

Appendix of: Why and when do (Czech) judges dissent: a quantitative empirical analysis of the Czech Constitutional Court

ŠTĚPÁN PAULÍK, Humboldt Universität zu Berlin, stepan.paulik.1@hu-berlin.de

GOR VARTAZARYAN, Charles University, gorike2000@gmail.com

Contents

1 Bayesian framework	2
2 Model 1: Effect of presence of dissenting opinion on the length of majority argumentation	3
2.1 Caselaw partitioning classification	3
2.2 Statistical model	3
3 Model 2: Effect of workload on the dissenting behavior	5
3.1 Model diagnosis	5

Contents

This is an appendix, in which we diagnose and compare the models from our main article and explain our choices in higher detail.

In this article, we utilize Bayesian regression analyses. In other words, we rely on quantitative methods of social science research. That is not without its limits. To not rely on one methodological approach, we set this article into a broader mixed effects effort. We conducted semi-structured interviews with the CCC judges to gain deeper understanding of the judicial behavior, to paint more details to our quantitative effort, and to help with further development of our theory. Parts of the interviews mirror our findings in this article. To name an example, in the interviews, we inquire into the effect of the coalitions at the CCC, whose effect we measure in the last part of this article. In terms of the output of our effort, we divide it into two articles: one utilizing the quantitative methods, the other utilizing qualitative methods.

1 Bayesian framework

Without delving too much into the Bayesian versus frequentist statistics, we opt for the Bayesian framework for it, we believe, reflects better our understanding of probability and scientific inquiry. There are two major differences in understanding of concepts between the two approaches towards statistics: that of role of prior knowledge and that of probability.

In the frequentist framework, prior knowledge does not play too much of a role and the inference is shaped solely by the observed data, whereas in the Bayesian framework prior knowledge is updated with new data to form new posterior conclusions. In other words, the Bayesian statistician concerns themselves not with the uncertainty of the data but also with how it fits into his prior knowledge.

That is reflected in different understandings of probability. The frequentist understanding of probability refers to the long-run relative frequency of a repeatable event. In other words, the main concern of frequentist statistics is what would the frequency of any event be if we could repeat it as many times as possible. The Bayesian probability measures the relative plausibility of an event (Johnson, Ott, and Dogucu 2022).

Science in general is based on the frequentist framework. The typical quantitative studies are driven by finding a low enough p-value, i.e., the measure of probability of having observed data as or more extreme than the observed data if in fact the original null hypothesis is incorrect. In simple terms, the search for statistical significance is a search for data so unlikely to have occurred due to chance, even if we could gather them again and again.

The Bayesian framework rather than measuring the uncertainty about observed data measures the uncertainty of the parameters of interests, given the observed data and our prior knowledge. In simple terms, the Bayesian statistician puts into doubt their conclusions about parameters of a certain model, given the observed data and their prior knowledge. Mathematically, the uncertainty is reflected in the fact that the posterior parameters are drawn from a posterior distribution of the model and are just an approximation of thereof in the form of probability density function rather than a single value. The posterior distribution of a parameter comes from simulating in our case 40000 (4 chains*10000 simulations) possible posterior models via the Monte Carlo Markov Chain simulations.

The Lee Epstein et al. original study employs the frequentist framework and hence they report statistically significant relationships between the variables of interest. Our approach will be Bayesian. We will firstly see whether the employed models actually make sense for the data by running posterior predictive checks and afterwards we will draw the parameters from the posterior distribution to see whether their distribution indeed looks similar to the ones reached by Epstein et al.

2 Model 1: Effect of presence of dissenting opinion on the length of majority argumentation

2.1 Caselaw partitioning classification

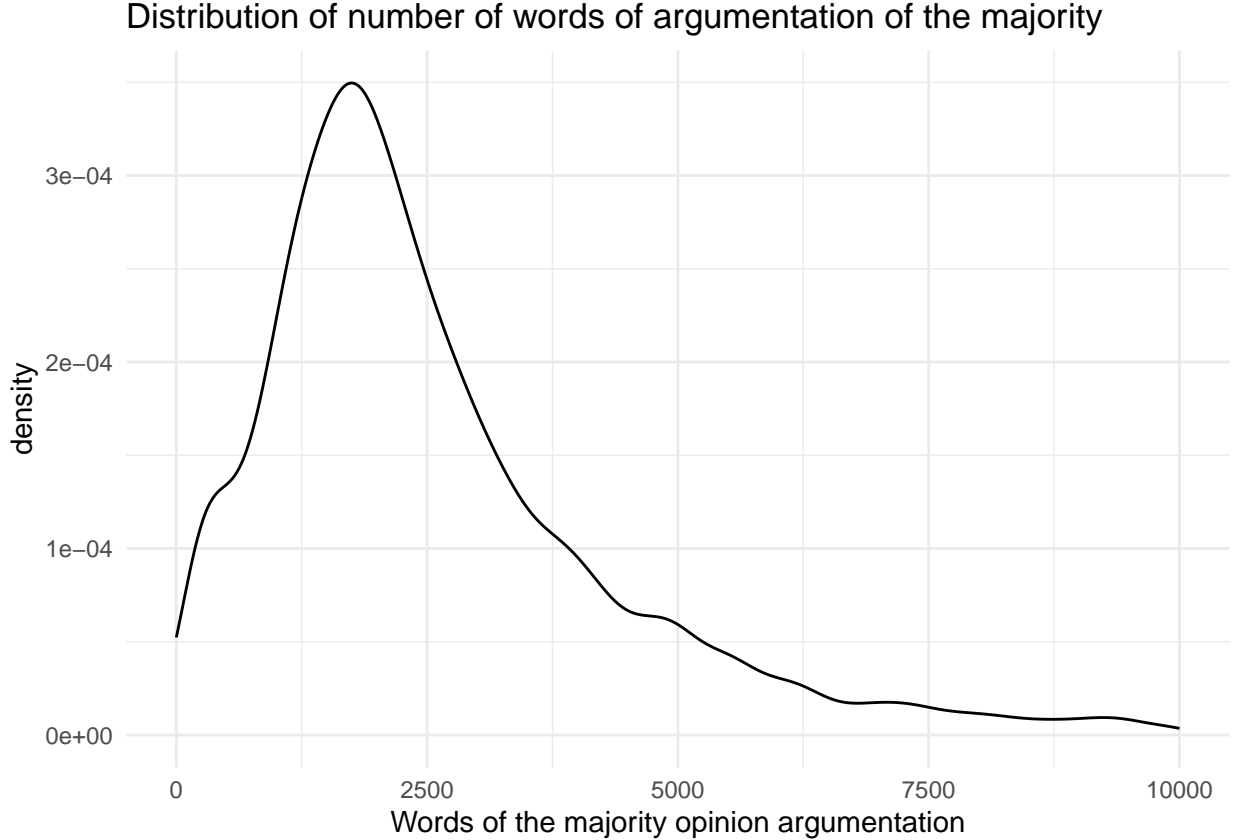
2.2 Statistical model

2.2.1 Model specification

We opted for a completely pooled model as the data did not contain any inherent structure (there were no clusters). At first glance, we assumed that

$$Y|\lambda \sim Pois(\lambda)$$

because our outcome variable of interest is a discrete count and the density plot of the length of court argumentation suggests so.



However, as the posterior checking revealed, the Y was actually overdispersed and the Poisson regression was not able to capture the overdispersion. Therefore, we instead opted for the Negative Binomial model, which allows for relaxing the assumption of equality of variance of Y to its expected value. Thus, the explanatory variable, the number of words of argumentation of the CCC Y

$$Y_{words}|\mu, r \sim NegBin(\mu, r)$$

As for the priors, we based the priors on the Epstein results as well as a cursory exploratory peak into the data. All our priors follow a normal distribution, the intercept being centered around the population mean. The remaining priors were kept uninformative, because we simply have no previous knowledge on the CCC.

2.2.2 Model diagnosis

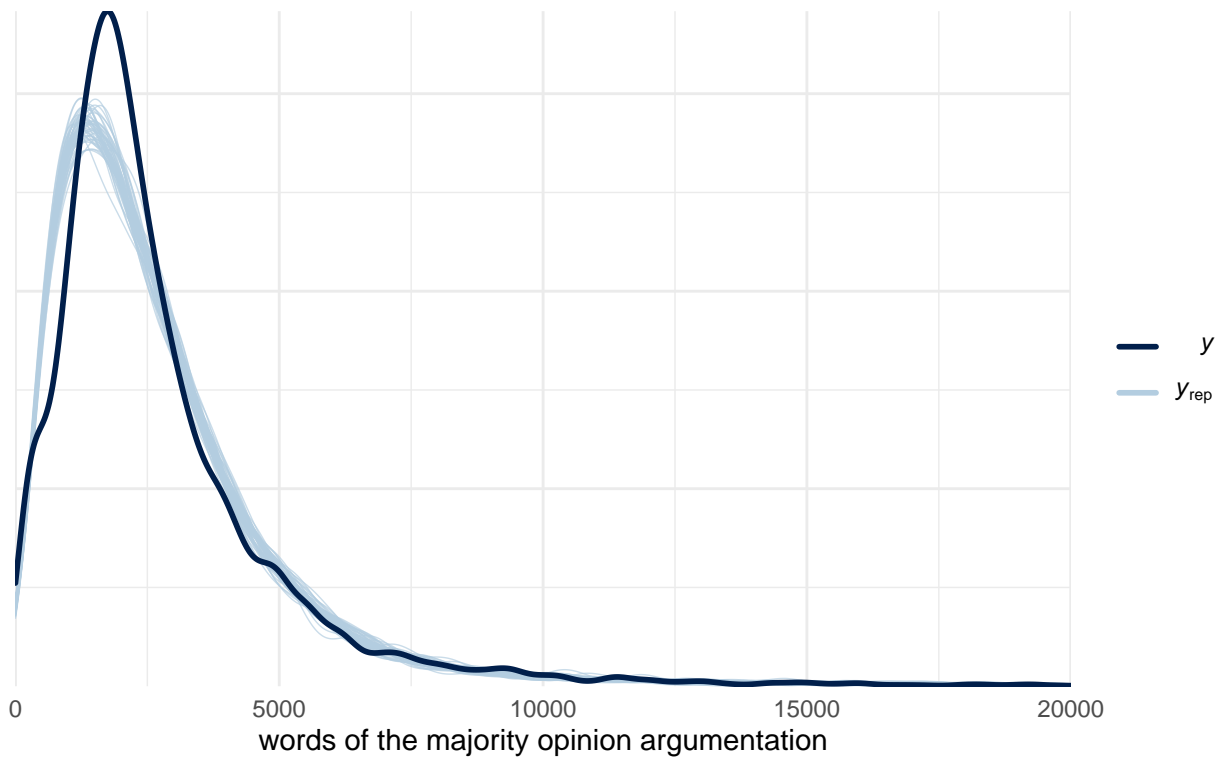
We ran the model via Stan with 4 Monte Carlo Markov Chains (MCMC) of 20000 iterations each, the first 10000 warm up iterations being discarded. The trace plot shows that the chains were

stable and probed plausible parameter values, the density plots of the MCMC show that all 4 chains exhibited similar behavior, and the autocorrelation between the iterations always dropped quickly and that the chains were moving around the potential parameter values quickly.

The posterior diagnosis confirms that although the simulations are not perfect, they do reasonably capture the features of the observed number of words of court arguments. In other words, we selected the correct model and the priors are not too off either. Thus, our Negative Binomial regression assumptions are reasonable.

Posterior predictive check of the Negative Binomial model

Dependence of length of majority opinion argumentation on presence of dissenting opinion



3 Model 2: Effect of workload on the dissenting behavior

3.1 Model diagnosis

We tried two models, a completely pooled and hierarchical model clustered around the judges. The main difference between the two models is that the former model completely ignores individual intercept. The latter allows for differentiating intercepts between the groups (in our case the individual judges) and the global intercept. The global parameter of interest is then informed both

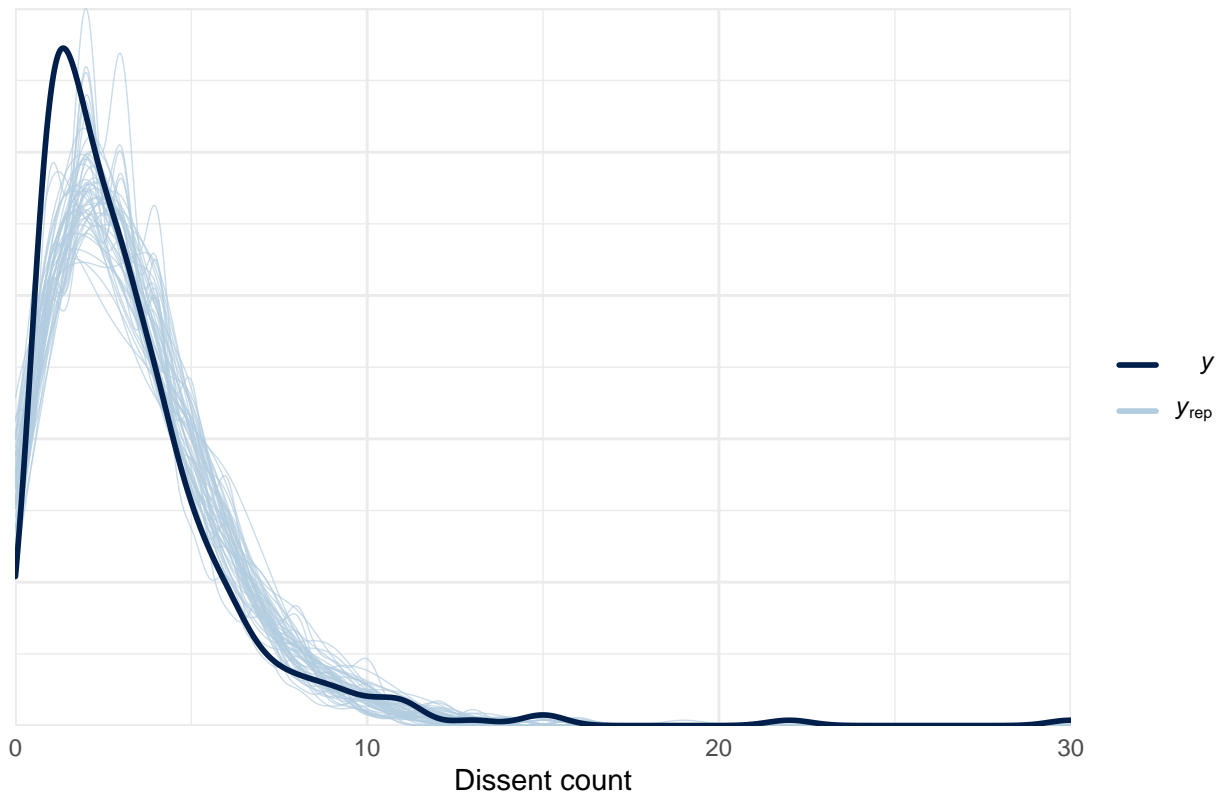
by the global trends as well as the individual intercepts. That can usually lead to higher accuracy in case of structured or time series data at the cost of higher computational expenses.

We ran both the models via Stan with 4 Monte Carlo Markov Chains (MCMC) of 20000 iterations each, the first 10000 warm up iterations being discarded. We did a diagnosis of all the models. In all cases, the trace plots show that the chains were stable and probed plausible parameter values, the density plots of the MCMC show that all 4 chains exhibited similar behavior, and the autocorrelation between the iterations always dropped quickly and that the chains were moving around the potential parameter values quickly.

We now compare the pooled against the hierarchical models. The former model got the posterior prediction 1.2 of the number of dissents per judge wrong (0.84 standard deviations off), whereas the former model got the posterior prediction wrong only by 0.94 of the number of dissents per judge (0.7 standard deviations off), with 99 % of the predictions falling within the 95 % posterior credible interval. 6-fold cross-validated check reveals that neither of the models overfitted. Thus, while the hierarchical model is slightly more computationally expensive, it yields better results.

Model 3: Collegiality costs of dissenting at the CCC ## Model diagnosis We ran the model via Stan with 4 Monte Carlo Markov Chains (MCMC) of 20000 iterations each, the first 10000 warm up iterations being discarded. We did diagnosis of all the models. The trace plots reveal that the chains were stable and probed plausible parameter values, the density plots of the MCMC show that all 4 chains exhibited similar behavior, and the autocorrelation between the iterations always dropped quickly and that the chains were moving around the potential parameter values quickly. The posterior predictive check again reveals that our posterior model reasonably captures the underlying data.

Posterior predictive check of the start and end of terms model



Johnson, Alicia A., Miles Q. Ott, and Mine Dogucu. 2022. *Bayes Rules!: An Introduction to Applied Bayesian Modeling*. CRC Press. <https://books.google.com?id=pISQzgEACAAJ>.