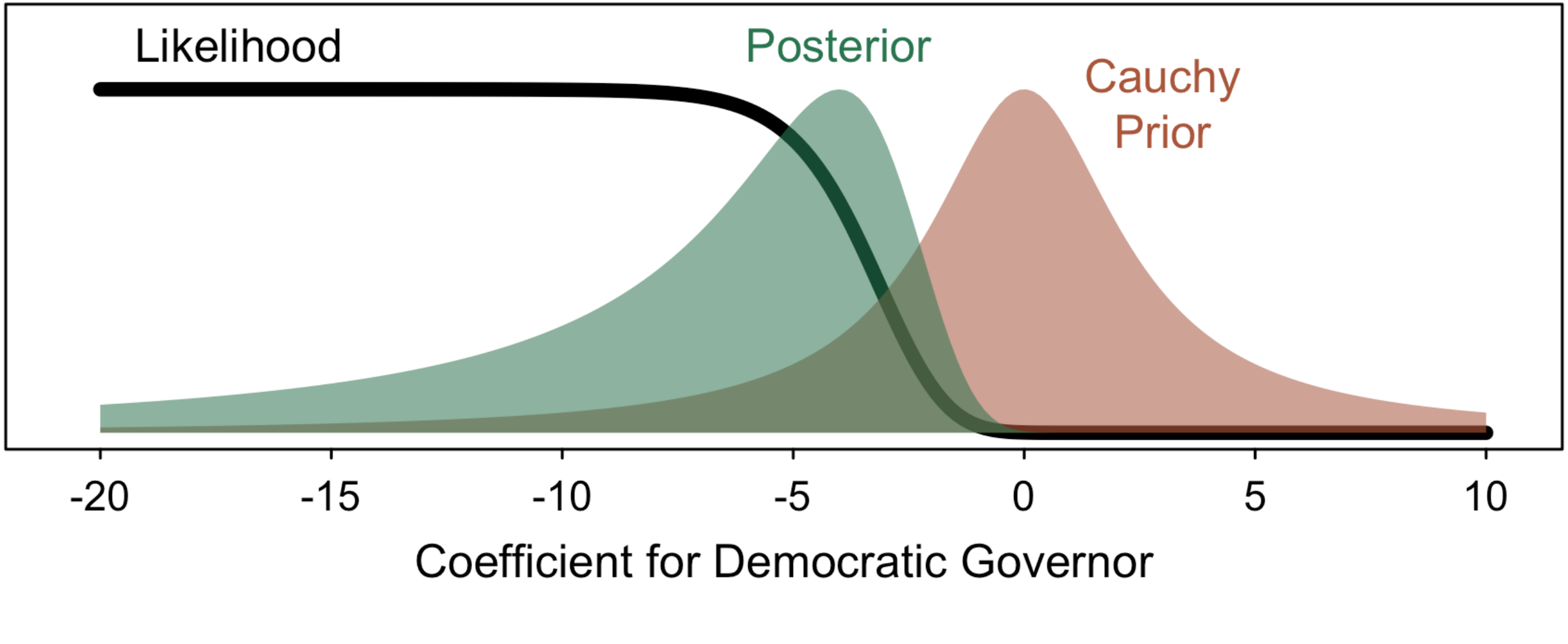
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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,

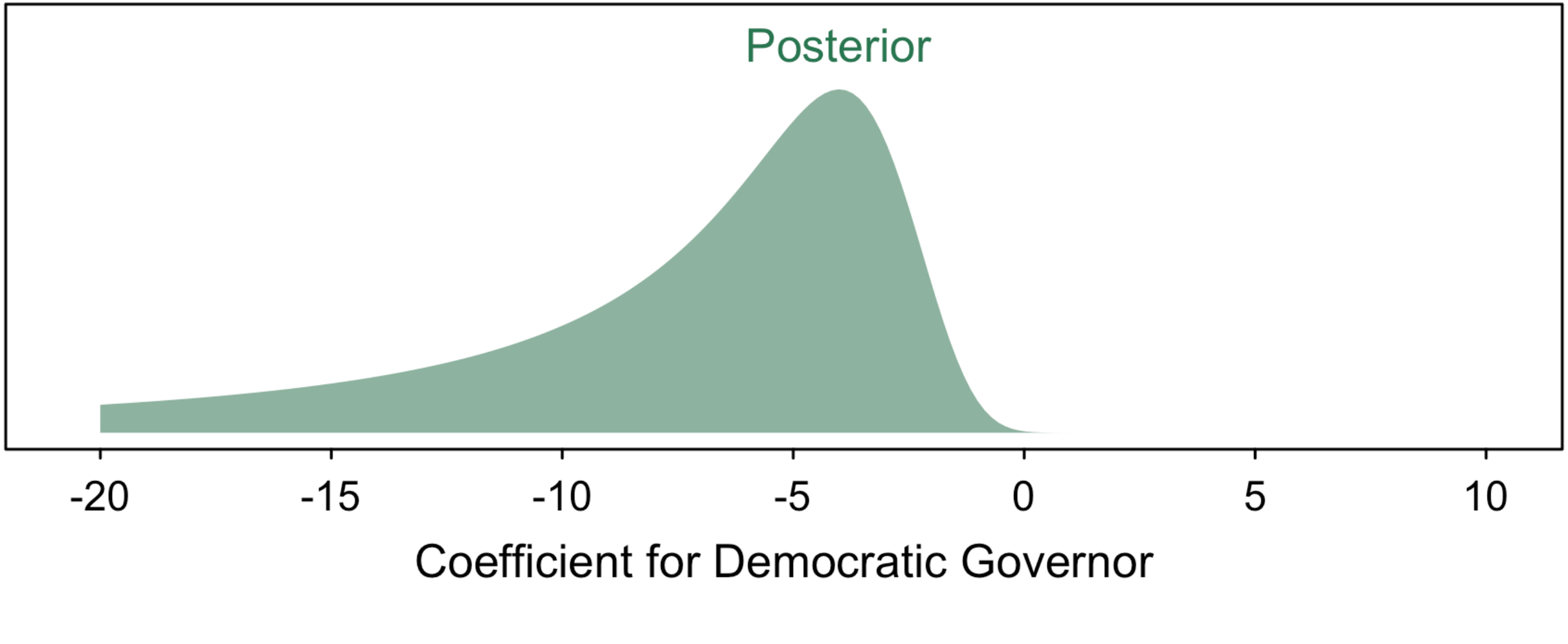
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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,

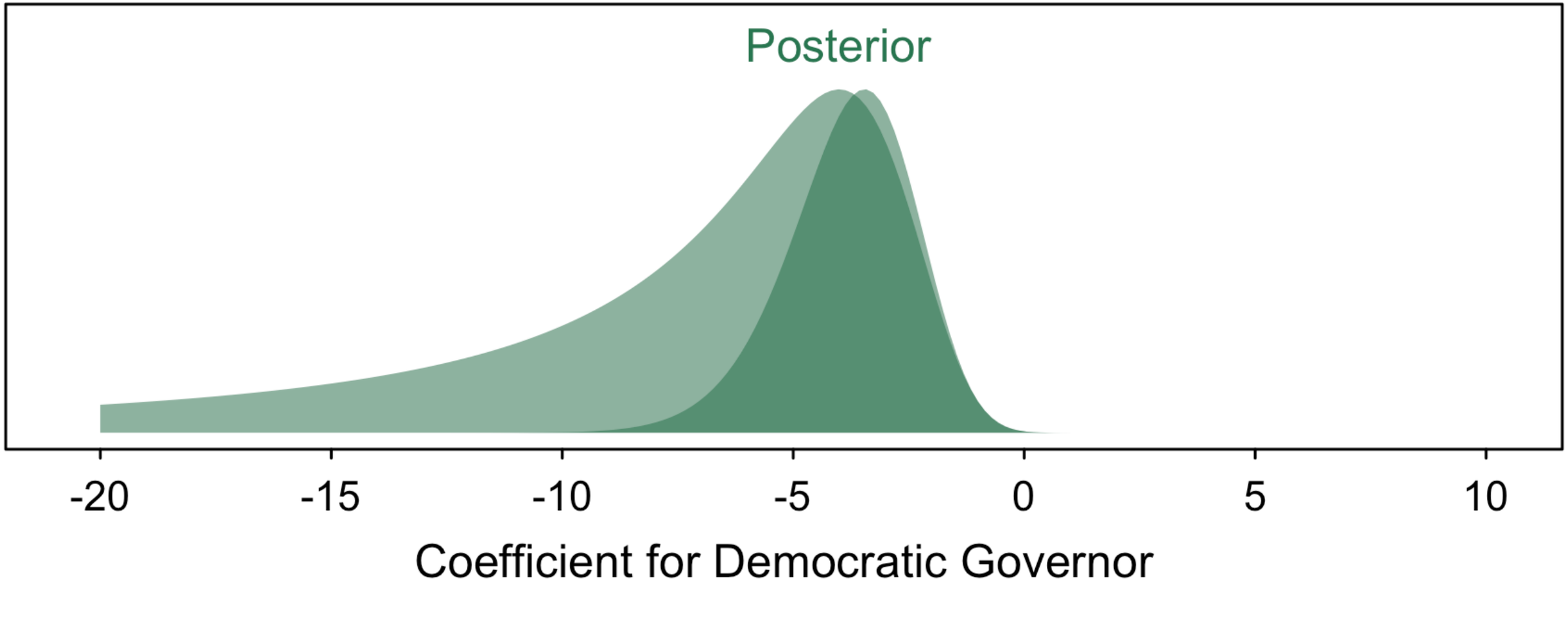
the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



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.

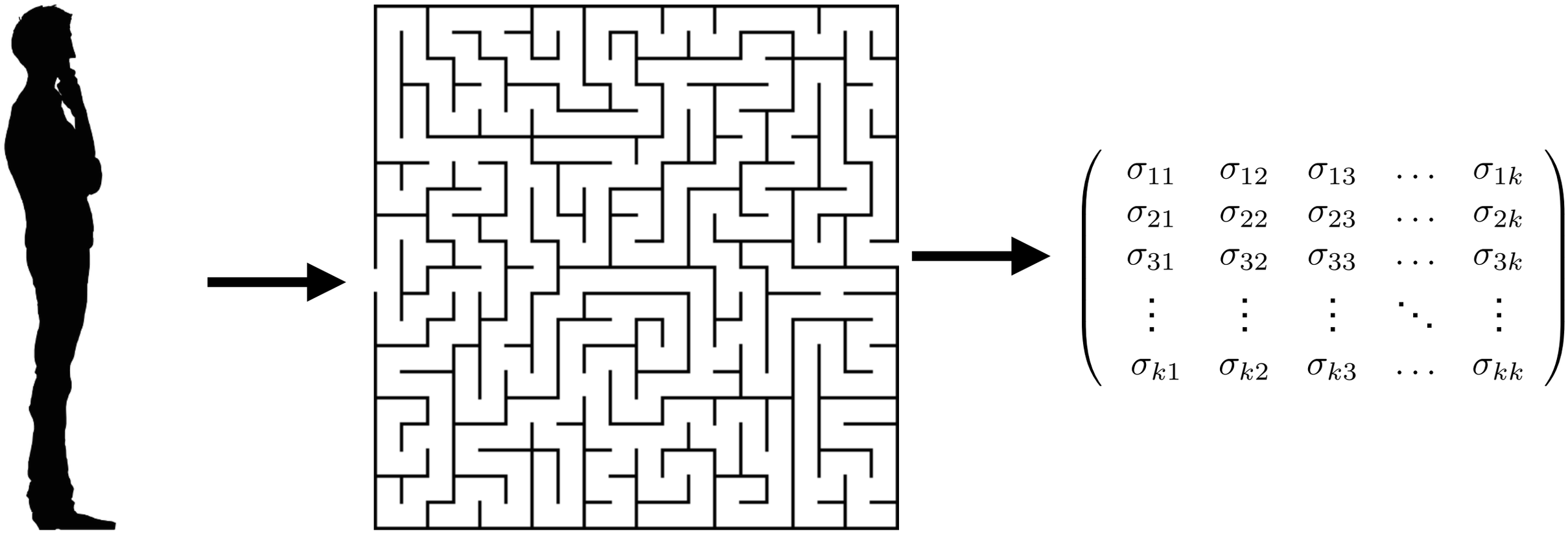
Key Concepts

for Choosing a Good Prior

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

Prior Predictive Distribution

$$p(y_{new}) = \int_{-\infty}^{\infty} p(y_{new}|\beta)p(\beta)d(\beta)$$



$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2k} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \cdots & \sigma_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \sigma_{k3} & \cdots & \sigma_{kk} \end{pmatrix}$$

simplify

We Already Know Few Things

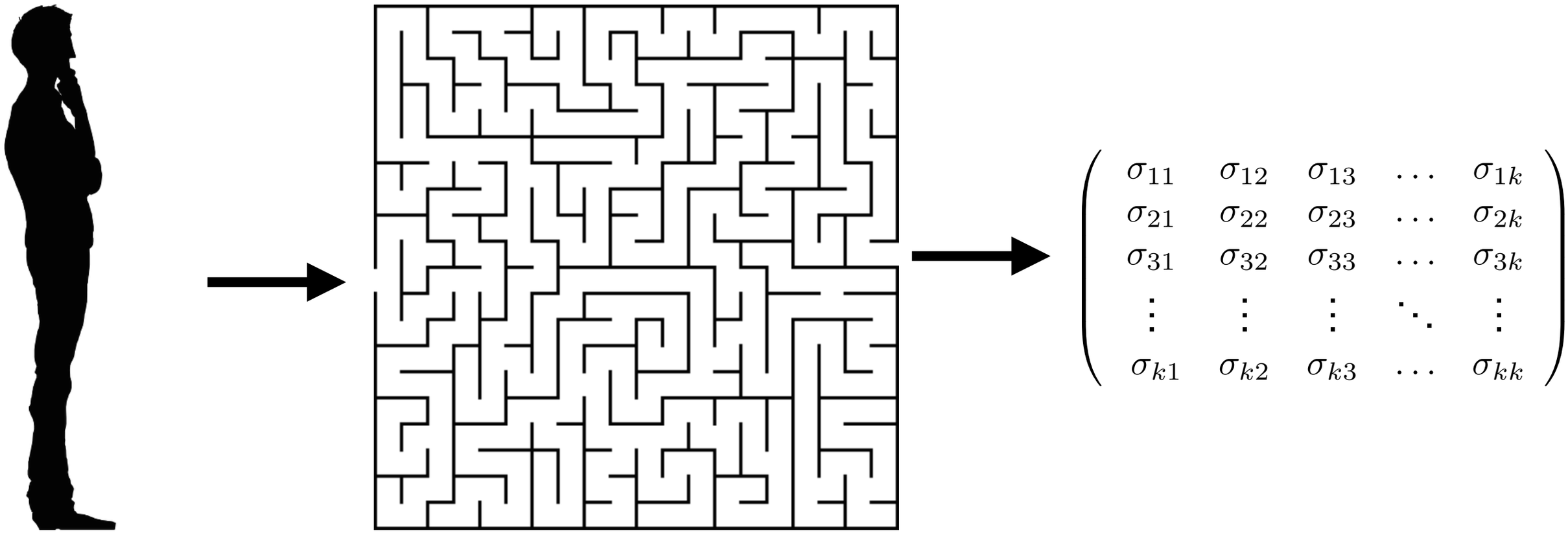
$$\beta_1 \approx \hat{\beta}_1^{mle}$$

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

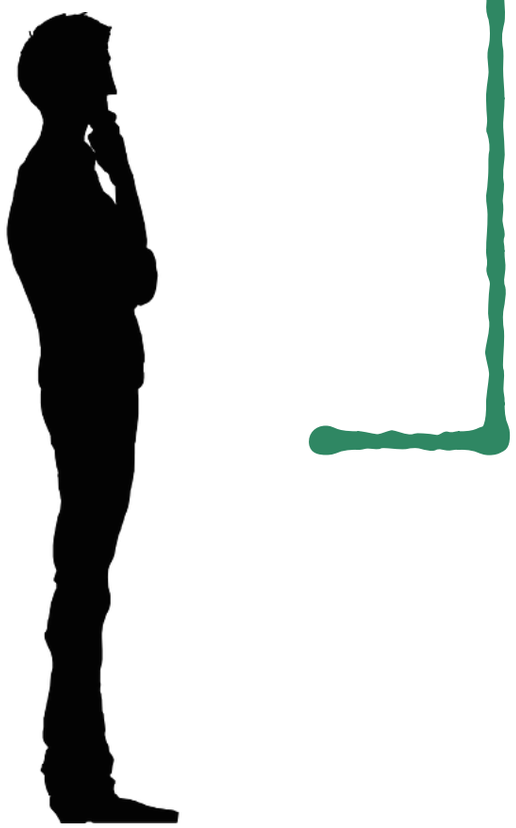
$$\vdots$$

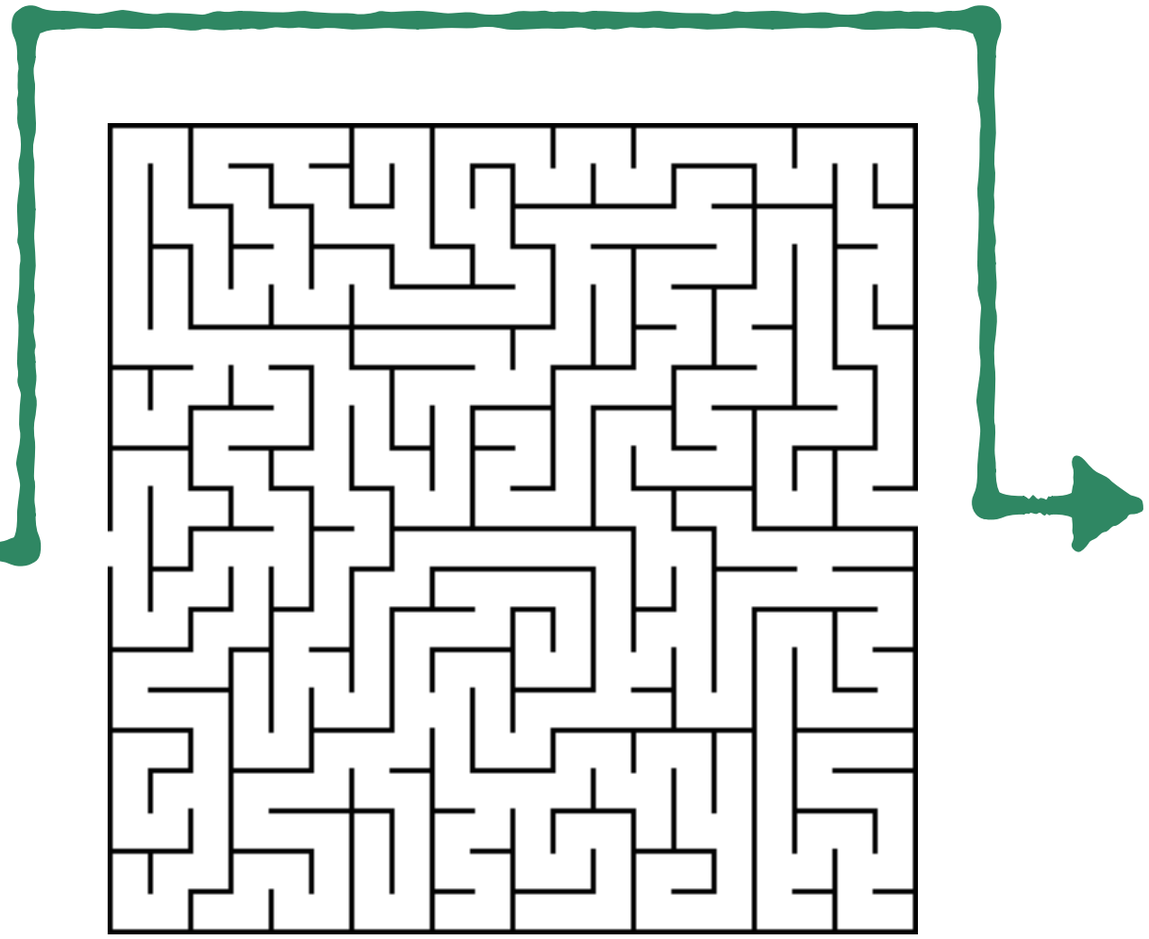
$$\beta_k \approx \hat{\beta}_k^{mle}$$

$$\beta_s < 0$$



$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2k} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \cdots & \sigma_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \sigma_{k3} & \cdots & \sigma_{kk} \end{pmatrix}$$





$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2k} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \cdots & \sigma_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \sigma_{k3} & \cdots & \sigma_{kk} \end{pmatrix}$$

Partial Prior Predictive Distribution

$$p^*(y_{new}) = \int_{-\infty}^{0} p(y_{new}|\beta_s, \hat{\beta}_{-s}^{mle}) p(\beta_s|\beta_s \leq 0) d(\beta_s)$$

1. Choose a prior distribution $p(\beta_s)$.
2. Estimate the model coefficients $\hat{\beta}^{mle}$.
3. For i in 1 to n_{sims}, do the following:
   (a) Simulate $\tilde{\beta}_s^{[i]} \sim p(\beta_s)$.
   (b) Replace $\hat{\beta}_s^{mle}$ in $\hat{\beta}^{mle}$ with $\tilde{\beta}_s^{[i]}$, yielding the vector $\tilde{\beta}^{[i]}$.
   (c) Calculate and store the quantity of interest $\tilde{q}^{[i]} = q\left(\tilde{\beta}^{[i]}\right)$.
4. Keep only the simulations in the direction of the separation.
5. Summarize the simulations $\tilde{q}$ using quantiles, histograms, or density plots.
6. If the prior is inadequate, then update the prior distribution $p(\beta_s)$.

Example

Nuclear Weapons and War

Table 1. The Ratio of the Probability of War in a Nonnuclear Dyad to That in a Nuclear Dyad According to (1) GEE and (2) Firth Logit, Including and Excluding the Kargil War, along with 95 Percent Confidence Intervals.[a]

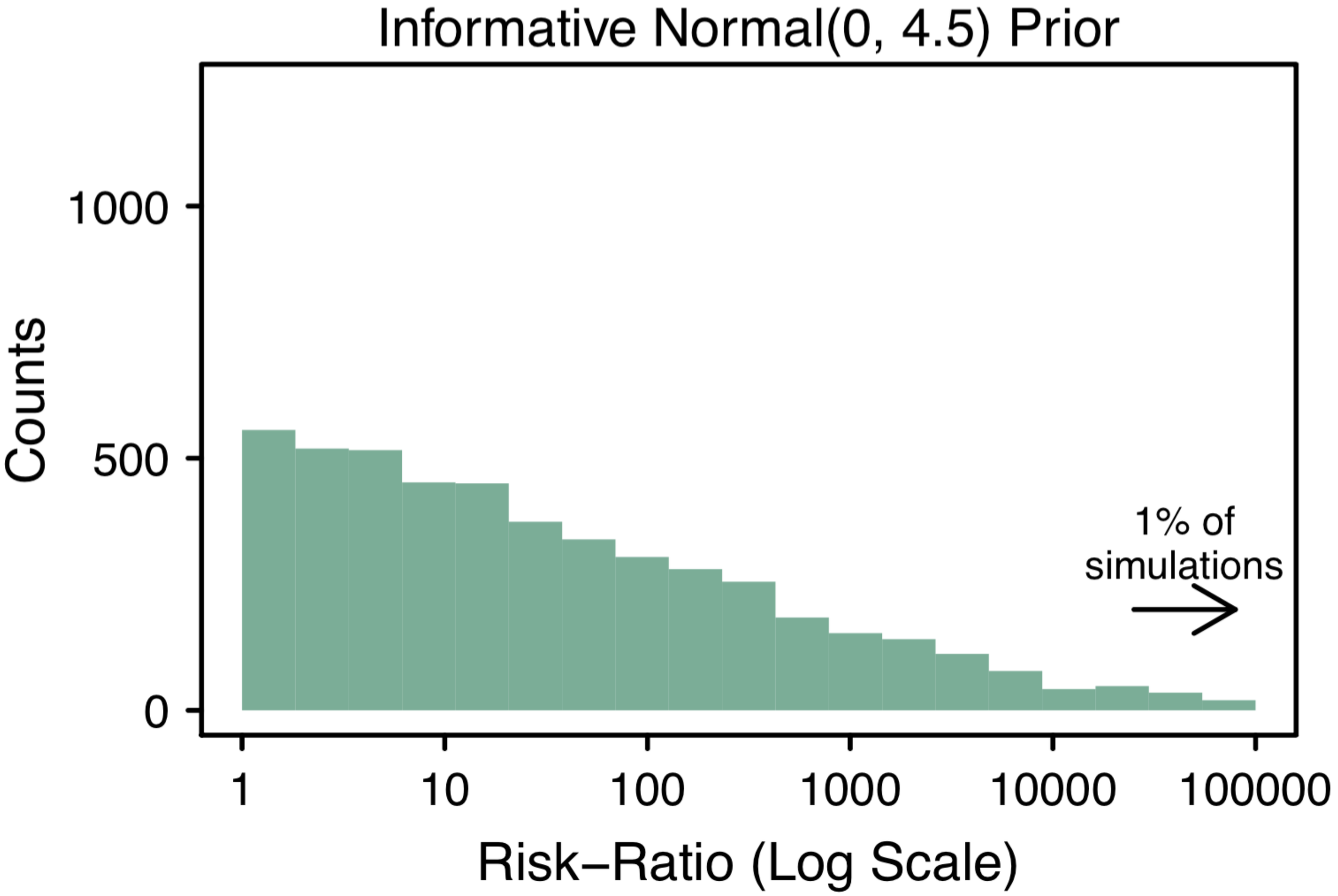
	(1) GEE		(2) Firth logit	
	Estimate	95 percent confidence interval	Estimate	95 percent confidence interval
Kargil excluded	2,717,000	[893,000, 8,531,000]	1.606	[0.088, 30.079]
Kargil included	0.693	[0.545, 9.563]	0.471	[0.077, 2.985]

Note: GEE = generalized estimating equation. All other covariates are held constant at their median values. An estimate of 1 indicates equal probability of war in nuclear and nonnuclear dyads. Ratios are simulated as recommended by King, Tomz, and Wittenberg (2000) and are based on 1,000 simulations.
[a]All models use the same battery of control variables used by Rauchhaus: contiguity, distance between states, state capabilities, presence of an alliance, a dummy for major powers, democracy, economic interdependence, and intergovernmental organization membership. Full regression tables for all models run are included in the online appendix.

Table 1. The Ratio of the Probability of War in a Nonnuclear Dyad to That in a Nuclear Dyad According to (1) GEE and (2) Firth Logit, Including and Excluding the Kargil War, along with 95 Percent Confidence Intervals.[a]

	(1) GEE		(2) Firth logit	
	Estimate	95 percent confidence interval	Estimate	95 percent confidence interval
Kargil excluded	2,717,000	[893,000, 8,531,000]	1.606	[0.088, 30.079]
Kargil included	0.693	[0.545, 9.563]	0.471	[0.077, 2.985]

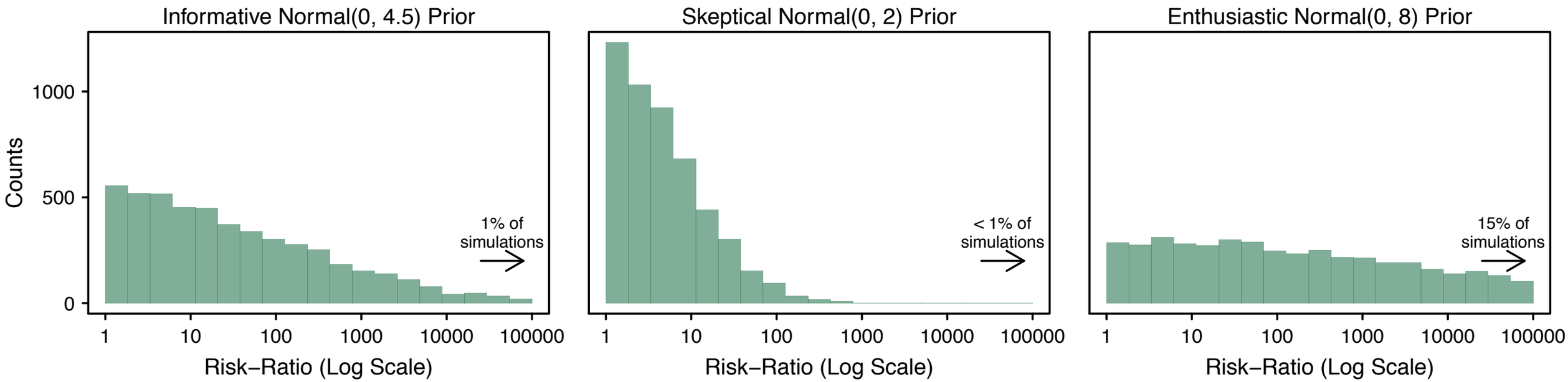
Note: GEE = generalized estimating equation. All other covariates are held constant at their median values. An estimate of 1 indicates equal probability of war in nuclear and nonnuclear dyads. Ratios are simulated as recommended by King, Tomz, and Wittenberg (2000) and are based on 1,000 simulations.
[a]All models use the same battery of control variables used by Rauchhaus: contiguity, distance between states, state capabilities, presence of an alliance, a dummy for major powers, democracy, economic interdependence, and intergovernmental organization membership. Full regression tables for all models run are included in the online appendix.

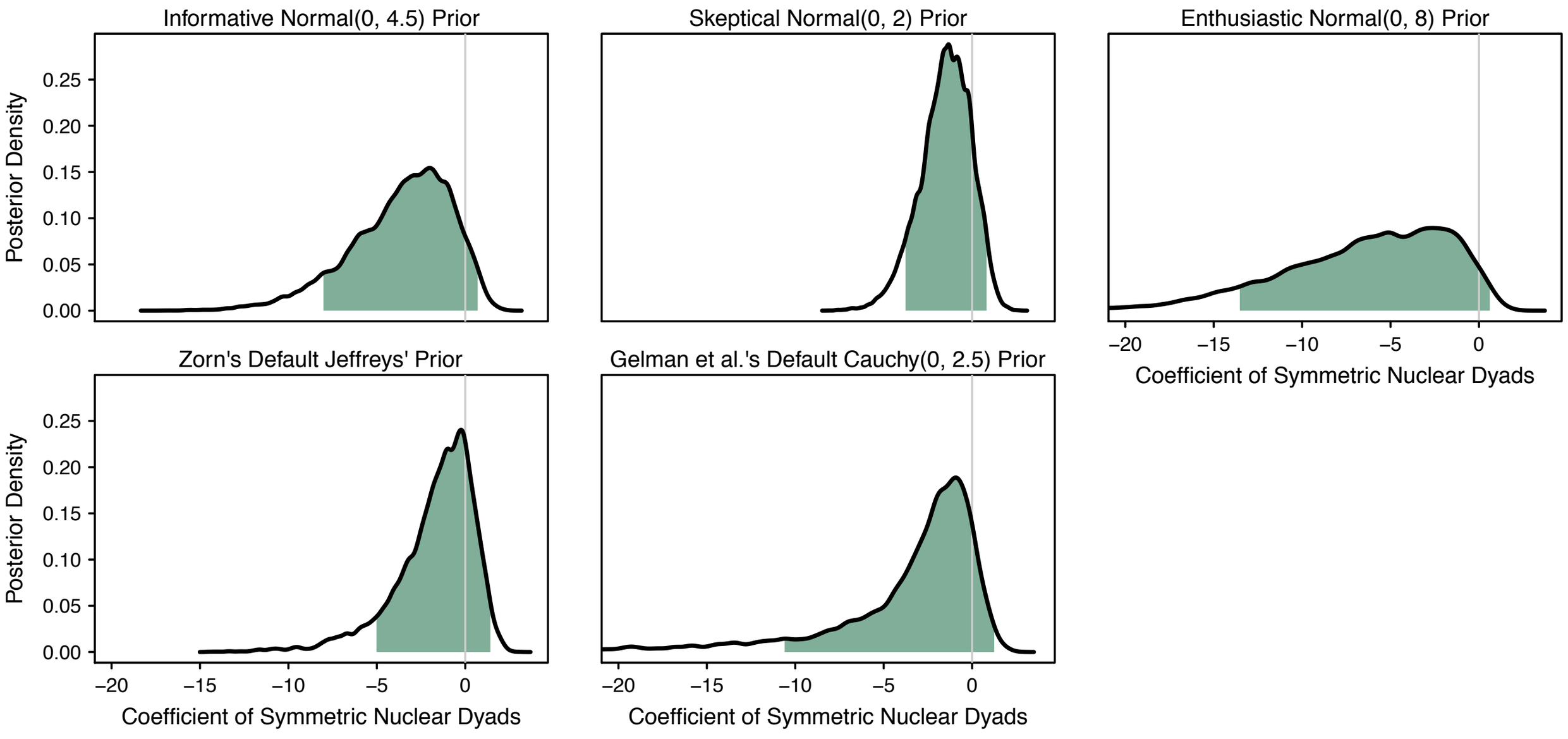

Informative Normal(0, 4.5) Prior
Counts
1000
500
0
1% of
simulations
1
10
100
1000
10000
100000
Risk−Ratio (Log Scale)

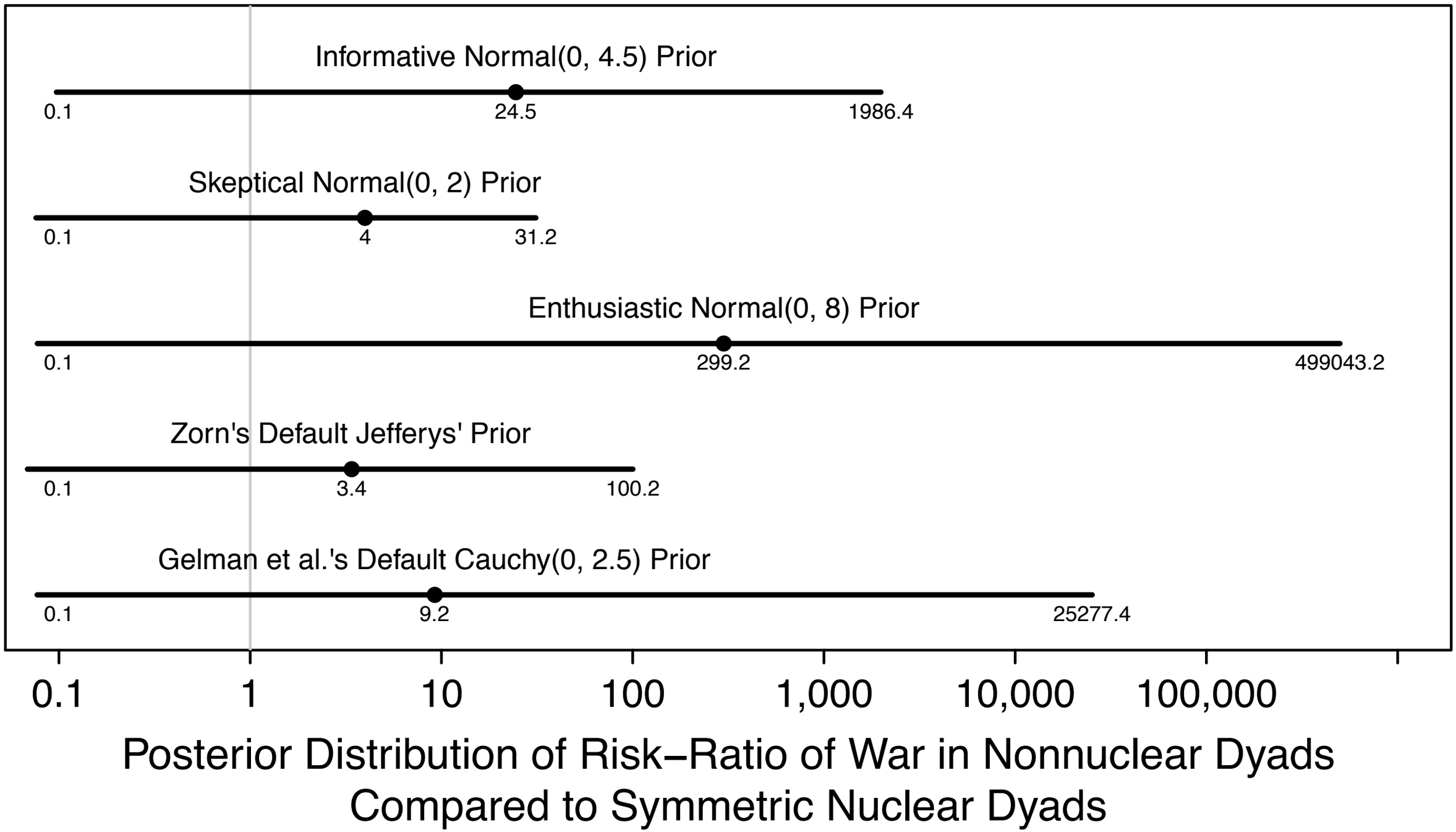
The prior matters,

so ***robustness checks***

are critical.


Informative Normal(0, 4.5) Prior
Skeptical Normal(0, 2) Prior
Enthusiastic Normal(0, 8) Prior
Counts
1000
500
0
1% of simulations
< 1% of simulations
15% of simulations
1
10
100
1000
10000
100000
Risk−Ratio (Log Scale)


Informative Normal(0, 4.5) Prior
Skeptical Normal(0, 2) Prior
Enthusiastic Normal(0, 8) Prior
Zorn's Default Jeffreys' Prior
Gelman et al.'s Default Cauchy(0, 2.5) Prior
Posterior Density
0.25
0.20
0.15
0.10
0.05
0.00
−20
−15
−10
−5
0
Coefficient of Symmetric Nuclear Dyads


Informative Normal(0, 4.5) Prior
0.1
24.5
1986.4
Skeptical Normal(0, 2) Prior
0.1
4
31.2
Enthusiastic Normal(0, 8) Prior
0.1
299.2
499043.2
Zorn's Default Jefferys' Prior
0.1
3.4
100.2
Gelman et al.'s Default Cauchy(0, 2.5) Prior
0.1
9.2
25277.4
0.1
1
10
100
1,000
10,000
100,000
Posterior Distribution of Risk–Ratio of War in Nonnuclear Dyads
Compared to Symmetric Nuclear Dyads

Software

for Choosing a Good Prior

separation

(on GitHub)

crain.co/example

```r
# install packages
devtools::install_github("carlislerainey/compactr")
devtools::install_github("carlislerainey/separation")

# load packages
library(separation)
library(arm)  # for rescale()

# load and recode data
data(politics_and_need)
d <- politics_and_need
d$dem_governor <- 1 - d$gop_governor
d$st_percent_uninsured <- rescale(d$percent_uninsured)

# formula to use throughout
f <- oppose_expansion ~ dem_governor +
  percent_favorable_aca + gop_leg +
  st_percent_uninsured + bal2012 +
  multiplier + percent_nonwhite +
  percent_metro
```

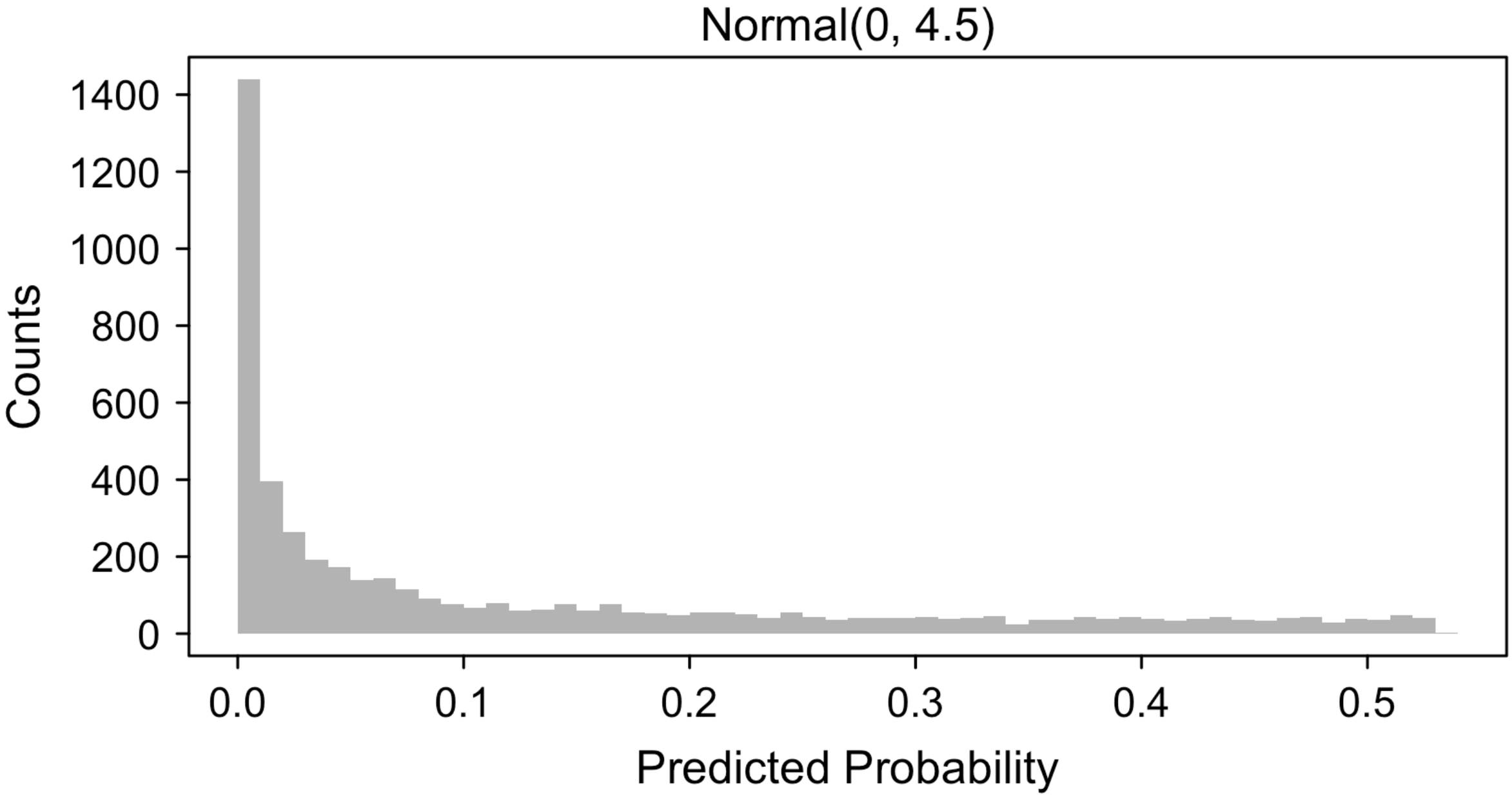
Workflow

1. Calculate the PPPD: `calc_pppd()`
2. Simulate from the posterior: `sim_post_*()`
3. Calculate quantities of interest: `calc_qi()`

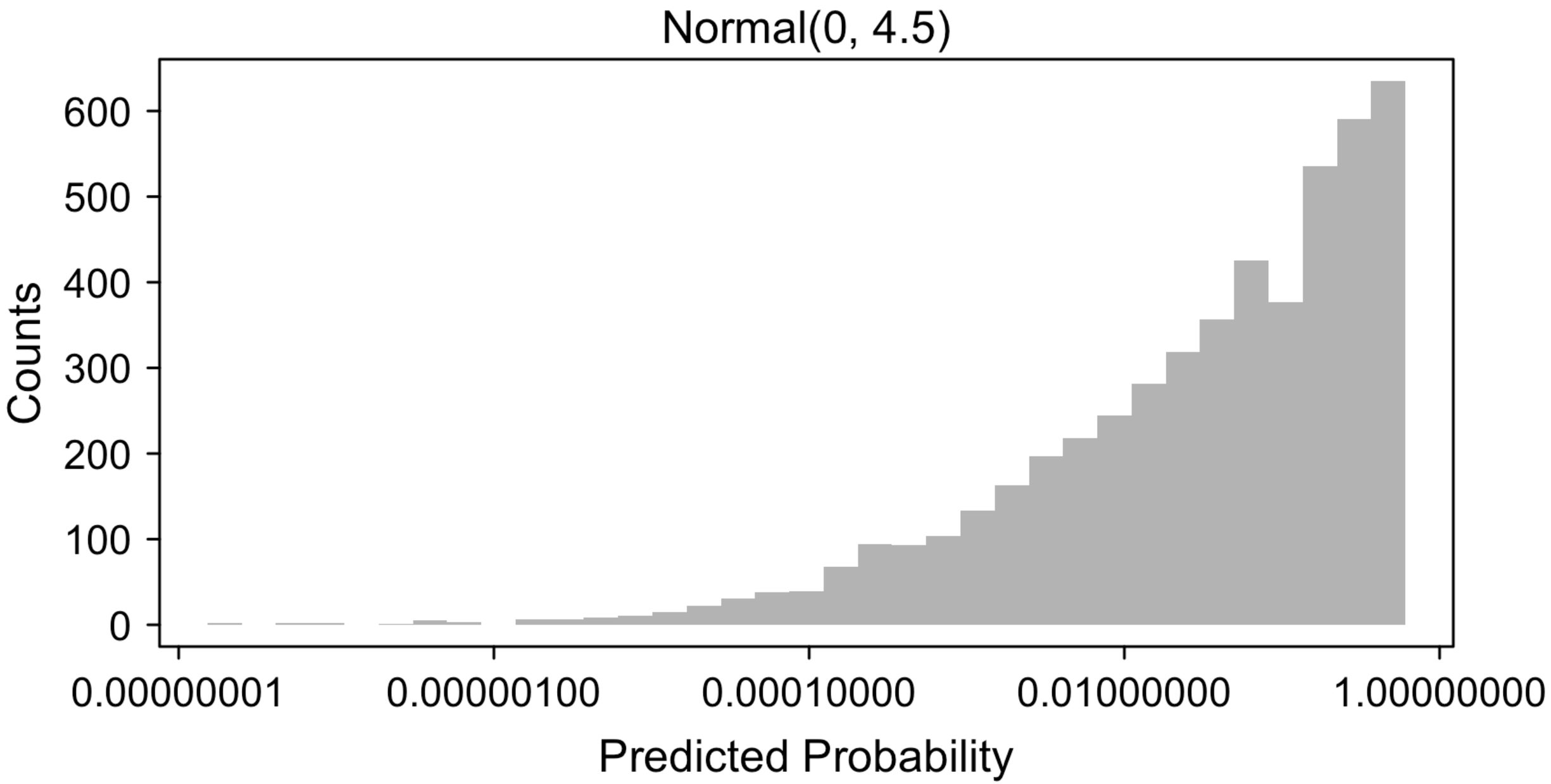
calc_pppd()

```r
# informative prior
prior_sims_4.5 <- rnorm(10000, 0, 4.5)
pppd <- calc_pppd(formula = f,
                  data = d,
                  prior_sims = prior_sims_4.5,
                  sep_var_name = "dem_governor",
                  prior_label = "Normal(0, 4.5)")
```

```
plot(pppd)
```


Normal(0, 4.5)
Counts
1400
1200
1000
800
600
400
200
0
0.0
0.1
0.2
0.3
0.4
0.5
Predicted Probability

```
plot(pppd, log_scale = TRUE)
```



sim_post_normal()

sim_post_gelman()

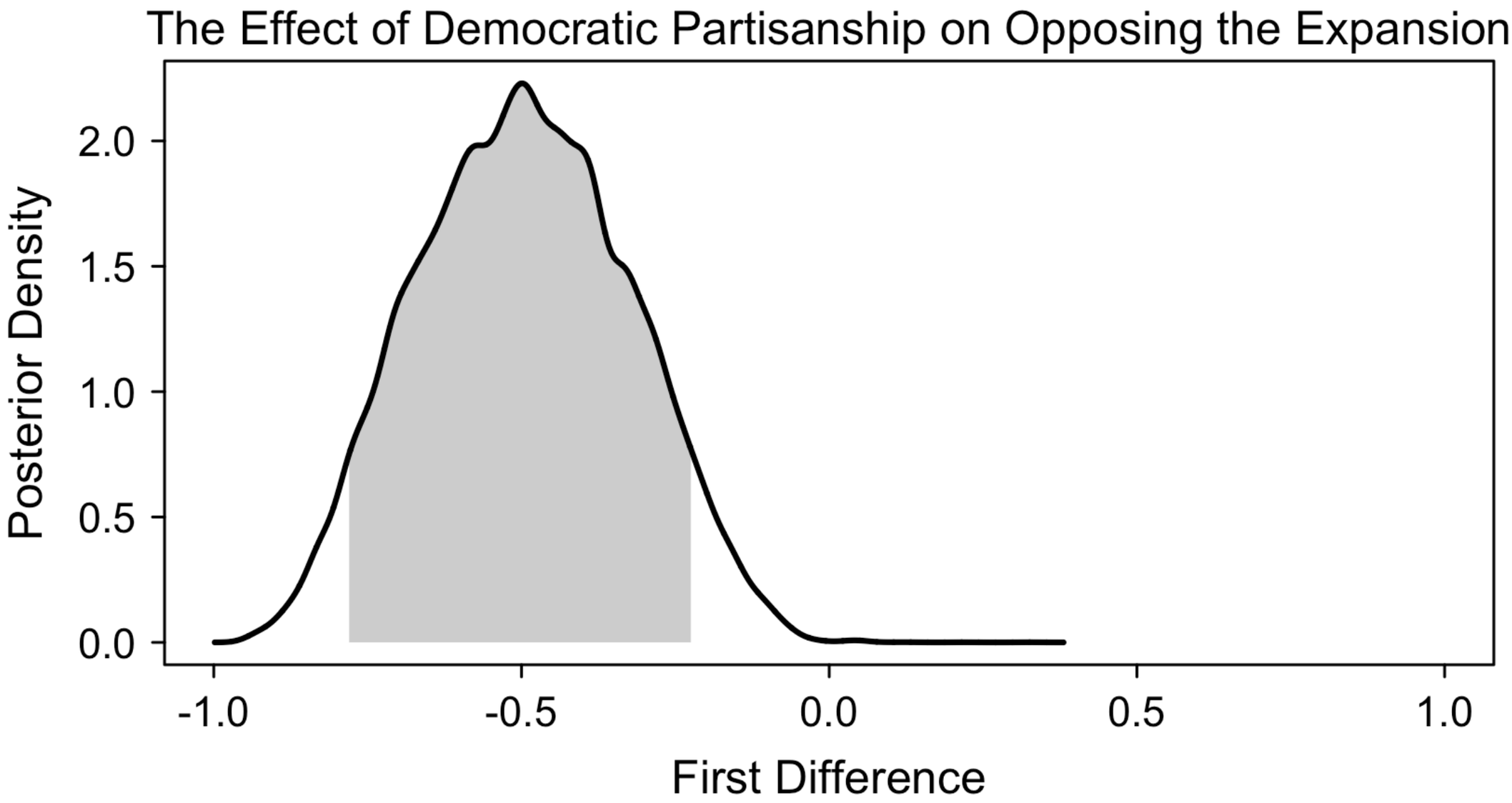
sim_post_jeffreys()

```r
# mcmc estimation
post <- sim_post_normal(f, d, sep_var = "dem_governor",
                        sd = 4.5,
                        n_sims = 10000,
                        n_burnin = 1000,
                        n_chains = 4)
```

calc_qi()

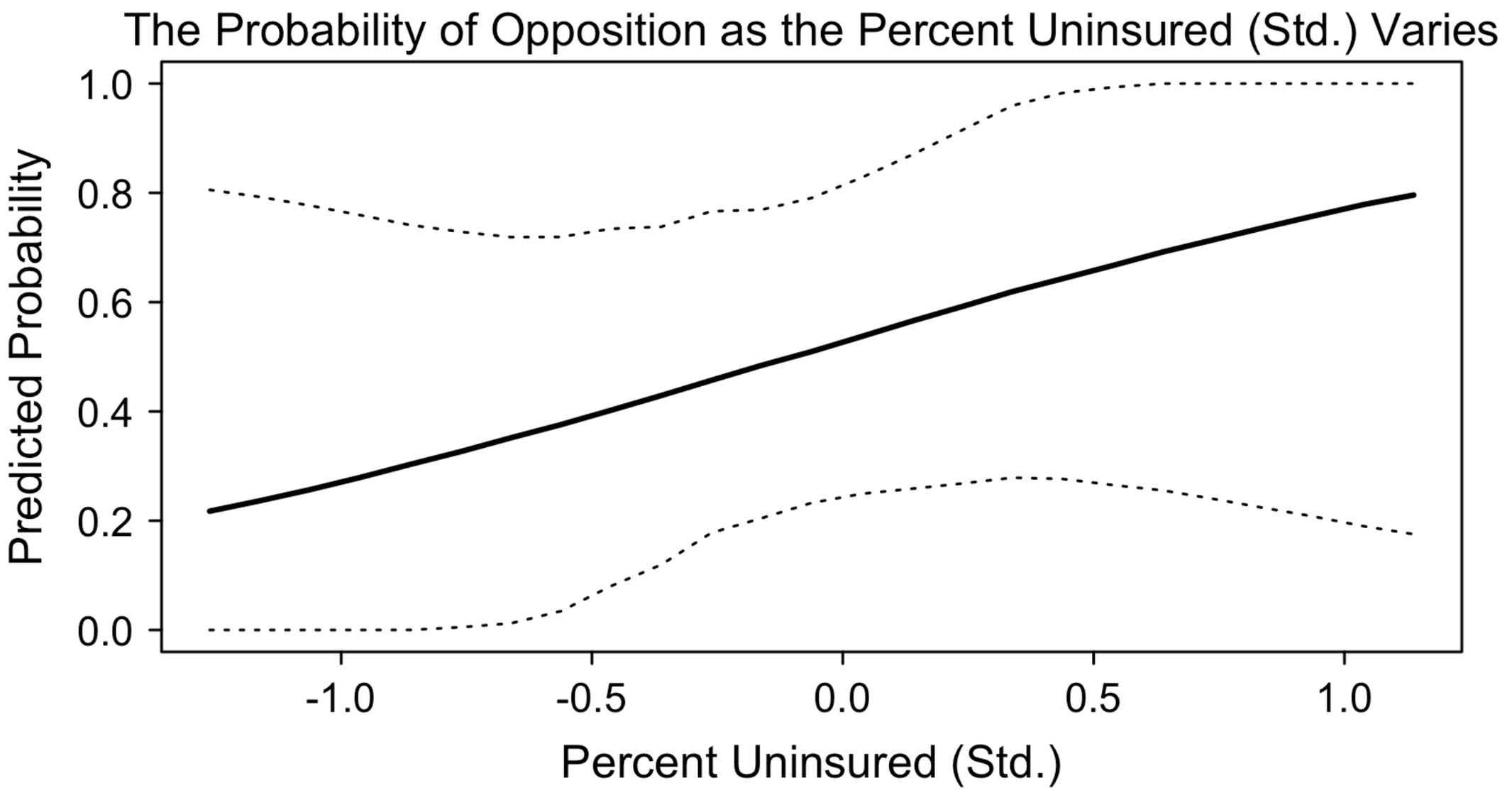
```
# compute quantities of interest

## dem_governor
X_pred_list <- set_at_median(f, d)
x <- c(0, 1)
X_pred_list$dem_governor <- x
qi <- calc_qi(post, X_pred_list, qi_name = "fd")
```

```
plot(qi, xlim = c(-1, 1),
     xlab = "First Difference",
     ylab = "Posterior Density",
     main = "The Effect of Democratic Partisanship on
Opposing the Expansion")
```



```r
## st_percent_uninsured
X_pred_list <- set_at_median(f, d)
x <- seq(min(d$st_percent_uninsured),
         max(d$st_percent_uninsured),
         by = 0.1)
X_pred_list$st_percent_uninsured <- x
qi <- calc_qi(post, X_pred_list, qi_name = "pr")
```

```
plot(qi, x,
     xlab = "Percent Uninsured (Std.)",
     ylab = "Predicted Probability",
     main = "The Probability of Opposition as the
Percent Uninsured (Std.) Varies")
```



15 lines

Conclusion

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

The prior matters **in practice.**

The prior matters

in theory.

The **partial prior predictive distribution** simplifies the choice of prior.

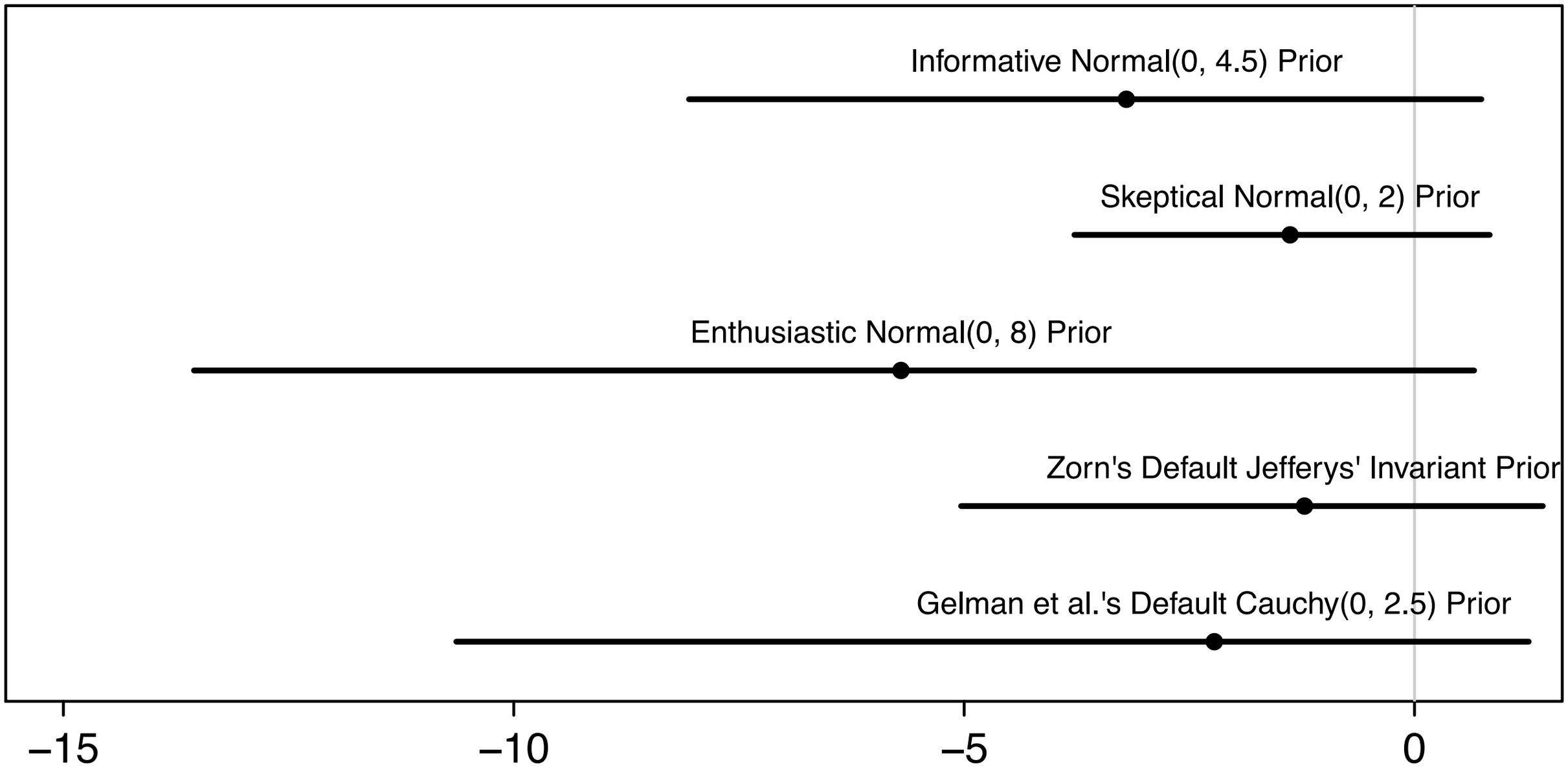
Software makes choosing a prior, estimating the model, and interpreting the estimates easy.

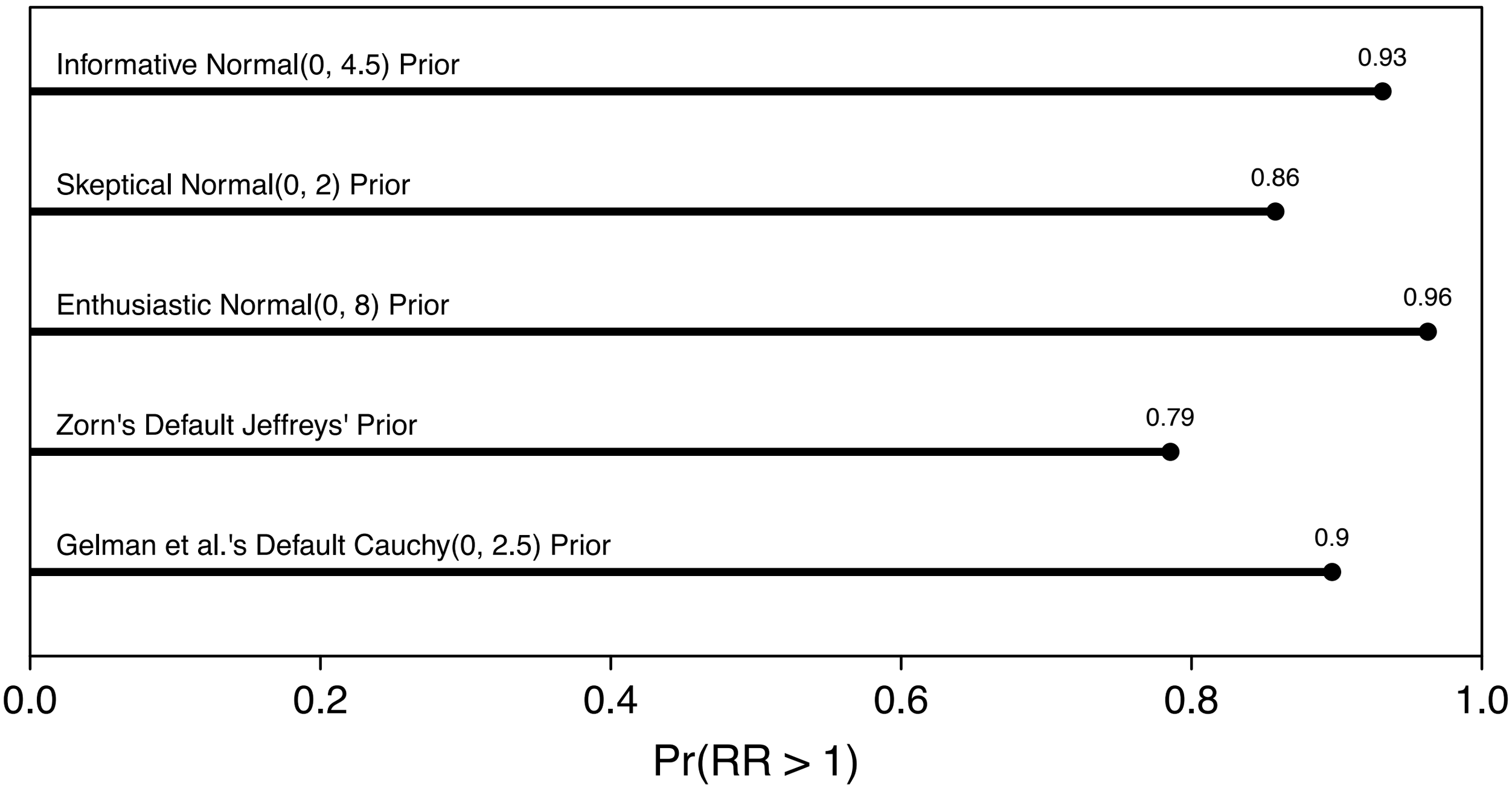
What should ***you*** do?

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.

Questions?

Appendix

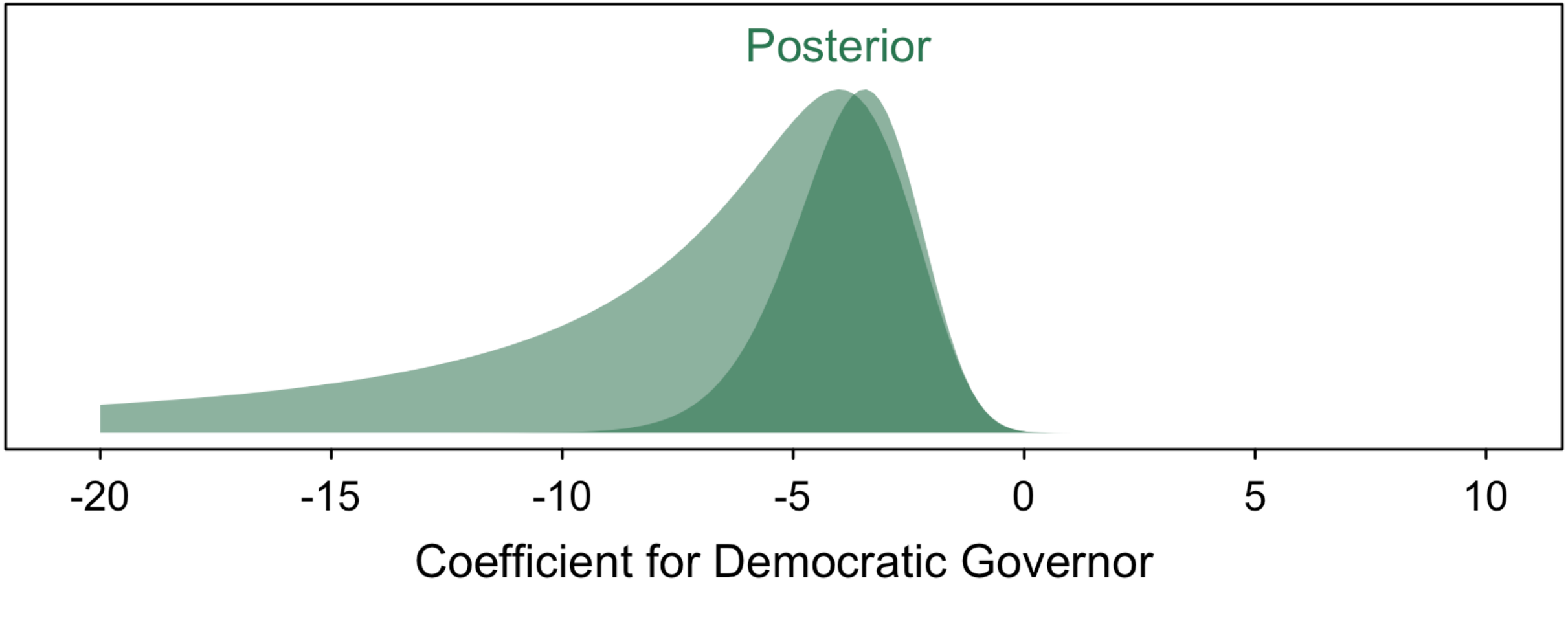

Informative Normal(0, 4.5) Prior
Skeptical Normal(0, 2) Prior
Enthusiastic Normal(0, 8) Prior
Zorn's Default Jefferys' Invariant Prior
Gelman et al.'s Default Cauchy(0, 2.5) Prior
−15
−10
−5
0
Posterior Median and 90% HPD for
Coefficient of Symmetric Nuclear Dyads


Informative Normal(0, 4.5) Prior
0.93
Skeptical Normal(0, 2) Prior
0.86
Enthusiastic Normal(0, 8) Prior
0.96
Zorn's Default Jeffreys' Prior
0.79
Gelman et al.'s Default Cauchy(0, 2.5) Prior
0.9
0.0
0.2
0.4
0.6
0.8
1.0
Pr(RR > 1)

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 $p(\beta_s|y) \propto p(\beta_s|\sigma)$.



Normal(0, 2.25)

Normal(0, 4.5)

Normal(0, 9)

Pr(Research Hyp.) = 0.99

Pr(Research Hyp.) = 1

Pr(Research Hyp.) = 1

Pr(Research Hyp.) = 0.99

Pr(Research Hyp.) = 0.99

Coefficient for Percent Uninsured

10

5

0

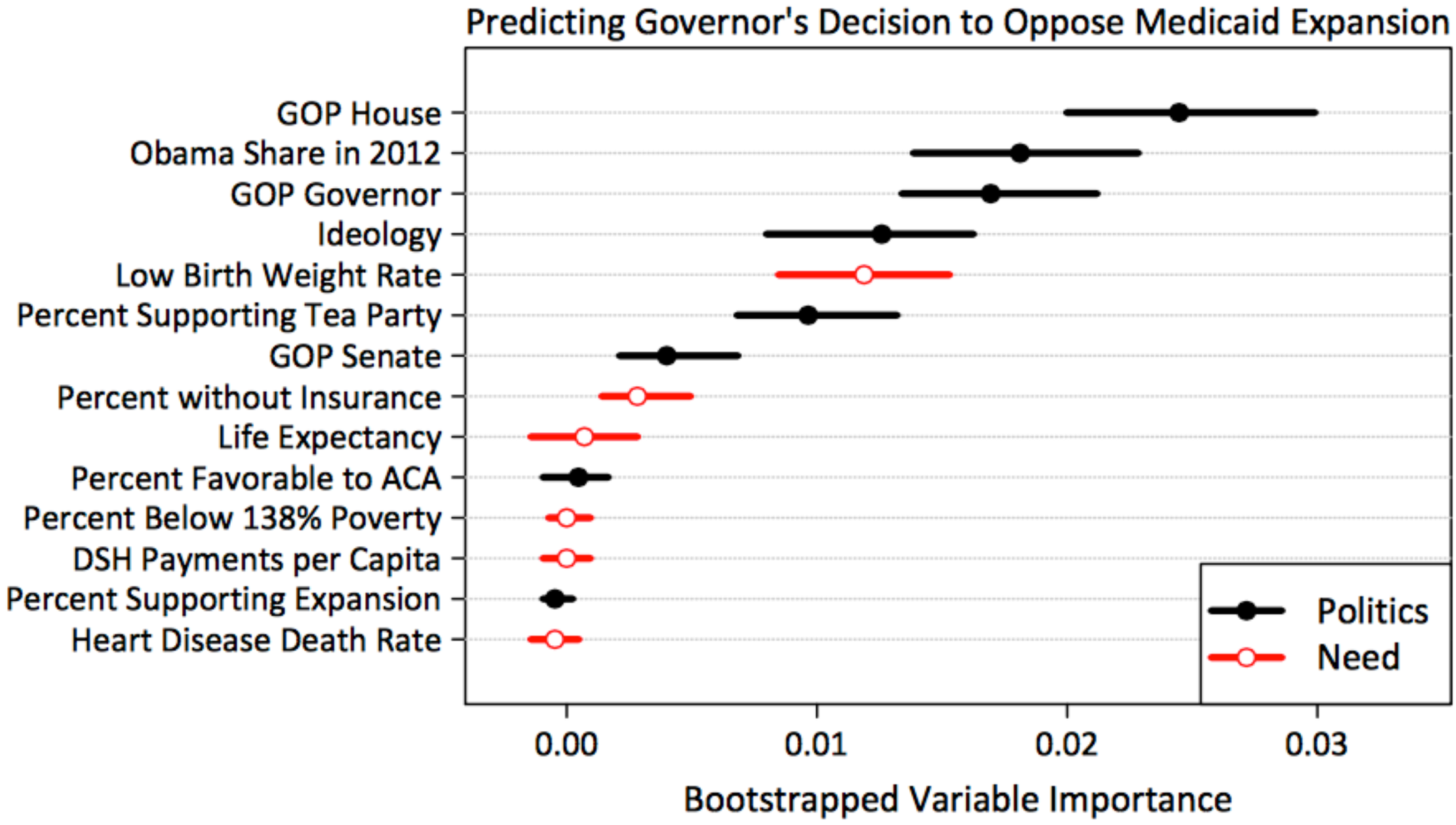
-5

-30 -25 -20 -15 -10 -5 0

Coefficient for Democratic Governor

Jeffreys' Prior

Gelman et al.'s Prior


Predicting Governor's Decision to Oppose Medicaid Expansion
GOP House
Obama Share in 2012
GOP Governor
Ideology
Low Birth Weight Rate
Percent Supporting Tea Party
GOP Senate
Percent without Insurance
Life Expectancy
Percent Favorable to ACA
Percent Below 138% Poverty
DSH Payments per Capita
Percent Supporting Expansion
Heart Disease Death Rate
Politics
Need
0.00
0.01
0.02
0.03
Bootstrapped Variable Importance

Hypothesis	$Pr(H_r \mid \text{data})$	**Evidence**
Gubernatorial Partisanship	> 0.99	Strong Evidence in Favor
Public Opinion	0.57	Ambiguous Evidence
Legislative Partisanship	0.97	Strong Evidence in Favor
Need	0.25	Ambiguous Evidence

Hypothesis	$Pr(H_r \mid \text{data})$	**Evidence**
The effect of gubernatorial partisanship is larger than the effect of need.	> 0.99	Strong Evidence in Favor
The effect of public opinion is larger than the effect of need.	0.75	Ambiguous Evdience
The effect of legislative partisanship is larger than the effect of need.	0.96	Strong Evidence in Favor

Hypothesis	Variable	Expectation	Estimate	$Pr(H_r \mid data)$	Evidence
Public Opinion	**Percent Favorable to ACA**	**-**	**-0.21**	**0.57**	**Ambiguous Evidence**
	Obama's 2012 Vote Share	-	-0.82	0.76	Ambiguous Evidence
	Obama Victory in 2012	-	-1.85	0.94	Moderate Evidence in Favor
	State Ideology	+	2.33	0.92	Moderate Evidence in Favor
	Percent Supporting Medicaid Expansion	-	1.01	0.17	Ambiguous Evidence
	Percent Supporting Tea Party	+	1.09	0.82	Ambiguous Evidence
Legislative Composition	**GOP Controls Both House and Senate**	**+**	**2.31**	**0.97**	**Strong Evidence in Favor**
	GOP House	+	6.48	1.00	Strong Evidence in Favor
	GOP Senate	+	2.17	0.97	Strong Evidence in Favor
	GOP House GOP Senate (as separate variables in the model)	+ +	6.44 0.21	0.98 0.55	Strong Evidence in Favor Ambiguous Evidence
Need	**Percent Without Health Insurance**	**-**	**0.91**	**0.25**	**Ambiguous Evidence**
	DSH Payments per Capita	-	1.21	0.10	Moderate Evidence Against
	Percent Below 138% Poverty	-	0.53	0.35	Ambiguous Evidence
	Low Birth Weight	-	2.76	0.02	Strong Evidence Against
	Heart Disease Death Rate	-	1.23	0.09	Moderate Evidence Against
	Life Expectancy	+	-1.50	0.09	Moderate Evidence Against

Theorem 1. *For a monotonic likelihood $p(y|\beta)$ increasing [decreasing] in β_s, proper prior distribution $p(\beta|\sigma)$, and large positive [negative] β_s, the posterior distribution of β_s is proportional to the prior distribution for β_s, so that $p(\beta_s|y) \propto p(\beta_s|\sigma)$.*

Proof. Due to separation, $p(y|\beta)$ is monotonic increasing in β_s to a limit $\underline{\mathscr{L}}$, so that $\lim_{\beta_s \to \infty} p(y|\beta_s) = \underline{\mathscr{L}}$. By Bayes' rule,

$$p(\beta|y) = \frac{p(y|\beta)p(\beta|\sigma)}{\int_{-\infty}^{\infty} p(y|\beta)p(\beta|\sigma)d\beta} = \frac{p(y|\beta)p(\beta|\sigma)}{\underbrace{p(y|\sigma)}_{\text{constant w.r.t. } \beta}}.$$

Integrating out the other parameters $\beta_{-s} = \langle \beta_{cons}, \beta_1, \beta_2, ..., \beta_k \rangle$ to obtain the posterior distribution of β_s,

$$p(\beta_s|y) = \frac{\int_{-\infty}^{\infty} p(y|\beta)p(\beta|\sigma)d\beta_{-s}}{p(y|\sigma)}, \tag{1}$$

and the prior distribution of β_s,

$$p(\beta_s|\sigma) = \int_{-\infty}^{\infty} p(\beta|\sigma)d\beta_{-s}.$$

Notice that $p(\beta_s|y) \propto p(\beta_s|\sigma)$ iff $\frac{p(\beta_s|y)}{p(\beta_s|\sigma)} = k$, where the constant $k \neq 0$.Thus,

$$p(\beta_s|\sigma) = \int_{-\infty}^{\infty} p(\beta|\sigma) d\beta_{-s}.$$

Notice that $p(\beta_s|y) \propto p(\beta_s|\sigma)$ iff $\frac{p(\beta_s|y)}{p(\beta_s|\sigma)} = k$, where the constant $k \neq 0$.Thus, Theorem 1 implies that

$$\lim_{\beta_s \to \infty} \frac{p(\beta_s|y)}{p(\beta_s|\sigma)} = k$$

Substituting in Equation 1,

$$\lim_{\beta_s \to \infty} \frac{\frac{\int_{-\infty}^{\infty} p(y|\beta)p(\beta|\sigma)d\beta_{-s}}{p(y|\sigma)}}{p(\beta_s|\sigma)} = k.$$

Multiplying both sides by $p(y|\sigma)$, which is constant with respect to β,

$$\lim_{\beta_s \to \infty} \frac{\int_{-\infty}^{\infty} p(y|\beta)p(\beta|\sigma)d\beta_{-s}}{p(\beta_s|\sigma)} = kp(y|\sigma).$$

Setting $\int_{-\infty}^{\infty} p(y|\beta)p(\beta|\sigma)d\beta_{-s} = p(y|\beta_s)p(\beta_s|\sigma)$,

$$\lim_{\beta_s \to \infty} p(\beta_s|\sigma)$$

Substituting in Equation 1,

$$\lim_{\beta_s \to \infty} \frac{\frac{\int_{-\infty}^{\infty} p(y|\beta)p(\beta|\sigma)d\beta_{-s}}{p(y|\sigma)}}{p(\beta_s|\sigma)} = k.$$

Multiplying both sides by $p(y|\sigma)$, which is constant with respect to β,

$$\lim_{\beta_s \to \infty} \frac{\int_{-\infty}^{\infty} p(y|\beta)p(\beta|\sigma)d\beta_{-s}}{p(\beta_s|\sigma)} = kp(y|\sigma).$$

Setting $\int_{-\infty}^{\infty} p(y|\beta)p(\beta|\sigma)d\beta_{-s} = p(y|\beta_s)p(\beta_s|\sigma)$,

$$\lim_{\beta_s \to \infty} \frac{p(y|\beta_s)p(\beta_s|\sigma)}{p(\beta_s|\sigma)} = kp(y|\sigma).$$

Canceling $p(\beta_s|\sigma)$ in the numerator and denominator,

$$\lim_{\beta_s \to \infty} p(y|\beta_s) = kp(y|\sigma).$$