Summer School JGU Mainz — Advanced Methods in Behavioral Economics, 2021

Dr. Carina I. Hausladen

September 24, 2021

What is a moderate interpretation of the text? Halfway between what it really means and what you'd like it to mean?

U.S. Supreme Court Justice Antonin Scalia

4

Text Classification of Ideological Direction in Judicial Opinions

Carina I. Hausladen, Marcel H. Schubert, Elliott Ash

Study to be Replicated

Landes, William M., and Richard A. Posner.
"Rational Judicial Behavior: A Statistical Study."
Journal of Legal Analysis 1.2 (2009): 775-831.

Empirical Analysis to be Replicated

$$FrCon_i = \beta_0 + \beta_1 X_i + \omega_i$$

where X_i includes:

- party of the appointing president
- fraction of republican senators at time of initial appointment
- term of judge's confirmation of the supreme court
- gender
- race
- district court at which the judge served prior to her appointment
- fraction of economic votes
- fraction of miscellaneous votes
- dummy variables denoting the judge's circuit

Political Ideology Labels

$$FrCon_i = \beta_0 + \beta_1 X_i + \omega_i$$

Political Ideology Labels

Labelling the Complete Set of Available Opinions
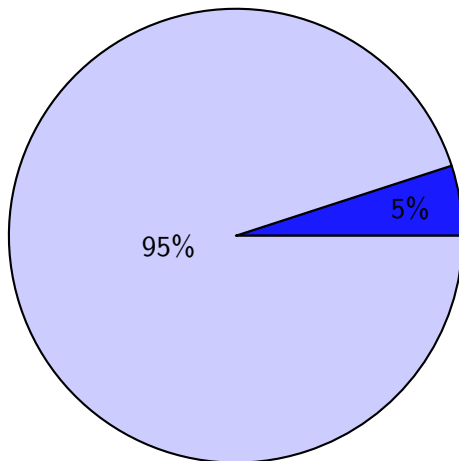
Supervised Classification

The parameters varied include:

- subset of data
- text input
- text preprocessing
- feature engineering
- classification models

Best Configuration

- feature representation: tf-idf weighted bigrams
- classifier: ridge classifier
- calibration: isotonic calibration (non-parametric)
- F1 score: **67%**
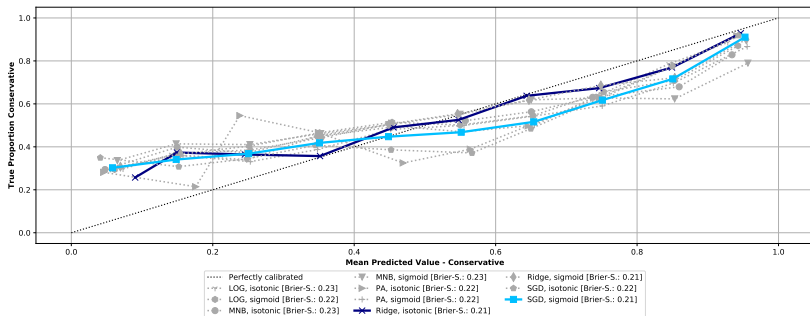
Figure: Reliability Curve

Predictions
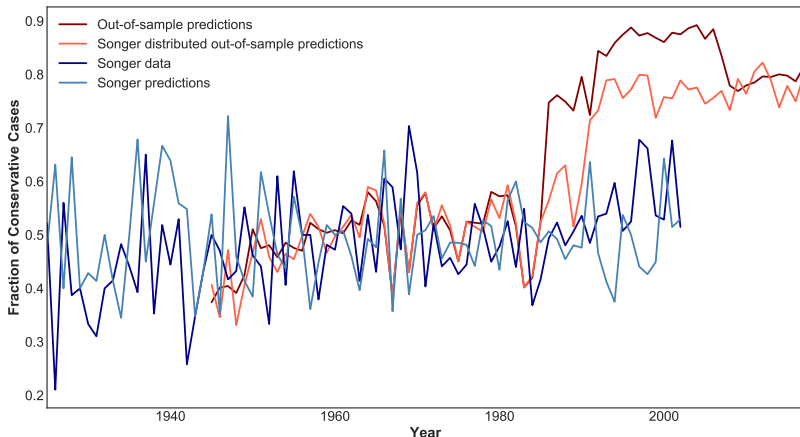
Figure: Fraction of Conservative Cases, each Calculated for Actual as well as Predicted Case Directionality, Plotted by Year

Replication

Table: Court of Appeals Votes by Subject Matter and Ideology for 538 Court of Appeals Judges; Years: 1925–2002

	Crim	Civ Rts	First	Due Proc	Priv	Labor	Econ	Misc	Total
Conservative	6823	2721	566	461	117	1351	9361	525	21925
Liberal	1876	1766	477	201	67	1922	9884	559	16752
Mixed	635	460	89	51	13	420	1775	22	3465
Other	5321	210	102	79	3	179	6047	958	12899
Total	14655	5157	1234	792	200	3872	27067	2064	55041

(a) Original by Landes et al. (2009)

	Crim	Civ Rts	First	Due Proc	Priv	Labor	Econ	Misc	Total
Conservative	7217	2647	397	412	83	1397	11084	478	23715
Liberal	1911	1755	379	176	38	0	10375	596	15230
Mixed	613	473	86	48	9	423	1689	31	3372
Other	5652	212	40	24	3	2232	5177	945	14285
Total	15393	5087	902	660	133	4052	28325	2050	56602

(b) Replication

Regression, 1925–2002, Criminal Cases

	Landes (2009)	replicated	multi.clus	vote	multi.clus.pred	vote.pred
	(1)	(2)	(3)	(4)	(5)	(6)
RepPres	0.056^{**}	0.077^{***}	0.077^{***}	0.051^{**}	0.038	0.005
	(4.220)	(3.634)	(3.811)	(3.022)	(1.734)	(0.829)
SenRep	-0.076	-0.151	-0.151	0.010	-0.020	0.078^{**}
	(1.090)	(-1.399)		(0.141)	(-0.542)	(2.844)
YrAppt	-0.001^{***}	-0.00001	-0.00001	-0.0003	0.001^{**}	-0.001^{**}
	(3.390)	(-0.023)	(-0.032)	(-0.601)	(2.876)	(-2.709)
Gender	-0.014	-0.019	-0.019	0.010	-0.023^{*}	-0.012
	(0.710)	(-0.740)	(-0.876)	(0.545)	(-2.219)	(-1.750)
Black	-0.057^{*}	-0.091^{*}	-0.091	-0.081^{**}	-0.020	-0.027
	(2.060)	(-1.814)	(-1.047)	(-2.717)	(-0.257)	(-1.697)
DistrictCourt	0.001	-0.001	-0.001	0.0003	-0.001	0.001
	(0.140)	(-0.817)	(-0.390)	(0.360)	(-0.346)	(1.917)
circuit FE	yes	yes	no	no	no	no
circuit-year FE	no	no	yes	yes	yes	yes
Observations	523	498	498	13543	498	13543
R^2	0.240	0.084	0.084	0.019	0.052	0.014

Note:t-statistics in parentheses.

Methods to Test Robustness

- Multi-way error clustering (non-nested) (Cameron et al. 2012) allows standard errors to be correlated (1) across a judge within multiple years and (2) across circuits within one year.
- The vote-level regression (dv is binary)
  - allows to control for case characteristics.
  - excludes the time dimension. Therefore, correlations between the dv and ivs over time are no longer problematic.

Methods to Test Robustness: Extreme Bounds Analysis

- e.g. `ExtremeBounds` by Hlavac (2016)
- Which determinants are robustly associated with the dv?
- How robust are the results w.r.t. to the inclusion or exclusion of ivs?
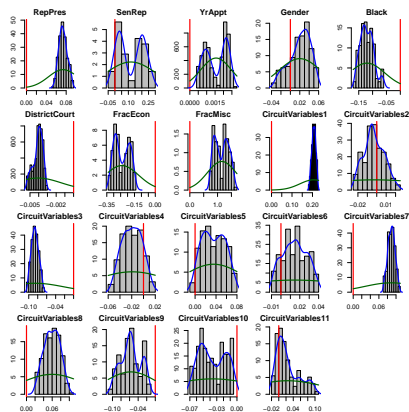
Conclusion

- Landes et al. (2009)'s empirical analysis could successfully be replicated based on the 5% Songer sample.
- Subjective rather than objective factors (such as the law) influence a judge's decision making.
- The classifier is predictive (F1 Score = .67).
- However, predictions might not be reliable after the year of 1980.

Let's Discuss!

- How would you assess whether the share of conservative decision-making increased after 1980?
- Which deep learning approaches would you use to increase predictive performance?
- You could use the classifier to predict the political ideology lower courts' opinions, which are understudied in the literature.
- How would you account for prediction errors in a regression analysis (predictions as dv instead of iv)? (Fong et al. 2020)

References I

∗ Cameron, A. Colin et al. (2012). "Robust Inference with Multiway Clustering". In: *Journal of Business & Economic Statistics*.

∗ Fong, Christian et al. (2020). "Machine Learning Predictions as Regression Covariates".

∗ Hlavac, Marek (2016). "Extremebounds: Extreme Bounds Analysis in R". In: *Hlavac, Marek.(2016)." ExtremeBounds: Extreme Bounds Analysis in R," Journal of Statistical Software* 72.

   Landes, William M. et al. (2009). "Rational Judicial Behavior: A Statistical Study". In: *Journal of Legal Analysis* 1.2, pp. 775–831.

Thanks!

Let's keep in touch!

Dr. Carina I. Hausladen
Postdoctoral Researcher
carina.hausladen@gess.ethz.ch

ETH Zurich
Computational Social Science COSS
Stampfenbachstrasse 48
8092 Zurich