Online Appendix to:

Why Are the Affluent Better Represented Around the World?

Noam Lupu and Zach Warner

May 9, 2020

A.1 Data sources and coding rules

Table A1: Data sources

Variable	Source
Δ EMD	Lupu and Warner (2020) variable emd_diff.
Foreign cap. depend.	World Bank (2019), foreign direct investment, net inflows, balance of payments in current US dollars (variable BX.KLT.DINV.CD.WD). Gathered using the R package WDI and logged.
GDP (logged)	World Bank (2019), GDP per capita in constant 2010 US dollars (variable NY.GDP.PCAP.KD). Gathered using the R package WDI and logged.
HDI	Quality of Government Standard Dataset, January 2019 version (Teorell et al. 2019). Variable undp_hdi, originally provided by the UNDP's Human Development Report.
Income inequality	World Bank (2019), GINI index, World Bank estimate (variable SI.POV.GINI). Gathered using the R package WDI.
Trade openness	World Bank (2019), trade as a percentage of gross domestic product (variable NE.TRD.GNFS.ZS). Gathered using the R package WDI.
Age of democracy	Boix et al. (2013), data version 3.0, variable democracy_duration.
Disproportionality	Gandrud (2019), variable disproportionality. Gathered using the R package devtools via `http://bit.ly/Ss6zDO`.
Party institutionalization	The Database of Political Institutions (Cruz et al. 2016), version DPI2015, variable partyage.
Clientelism	V-Dem, data version 7.1 (Coppedge et al. 2017). Variable v2psprlnks, inverted so that higher values indicate more clientelistic and less programmatic linkages.
Corruption	V-Dem, data version 7.1 (Coppedge et al. 2017). Variable v2x_corr.

Table A1: Data sources (continued)

Variable	Source
Government ideology	Chapel Hill Expert Survey, 1999-2014 Trend File, version 1.1 (Bakker et al. 2015). Variable seat divided by the sum of seat for a given country-year gave the legislative proportion for a given party, while lrgen gave the party's ideology. Parties in government were chosen using the govt variable, values "in government" or ".5" (in government for part of the year). We then imputed missing years for which CHES data were available. We then supplemented with Manifesto Project data, version 2018b (Volkens et al. 2018), using variable ideology and manually selecting parties in government using secondary sources. We then supplemented with data from Baker and Greene (2011), updated through 2018 in the 8 January 2019 data version. Here again we used variable ideology and manually selected parties in government using secondary sources.
% female legislators	Scraped from the Inter-Parliamentary Union website, now available through Parline (Inter-Parliamentary Union 2019).
Civil society	V-Dem, data version 7.1 (Coppedge et al. 2017). Variable v2x_cspart.
Pol. donation restrictions	V-Dem, data version 7.1 (Coppedge et al. 2017). Variable v2eldonate.
Trade union density	Trade union density rate (percentage), downloaded from ILOSTAT (International Labour Organization 2019) on 27 April 2019.
Compulsory voting	V-Dem, data version 7.1 (Coppedge et al. 2017). Variable v2elcomvot, recoded into a binary variable by setting all values greater than 1 to 1, to reflect any legal requirement to vote.
Cross-cuttingness	Data from Selway (2011), August 2013 version, variable RaIC.
Turnout	V-Dem, data version 7.1 (Coppedge et al. 2017). Variable v2elvaptrn, divided by 100 so as to indicate proportions.

A.2 Models used in the main analysis

The following list gives the name and description for each model studied in our machine learning task, as given in the R package caret (Kuhn 2008). We chose these models for their diversity of underlying approach.

1. avNNet: Model Averaged Neural Network
2. bagEarthGCV: Bagged MARS using gCV Pruning
3. dnn: Stacked AutoEncoder Deep Neural Network
4. glm: Generalized Linear Model
5. glmboost: Boosted Generalized Linear Model
6. glmnet: glmnet
7. knn: k-Nearest Neighbors
8. mlp: Multi-Layer Perceptron
9. nnet: Neural Network
10. pcaNNet: Neural Networks with Feature Extraction
11. ppr: Projection Pursuit Regression
12. rf: Random Forest
13. rvmLinear: Relevance Vector Machines with Linear Kernel
14. xgbLinear: eXtreme Gradient Boosting

A.3 Model performance

Table A2: Predictive performance

Model	RMSE	Model	RMSE
avNNet	0.98	mlp	1.03
	(0.16)		(0.19)
bagEarthGCV	0.96	nnet	0.98
	(0.18)		(0.17)
dnn	1.04	pcaNNet	0.99
	(0.19)		(0.18)
glm	0.95	ppr	1.04
	(0.17)		(0.15)
glmboost	0.96	rf	0.93
	(0.16)		(0.21)
glmnet	0.95	rvmLinear	0.96
	(0.17)		(0.17)
knn	0.95	xgbLinear	1.03
	(0.18)		(0.22)

Values in parentheses indicate standard deviations. Note that RMSE is on the scale of the rescaled dependent variable, which ranges over $[-3, 5]$.

A.4 Additional partial dependence plots

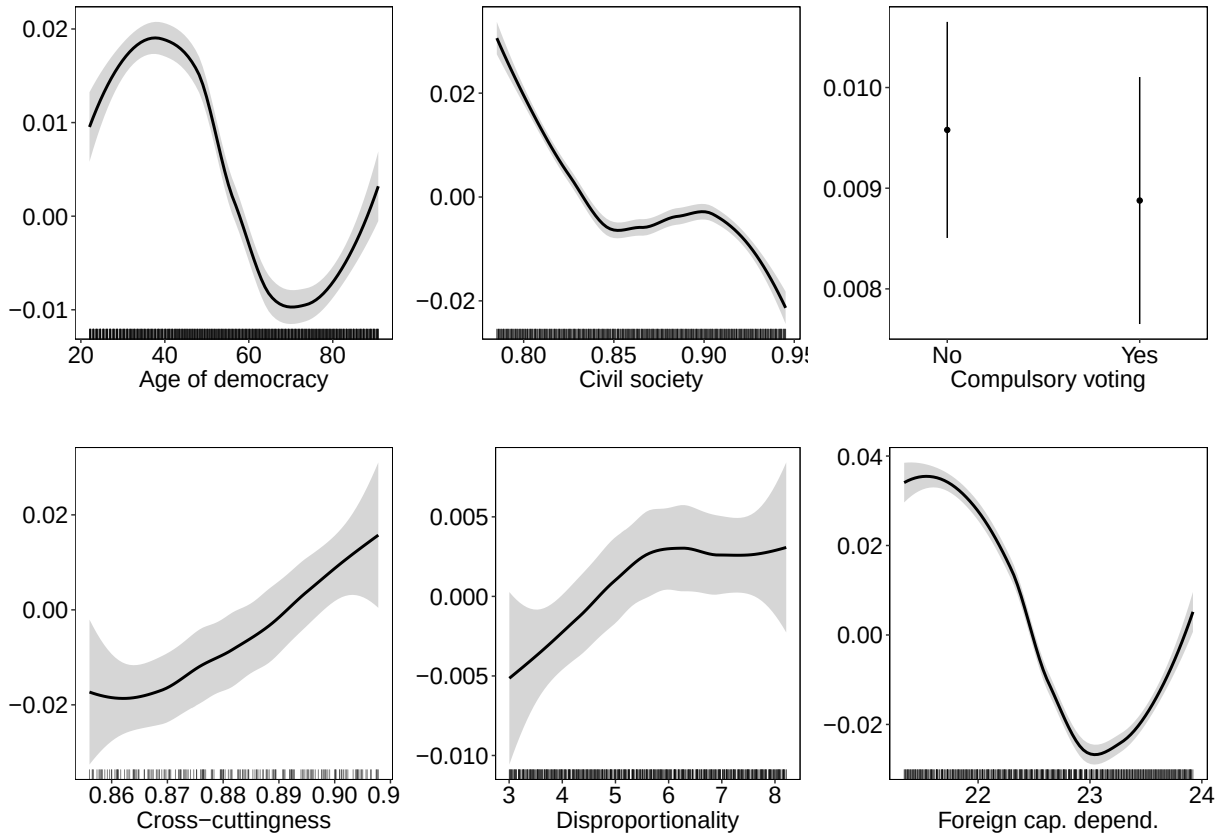

Figure A1: Partial dependence plots. Each panel provides the predicted change in unequal representation as a predictor is moved across its inter-quartile range. Lines represent loess fits, with 95% confidence intervals in gray, computed from random forest predictions across all imputation replicates. Rug plots are also provided along the x axis to indicate support in the underlying data for these predictions. Note the differing axes in each panel.

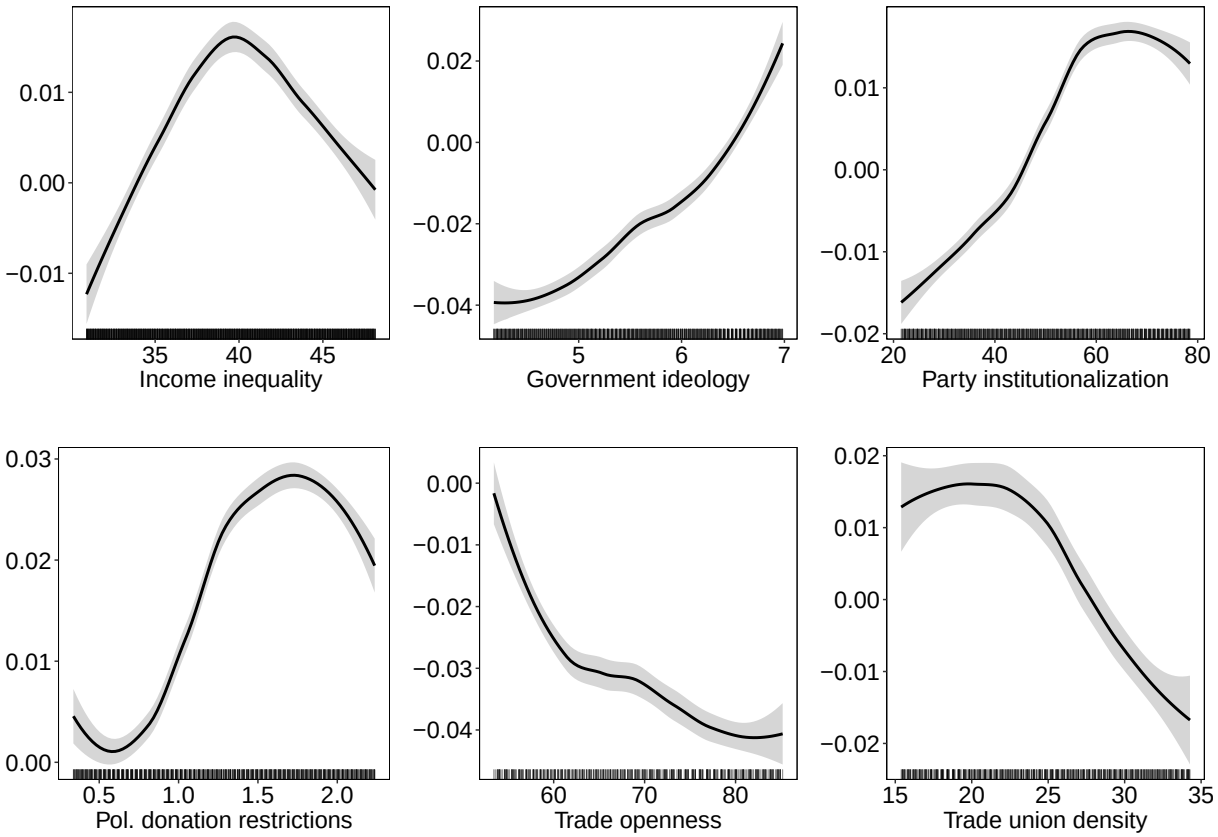

Figure A1 (continued): Partial dependence plots. Each panel provides the predicted change in unequal representation as a predictor is moved across its inter-quartile range. Lines represent loess fits, with 95% confidence intervals in gray, computed from random forest predictions across all imputation replicates. Rug plots are also provided along the x axis to indicate support in the underlying data for these predictions. Note the differing axes in each panel.

References

Baker, Andy, and Kenneth F. Greene. 2011. "The Latin American Left's Mandate: Free-Market Policies and Issue Voting in New Democracies." *World Politics* 63 (1): 43-77.

Bakker, Ryan, Erica Edwards, Liesbet Hooghe, Seth Jolly, Jelle Koedam, Filip Kostelka, Gary Marks, Jonathan Polk, Jan Rovny, Gijs Schumacher, Marco Steenbergen, Milada Vachudova, and Marko Zilovic. 2015. "1999-2014 Chapel Hill Expert Survey Trend File.".

Boix, Carles, Michael K. Miller, and Sebastian Rosato. 2013. "A Complete Data Set of Political Regimes, 1800-2007." *Comparative Political Studies* 46 (12): 1523-1554.

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Joshua Krusell, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Moa Olin, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Constanza Sanhueza Petrarca, Johannes von Römer, Laura Saxer, Brigitte Seim, Rachel Sigman, Jeffrey Staton, Natalia Stepanova, and Steven Wilson. 2017. "V-Dem Country-Year Dataset v7.1." Varieties of Democracy (V-Dem) Project.

Cruz, Cesi, Philip Keefer, and Carlos Scartascini. 2016. "Database of Political Institutions Codebook, 2015 Update (DPI2015)." Inter-American Development Bank.

Gandrud, Christopher. 2019. "Gallagher Electoral Disproportionality Data." Available at http://bit.ly/Ss6zDO. Last accessed November 5, 2019.

Inter-Parliamentary Union. 2019. "Parline." Available at https://data.ipu.org/women-ranking. Last accessed November 5, 2019.

International Labour Organization. 2019. "ILOSTAT." Available at https://ilostat.ilo.org. Last accessed November 5, 2019.

Kuhn, Max. 2008. "Building Predictive Models in R using the caret Package." *Journal of Statistical Software* 28 (5): 1-26.

Lupu, Noam, and Zach Warner. 2020. "Affluence and Congruence: Unequal Representation Around the World." Unpublished manuscript.

Selway, Joel Sawat. 2011. "Electoral Reform and Public Policy Outcomes in Thailand: The Politics of the 30-Baht Health Scheme." *World Politics* 63 (1): 165 - 202.

Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, and Richard Svensson. 2019. "The Quality of Government Standard Dataset, version Jan19." University of Gothenburg: The Quality of Government Institute.

Volkens, Andrea, Werner Krause, Pola Lehmann, Theres Matthieß, Nicolas Merz, Sven Regel, and Bernhard Weßels. 2018. "The Manifesto Data Collection." The Manifesto Project (MRG/CMP/MARPOR). Version 2018b. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).

World Bank. 2019. “Data Catalog.” Available at http://datacatalog.worldbank.org. Last accessed November 5, 2019.