tl;dr

When facing separation, the prior you use matters a lot, so you need to choose a reasonable one and/or assess the robustness of your results to alternative priors.

Dealing with Separation in Logistic Regression Models

Carlisle Rainey

paper, code, and data at crain.co

The Problem

When facing separation (i.e., perfect prediction with a logistic regression model), researchers must regularize infinite coefficient estimates.

The literature offers two default priors:

1. Jeffreys Invariant Prior (Zorn 2005)
2. Cauchy Prior (Gelman et al. 2008)

Point 1: The default prior you use matters a lot.

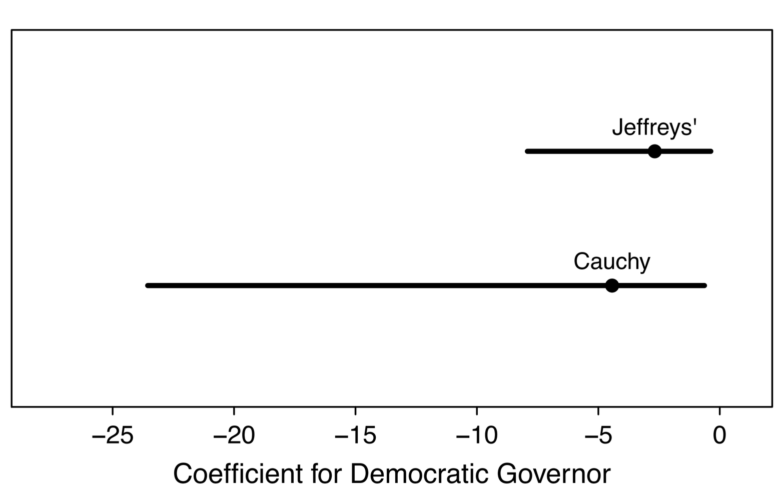
An Application: Barrilleaux and Rainey (2014)

Does a governor's partisanship have a larger effect on her decision to oppose the Medicaid expansion than the level of need in her state?

Maximum likelihood estimates

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]

Bayesian Estimates

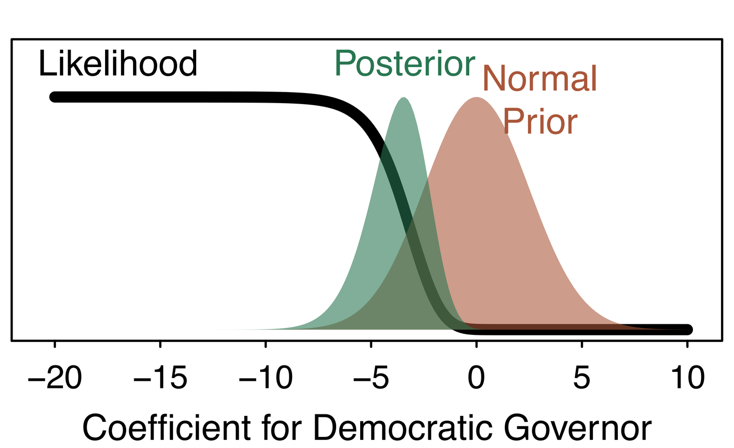


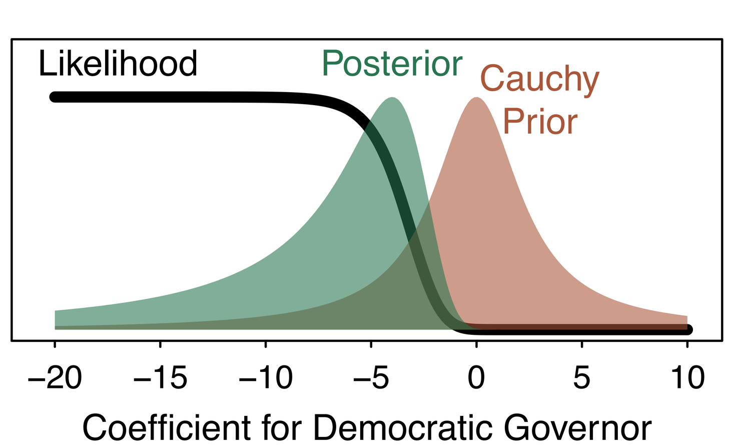
How does a Cauchy prior change the the inferences from Jeffreys prior?

- The posterior for mode is about 50% larger.
- The 90% equal-tailed credible interval is about 250% wider.
- The risk-ratio is 43 *million* times larger.

A Theorem

For a monotonic likelihood increasing [decreasing] in the coefficient of interest, a proper prior distribution, and large positive [negative] value of the coefficient of interest, the prior distribution *determines* the posterior distribution.





Point 2: You need to choose a reasonable prior and perform robustness checks.

The Partial Prior Predictive Distribution

"As I see it, the fundamental problem facing econometrics is how adequately to control the whimsical character of inference, how sensibly to base inferences on opinions when facts are unavailable."

—Leamer (1983)

Prior predictive distribution

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

Just integrate out the coefficients from the prior model and see if the data (or a function of the data) have a reasonable distribution given the prior information. But the prior predictive distribution has too many dimensions to work with for most problems (e.g, eight variables → about 100 million combinations to consider).

Things we know

1. The maximum likelihood estimates for the coefficients of the other (i.e., overlapping) variables are *approximately* correct.
2. The coefficient for the separating variable is likely in the direction of the separation.

Partial prior predictive distribution

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

This is much easier to work with. It allows the researcher to focus on a single dimension and in a single direction (i.e., "how large should the effect be?").

Things We Need

1. *Shrinks toward zero.* The prior distribution largely drives the inferences in the direction of the separation. The researcher simply needs to choose the amount of shrinkage.
2. *Allows plausible effects.* The prior distribution should assign realistic prior probabilities to estimates that are a priori plausible according to the researcher's prior information.
3. *Rules out implausible effects.* The prior distribution should assign essentially no prior probability to estimates that are a priori implausible according to the researcher's prior information.

Where does this prior information come from? Perhaps other quantitative studies on similar topics, detailed studies of particular cases, or theoretical arguments.

Some Software

On GitHub

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

Workflow

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

An Application: Bell and Miller (2014)

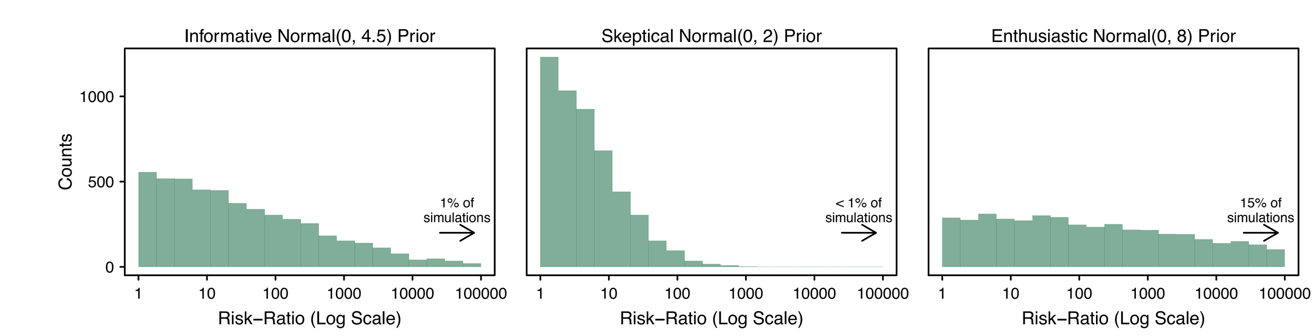
The Debate

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]

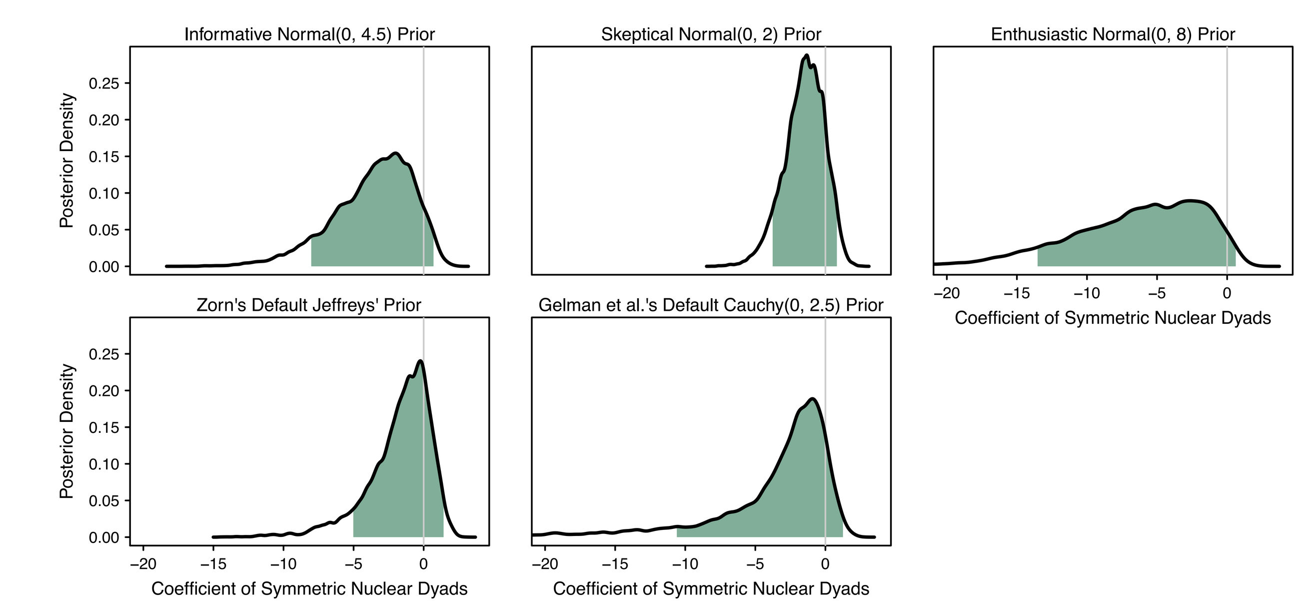
	(1) GEE		(2) Firth logit	
	Estimate	95 percent confidence interval	Estimate	95 percent confidence interval
Kargil excluded	2,717,000	[893,000, 8,531,000]	1.606	[0.088, 30.079]
Kargil included	0.693	[0.545, 9.563]	0.471	[0.077, 2.985]

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.

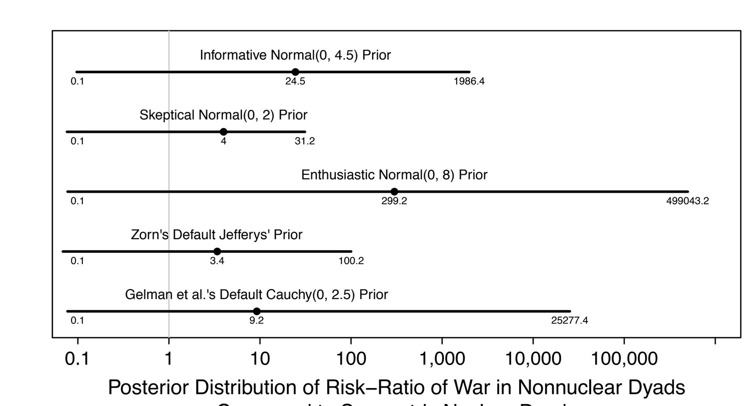
The Priors



The Posteriors



The Risk Ratios



Summary

1. The two default priors produce very different inferences.
2. Zorn's (2005) *default* prior suggests the least amount of evidence for Rauchhaus' hypothesis and the smallest effect—even smaller than the *skeptical* prior.
3. The enthusiastic prior suggests that risk-ratios as large as 500,000 might be reasonable, while the skeptical prior essentially rules out risk-ratios larger than 30.

What should you do?

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