QM - Data essay

Francesco Catalfamo

2024-12-11

Contents

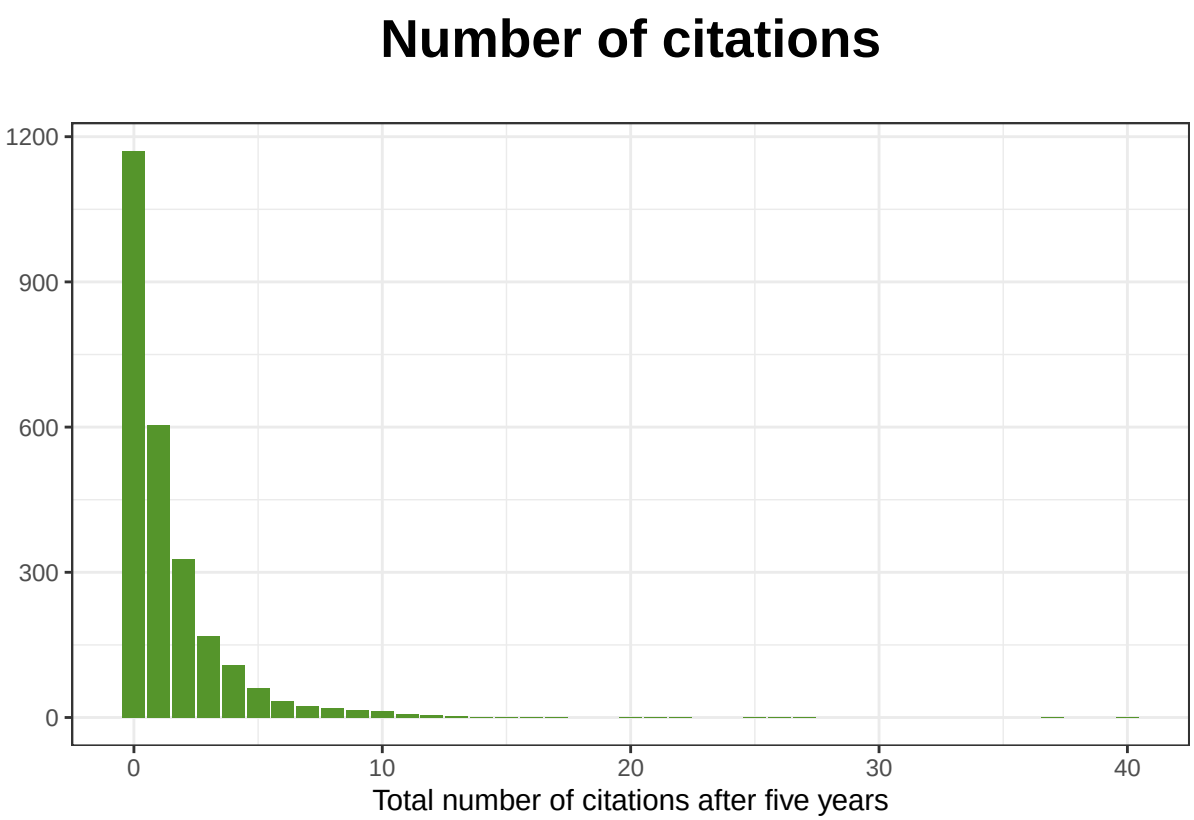
1 Intro

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

1.1 Data description

The variable *citation_5yr_diff_circ_2022* is the dependent variable, measuring the number of times appellate court decisions are cited over five years after publication. Citations are a key indicator of a decision's influence and authority in the legal field. A higher citation count suggests greater impact and authority. The histogram in *Figure 1* shows a right-skewed distribution, with most decisions receiving few or no citations, while a few have high citation counts. This skewness indicates uneven influence, with only a minority of cases having significant legal impact. Given this, a negative binomial regression is used, as it is suited for count variables with overdispersion.

Figure 1: Distribution of the judicial citation



The two main independent variables used in the analysis that will be presented in the following pages are *femaleOA* and *OAminority*, that is, the two dichotomous variables that indicate the gender and ethnic origin of the judge who issued the sentence.

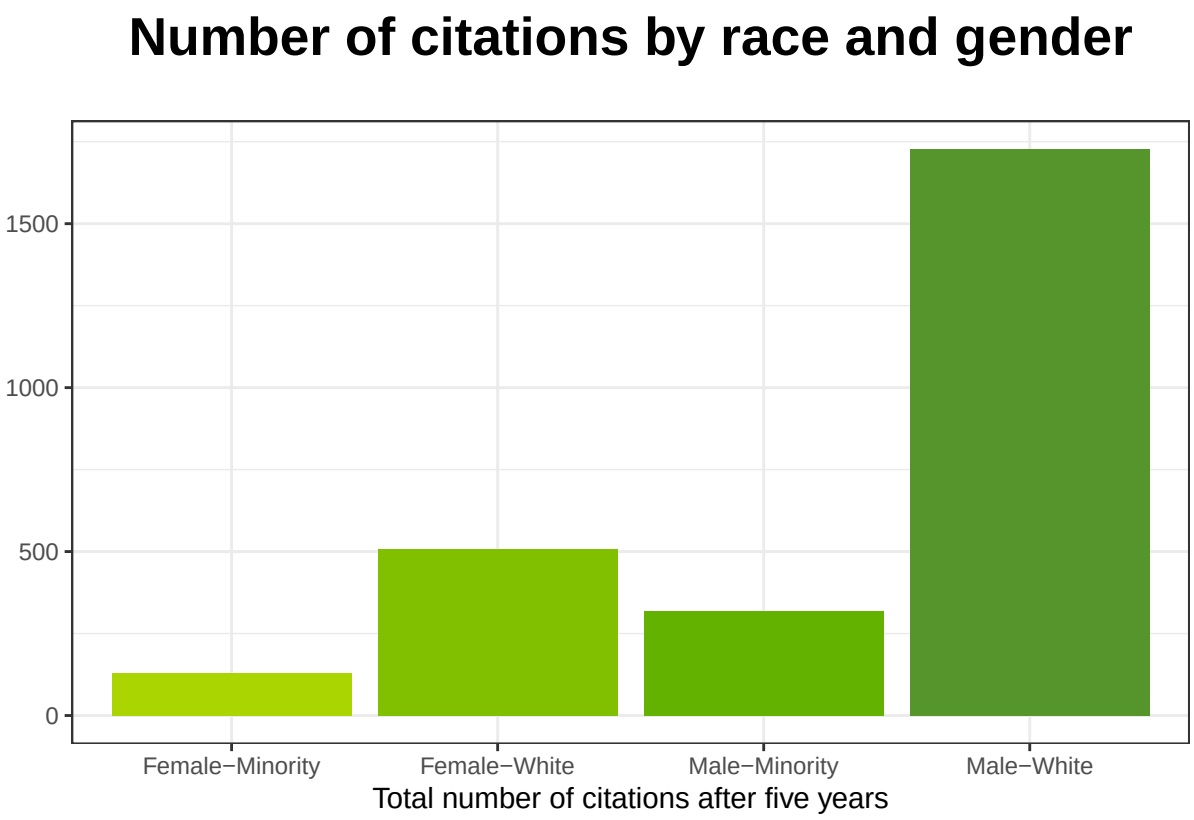
The variable *femaleOA* is a dummy variable indicating the gender of the judge who authored each of the decision in the dataset. It is coded as 1 if the judge is female and 0 if the judge is male. Looking at the values contained in the database, it is possible to observe that there are 2045 male judges and 637 female judges. Considering these values, it is possible to argue that, at least within our data set, the judicial system being analyzed is still mostly composed of men, given their clear predominance.

The variable *OAminority* is a dummy variable indicating whether the judge belongs to an ethnic minority. It is coded as 1 if the judge is non-white and 0 if the judge is white. Looking at the values contained in the database, it is possible to observe that there are 2426 white judges and 478 non-white judges. Taking these values into consideration, it is possible to observe how the judicial system appears to be composed of a clear prevalence of white judges rather than judges belonging to ethnic minorities.

Consequently, by observing the values reported in *Figure 2* below it is possible to verify the distribution of

the number of citations for both the main independent variables, namely gender and ethnic group. Looking at the graph it is possible to notice the high presence of white-male judges, who are more than all the other three categories added together. On the contrary, the group that appears to be at the center of the analysis, namely women belonging to ethnic minorities, appears to be the one with the least representation within the data set.

Figure 2: Distribution of race and gender



In addition to these two variables, the decision made was to include their interaction in order to capture the combined effect of gender and minority status on judicial citations. This approach is crucial because the intersection of these characteristics can lead to unique challenges or biases that are not accounted for by their individual effects. For example, female judges from minority backgrounds may face compounded disadvantages that further amplify their underrepresentation in citations. Including this interaction allows for a more nuanced analysis of whether these overlapping identities create distinct patterns of influence, ensuring that the model accurately reflects the complexity of demographic impacts on legal authority and perceived judicial credibility.

Table 1: Summary statitics of the control variables

Variable	Minimum	Median	Maximum
Quality of decision	0	4	54
Elite school	0	1	1
Years in court (log)	0	3	4

Moreover, with the purpose of ensuring robust and unbiased results, the decision was made to include some key control variables, *OAls_elite*, *authorities_deep* and *OAtenure_ln* that address the factors that influence court citations. The key descriptive statistics for these three variables are reported in *Table 1* at the end of this section.

Reliance on precedent is a key indicator of legal quality, with decisions based on precedent often seen as more rigorous. The variable *authorities_deep* controls for the extent of reliance on precedent, helping researchers distinguish between citations driven by legal reasoning and those influenced by a judge's demographics. It captures the depth of decision-making, as well-founded decisions are typically more persuasive and cited more. By controlling for precedent reliance, analyses avoid skewing results, allowing for a clearer understanding of how demographic factors affect citation patterns. Judges from elite law schools are often seen as more authoritative, influencing how frequently their decisions are cited due to the prestige of their education. The *OAls_elite* variable helps account for this citation advantage, separating it from the quality of their decisions or demographic factors. By controlling for educational background, the analysis can better assess how demographic traits, such as the main independent variables, impact citation patterns without the influence of perceived educational prestige. Judicial tenure influences a judge's experience and authority, often making their decisions more likely to be cited. The *OAtenure_ln* variable accounts for the effect of professional experience, represented by the logged years a judge has served on the Court of Appeals. This control separates the impact of tenure from demographic factors, ensuring that citation patterns reflect experience, not just demographic characteristics. Including tenure reduces bias and provides a clearer understanding of how demographics influence citation rates, independent of a judge's expertise and seniority.

‘‘‘

Including these variables ensures the model accounts for factors influencing citation patterns beyond demographics, minimizing omitted variable bias, strengthening causal inference, and providing a more accurate assessment of how gender and minority status impact judicial influence. This approach isolates the hypothesized effects, ensuring observed differences are attributable to the judges' demographics, not other factors. Based on this, the model presented in the following pages is designed as follows:

$$citation = \beta_0 + \beta_1 * female + \beta_2 * minority + \beta_3 * female * minority + \beta_4 * authorities + \beta_5 * elite + \beta_6 * tenure$$

2 Analysis

Before we can present the models, it is necessary to explain why the chosen regression type is *Negative Binomial* and not *Poisson*. After testing both models, a Likelihood Ratio Test (overdispersion) was applied to decide between the Poisson and Negative Binomial specifications. The results show how the Negative Binomial model is more efficient in representing the data used in the regression.

cat("\begin{table}[ht] \centering \resizebox{\textwidth}{!}{", readLines("table.tex"), "} \end{table}")

2.0.1 Hypothesis 1

2.0.2 Hypothesis 2

2.0.3 Hypothesis 3

3 Simulation (aka scenario)

To assess gender and ethnic differences as outlined in the hypotheses, it is necessary to control for the expected values from previous models based on the same personal characteristics. This requires four simulations to evaluate the differences across the three hypotheses. The first simulation will focus on women (female = 1) from ethnic minorities (non-white = 1). The results will be used to calculate the first differences between simulations involving white men, white women, and non-white men.

3.0.1 Hypothesis 1

- Intercept: 1

Table 2: Results of the negative binomial models

	Dependent variable:	
	Number of citation	
Constant	0.409^{***}	0.0004
	(0.038)	(0.105)
Female	-0.046	-0.153^{*}
	(0.080)	(0.079)
Ethnic minority	0.112	0.060
	(0.094)	(0.092)
Years in Court		-0.001
		(0.034)
Quality of decision		0.065^{***}
		(0.006)
Elite school		0.100
		(0.061)
Female*Ethnic minority	-0.433^{**}	-0.285
	(0.187)	(0.183)
Observations	2,373	2,373
Log Likelihood	$-3{,}939.913$	$-3{,}878.208$
θ	0.642^{***} (0.032)	0.717^{***} (0.038)
Akaike Inf. Crit.	7,887.826	7,770.415
Note:		*p<0.1; **p<0.05; ***p<0.01

- Female: 1
- Minority: 1
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 1*1

Compare

- Intercept: 1
- Female: 0
- Minority: 0
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 0*0

3.0.2 Hypothesis 2

- Intercept: 1
- Female: 1
- Minority: 1
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 1*1

Compare

- Intercept: 1
- Female: 1
- Minority: 0
- Authorities: median
- Elite: median

- Tenure: median
- Interaction: 1*0

3.0.3 Hypothesis 3

- Intercept: 1
- Female: 1
- Minority: 1
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 1*1

Compare

- Intercept: 1
- Female: 0
- Minority: 1
- Authorities: median
- Elite: median
- Tenure: median
- Interaction: 0*1

4 Robustness check

5 Conclusion

Figure 3: First difference