Legal Research as a Collective Enterprise:
An Examination of Data Availability in Empirical Legal Scholarship

Abigail A. Matthews and Jason Rantanen

Description of Variables Used in Analyses

PublicationName
String variable of journal names. Examples include: Econometrica, Iowa Law Review, Journal of Politics, New York University Law Review.

discipline
Numeric variable identifying the three disciplines and law-social science interdisciplinary journals in the sample. Coded as:

1 Law
2 Political Science
3 Economics
4 Interdisciplinary

avail
Binary measure of data availability, equals one if the article indicate that the final, compiled dataset is available and zero otherwise.

Year
Is the year the article was published. For the student-edited law journals, the years are 2010–2022. For the political science, economics, and interdisciplinary journals, the publication years are 2018–2022.

DataPolicy
DataPolicy is the journal's data policy. This is a categorical variable coded as:

0 No Policy
1 Voluntary
2 Mandatory

coauthor
Coauthor is a binary variable coded zero if the article is not coauthored (or solo-authored) and one if it is coauthored.

phd
PhD is a binary variable coded one if at least one of the authors has a PhD, zero otherwise.

peerreview
Peer review is a binary variable used for the student-edited law journals to indicate whether the journal employs elements of peer review in their submission process. Student-edited peer review is when the journal seeks input from legal scholars but often offer the peer scholar a very short period in which to review the submission (typically 24-48 hours, based on one author's experience). Our understanding is that in these cases, the student editors make the ultimate decision. Peer review is coded 1 if the journal uses peer review and zero if it does not.

qualitative
Qualitative is a binary variable coded 1 if the empirical study purely qualitative, zero otherwise. This variable is only used for the student-edited law journals.

summarystats
Summary Statistics is a binary variable coded 1 if the empirical study uses numbers to describe or summarize the data, zero otherwise. Examples include averages, means, medians, modes, per cents, tables or graphs of the data. A study that includes statistical analysis may also include summary statistics. This variable is only used for the student-edited law journals.

statisticalanalysis
Statistical analysis is a binary variable coded 1 if the empirical study applies any sort of inferential methods, zero otherwise. Statistical inferential data often (1) test hypotheses (e.g., Is the population mean greater than or less than a particular value? Are the means of two or more populations different from each other?); (2) have confidence intervals (measure uncertainty of sample error such as margin of error in election polls); and/or (3) has regression analysis (analyzes the relationship between a set of independent variables and a dependent variable). This variable is only used for the student-edited law journals.

datalocation1
Datalocation1is a categorical variable identifying where the article indicates the reader can find the data. This does not include whether the data is available on request from the author. For that information, *see* AuthorData variable. This variable is only used for the student-edited law journals.

1 Journal Website
2 Other Website
3 Appendix

subject1
subject1is a string variable identifying the subject of the article. The possible choices include:

- Alternative dispute resolution, including Arbitration, Mediation,
- Business Law, including Corporate Law; Antitrust Law; Stock Exchange; Bankruptcy; Commercial Law; Business Law; Securities
- Civil Procedure, including Evidence
- Civil Rights, including Civil Justice Access; Policing

- Constitutional Law
- Contract Law, including Government Contracts
- Criminal Law
- Employment Law, including Labor; Works Compensenation
- Environmental Law, including Energy Law, Water Law
- Family Law
- Health Law
- Immigration
- Intellectual Property, including Arts & Entertainment; Sports Law; Public Domain
- International Law, including Maritime Law, Admiralty Law
- Judges and Judicial Decisionmaking, including Judiciary; Federal Judges; Judicial Decisionmaking; Judicial Nominations; Prosecutorial elections; Supreme Court
- Law and Economics, including Game Theory
- Law School, including Law School Admissions
- Legal Scholarship
- Privacy
- Professional Responsibility, including Ethics
- Property Law
- Race and Gender, including Critical Legal Studies
- Tax Law
- Tort Law, including Insurance Law; Products Liability
- Trusts and Estates
- Other

This variable is only used for the student-edited law journals.

Final_Article

Final_Article is a binary variable coded 3 if the publication is an essay or article. It is coded 2, not an article or essay, if the publication is a student notes or comments, review essays, tribute issues/festschrifts, forwards or responses. An essay or article in a symposium issue is considered an essay or article for purposes of this study.

Final_Empirical

Final_Empirical is a categorical variable to indicate whether an article involves a substantive original empirical study or if a major component of the article is an original empirical study. "Empirical" means research based on observations of the world. An *original* empirical study is a study that the authors performed, as contrasted with simply reporting the results of someone else's study, as authors might do when reviewing previous studies or relying on studies conducted by others. Mini-empirical claims, such as a Westlaw query and count in a footnote, are not substantive empirical studies for this definition. The key hallmarks of an empirical study are (1) claims about the world based on the collection of data and (2) a systematic methodology for collecting and analyzing the data. Empirical studies may include both quantitative (such as census data, case counts, applying statistics) and qualitative approaches (such as surveys and interviews). Final_Empirical is coded as follows:

1 Maybe
2 No
3 Yes

authordata
Authordata is a categorical variable indicating whether the article indicates the data is available upon request from the author. Coded 1 only when there's an express invitation to contact the author, often using language such as “data available on request.” This variable is only used for the student-edited law journals.

0 Not Available
1 Data Available from Authors
2 NA

Title
Title is a string variable of the article’s title.

Author
Author is a string variable of the author or authors of the article.

yearbins
Yearbins is a categorical variable that collapses the years an article was published into three categories, as follows:

0 2010–2014
1 2015–2018
2 2019–2022

name
Name is a string variable of an author’s name who has published more than one article in our sample from 2018–2022.

discipline1
Discipline1 is a binary variable coded 1 if the author published in a student-edited law journal, zero otherwise.

discipline2
Discipline2 is a binary variable coded 1 if the author published in a political science or economics journal, zero otherwise.

discipline3
Discipline3 is a binary variable coded 1 if the author published in an interdisciplinary journal, zero otherwise.

availcountint
When the author has published more than one article in our sample, 2018–2022, availcountint is a variable of the number of times an author makes their data available in interdisciplinary journals. Ranges from 0 to 4.

availtotalint
When the author has published more than one article in our sample, availtotalint is a variable of the total number of interdisciplinary articles the author has published from 2018–2022. Ranges from 1 to 1.

availcountSS
When the author has published more than one article in our sample, 2018–2022, availcountSS is a variable of the number of times an author makes their data available in social science journals. Ranges from 0 to 1.

availtotalSS
When the author has published more than one article in our sample, availtotalSS is a variable of the total number of social science articles the author has published from 2018–2022. Ranges from 1 to 3.

availcountlaw
When the author has published more than one article in our sample, 2018–2022, availcountlaw is a variable of the number of times an author makes their data available in student-edited law journals. Ranges from 0 to 2.

availtotallaw
When the author has published more than one article in our sample, availtotallaw is a variable of the total number of student-edited law journal articles the author has published from 2018–2022. Ranges from 1 to 6.

availcount
Availcount is a count of the number of articles in student-edited law journals that make their data available by year published.

availcountyeartotal
Availcountyeartotal is a count of the total number of articles in student-edited law journals by year published.

percentavail
Percentavail is the proportion of articles in student-edited law journals that make their data available by year divided by the total number of articles published in in student-edited law journals, multiplied by 100 to create a percent.