KELLOGG INSTITUTE
FOR INTERNATIONAL STUDIES

V-Dem
REGIONAL CENTER
North America

V-Dem Data Workshop

Kellogg Institute

January 24, 2025

Varieties of Democracy Overview

The team, conceptual scheme, varieties and definitions, major trends, original surveys and specific indicators, and types of data

Historical V-Dem

While I’m talking, download v14 of the data (if you haven’t already)

→v-dem.net, Data, Data Version 14

Recommended:

Country-Year: V-Dem Full+Others

Brief History

- Conversations and planning began in 2007
- Principal Investigators came together
- Most Project Managers recruited by 2010; questions written
- Pilot Study in 2011, followed by data collection on a few countries
- 2011ff: Development of database and website
- Two institutional homes at Gothenburg and Kellogg Institute
- Various waves of data collection as funding came in for countries

Brief History, continued

- RBJ grant: $5.8m over 6 years
- 2014-15: Partial data releases
- January 2016: First release of full dataset
- All data collection shifted to Gothenburg: “V-Dem Institute”
- 2018: Kellogg Institute designated “V-Dem Regional Center in North America”
- 2020 Cambridge book: *Varieties of Democracy: Measuring Two Centuries of Political Change*
- 2022 Cambridge book: *Why Democracies Develop and Decline*

The Global Team of Varieties of Democracy

Principal Investigators and Project Managers

Direct Democracy

Civil Society & Sovereignty

Media

Legislatures

Political Equality

Methodology

Parties & Party Systems

Civic & Academic Space

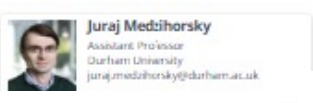
Historical Data

Elections

Evolutionary Theory

Measurement Methods

Subnational Governments



Statistical Computing

Prediction Modeling

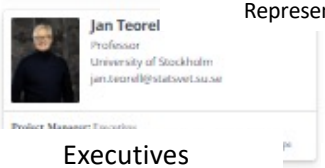
Formal & Descriptive Representation

Measurement Methods

Exclusion & State Capacity

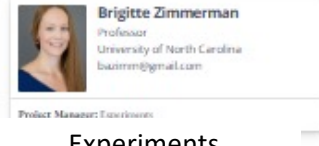
Civil Liberty

Judiciary



Executives

Computational Infrastructure



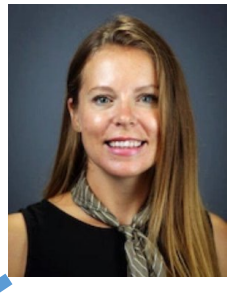
Experiments

8 Regional Centers

- North America (Kellogg Institute, Notre Dame)
- Latin America (Universidad Católica de Chile)
- Southern Europe (Universidade Nova de Lisboa)
- East Asia (Keio University in Tokyo)
- Eastern Europe and Russia (Institute of Government and Politics, University of Tartu, Estonia)
- Central Asia (American University of Central Asia)
- Southern Africa (University of Zambia)
- Balkans

The V-Dem Institute
In the Department of Political Science, University of Gothenburg, Sweden

Josephine Pernes
Deputy Director,
Executive Officer
& Director of Grants



Staffan I. Lindberg
Director and
Professor

Core Operations Team (10)
Evie Papada, Linnea Fox, Lisa Gastaldi, Ana Good God, Sandra Grahn, Sara Haug Andersson, Melina Liethmann, Natalia Natsika, Maria Verkovtseva

Research Team (2 Postdoctoral Research Fellows)
Fabio Angiolillo and Marina Nord

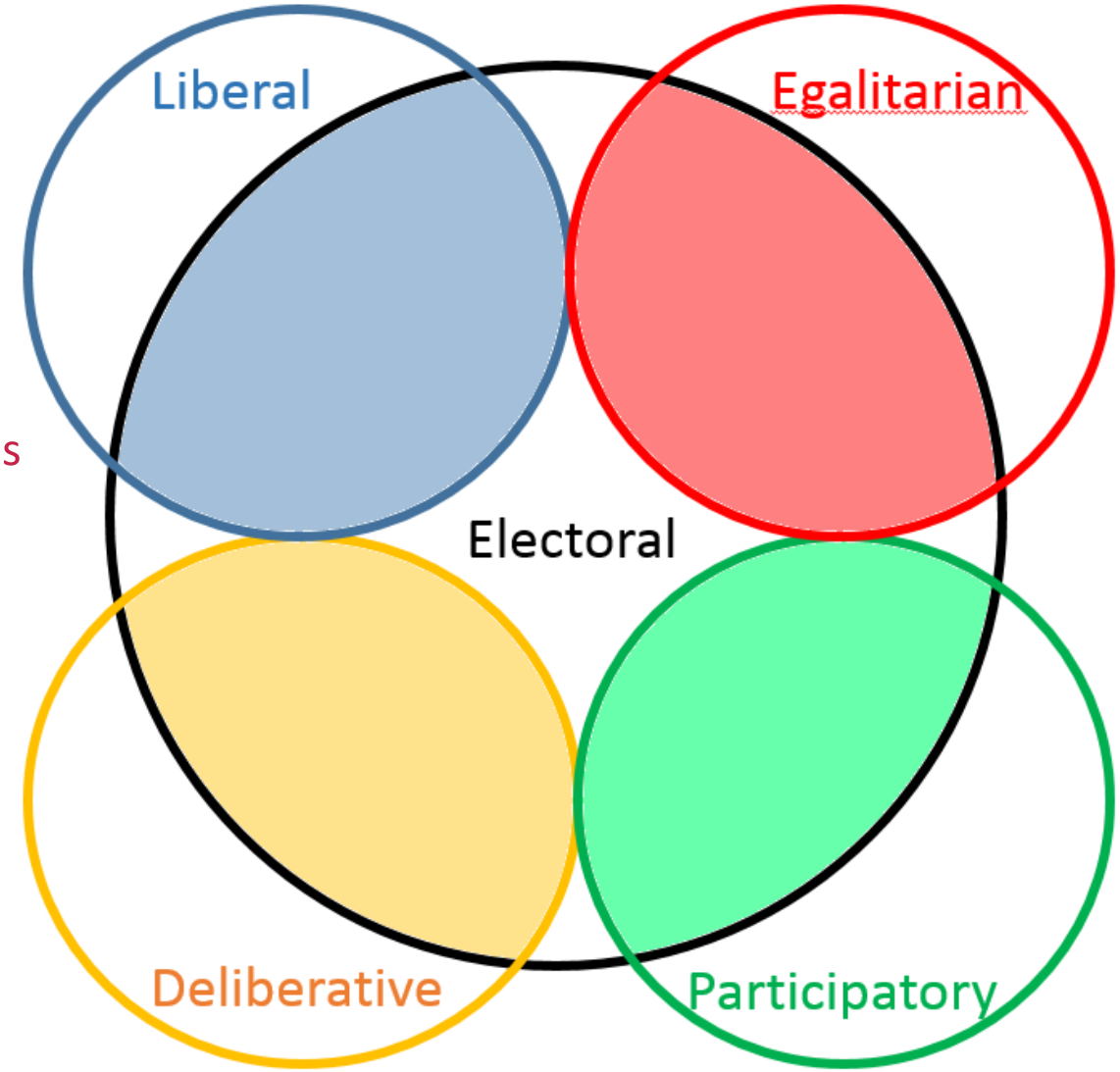
2 PhD students
2 undergraduate interns

6 other former postdocs
now working
all over the world

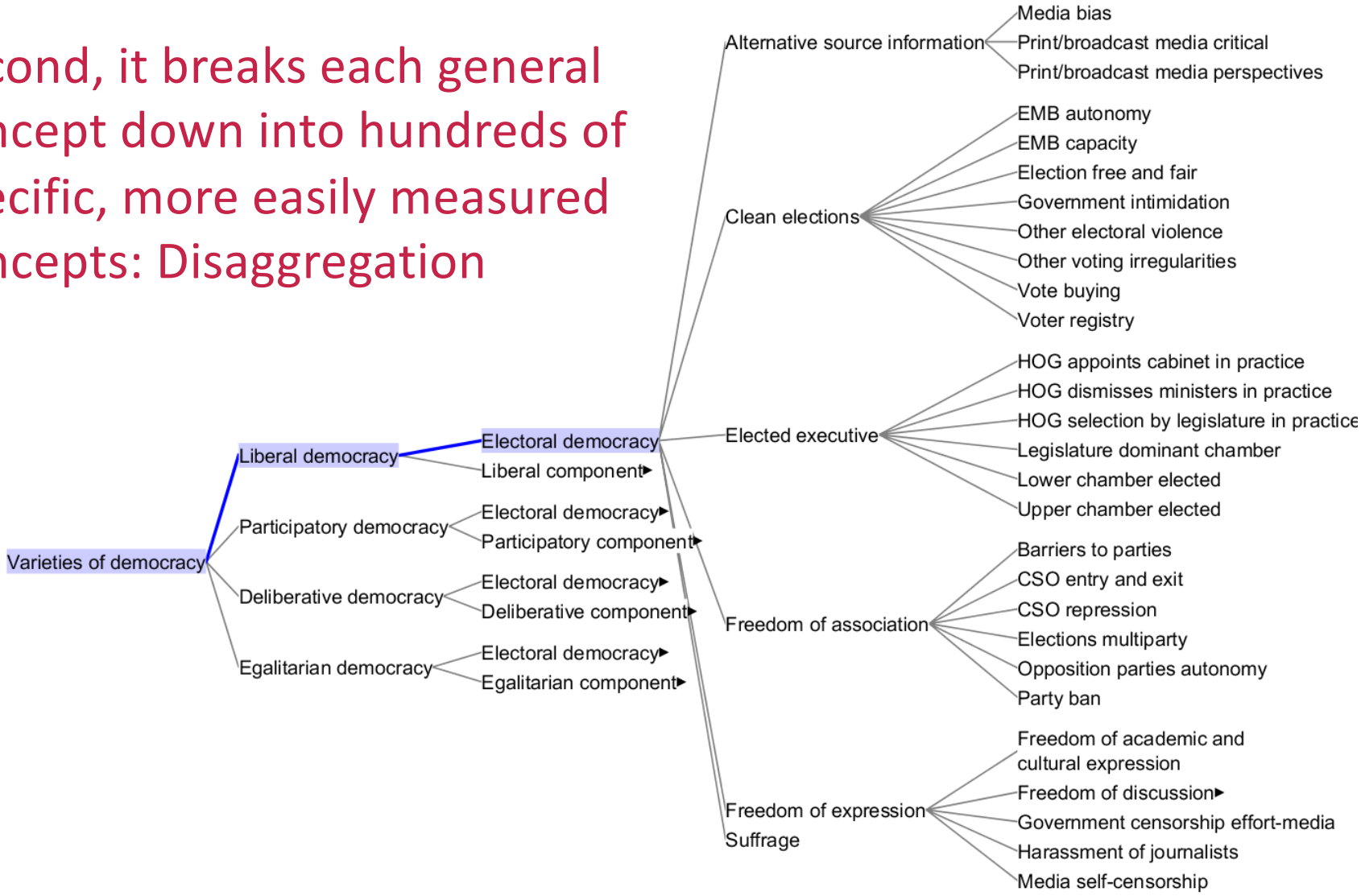
V-Dem Institute Research Associates (15)
Mostly former postdocs, now working
all over the world

What is distinctive about V-Dem data?

First, it captures multiple dimensions of democracy



Second, it breaks each general concept down into hundreds of specific, more easily measured concepts: Disaggregation



A sample question:
Election vote buying (C) (v2elvotbuy)

In this national election, was there evidence of vote and/or turnout buying?

Clarification: Vote and turnout buying refers to the distribution of money or gifts to individuals, families, or small groups in order to influence their decision to vote/not vote or whom to vote for. It does not include legislation targeted at specific constituencies, i.e., "porkbarrel" legislation.

- 0: Yes. There was systematic, widespread, and almost nationwide vote/turnout buying by almost all parties and candidates.
- 1: Yes, some. There were non-systematic but rather common vote-buying efforts, even if only in some parts of the country or by one or a few parties.
- 2: Restricted. Money and/or personal gifts were distributed by parties or candidates but these offerings were more about meeting an 'entry-ticket' expectation and less about actual vote choice or turnout, even if a smaller number of individuals may also be persuaded.
- 3: Almost none. There was limited use of money and personal gifts, or these attempts were limited to a few small areas of the country. In all, they probably affected less than a few percent of voters.
- 4: None. There was no evidence of vote/turnout buying.

Third, it taps the expertise of more than 4,200 experts in 180 countries around the world.

63 percent of the raw scores come from local experts.

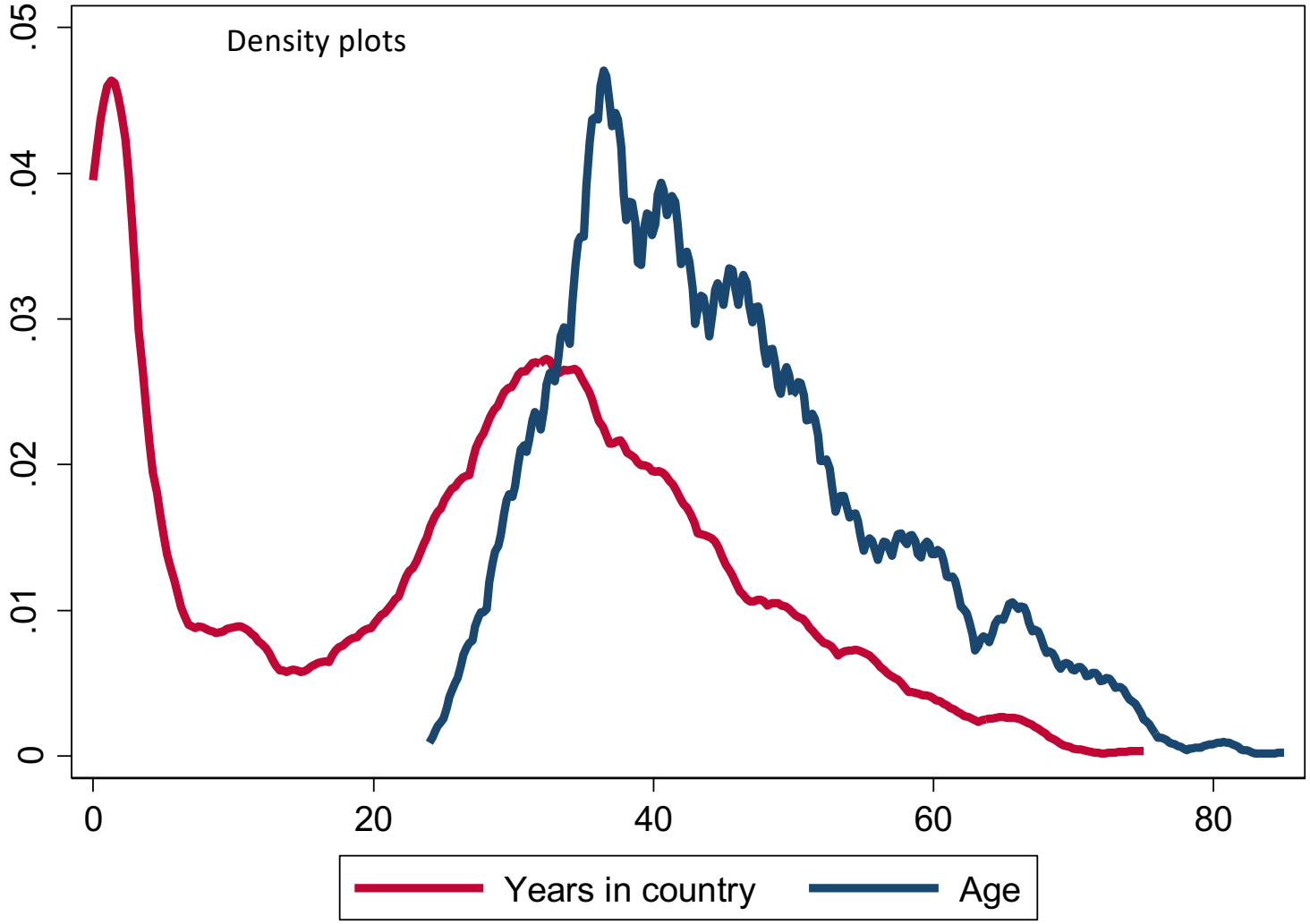
Who are the country experts?*

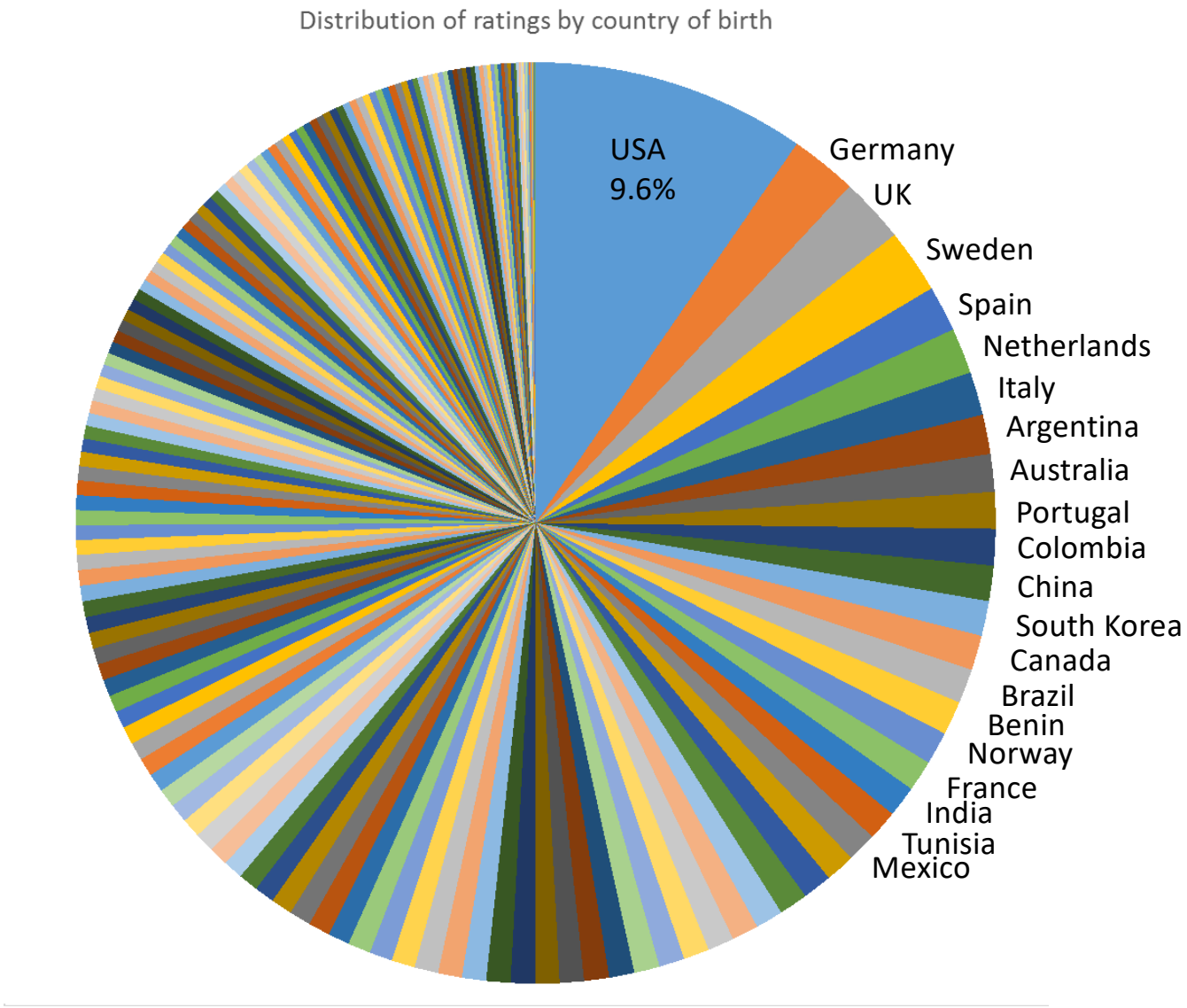
All are anonymous: identified only in administrative database that is separate from research database. But we can share summary information about them.

- 69% either born in or reside in the country they coded
- 67% not born in a Western country
- 94% have some graduate education
- 65% earned highest degree outside their country
- 64% academics, 23% private sector
- 27% women (v9)
- Hours spent on coding (v9): mean of 17, median of 10

*Based on v4 data, March 2015 unless otherwise noted

96.5% of the ratings are by people who have lived in the country they rated, usually for most of their lives.




Distribution of ratings by country of birth
USA
9.6%
Germany
UK
Sweden
Spain
Netherlands
Italy
Argentina
Australia
Portugal
Colombia
China
South Korea
Canada
Brazil
Benin
Norway
France
India
Tunisia
Mexico

Fourth,

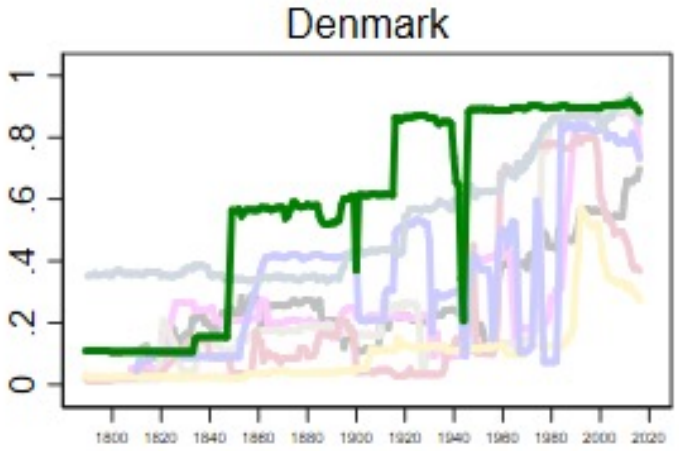
- Because we rely on the expertise of thousands of people from widely varying backgrounds,
- we go to great lengths to combine their scores in a way that maximizes comparability across countries and over time.*

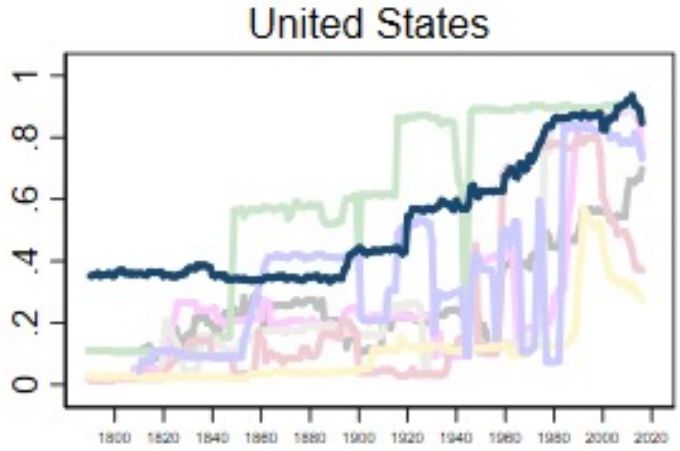
- As a by-product of this process, we also provide estimates of measurement uncertainty – Bayesian confidence intervals.

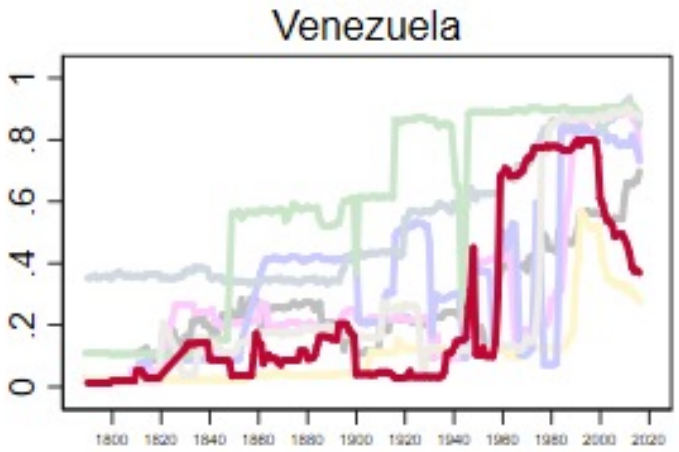
*Kyle L. Marquardt and Daniel Pemstein. October 2018. "IRT Models for Expert-Coded Panel Data," *Political Analysis*. DOI 0.1017/pan.2018.28

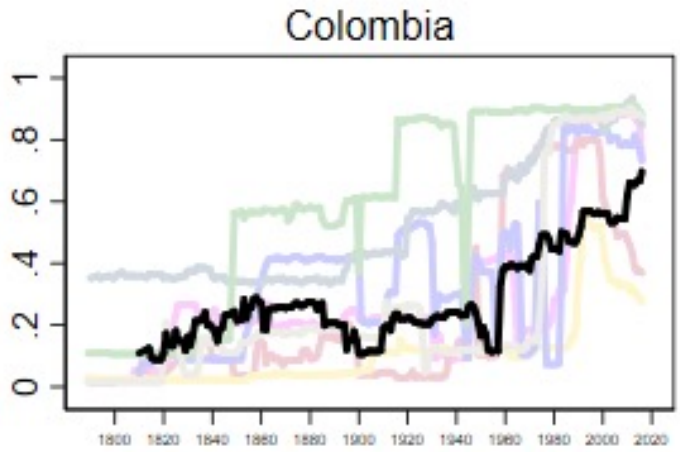
Fifth, V-Dem data now covers 1789-2018

Electoral Democracy Index









Thanks to our funders!

Some Descriptive Trends

Mostly from Carl Henrik Knutsen and Svend-Erik Skaaning, “The Ups and Downs of Democracy, 1789-2017,” draft chapter for Coppedge and Lindberg, “What Have We Learned about Democratization after 230 Years?” (book manuscript in progress).

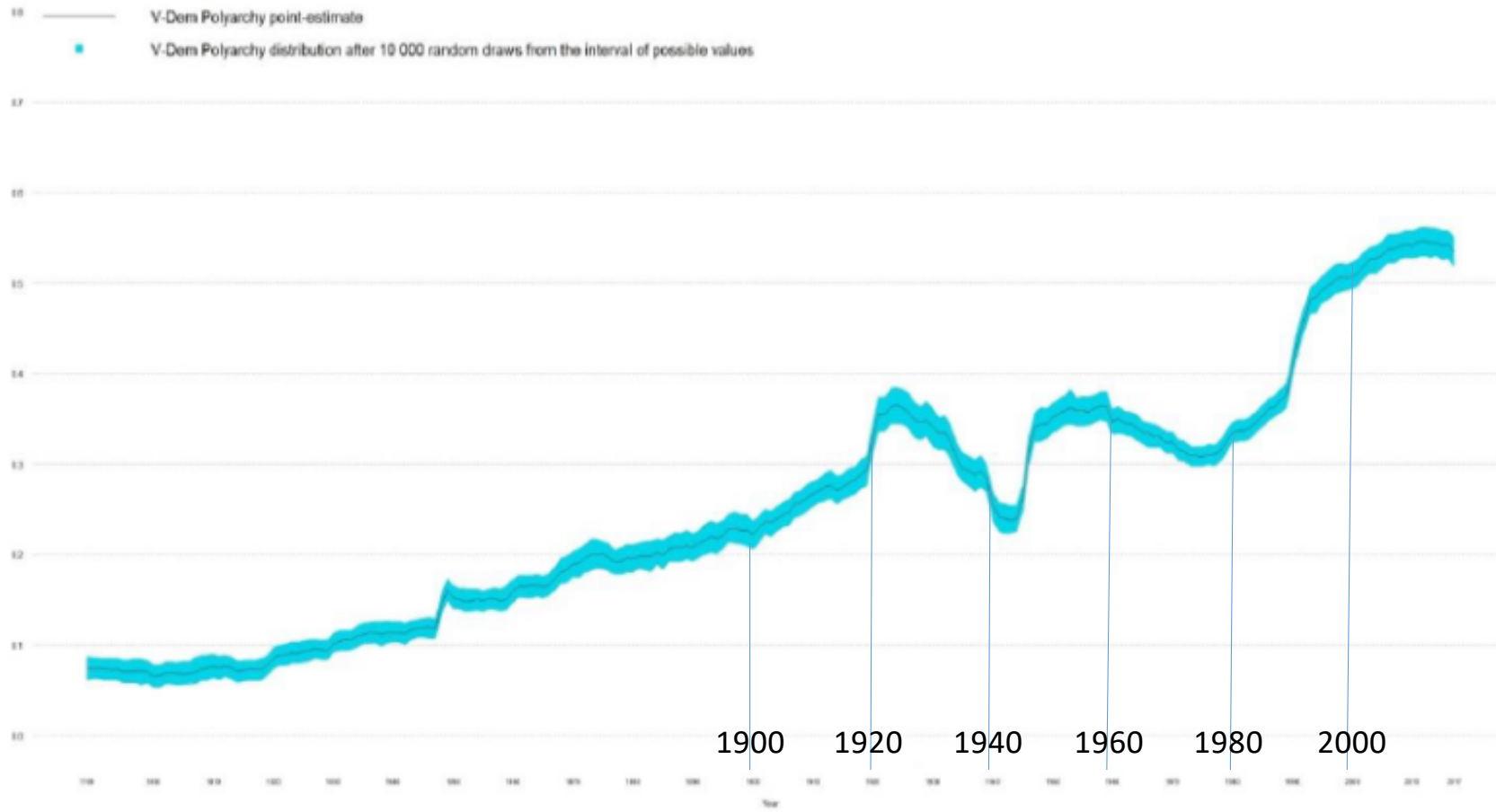
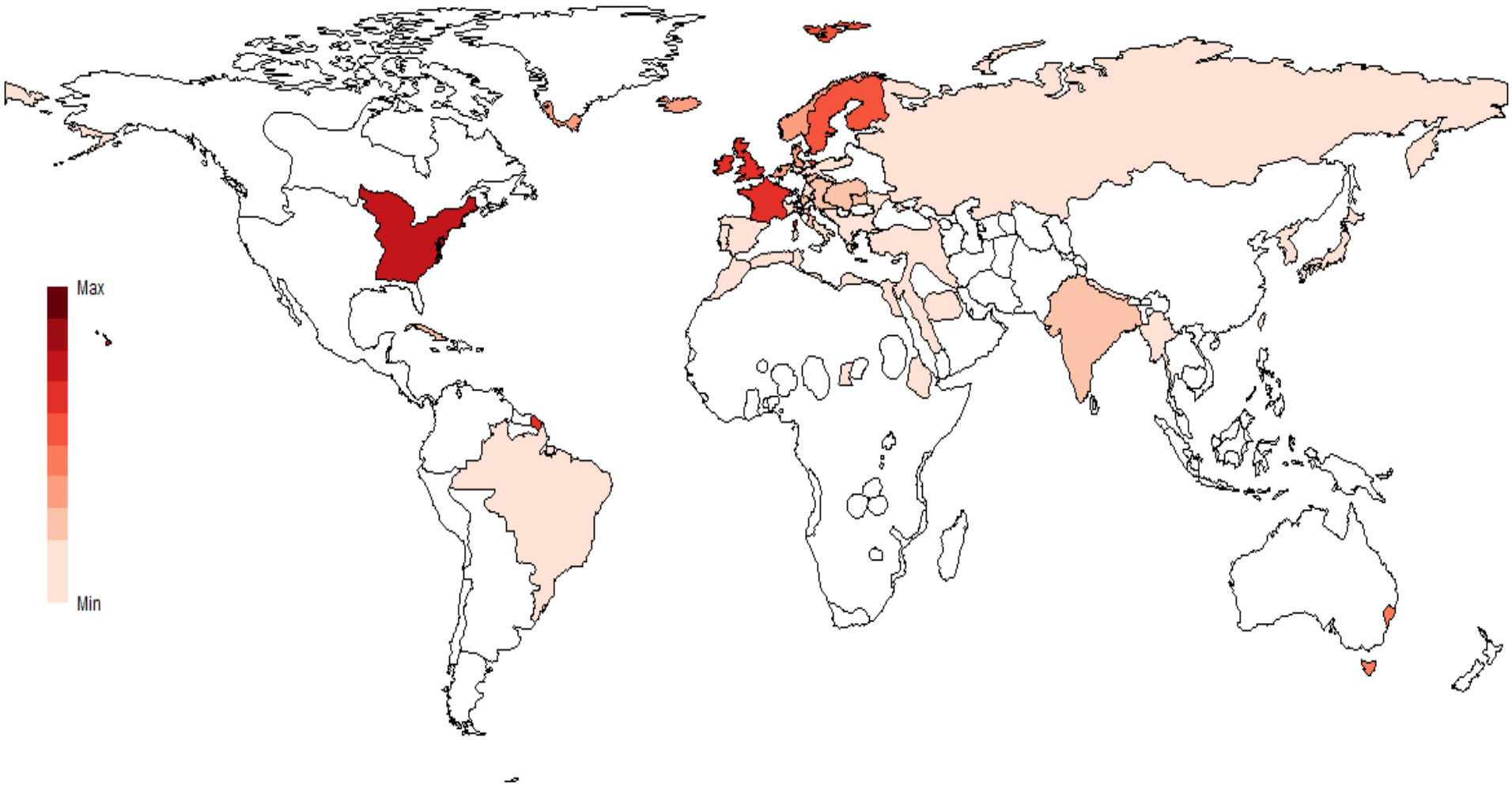


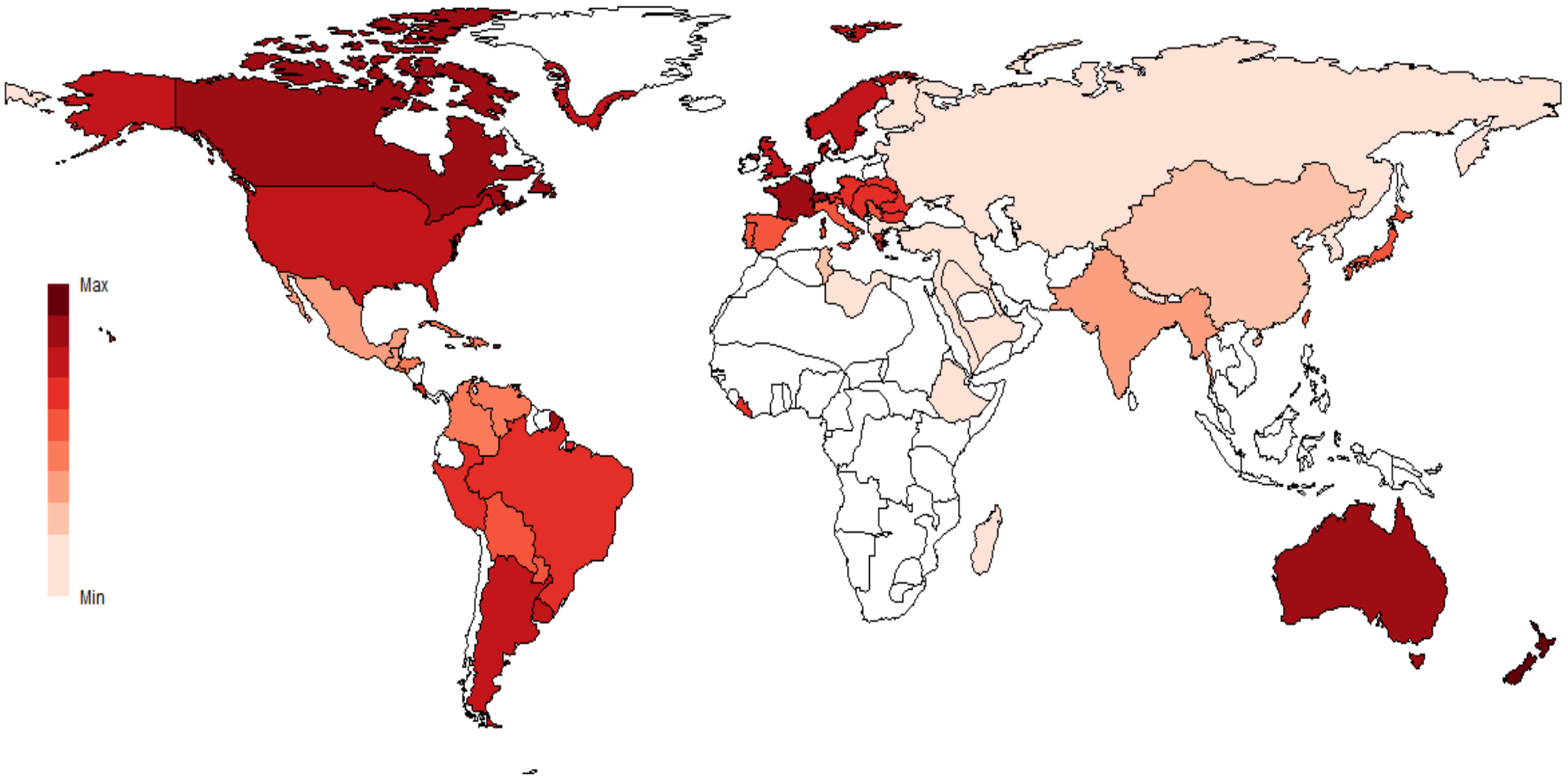
Figure 2 Global trend in Polyarchy from 1789-2017 for all independent countries

Notes: The black line represents best estimates and the blue uncertainty bounds incorporates global trends as calculated from 10,000 random draws from V-Dem dataset.

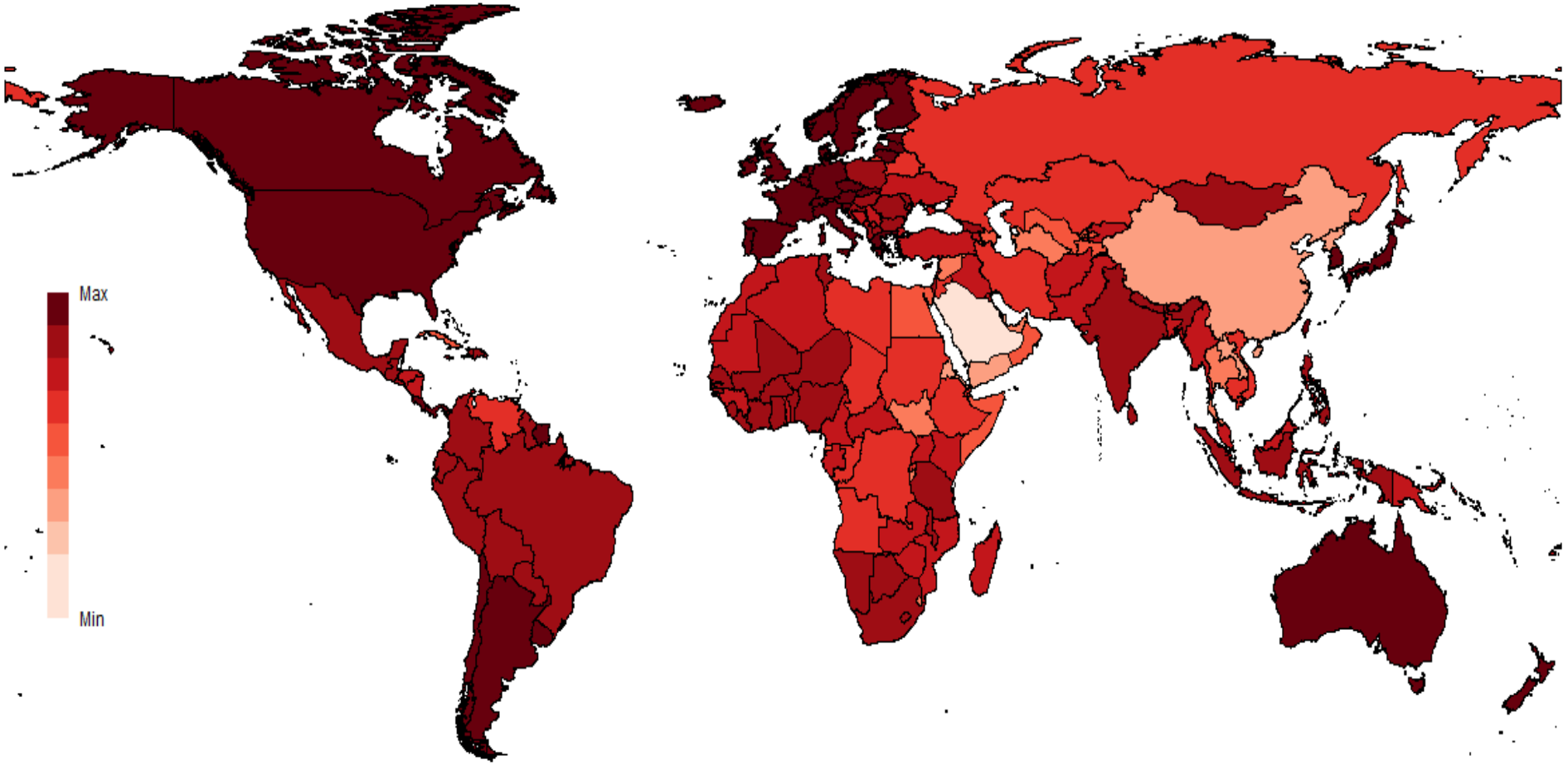
1789

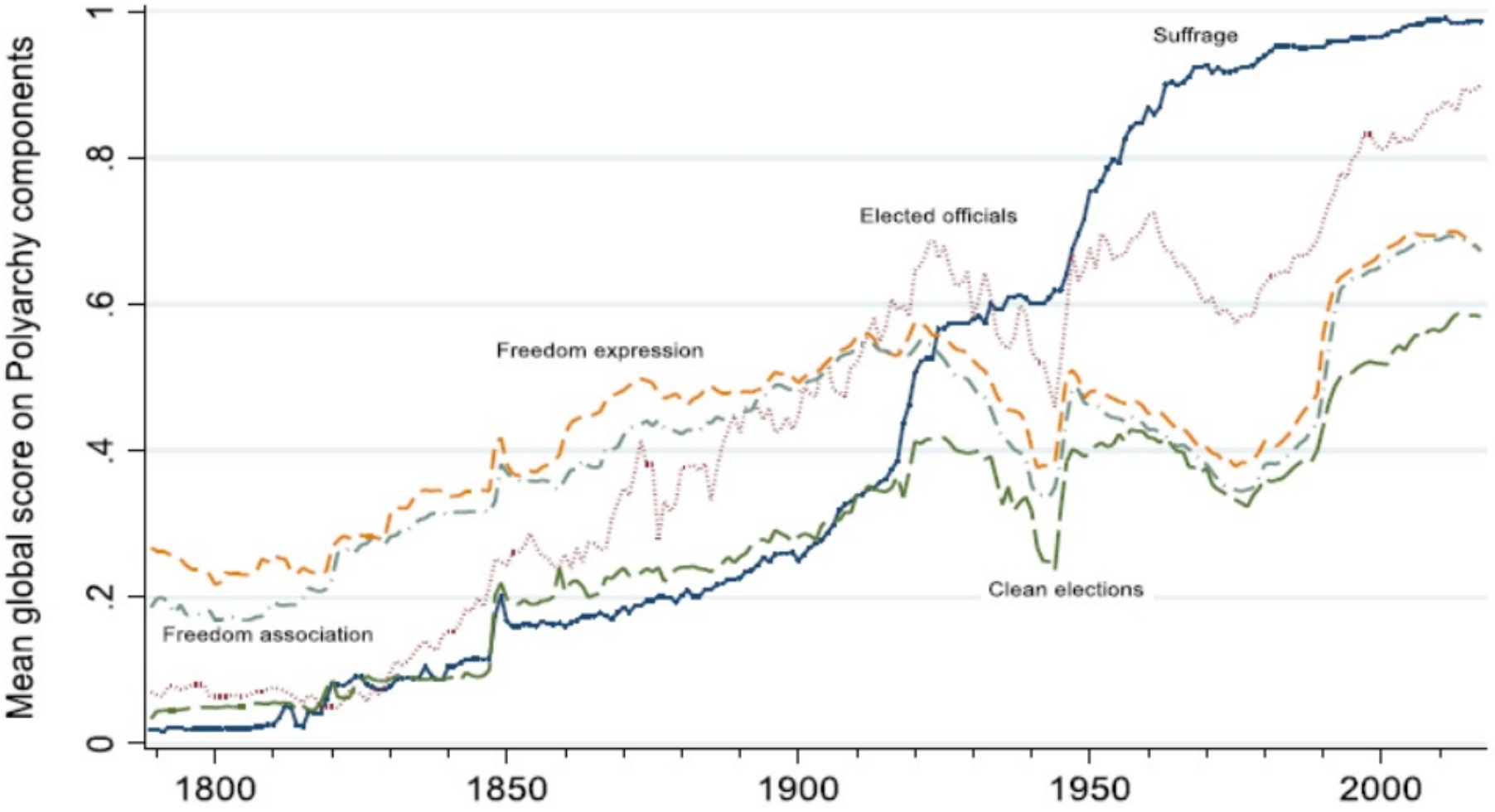


1921



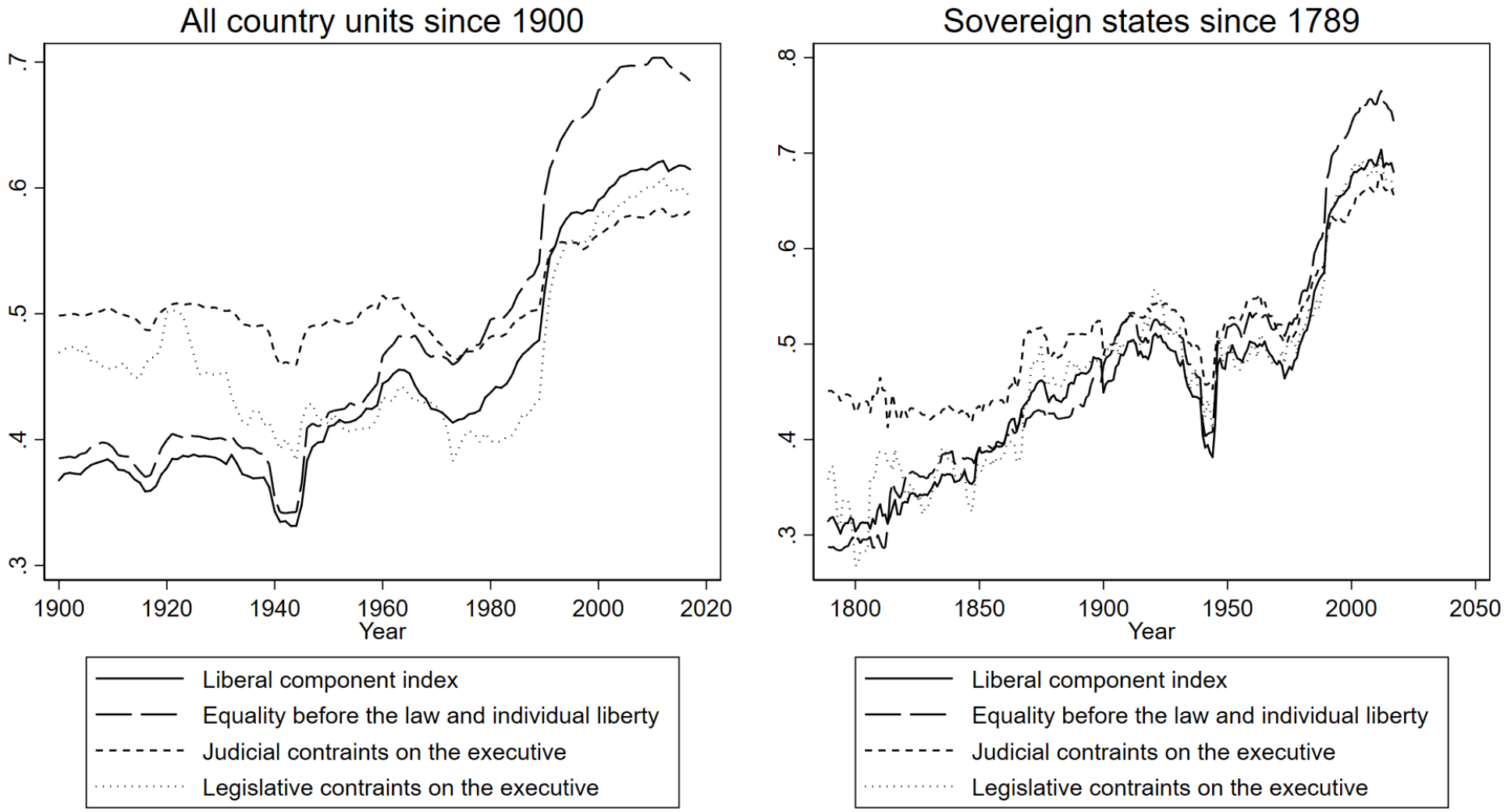
2017





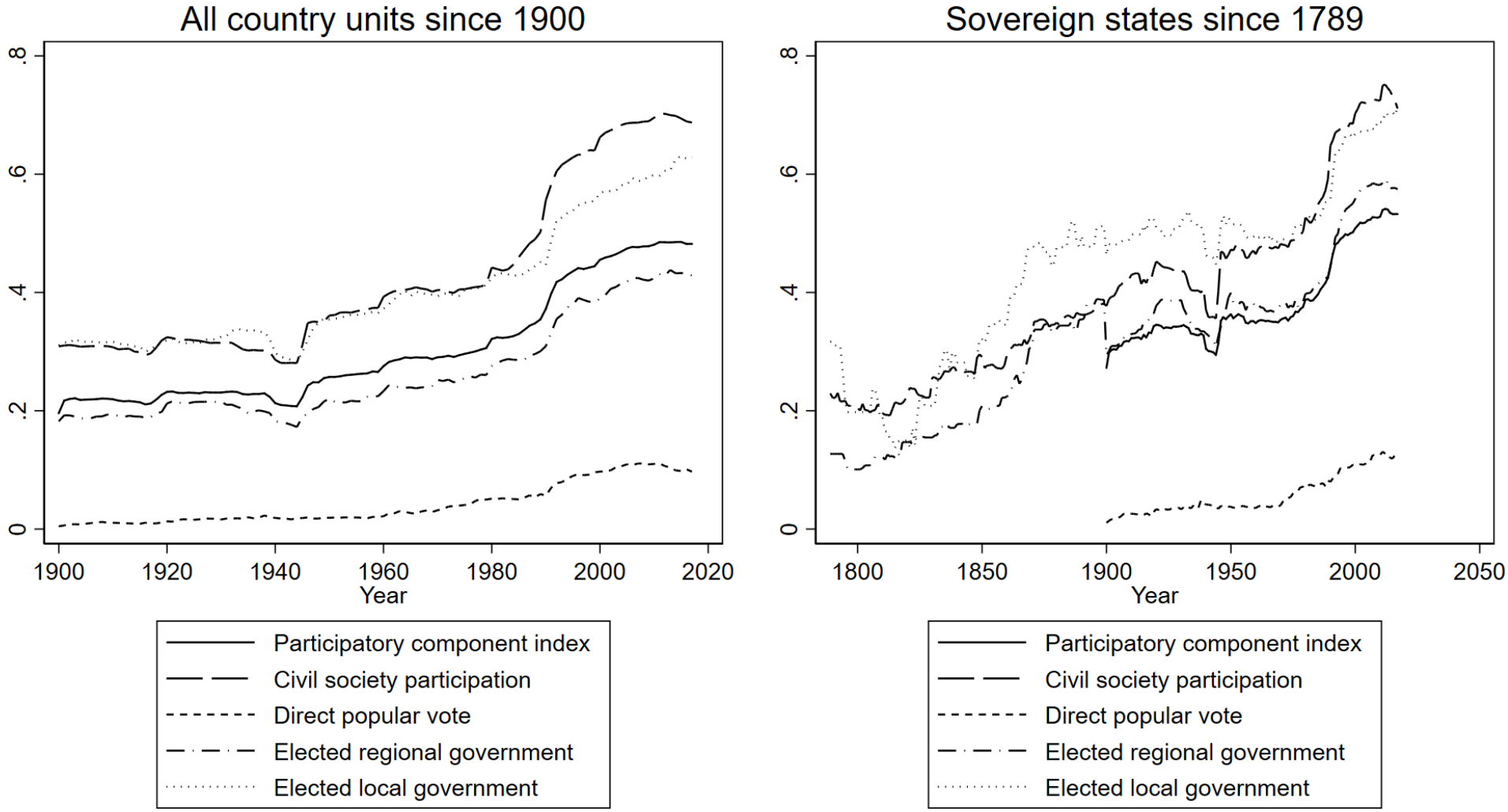
Mean Global Score of Polyarchy Components for all independent states

The liberal component and its subcomponents



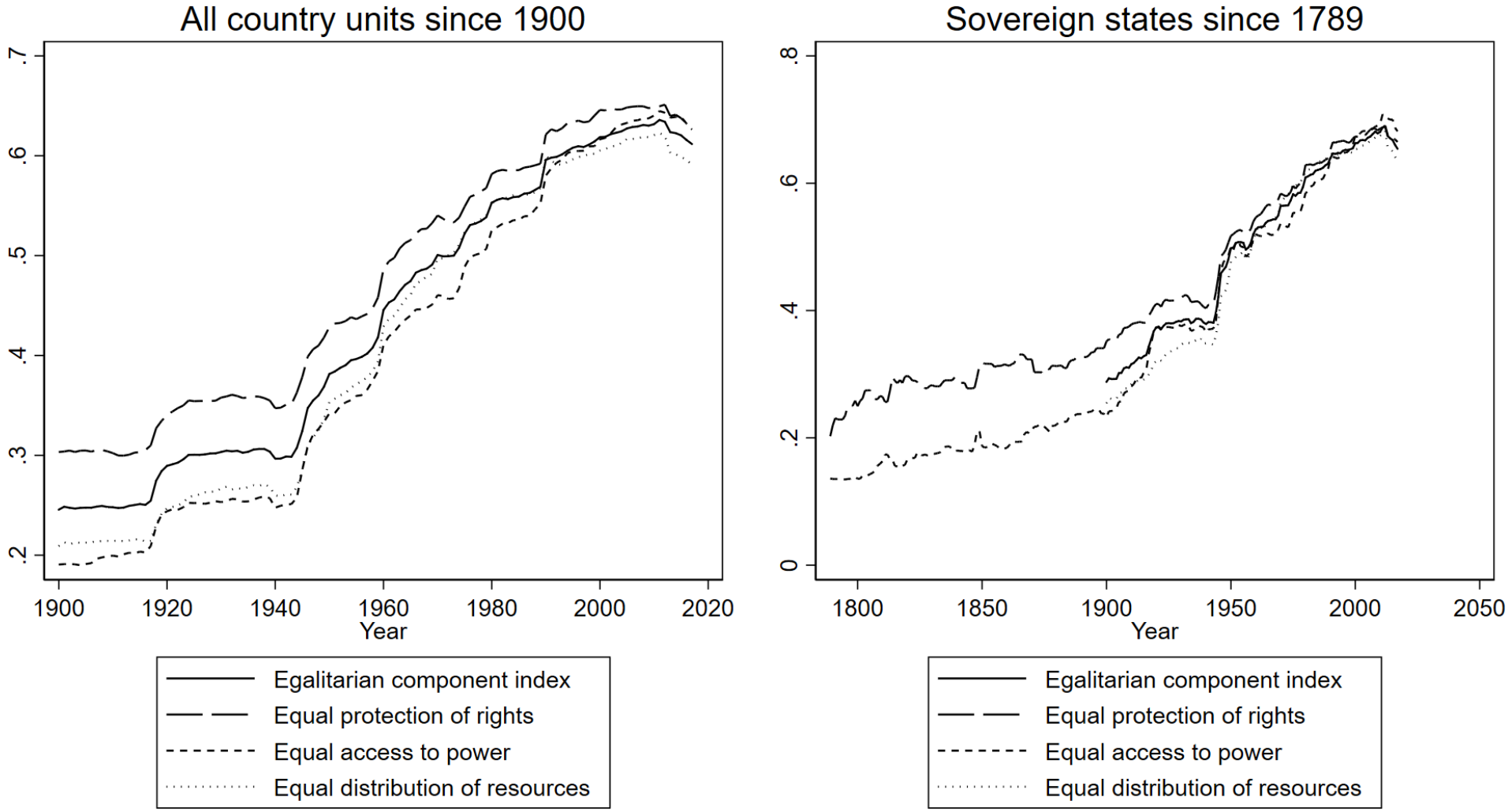
From Coppedge & 23 coauthors, *Varieties of Democracy: Measuring Two Centuries of Political Change* (Cambridge UP, forthcoming 2020).

The participatory component and its subcomponents

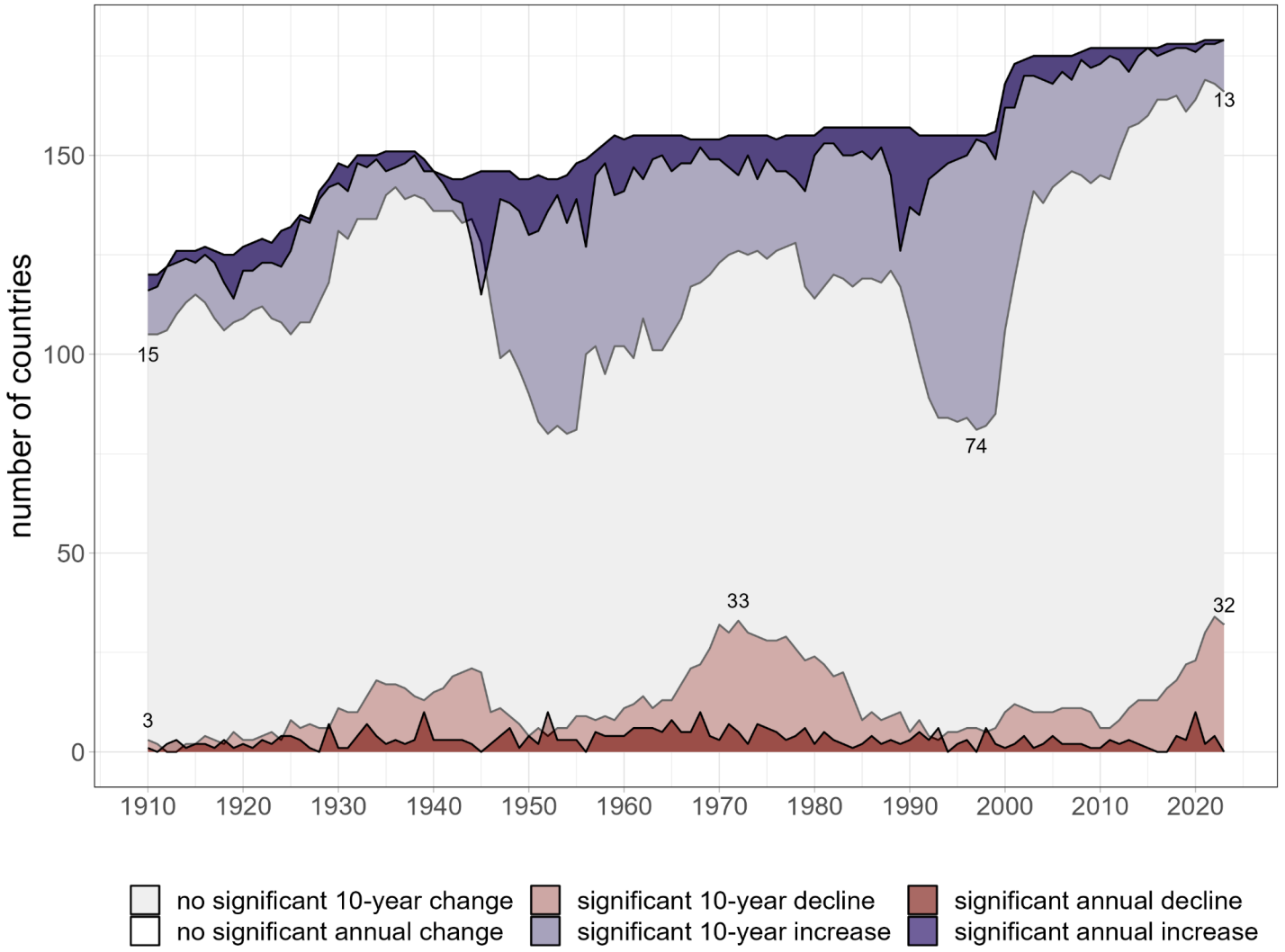


From Coppedge & 23 coauthors, *Varieties of Democracy: Measuring Two Centuries of Political Change* (Cambridge UP, forthcoming 2020).

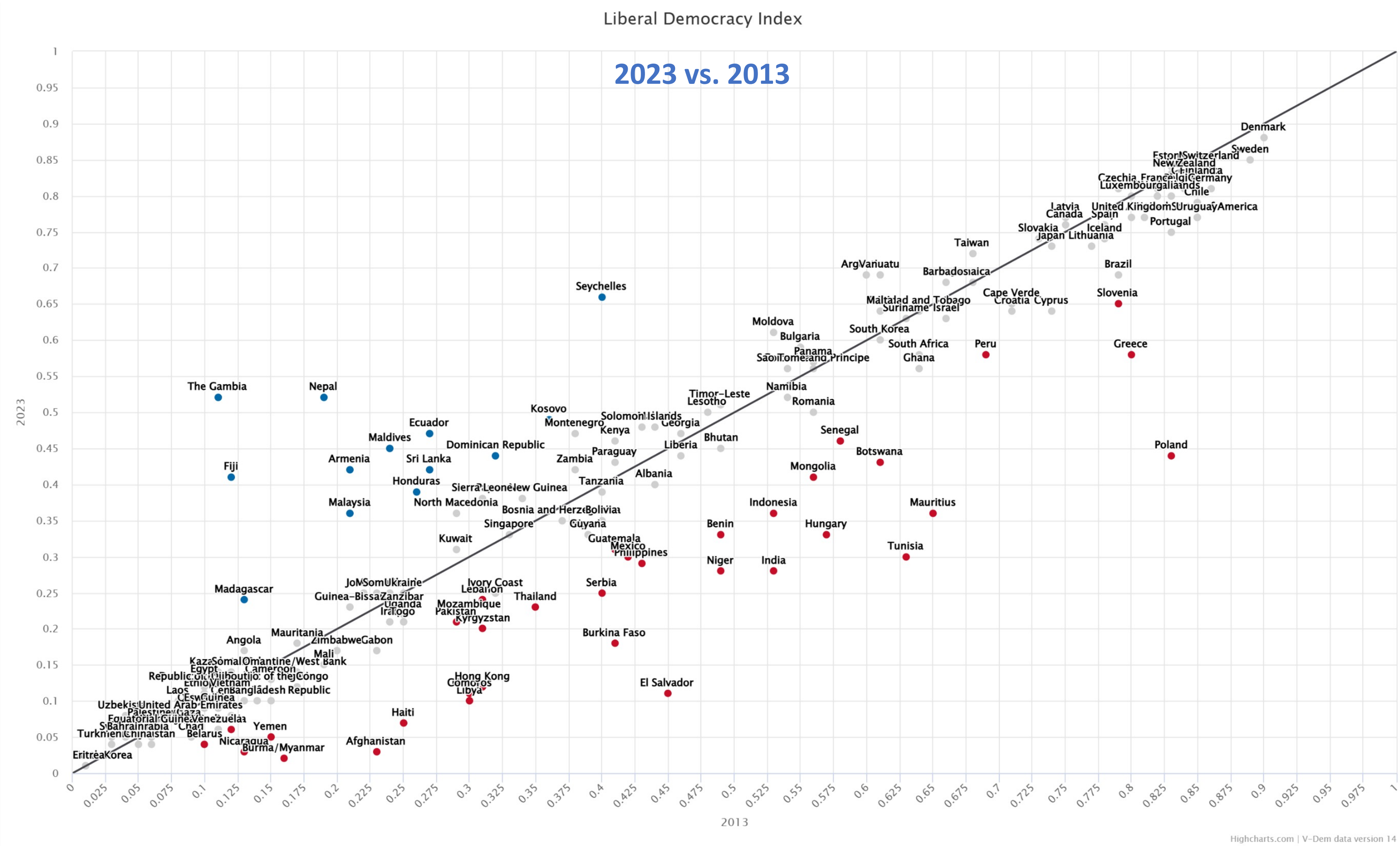
The egalitarian component and its subcomponents



From Coppedge & 23 coauthors, *Varieties of Democracy: Measuring Two Centuries of Political Change* (Cambridge UP, forthcoming 2020).



The widths of the color bands represent the number of countries in each category.
Categories reflect changes over 1 or 10 years that exceeded their country's 95% HPD confidence bounds.


Liberal Democracy Index
2023 vs. 2013
2023
2013
Highcharts.com | V-Dem data version 14
Denmark
Sweden
Chile
Portugal
Latvia
Canada
Spain
Slovakia
Iceland
Taiwan
Brazil
Barbados
Cape Verde
Croatia
Cyprus
Slovenia
Seychelles
Moldova
Bulgaria
South Korea
South Africa
Panama
Ghana
Peru
Greece
The Gambia
Nepal
Namibia
Timor-Leste
Lesotho
Romania
Kosovo
Montenegro
Kenya
Georgia
Ecuador
Maldives
Dominican Republic
Senegal
Bhutan
Liberia
Paraguay
Zambia
Botswana
Poland
Armenia
Sri Lanka
Fiji
Albania
Mongolia
Honduras
Tanzania
Malaysia
North Macedonia
Indonesia
Mauritius
Singapore
Guyana
Benin
Hungary
Kuwait
Guatemala
Mexico
Philippines
Tunisia
Niger
India
Madagascar
Ivory Coast
Lebanon
Serbia
Thailand
Mozambique
Pakistan
Kyrgyzstan
Mauritania
Angola
Mali
Burkina Faso
Hong Kong
Comoros
Libya
El Salvador
Laos
Haiti
Yemen
Belarus
Nicaragua
Burma/Myanmar
Afghanistan

Alternative forms of the data

This would be a good time to open the dataset.

Relative scale

- Variables with no suffix: v2svinlaut, etc.
- The mean of many draws from the measurement model output
- Accompanied by *_codelow and *_codehigh bounds of the 70% highest posterior density (HPD) interval
- Best for most analyses: continuous, interval-level estimates

Ordinalized version

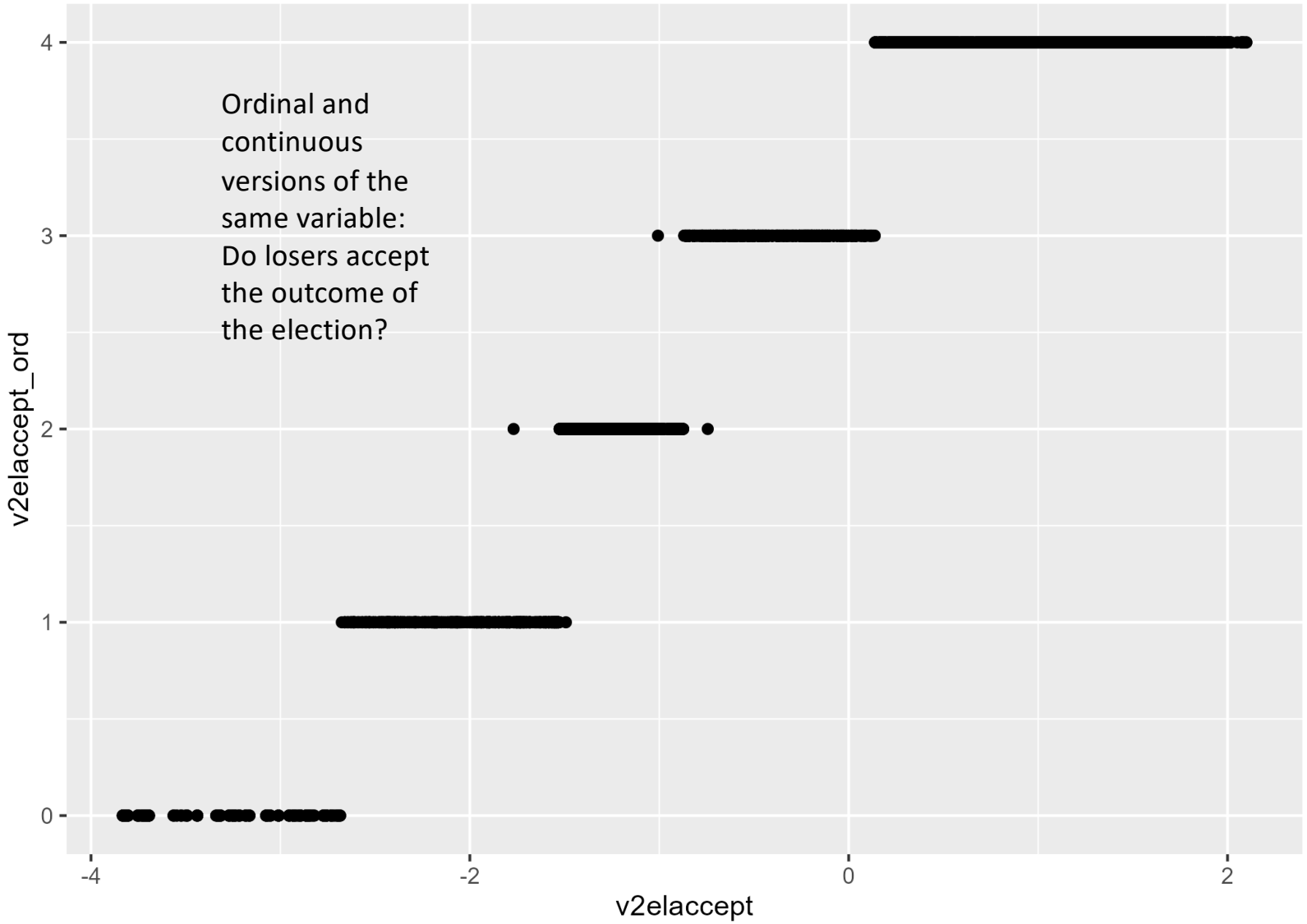
- Has the suffix *_ord
- The most probable original ordinal scale score (0, 1, 2, etc.) corresponding to the continuous MM estimates
- Includes *_ord_codelow and *_ord_codehigh HPD bounds, which are also integers.
- Appropriate if you need discrete indicators, for example for hazard rate models

Linearized Ordinal-Scale Posterior Prediction

- Also called “original scale” on the website
- Has the suffix *_osp and includes upper and lower bounds
- Intended to be the MM estimates rescaled to the original scale, but with degrees of closeness
- Calculated as a weighted average of each original score, weighted by the probability of that score.
- In line graphs, makes it easier to match scores to coding criteria.
- *Do not use in analyses*: not equal intervals; not necessarily closest to the most likely score.

K-chotomy classifications

- Have suffixes *3C, *4C, or *5C
- The relative scale values divided into 3, 4, or 5 ordinal categories
- Requested by those who want all variables recoded into the same number of categories
- Not recommended for most purposes (if at all)


Ordinal and continuous versions of the same variable: Do losers accept the outcome of the election?
v2elaccept_ord
4
3
2
1
0
-4
-2
0
2
v2elaccept

Current codebook entry

Question: Did losing parties and candidates accept the result of this national election within three months?

Responses:

0: None. None of the losing parties or candidates accepted the results of the election, or all opposition was banned.

1: A few. Some but not all losing parties or candidates accepted the results but those who constituted the main opposition force did not.

2: Some. Some but not all opposition parties or candidates accepted the results but it is unclear whether they constituted a major opposition force or were relatively insignificant.

3: Most. Many but not all opposition parties or candidates accepted the results and those who did not had little electoral support.

4: All. All parties and candidates accepted the results.

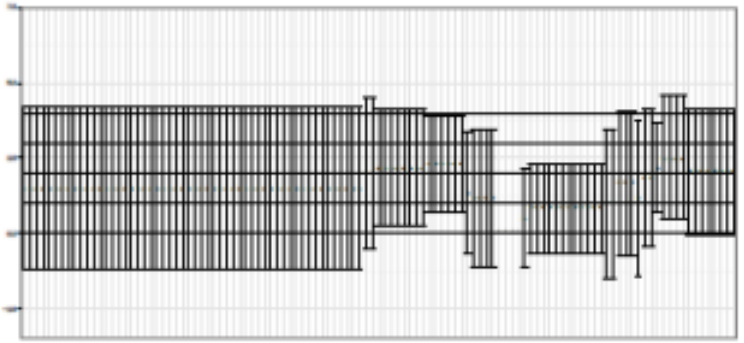
Proposed revision

Question: Did losing parties and candidates accept the result of this national election within three months?

Interval	Ordinal	Interpretation
-3.68 to -2.67	0	None. None of the losing parties or candidates accepted the results of the election, or all opposition was banned.
-2.67 to -1.54	1	A few. Some but not all losing parties or candidates accepted the results but those who constituted the main opposition force did not.
-1.54 to -0.88	2	Some. Some but not all opposition parties or candidates accepted the results but it is unclear whether they constituted a major opposition force or were relatively insignificant.
-0.88 to 0.14	3	Most. Many but not all opposition parties or candidates accepted the results and those who did not had little electoral support.
0.14 to 2.1	4	All. All parties and candidates accepted the results.

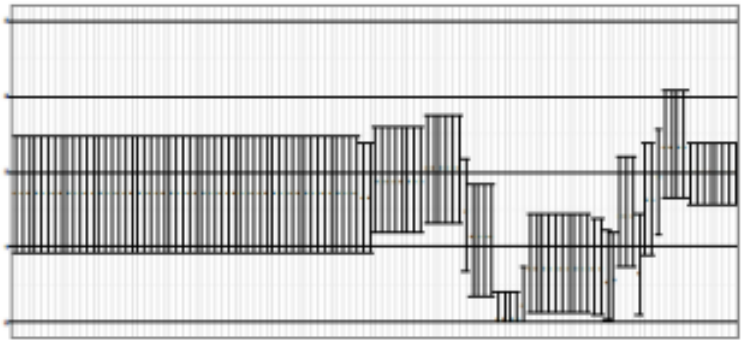
Four different versions

Figure 6: Longitudinal trends in freedom from political killings in Cambodia, 1900-2012

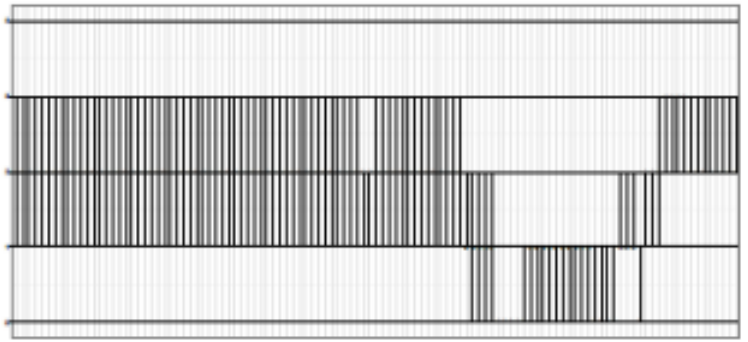


(a) Mean and standard deviation

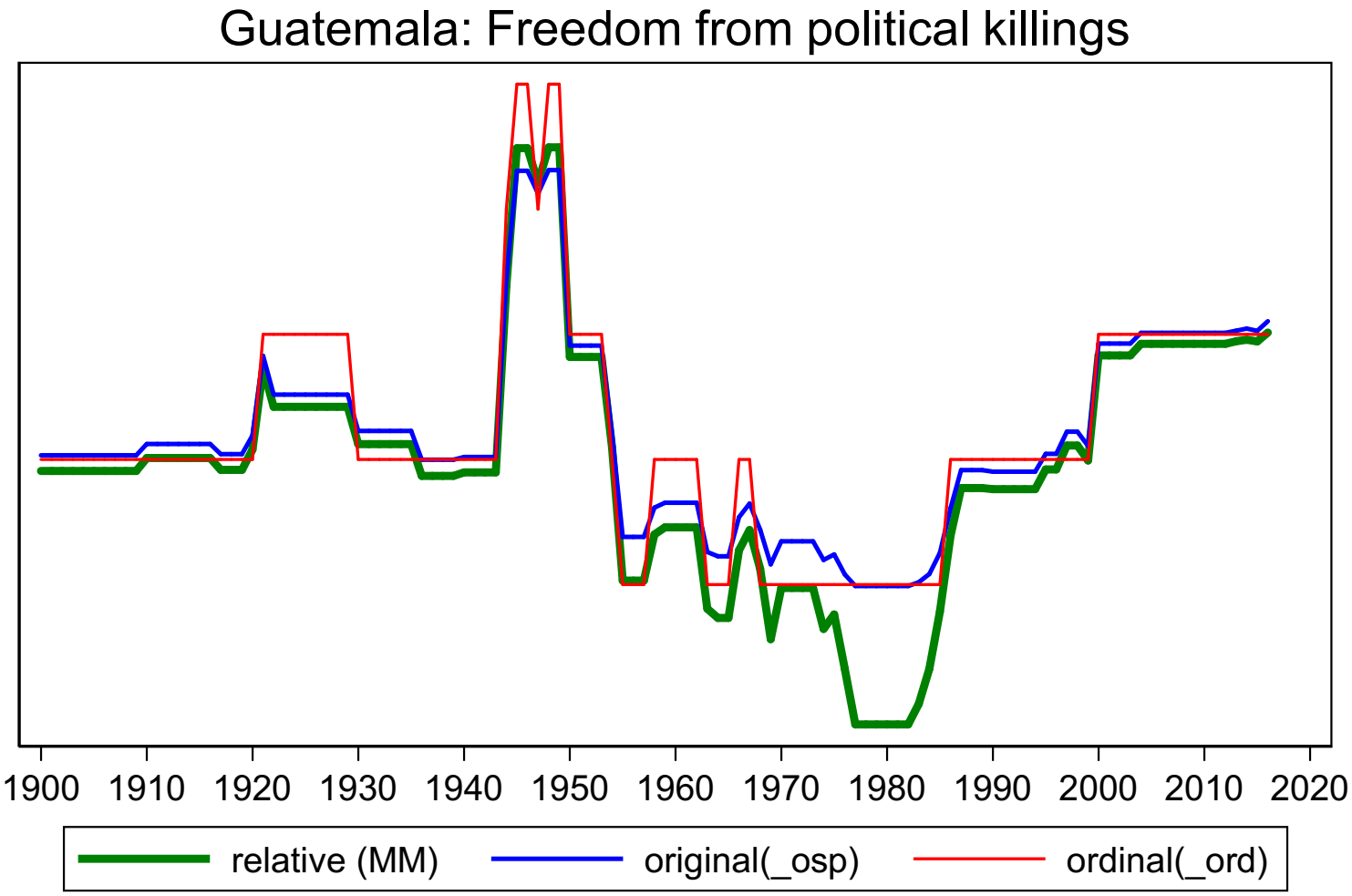
(b) Measurement model output



(c) Linearized original scale



(d) Ordinal scale

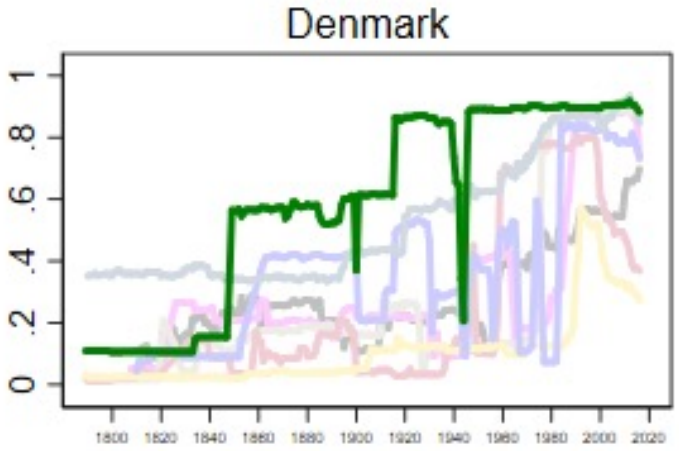


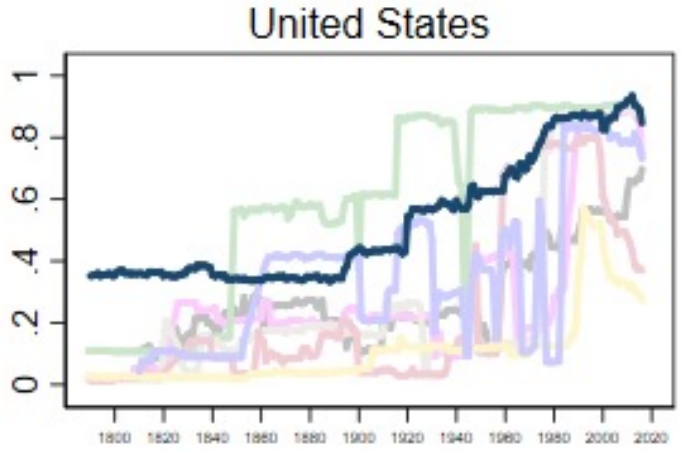
Constants have been subtracted from original and ordinal scales to maximize overlaps.

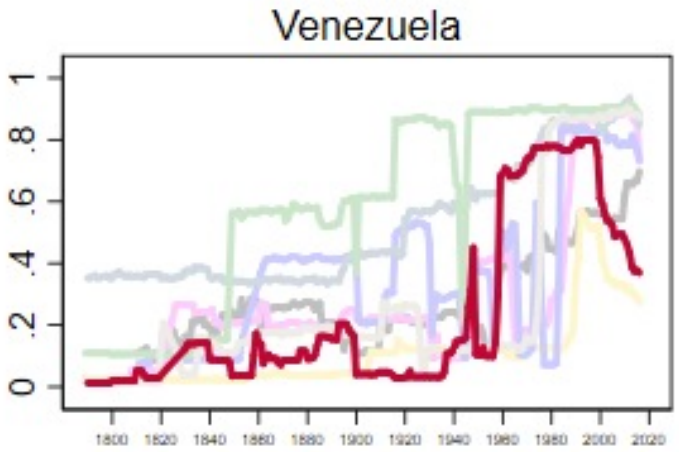
New Surveys

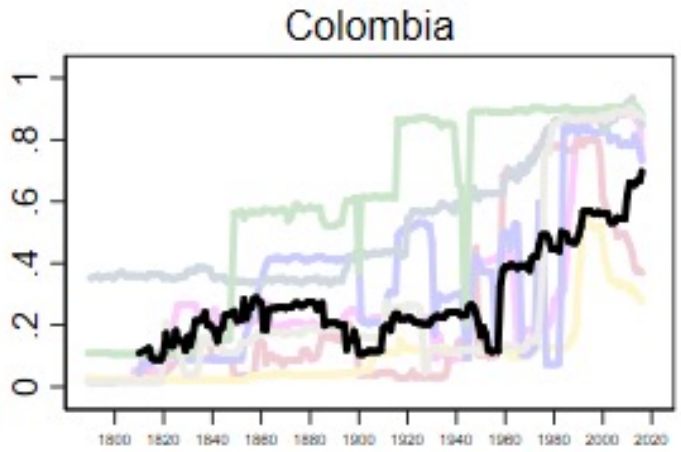
Historical V-Dem: covers 1789-2022

Electoral Democracy Index









A Swedish national data infrastructure clearinghouse that combines the data of

- Varieties of Democracy
- The Quality of Government Institute
- The Uppsala Conflict Data program

Demscore has secured full funding to continue for the foreseeable future.

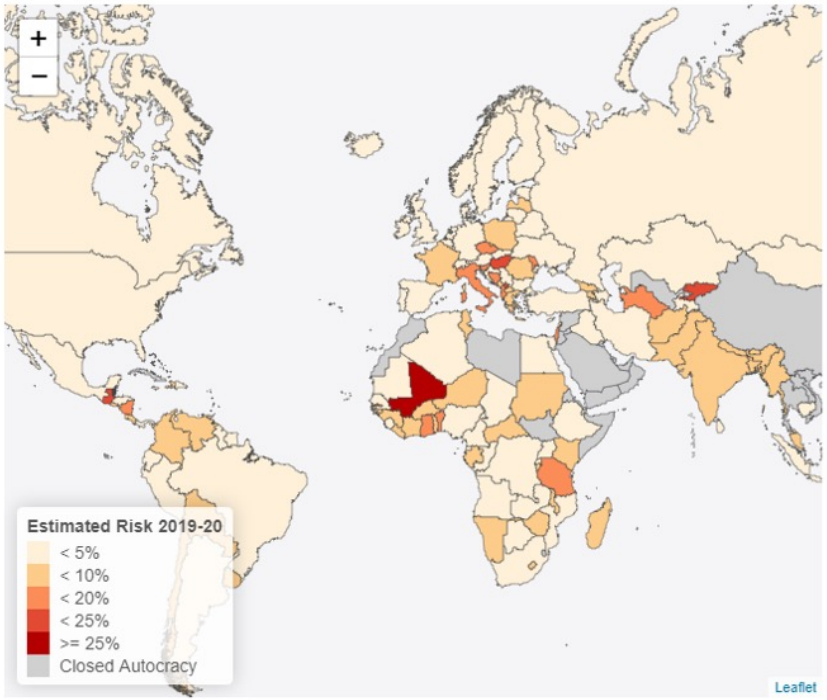
Case for Democracy

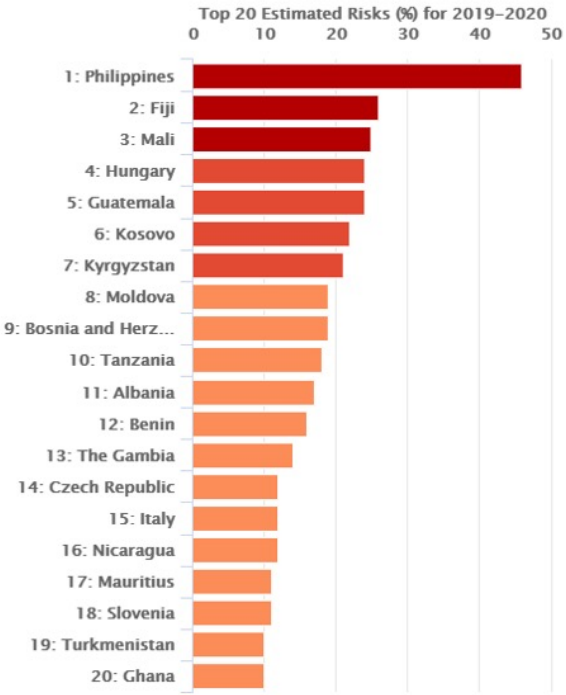
digital society project

Provides high-quality, publicly available, data describing the intersection between politics and social media in countries around the world:

- online censorship
- polarization and politicization of social media
- disinformation campaigns
- coordinated information operations
- foreign influence in and monitoring of domestic politics.
- candidate social media presence

V-Forecast





- DemSpace: forecasting openings, closings, and stability on six dimensions of democracy
- Predicting Adverse Regime Transitions (PART)

Some other spinoff projects

- Academic and Civic Space
- Regime Legitimation
- Pandemic Backsliding Project (PanDem)
- The Failing and Successful Sequences of Democratization (FASDEM)
- Varieties of Autocracy and Autocratization (V-Aut)
- V-Party: Data on orientations of political parties historically and around the world, with special attention to populism and anti-pluralism. See V-Party Explorer on the graphing page.
- Varieties of Indoctrination

Data Collection

Different modes of data collection

- 167 A indicators: centrally coded by PMs or RAs

Relatively objective and well documented by others

- David Altman: 38 direct democracy indicators
- Jeff Staton: 47 de jure judicial institutions
- Svend-Erik Skaaning: 1 suffrage indicator
- RAs in Sweden (81 variables):*
  - Recoding of NELDA, CCP, and other data
  - Original coding of certain characteristics of executives, legislatures, and elections
  - 9 of these are pre-coded before surveys go out

Different modes of data collection

- B data: coded by Country Coordinators
  - Relatively objective but hard to find
  - 27 indicators of characteristics of executives, legislators, and elections
  - A growing number of these are now being coded centrally
- D data: 86 indices or other variables calculated from V-Dem variables
- 24 Post-Survey Questions (coder characteristics, not public)
- E data: 179 indicators from other datasets (being reduced)

236 C indicators, from online surveys
These require subjective judgment

- 37 on Elections
- 27 on Civil liberties
- 24 on Executive
- 15 on Political parties
- 14 on Legislature
- 13 on Judiciary
- 11 on Media
- 10 on Civil society
- 7 on Political equality
- 7 on Deliberation
- 4 on Sovereignty
- 4 on Regime Legitimation
- 25 on Exclusion
- 38 on Digital Society

New!

+ 97 new indicators for Historical V-Dem

The Process

- In the first waves, we collected data in short waves throughout the year, as funding and technology permitted.
- Now we collect data once a year
  - “A” data updated mid-year
  - Coders recruited in the fall
  - All “C” ratings submitted in January
  - Measurement model run and cleaning done in February and March
  - Data released in March or April
  - Along with annual “Democracy Report”

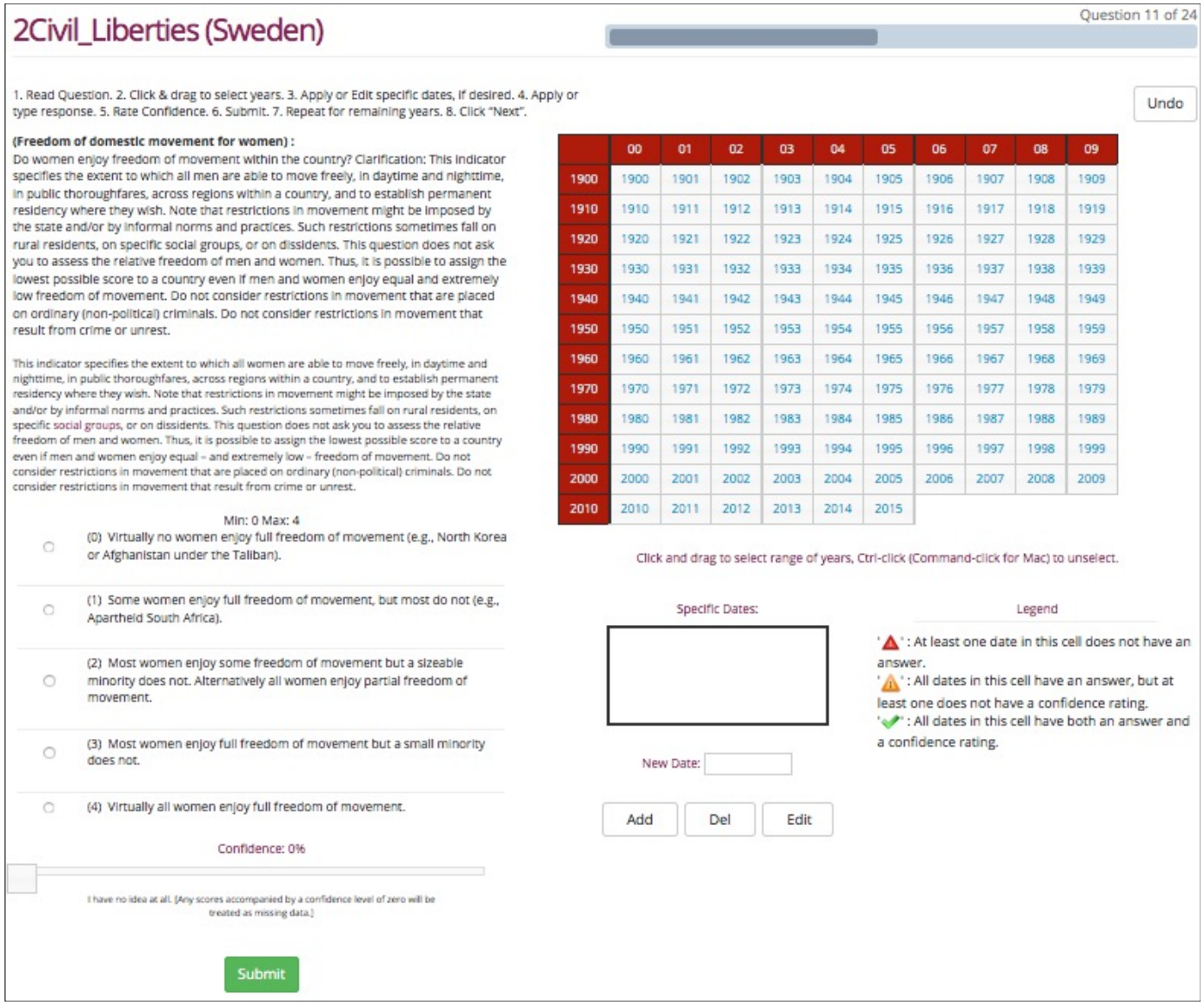
The country experts

- We have multiple thematic “surveys”
- Most country experts do multiple surveys; few can do all of them
- We aim for at least 5 country experts per question for each country-year
- We don’t publish data based on fewer than 3 country experts
- All country experts are anonymous
- After the survey, they are invited to answer “vignette” questions
- We have sometimes done survey experiments, as with forecasting in v9 and v10.

The surveys

- Available in English, Spanish, French, and Portuguese. (Arabic and Russian discontinued.)
- In the initial waves (through 2012), country experts answered each question for each year, 1900 (or first year of existence) to 2012.
- However, most election questions are coded only for election years.
- The survey interface contains a grid of years by decade that makes it easy to code spans of years at a time.
- Country experts have to assign 0-100% confidence to each rating.
- They have opportunities to type in comments.
- Since 2012, for continuing questions, we have done updates. Updates go back 10-15 years to help link up with earlier data.
- New questions go back varying lengths of time.

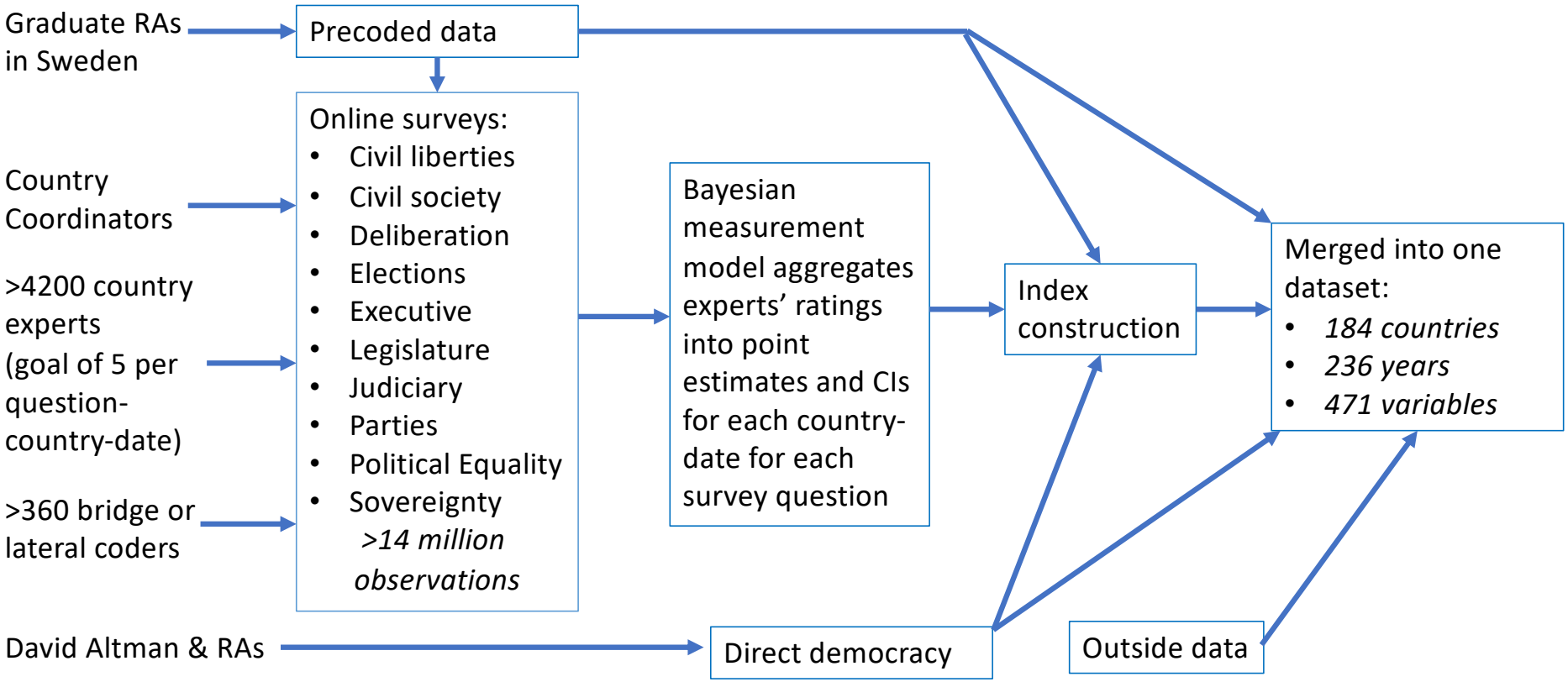
Screenshot of coding interface



Three kinds of coding

- Country experts
  - Typically code one country for 1900-present
  - They provide the within-country trends and much of the guidance about levels
- Lateral coders
  - Typically code several countries for one year (2012)
  - They improve the cross-national comparability
- Bridge coders
  - They code more than one country for the whole 1900-present period.
  - Most valuable but hard to recruit!

How do we generate V-Dem data?



Ongoing improvements

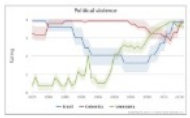
- More bridge coding
- Historical V-Dem (Teorell, Knutsen, Gerring, Skaaning, Ziblatt, Cornell)
  - Back to 1789 or 1800, wherever possible
  - One expert per country, chosen for historical expertise
- Vignettes (Seim, Glynn, Pemstein, Gerring)
  - The recommended way to anchor coder thresholds
  - This is being done for updates since 2016.
  - We have >50,000 vignette ratings so far.
  - It asks experts to rate a pair of hypothetical vignettes on several key variables in each survey they do.
  - It does not cover all past coders, but will help. Eventually we hope to have almost all past coders answer the vignettes.

Using the site: v-dem.net

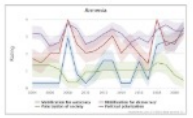
Online analysis tools, downloading, archive, reference materials available online

Fifteen different analysis tools are now available online!

Mapping Tool
Interactive tool that visualize data by creating a color-coded map to view the distribution of scores for an indicator around the world.

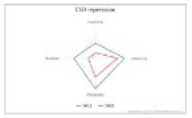


Variable Graph
Compares multiple countries for one indicator/or index. Select one indicator and multiple countries/regions. The data are aggregated by year.

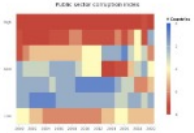


Country Graph
Compares multiple variables/indices for one country/region. Select one country/region and multiple indicators/indices. The data are aggregated by year.

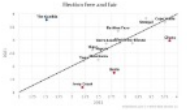
Country Radar Chart
This tool displays multiple variables and indices for one country/region in a radar chart. Select one country/region and multiple indicators/indices.



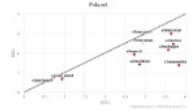
Variable Radar Chart
This tool displays multiple countries (three or more) for one indicator/index in a radar chart. Select one indicator/index and multiple countries.



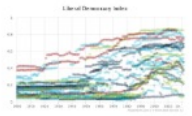
Heat Map
This tool displays one indicator/index on a heat map - a graphical representation of data where values are represented by colors.



Regional Comparison
This tool displays the development of a V-Dem indicator/index in a region between two selected years.



Thematic Comparison
Displays the development of a thematic category between two selected years.



Scatter Chart
Displays one indicator/index as a scatter plot.

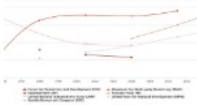
Contingency Tables
By utilizing novel sequencing methods, this tool shows sequencing relationships between indicators in a selected category with a comparison between country vs. region.

Low Hanging Fruit Tool
The tool is based on novel sequencing methods, data analysis and scenario modeling, and can explain which indicators should be developed in order to reach progress in a selected category.

V-Forecast
The Predicting Adverse Regime Transitions (PART) project uses V-Dem data and other sources to estimate the risk of adverse regime transitions during the next two years.

Demspace
The Democratic Space Barometer estimates the probability that a country will experience at least one **opening event** (shift towards more democratic governance) or at least one **closing event** (shift towards more autocratic governance) within a two-year window.

PanDem
The Pandemic Backsliding Project tracks state responses to Covid-19 and their potential effect on the overall quality of democracy within the country. The current version of the data reflects the situation between March 2020 and June 2021.



V-Party Explorer
Varieties of Party Identity and Organization (V-Party) examines the policy positions and organizational structures of political parties across the world.

Measurement Model

Measurement model team

Current team members

- Daniel Pemstein, Project Manager
- Kyle L. Marquardt, Research Fellow
- Lisa Gastaldi, Data Manager
- Johannes von Römer, Data Manager
- Nina Ilchenko, Data Manager

Former team members

- Eitan Tzelgov, former Research Fellow
- Yi-ting Wang, former Research Fellow
- Joshua Krusell, Data Manager
- Farhad Miri, Data Manager

Typical expert-rating projects

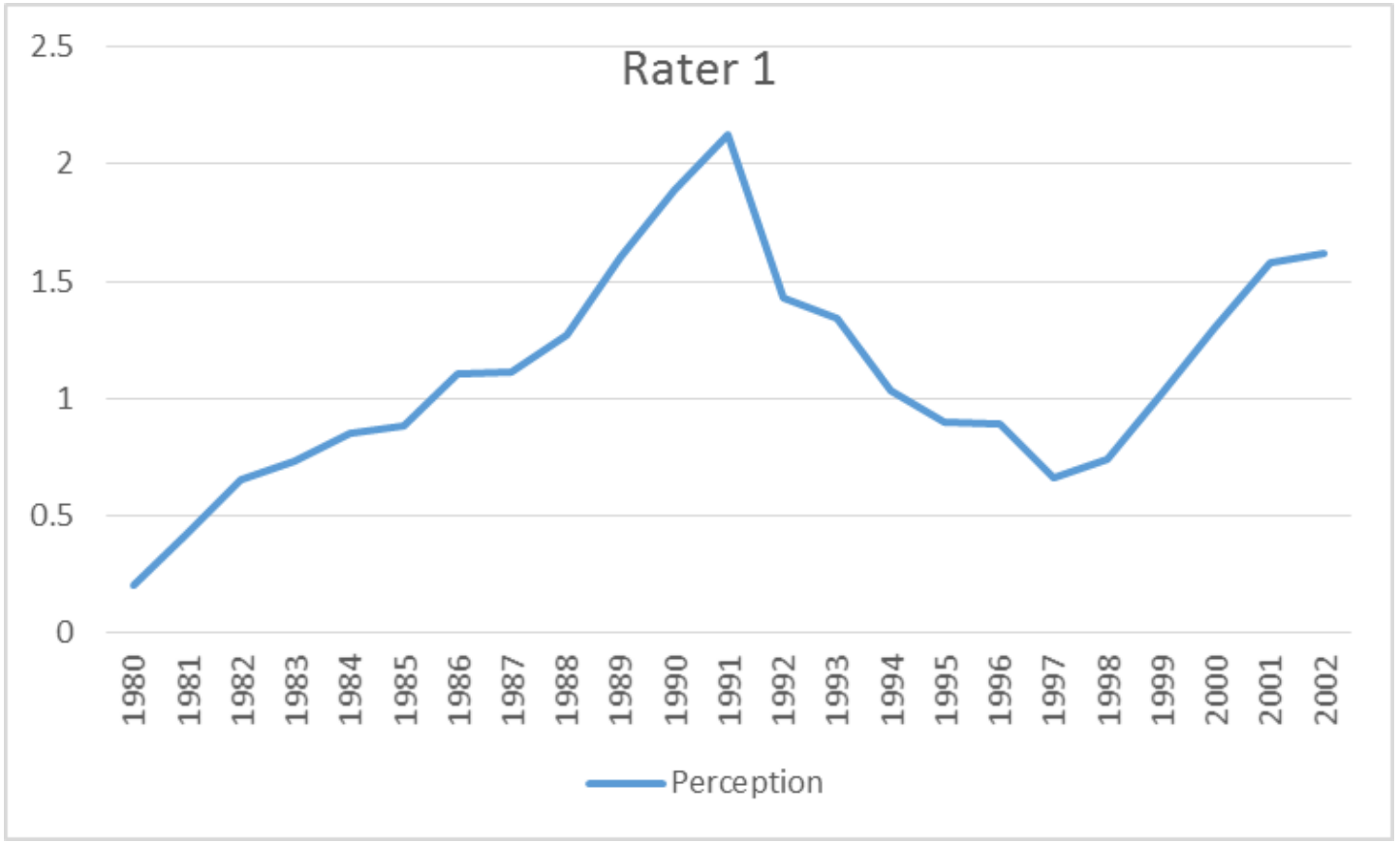
- Assume that experts rate without error
  - All interpret ordinal thresholds the same way: your “2” is the same as my “2”
    - even if they are coding different countries.
  - All experts are either
    - Perfectly skillful (when there is one expert)
    - or equally skillful (when there are multiple experts)
- V-Dem knows these are not safe assumptions.

Measurement challenges

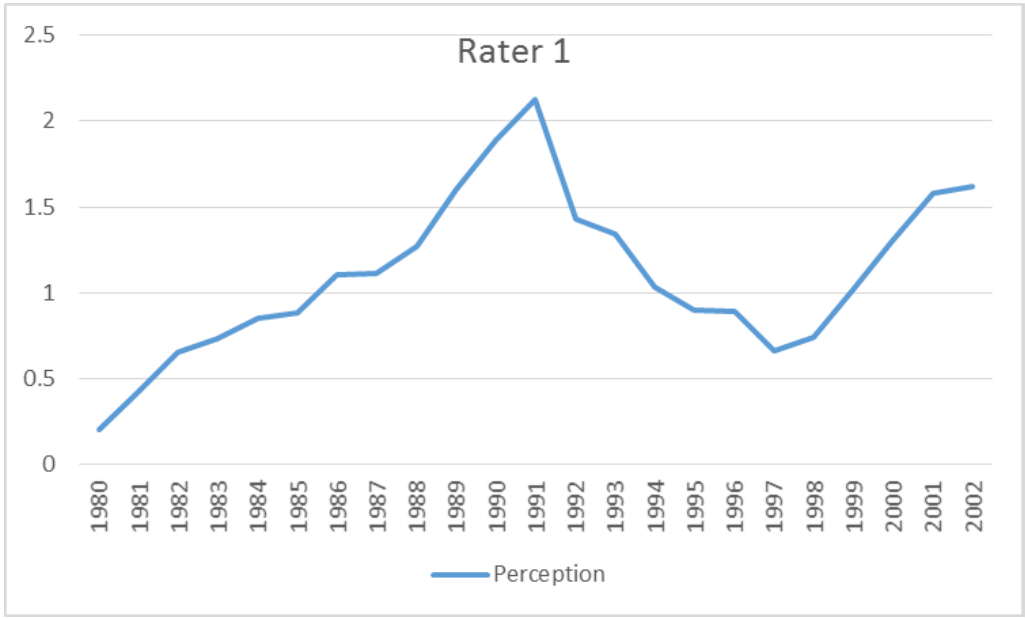
- Some coders are less reliable than others.
  - Differences in amount of knowledge
  - Differences in type of knowledge
  - Differences in diligence: time spent, care, precision
- Differential Item Functioning (DIF)
  - Which information is relevant for answering this question?
  - How should I interpret the thresholds between the ordinal scores?

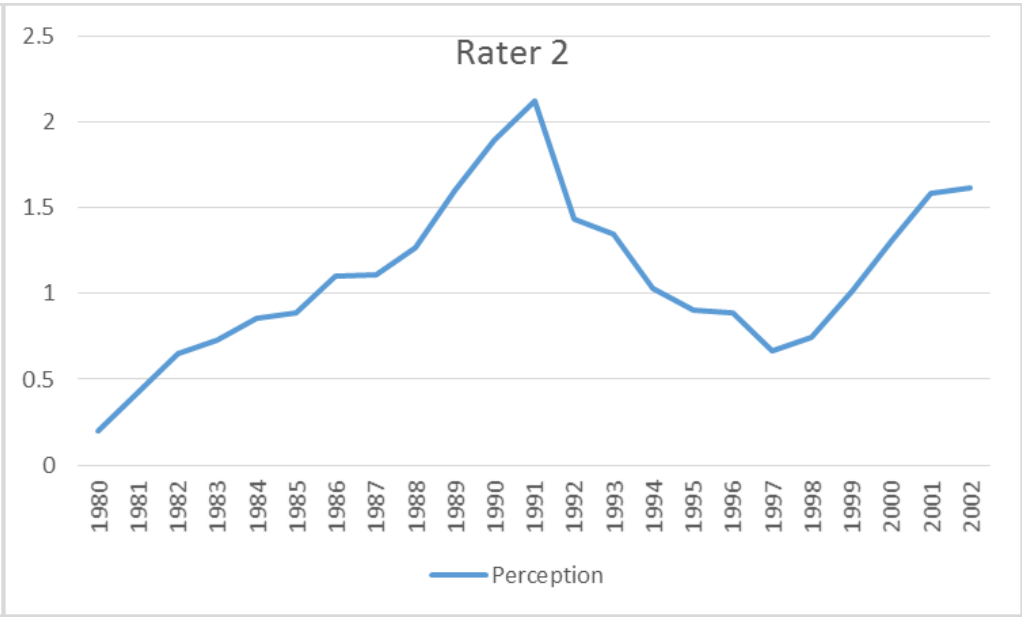
  - Coders of the same country interpret our ordinal scales differently.
  - Coders of different countries may interpret the scales differently.

We assume that coders/raters *perceive* a *continuous* underlying reality.

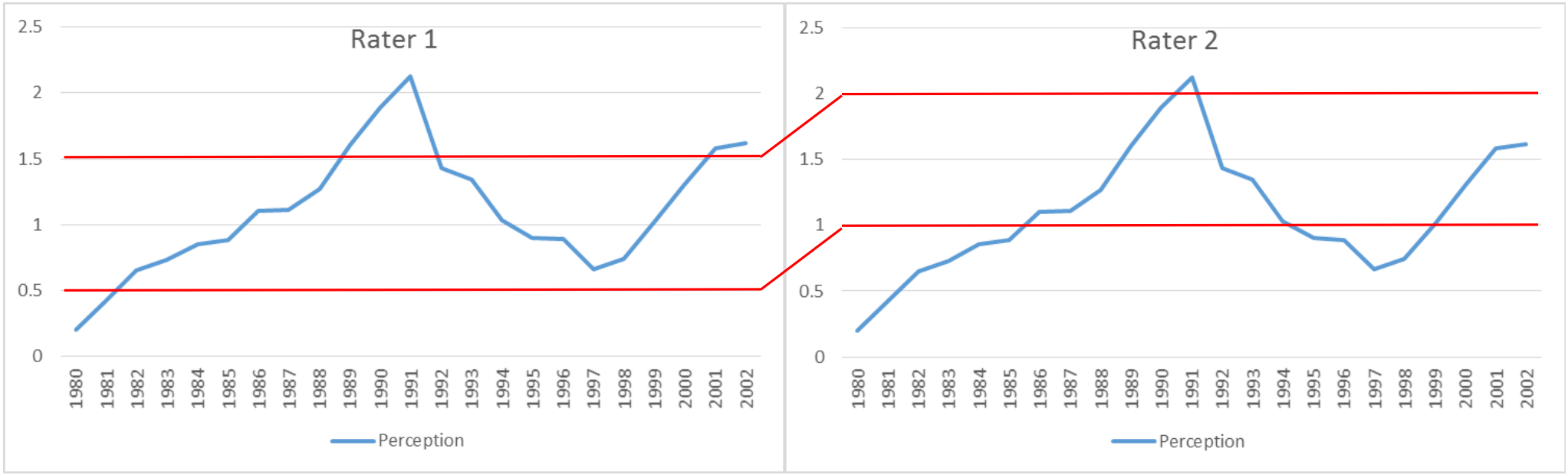


However, raters who perceive the same reality. . .

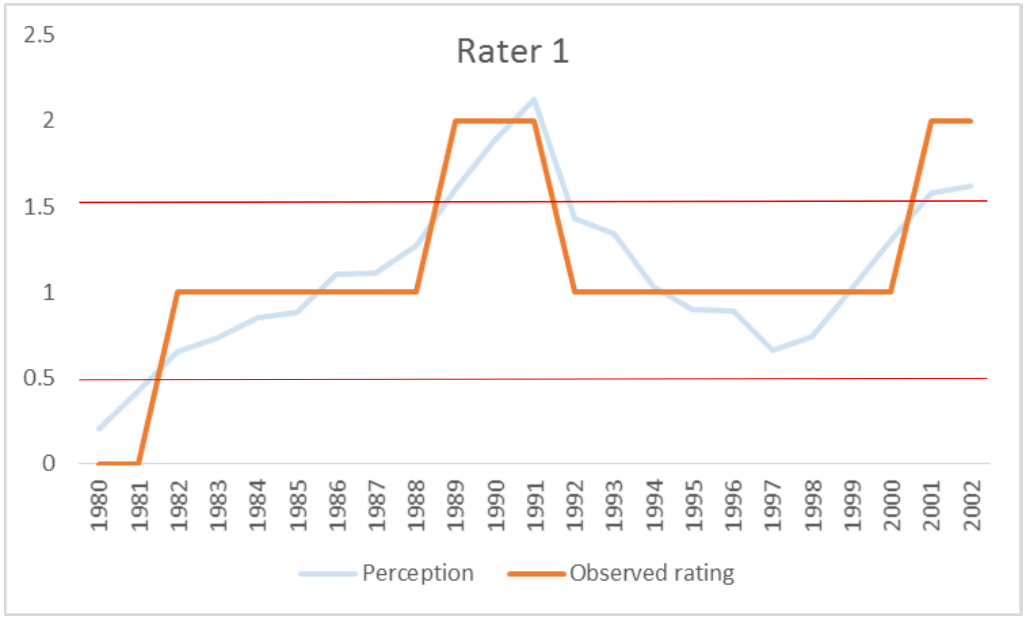


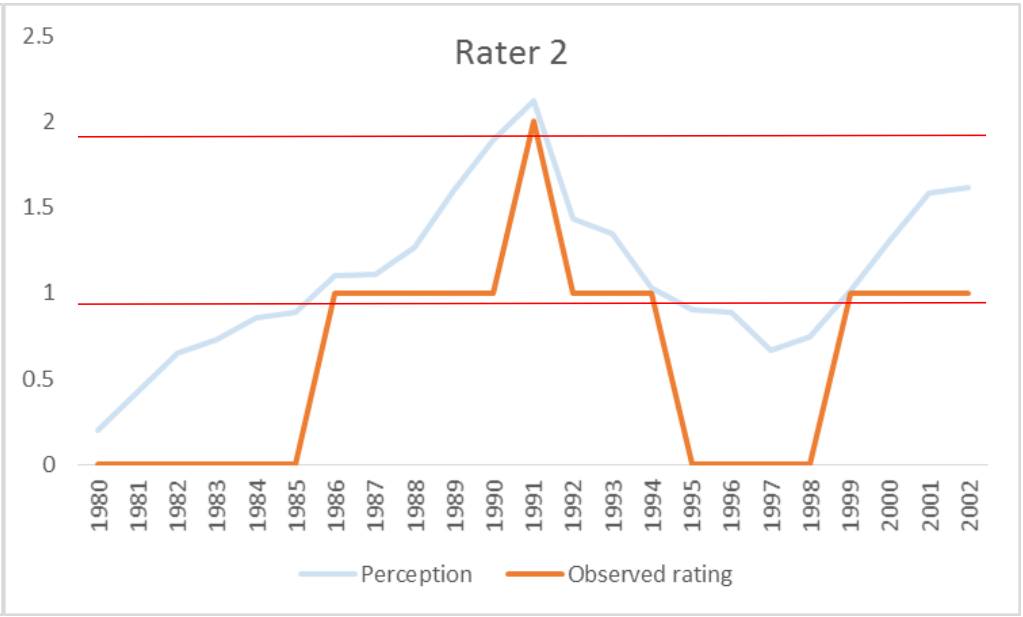


. . .but with different ordinal thresholds. . .

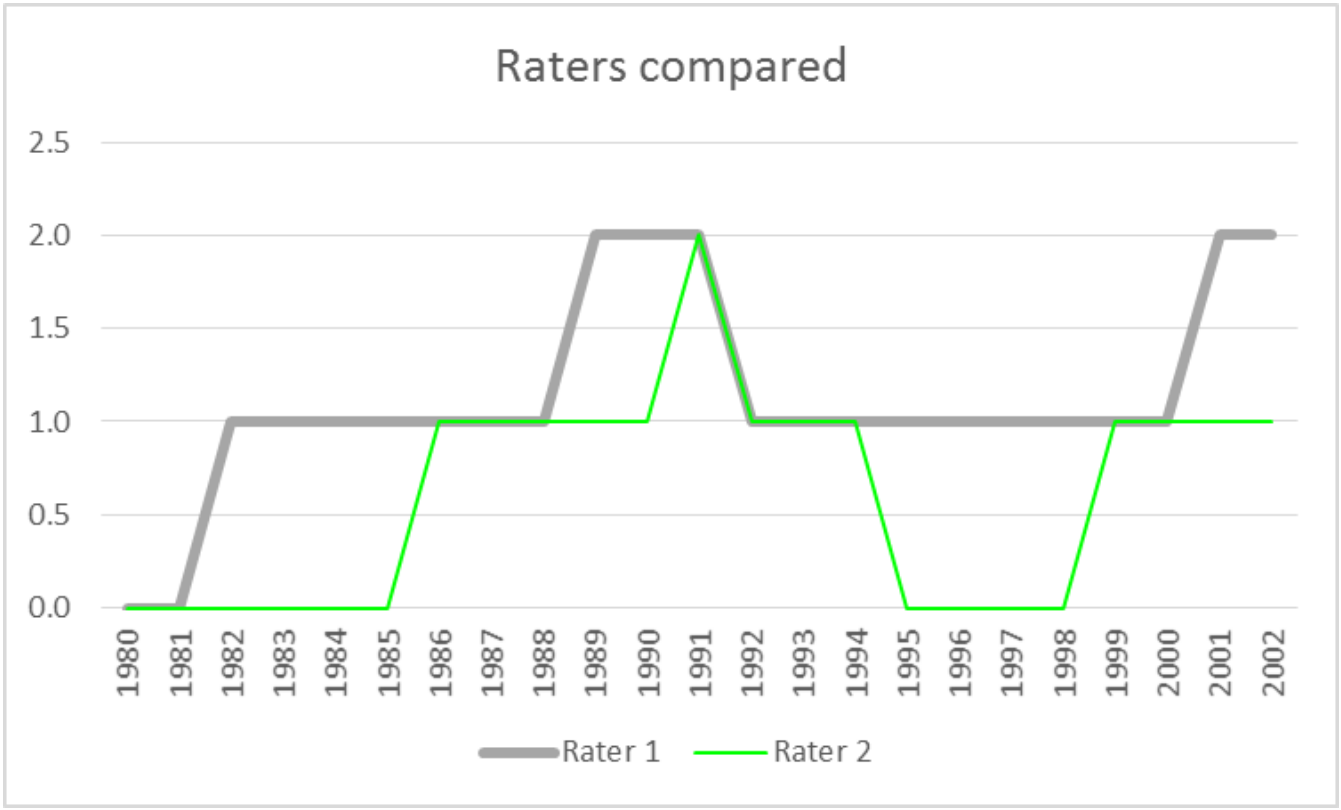


. . .can express their perceptions differently.





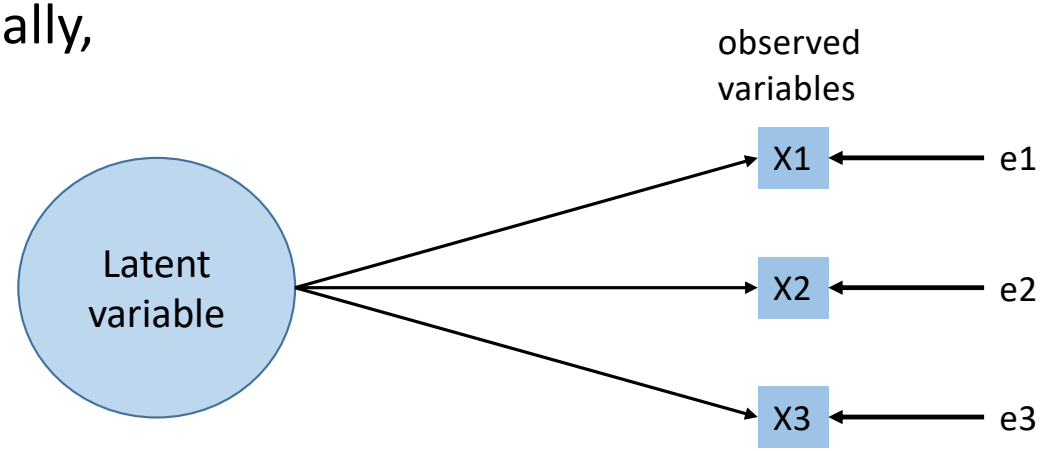
The result:


Raters compared
2.5
2.0
1.5
1.0
0.5
0.0
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
Rater 1
Rater 2

- It’s also possible that raters who ***agree*** on their observed ratings perceive different realities!
- So it’s very important to get good estimates of raters’ thresholds on each indicator.
- Dan Pemstein custom-designed a Bayesian Ordinal IRT measurement model to estimate these and other parameters. It now incorporates innovations by Kyle Marquardt and others as well.

Latent variables

- These are a class of models in which only some variables are observed (or “manifest”); others are unobserved (or “latent”).
- E.g., factor analysis, principal components
- Typically,

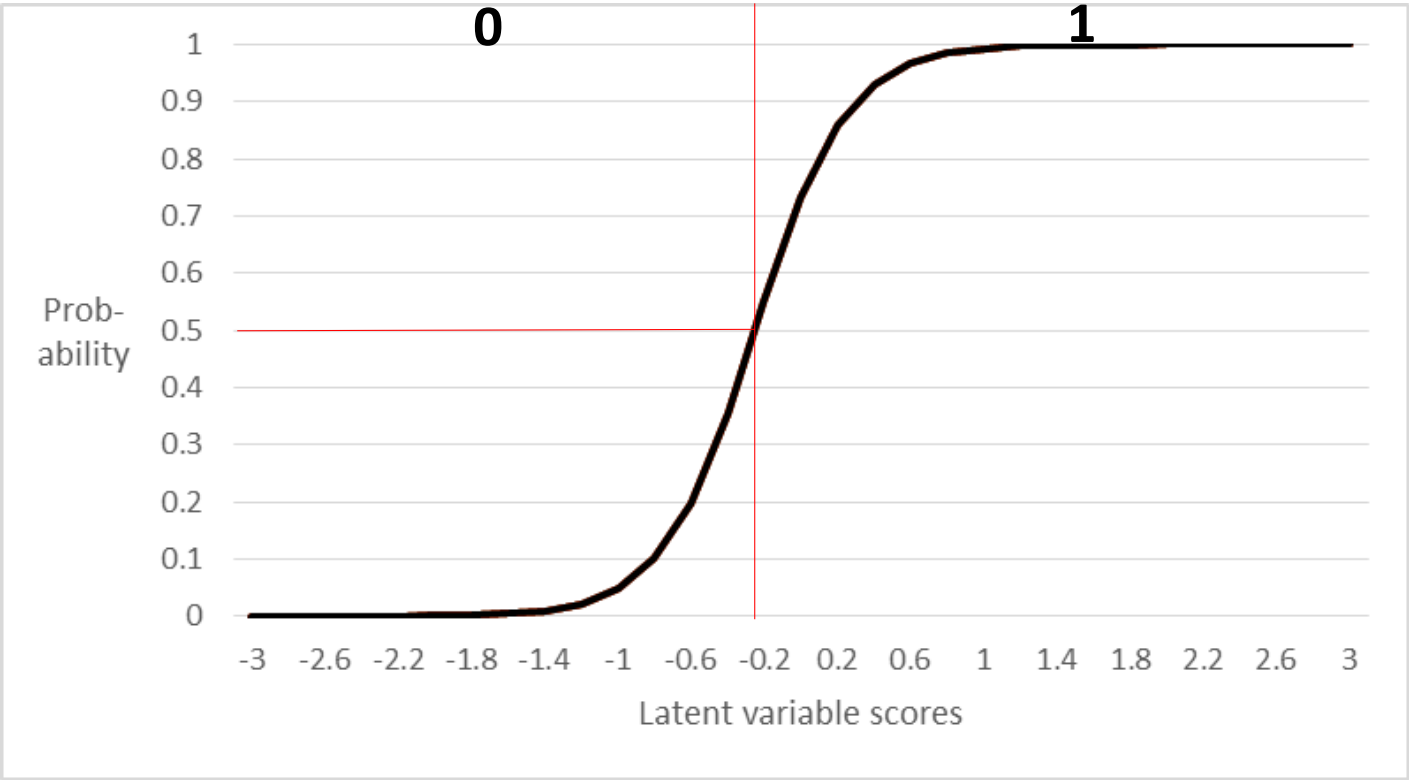


These can be estimated as a set of simultaneous of equations: one for X1, one for X2, one for X3.

The Item-Response Theory (IRT) Framework: A special type of latent-variable model

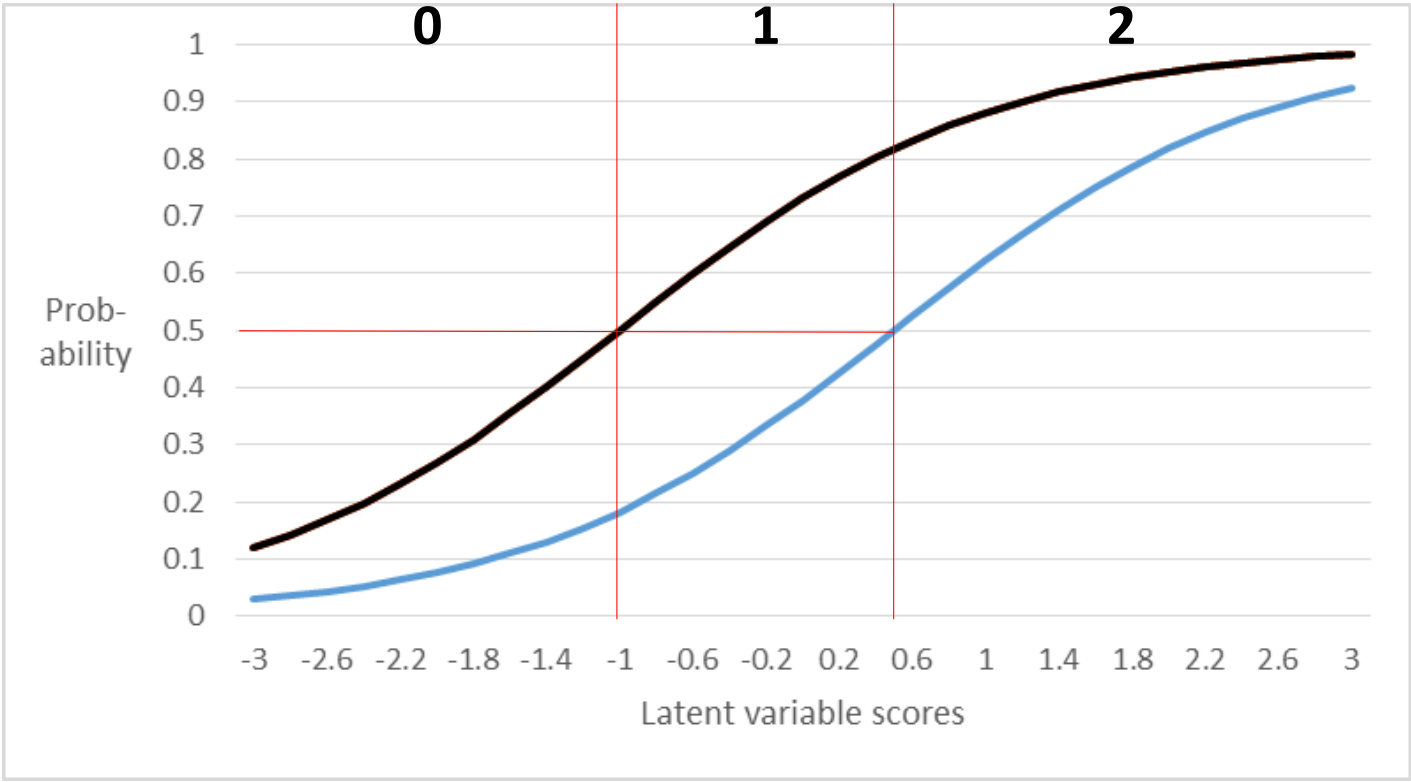
We can understand observed ordinal scores as being above or below a threshold on a latent variable.

The higher the threshold, the more democratic the perceived reality must be to earn a higher ordinal score from the coder.



Ordinal IRT: 2 thresholds dividing 3 levels

Here there are ranges of the latent variable that correspond to each ordinal score:
0: less than -1
1: -1 to 0.6
2: greater than 0.6



Two parameters in ordinal IRT models

- ***Difficulty*** is estimated by the thresholds on the latent variable that separate ordinal scores. Each level of an indicator (minus one) has its own threshold.
- ***Discrimination*** is a coefficient estimating how crisply the coder distinguishes between ordinal scores. It determines the slope of the S-curve.
  - Coders whose scores are correlated best with other coders' scores are treated as being more "discriminating," and therefore get more weight.
  - Less-discriminating coders get less weight.

[See IRT simulator]

Bayesian estimation

- In a country-year-coder*indicator dataset, most of the cells would be empty because experts code only a few surveys in one or a few countries.
- Bayesian estimation avoids making the heroic assumptions that would be necessary using frequentist methods with such a sparse dataset.
- It also – through the magic of resampling – gives us confidence bounds around our parameters, including the latent variable.

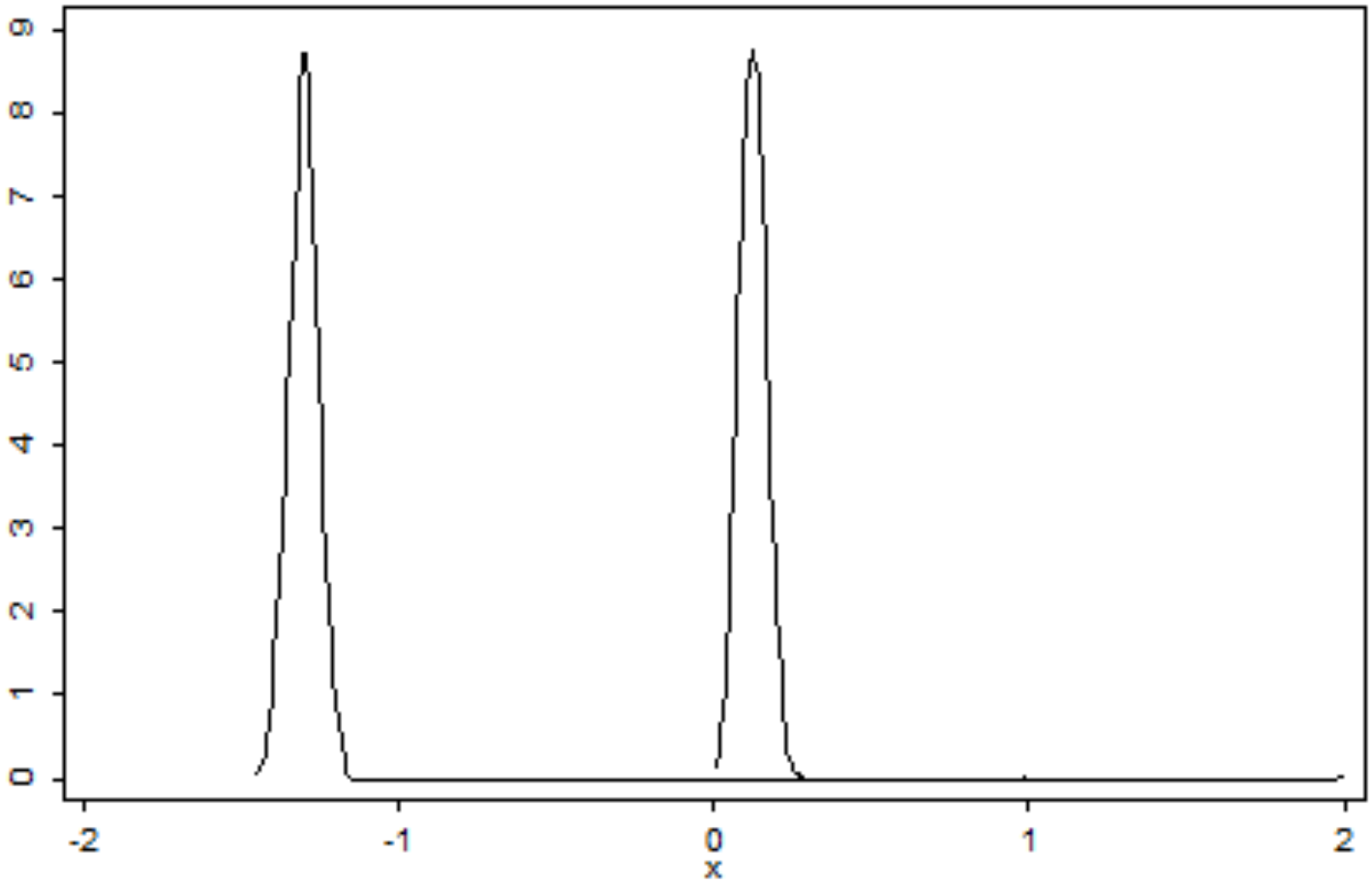
The model estimates difficulty thresholds, assuming

- Global mean thresholds are between -2 and 2 (uniformly distributed)
- The mean country thresholds are allowed to vary around the global thresholds, with a standard deviation of 0.2
- Coder thresholds are allowed to vary around their country's thresholds, with a standard deviation of 0.2

What this looks like:

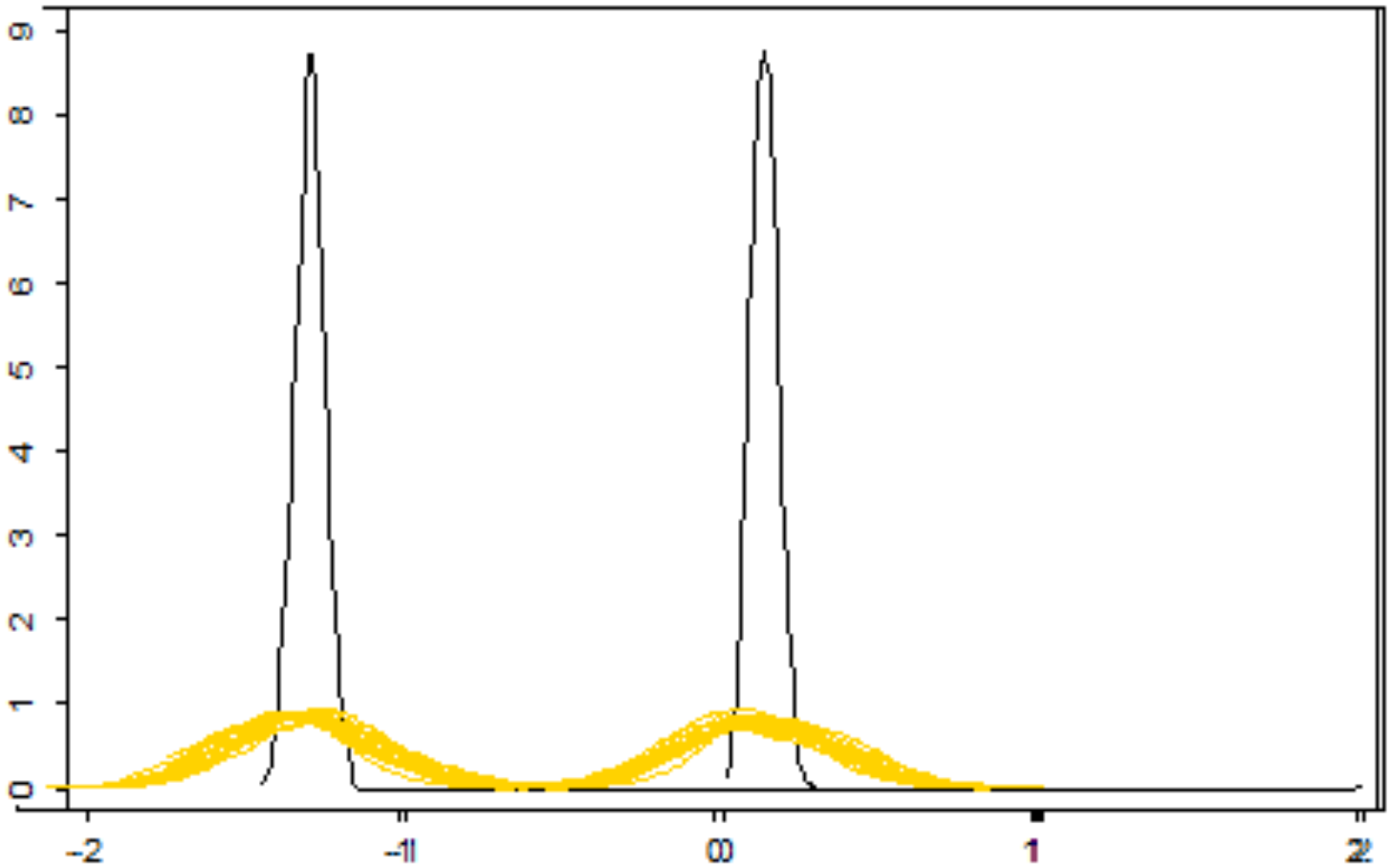
An example for v2svinlaut: International autonomy

Black: posteriors of global mean thresholds

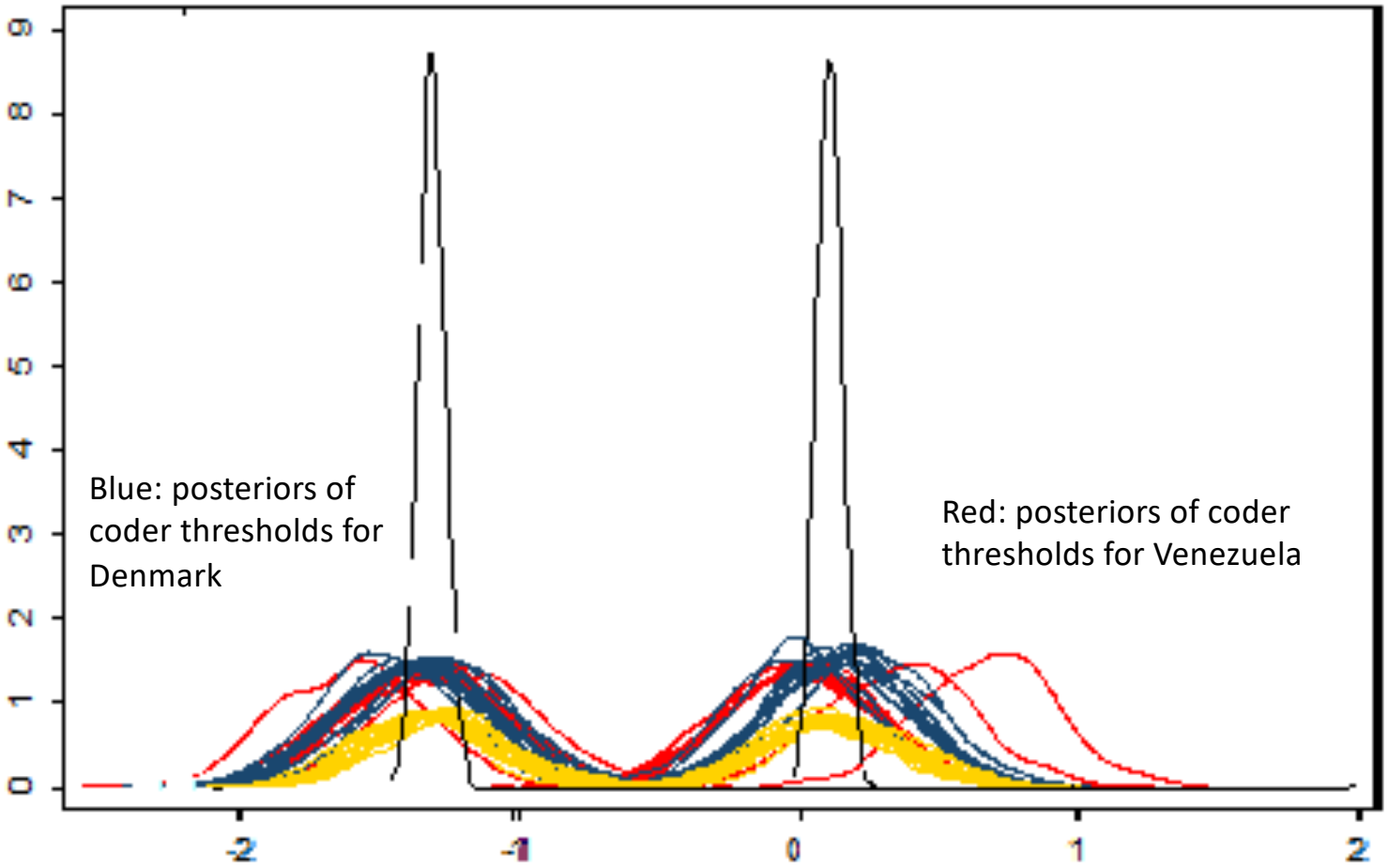


An example for v2svinlaut: International autonomy

Gold: 20 posteriors for all country thresholds



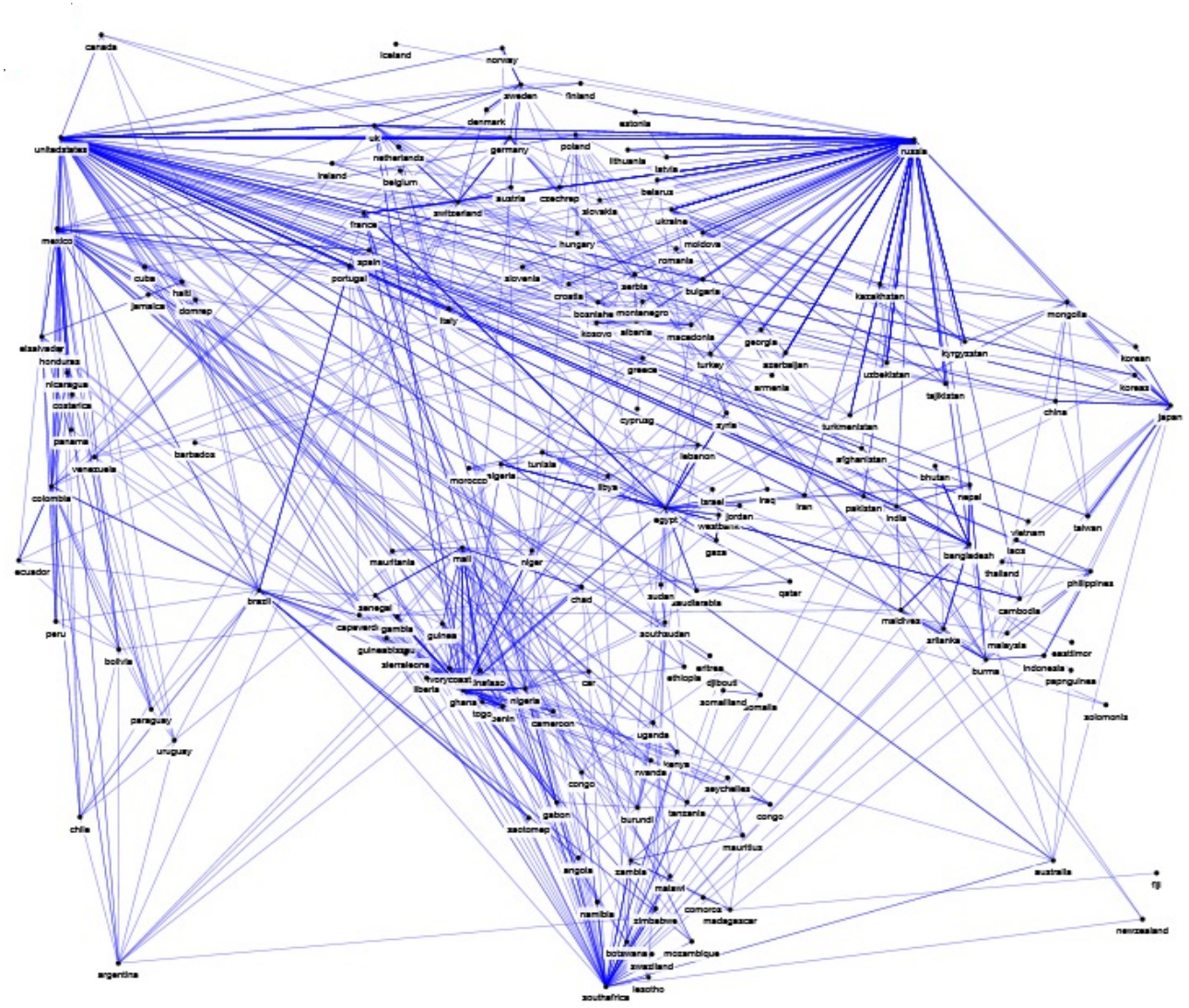
An example for v2svinlaut: International autonomy



Why these assumptions?

- It's a departure from the usual MCMC practice of weak priors, but much better than the typical expert-coding assumption that DIF is not an issue.
- It allows the lateral and bridge coding to help calibrate the thresholds.
- It helps especially with the countries that are not yet bridged, or not sufficiently bridged.
  - Ideally all countries would be connected, directly or indirectly, by a network of experts who have coded more than one country. This would enable us to compare a "3" in Gambia to a "3" in any other country.

This is the bridging as of March 2015 for the Elections survey.
The bridging we need is nearly complete.
Only 7 countries were not completely bridged.
Now they all are.
However, more would help.



Another issue

- Without a further assumption, latent variable estimates would be biased toward zero in some cases
  - Unbridged countries
  - Countries with invariant scores, which tell us nothing about what their coders' thresholds would be for other scores
- The result is that the Switzerlands of the world would be biased downward and the Saudi Arabias would be biased upward. The model just wouldn't "know" that a high score is really high and a low score is really low, so it would hedge its bets.

Fixed by an assumption about the latent variable:

- When the model does *not* have enough information about coders' thresholds for a country, the country gets an average of the coders' scores.*
- When the model *does* have enough information (which is most of the time!), this average is adjusted for the threshold estimates, as described above.

*Actually, the confidence-weighted average of the scores for all coders of that country-year, normalized with respect to scores for all country-years.

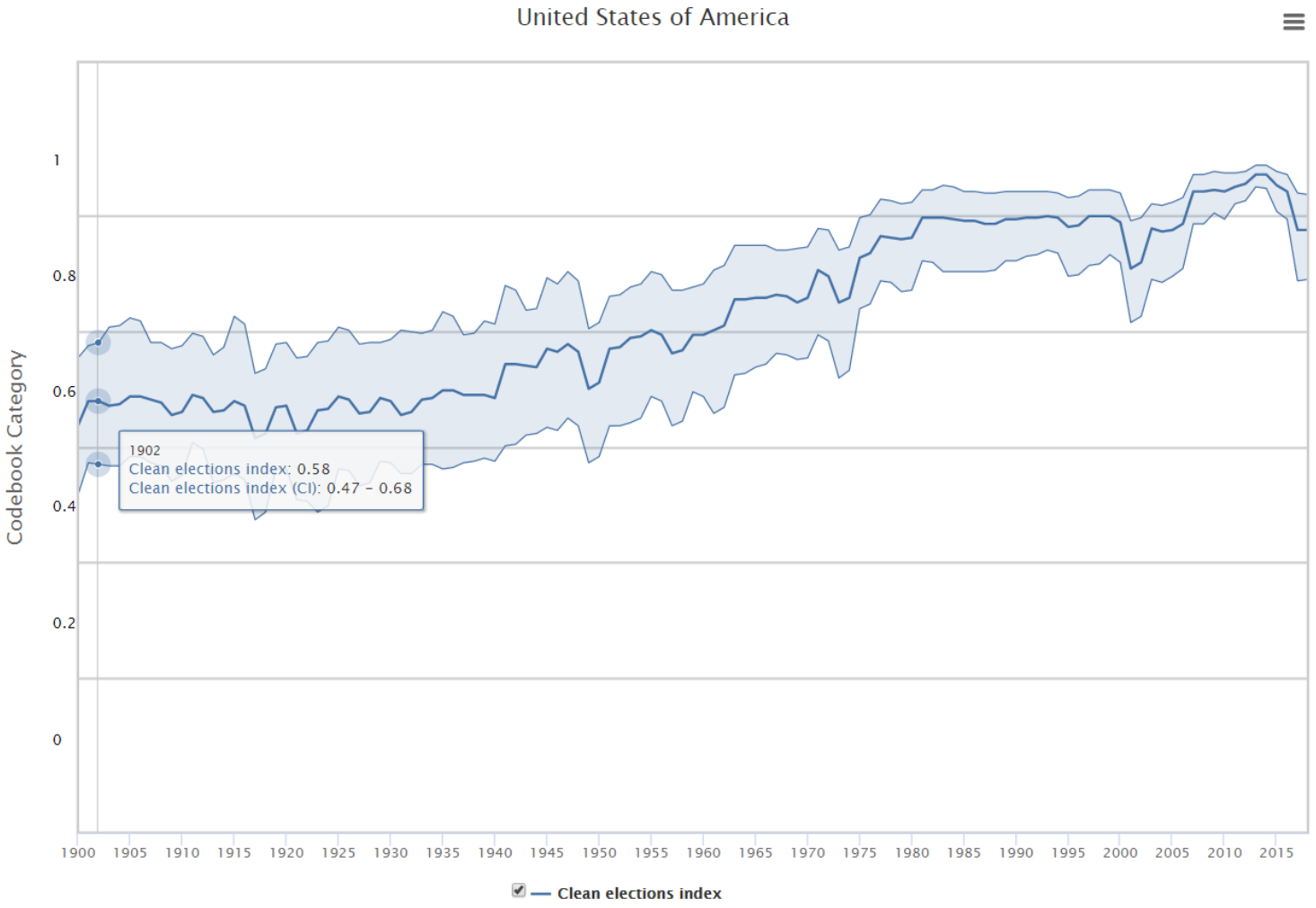
Temporal granularity

- Scores are not serially independent. However, the model does not assume that they are, so estimates are allowed to jump or fall suddenly when the data call for it.
- This falsely inflates the sample size, which would make us overconfident of the point estimates.
- Therefore, for the MM our observations are not country-years or country-days, but “regimes”: country-periods in which no coder changed his/her score or confidence for that country.
- This yields more conservative estimates.

Estimation

- Markov-chain Monte Carlo methods using Stan
- Iterative procedures identify the parameter estimates that best fit the observed data.
- We at first used high-performance computing hosted by the Center for Research Computing at Notre Dame, but now use machines in Sweden.
- Each variable is modeled separately. Originally it took 2 hours to several days for each variable; weeks to estimate all 156! However, due to recent improvements in the code, now takes just a few days to run everything.
- More detailed information is in Working Paper No. 21.

HPDs tend to be narrower in recent years.



Reliability

Based on Brigitte Seim (lead author), “Data Validation,” in Coppedge & 23 coauthors, *Varieties of Democracy: Measuring Two Centuries of Political Change* (Cambridge UP, forthcoming 2020).

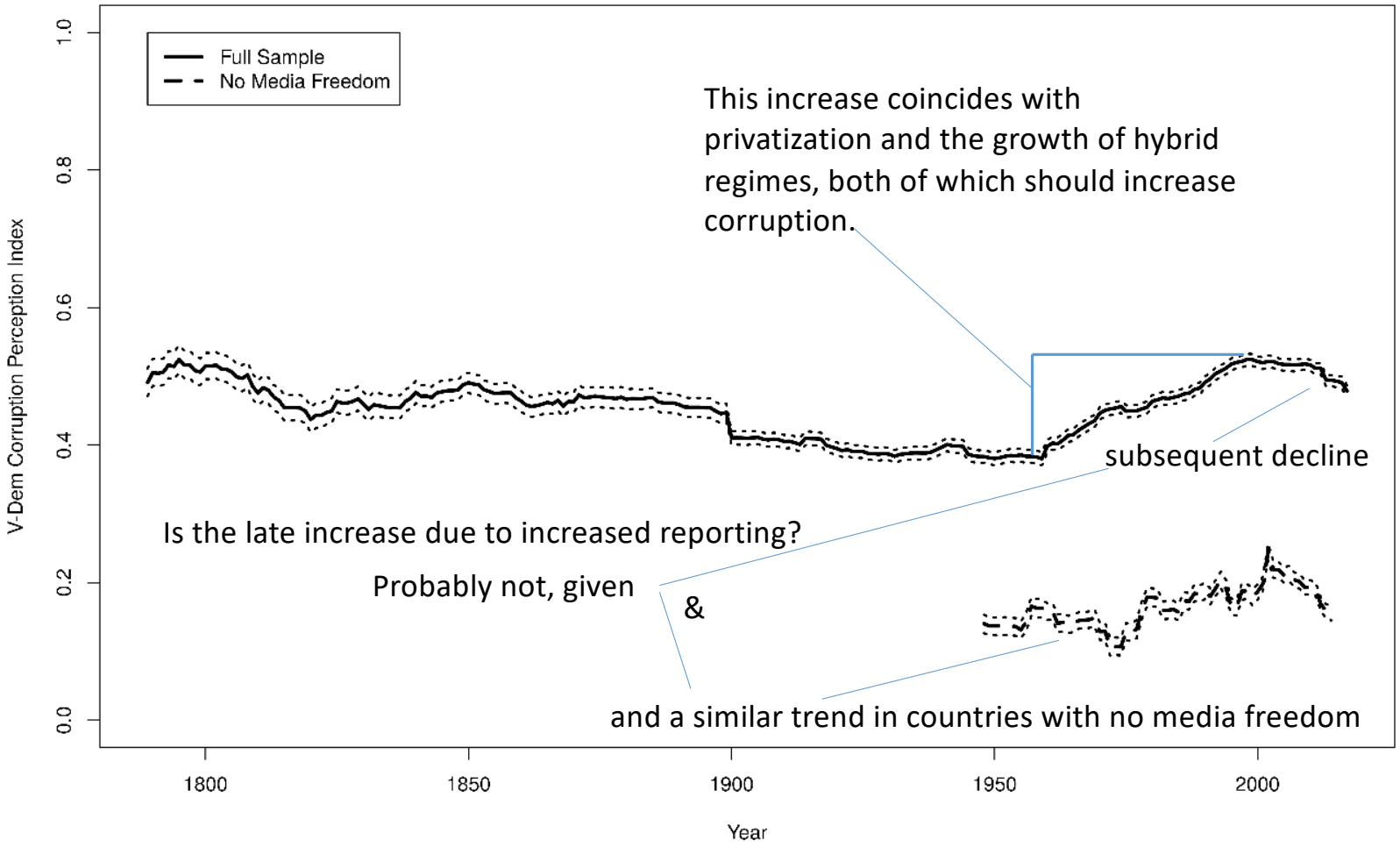
Validation strategies

- (I'm skipping over content validation and data generation assessment and am not reporting an interesting qualitative replication.)
- Convergent validity
  - Correlation with other indicators of similar concepts
    - Predictors of deviations from other concepts
  - Degree of convergence across coders
    - Predicted by coder traits?
    - Predicted by country traits?

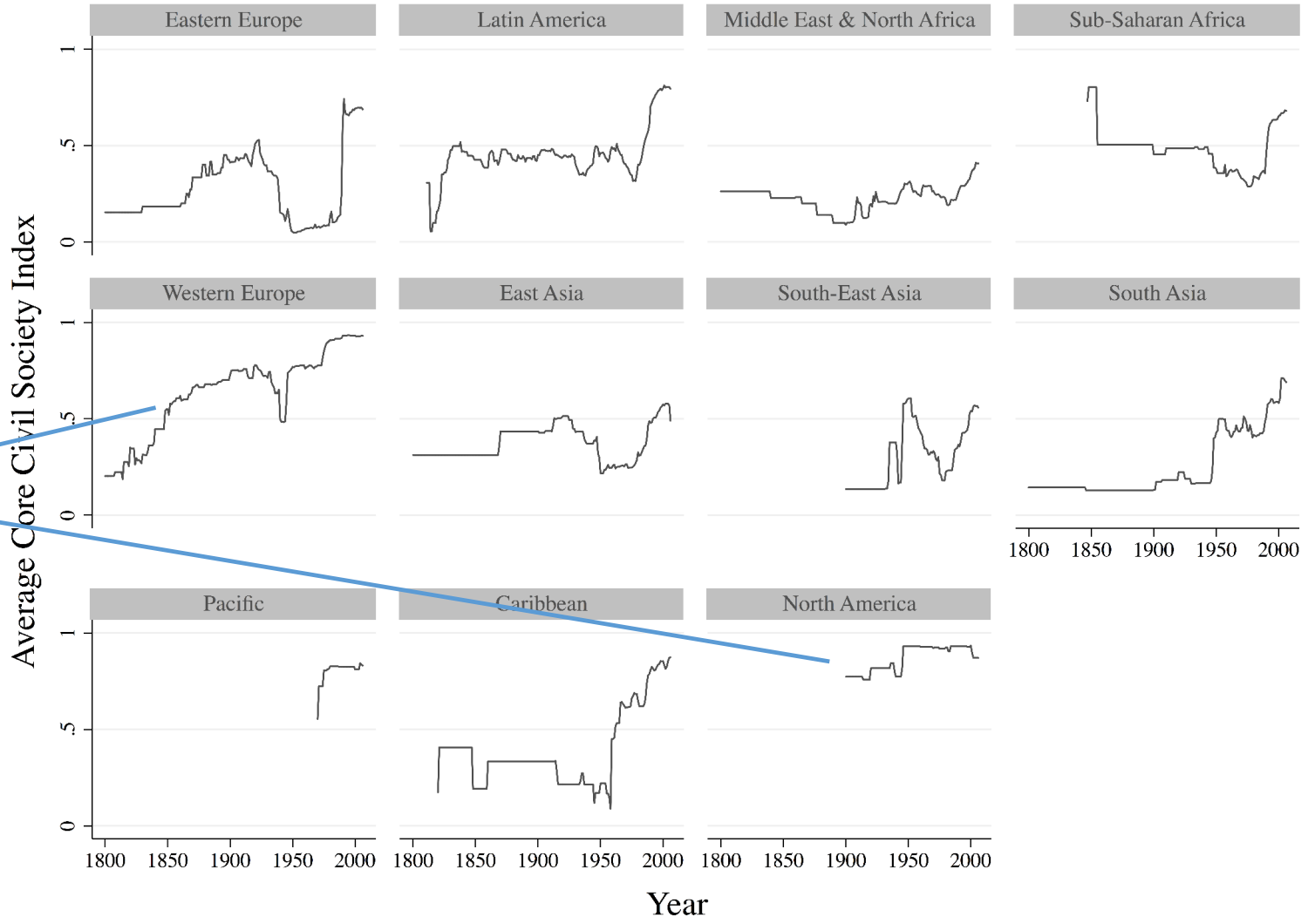
Applied to
*Polyarchy
*Corruption
*Core Civil Society Index

Corruption Trends

The long-term trend is very different from the dramatic increase we see in democracy.

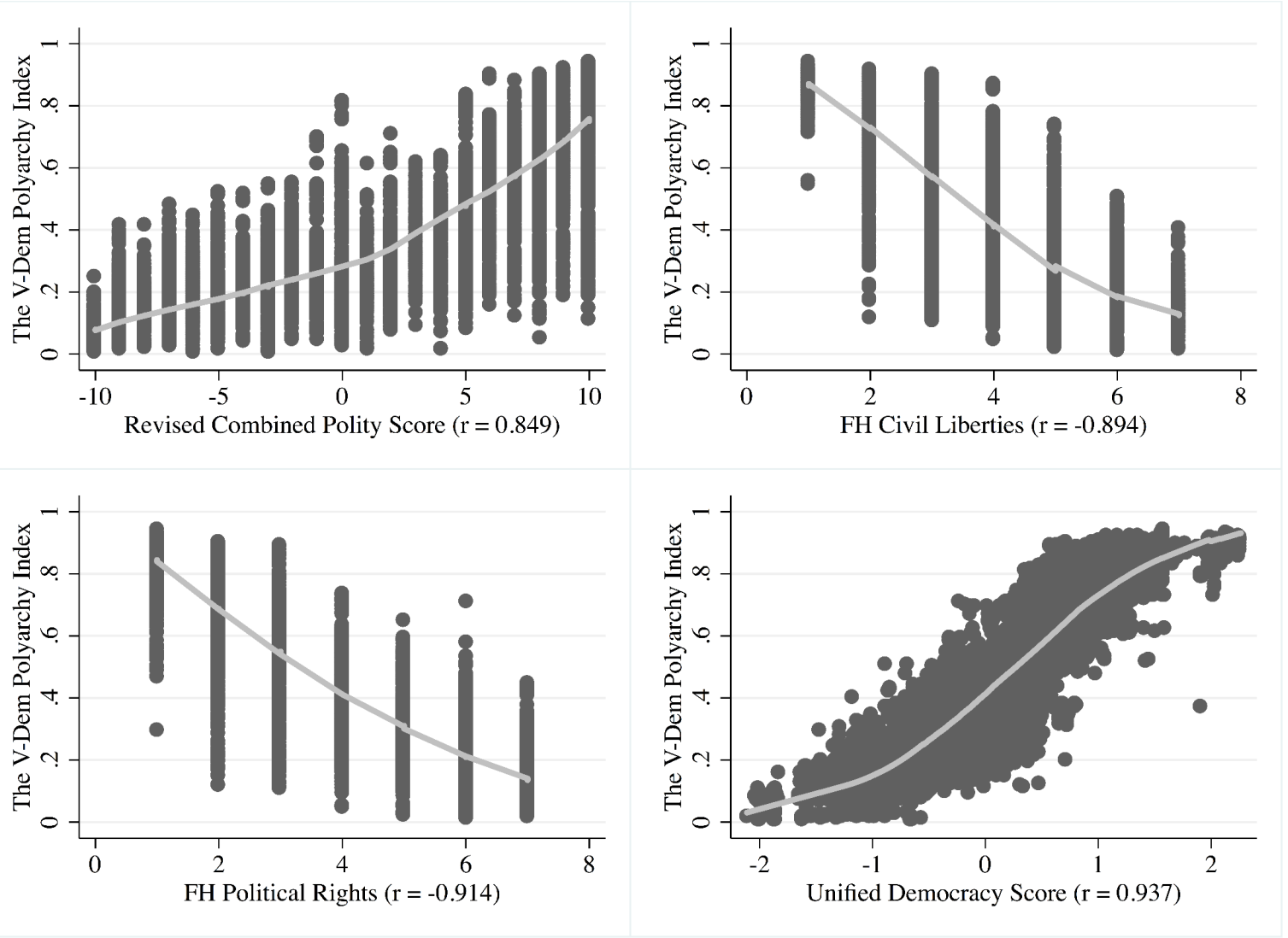


we see strong civil societies to be prevalent in places like Western Europe and North America compared to other regions of the world



On average, V-Dem agrees with other familiar measures of democracy.

However, there are some real differences.



Do coder characteristics explain deviations from other democracy measures? Not really.

	Deviation from Polity	Deviation from Freedom House	Deviation from UDS
Share female coders	**−0.084****	−0.047	−0.059
Average age of coders	0.009	−0.004	0.012
Average age^2	−0.000	−0.000	−0.000
Share of PhD coders	−0.060	−0.022	−0.047
Share of coders employed by government	−0.060	−0.066	−0.060
Share of coders born in country	0.068	0.054	0.061
Share of coders residing in country	−0.039	**−0.110*****	−0.053
Share of Western coders	0.048	0.009	0.031
Average free market support	0.004	**−0.035****	−0.009
Average conventional understanding	−0.019	−0.007	−0.028
Average alternative understanding	−0.026	−0.003	−0.024
Coder disagreement (=hard to code countries)	**0.223*****	**0.214*****	**0.250*****
No. of coders	**−0.002****	**−0.003*****	**−0.006****
No. of lateral coders	−0.003	0.003	−0.008

Bear in mind: N=529,367 to 930,161!

Do coder characteristics explain deviations of V-Dem PCI from the WGI corruption measure?
Not really.

	Deviation from WGI
Share of female coders	**0.052****
Average age of coders	−0.002
Average age of coders^2	0
Share of PhD coders	**−0.084****
Share of coders employed by government	−0.068
Share of coders born in country	−0.009
Share of coders residing in country	0.01
Average free market support	0.006
Average electoral democracy support	0.001
Average liberal democracy support	−0.005
Mean coder discrimination (beta)	0.004
Coder disagreement	**0.345****
No. of coders	**−0.008****

N = 54,235

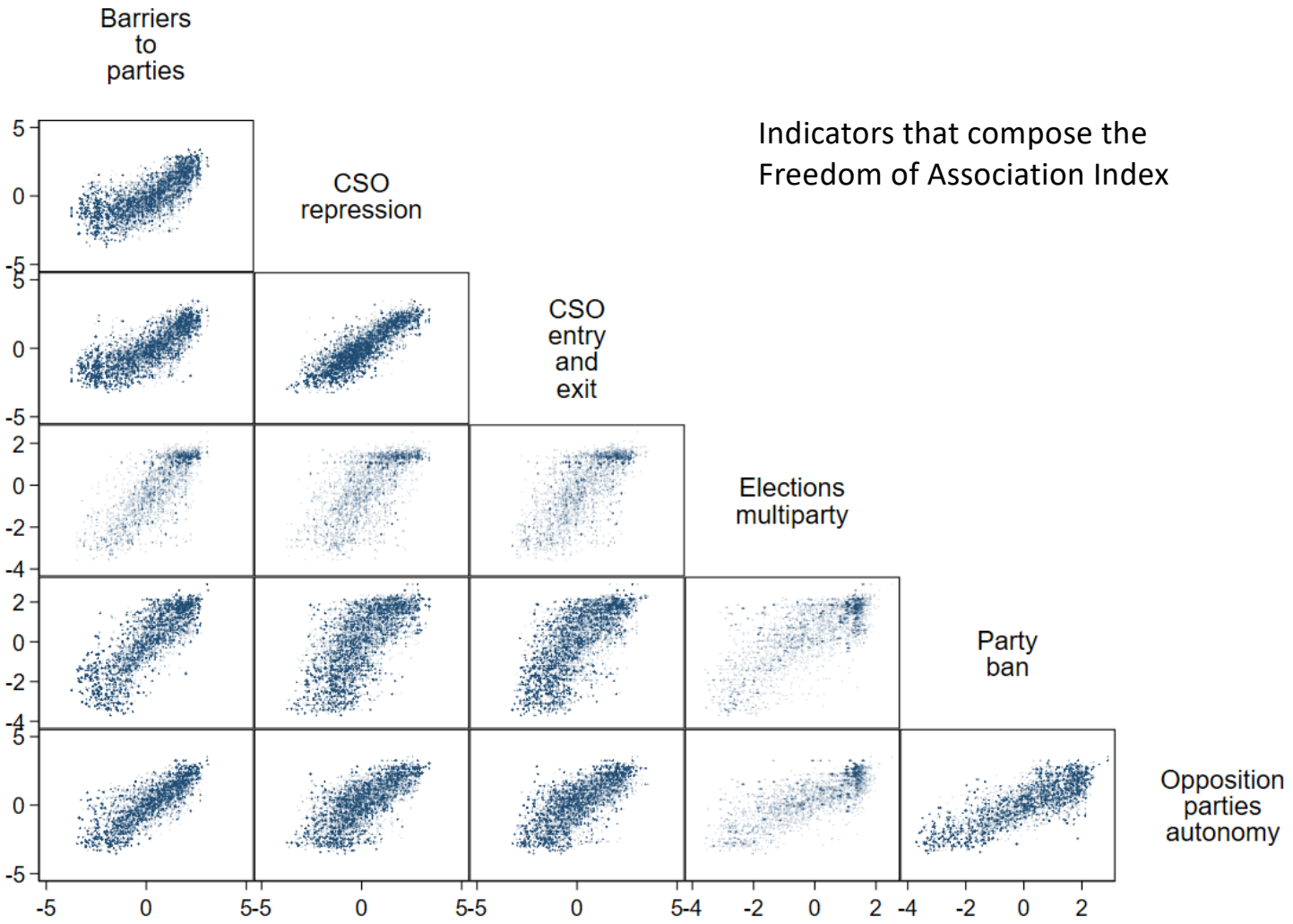
Conclusion about Validation

- V-Dem data appear to have few, if any, systematic biases compared with other measures of similar concepts.
- Any systematic biases are small.
- V-Dem measures are no worse than anyone else's.
- There are methodological reasons to expect that V-Dem's measures are more valid and reliable.

Index Construction

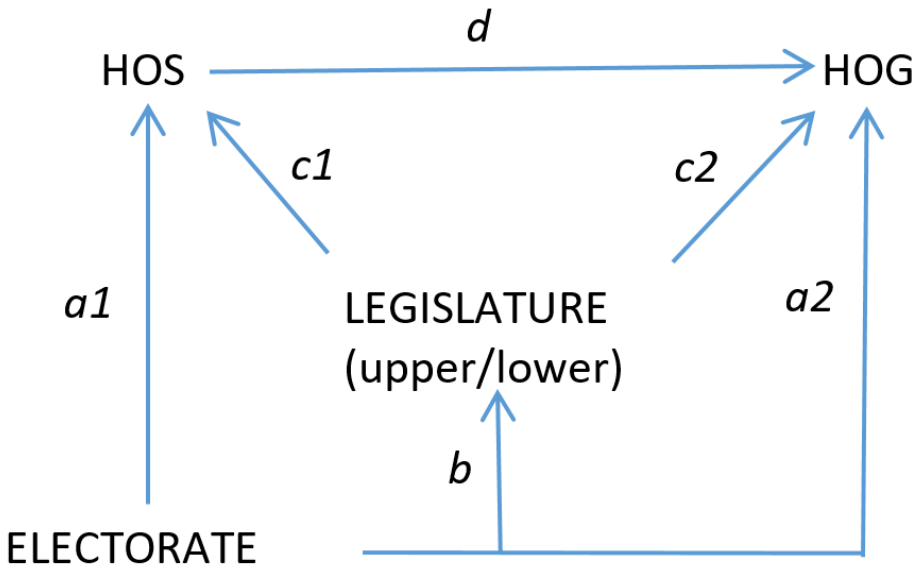
Some indices are based on factor analysis, when indicators are approximately unidimensional

We treat them as "substitutable" and "reflective" indicators.



A few indices are based on multidimensional indicators, which require more complex aggregation rules

For example, the
Elected Officials Index



Electoral Democracy Index

- Two versions are built from the same 5 indices or indicators
  - One is additive

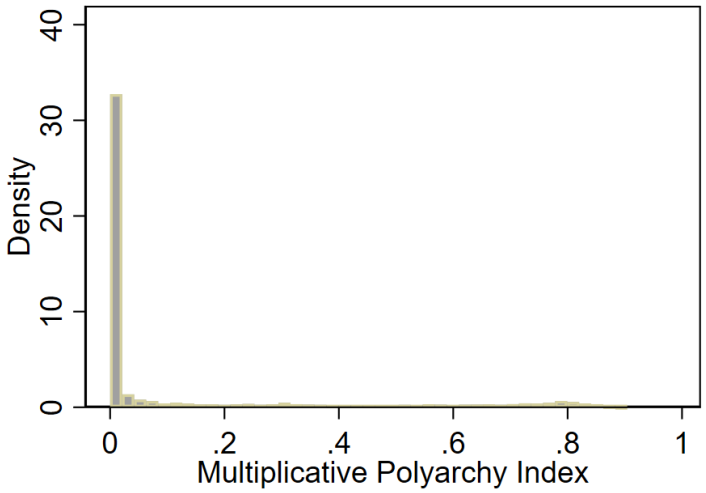
  API = [(Elected Officials + Suffrage) + 2 * (Clean Elections + Freedom of Association + Freedom of Expression and Alternative Information)]/8

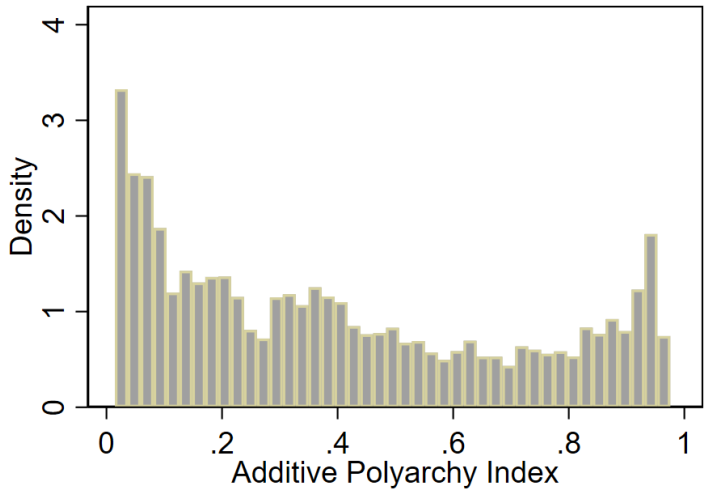
  - One is multiplicative

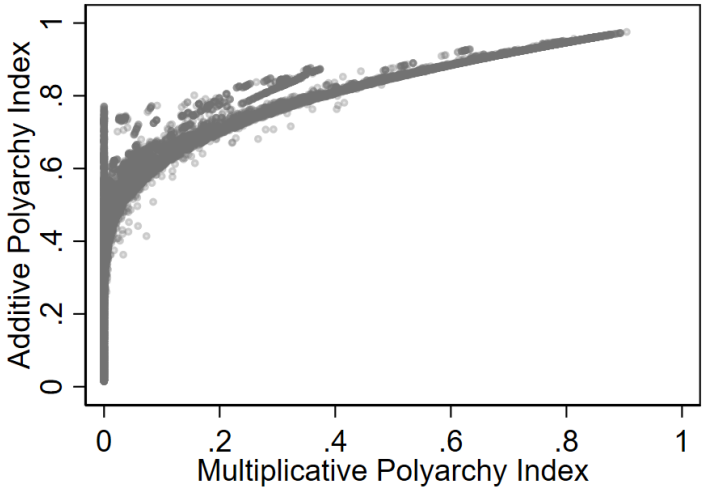
  MPI = Elected Officials * Clean Elections * Freedom of Association * Suffrage * Freedom of Expression and Alternative Sources of Information.

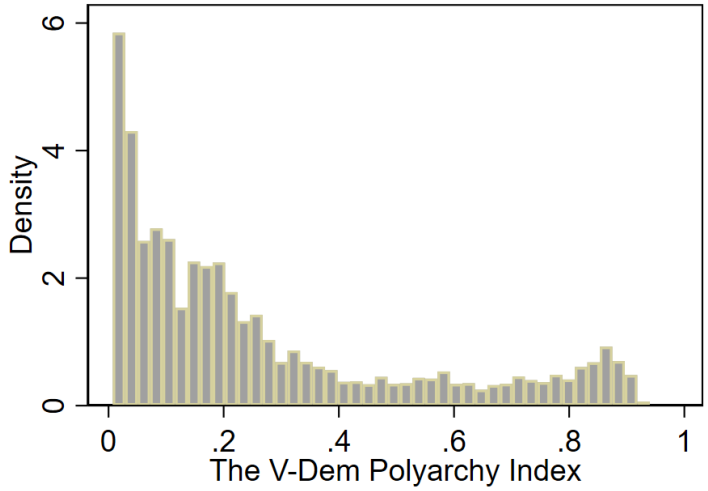
  - The EDI ("Polyarchy") is the average of these two:

  v2x_polyarchy = (API + MPI)/2


Density
0
10
20
30
40
0
.2
.4
.6
.8
1
Multiplicative Polyarchy Index

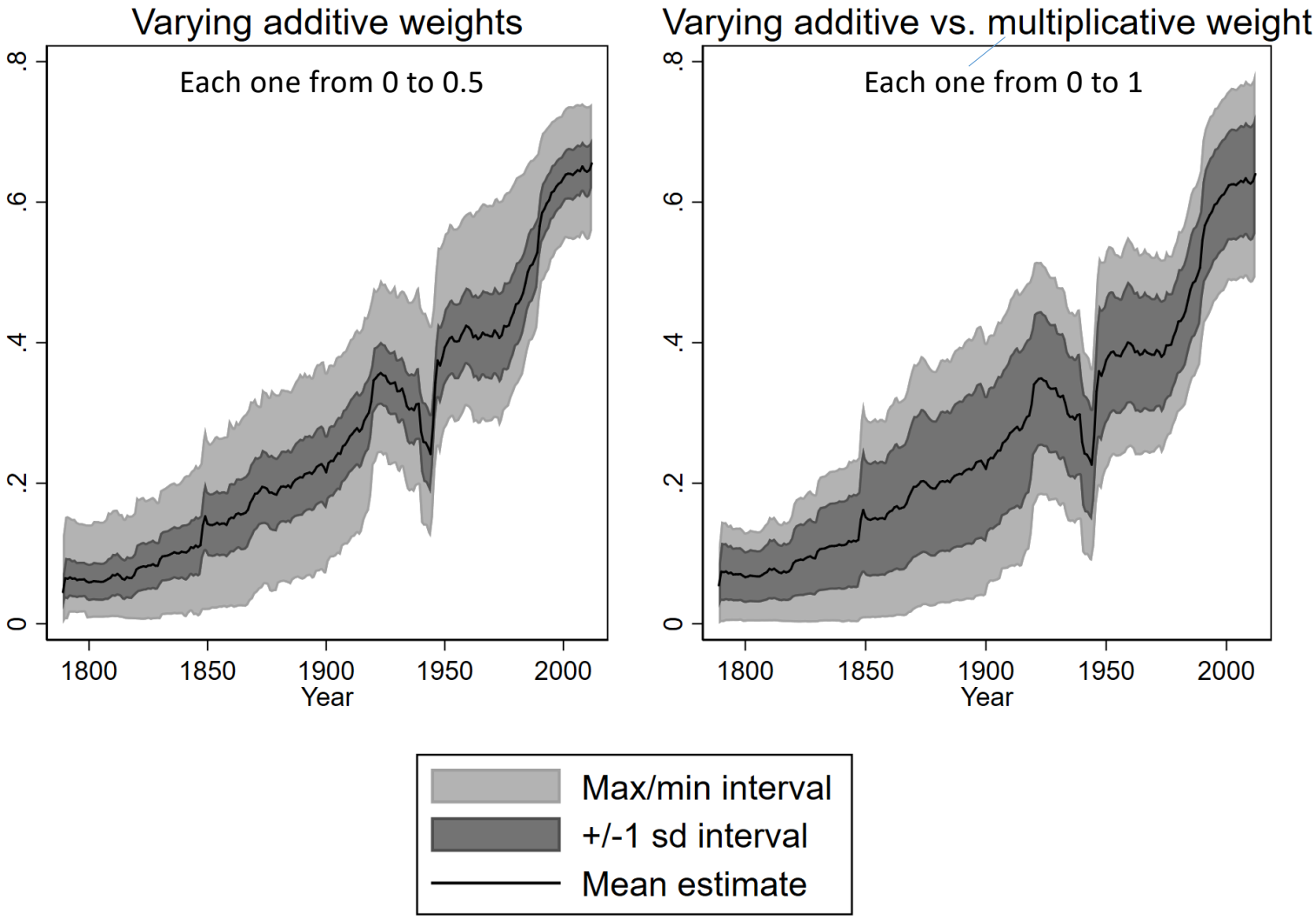

Density
0
1
2
3
4
0
.2
.4
.6
.8
1
Additive Polyarchy Index


Additive Polyarchy Index
0
.2
.4
.6
.8
1
0
.2
.4
.6
.8
1
Multiplicative Polyarchy Index


Density
0
2
4
6
0
.2
.4
.6
.8
1
The V-Dem Polyarchy Index

How robust is this formula to different additive and multiplicative weights?

These estimates are based on thousands of simulations.



HLIs, e.g., the Liberal Democracy Index

= .25*Polyarchy$^{1.58}$

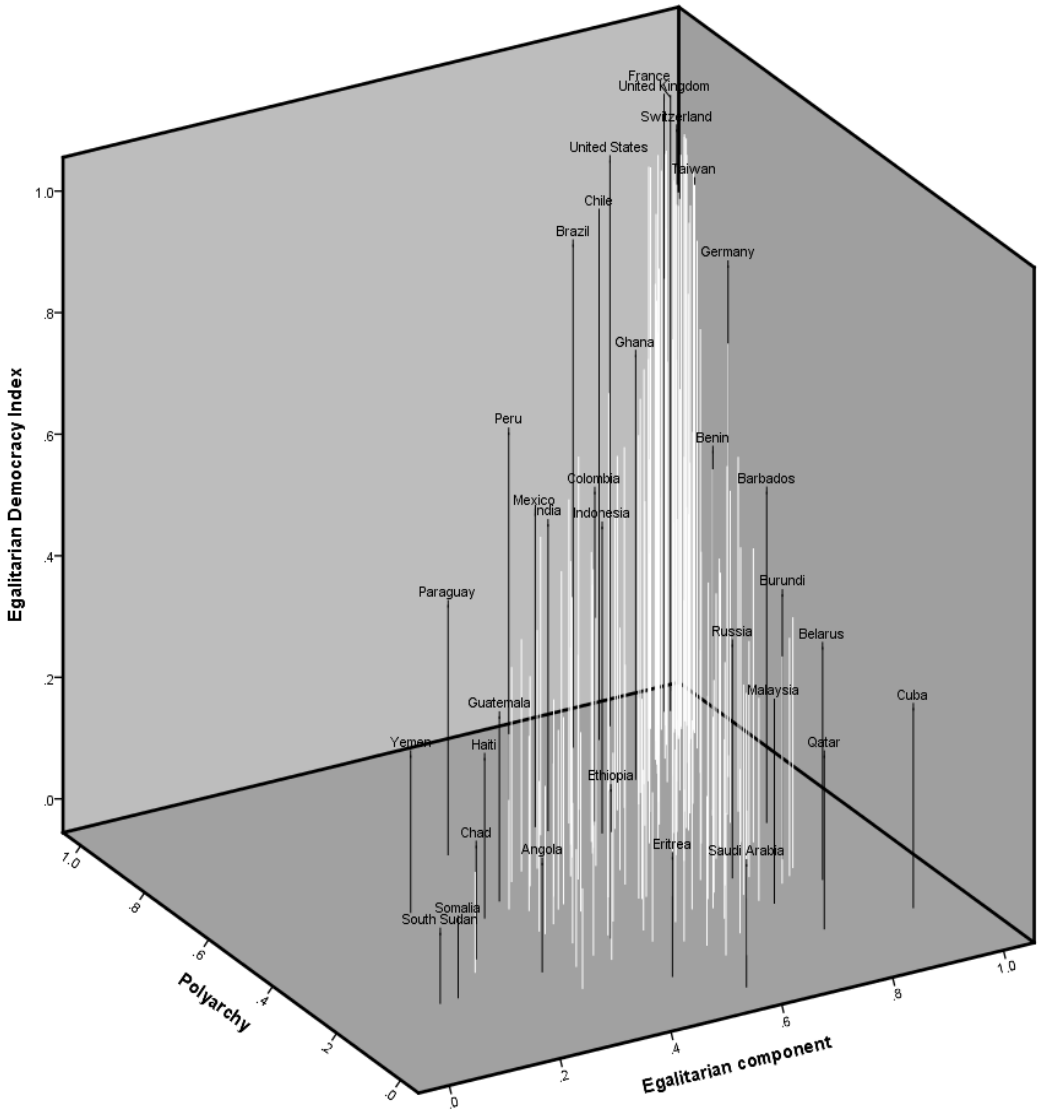
+ .25*Liberal component

+ .5*Polyarchy$^{1.58}$*Liberal component

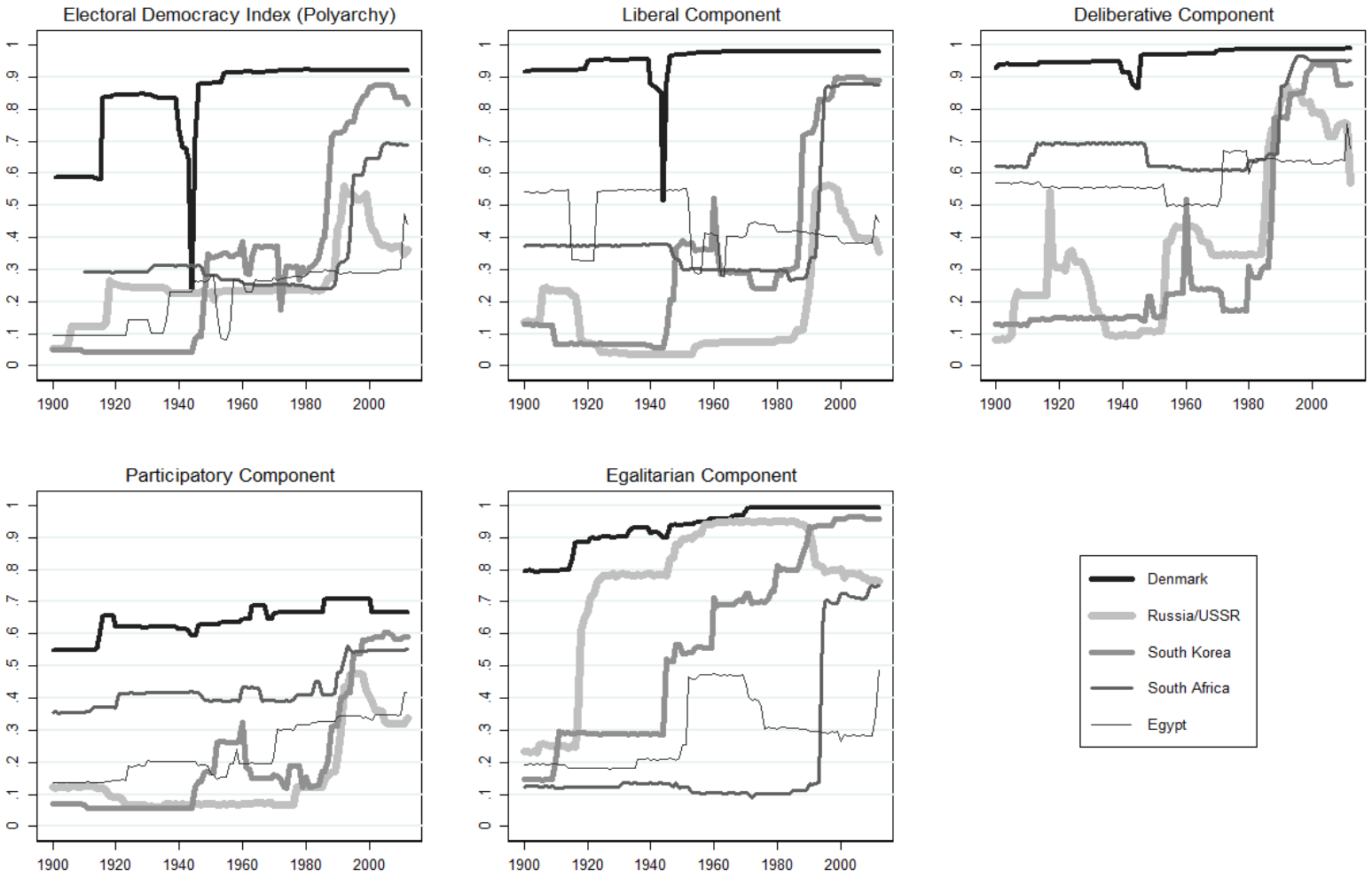
The exponent of 1.58 makes the index=0.5 when Polyarchy=0.5 and Liberal component=1.

We use the same formula for the three other democracy indices, too.

What this formula does



Components by Benchmark Countries, 1900-2012



Getting the most out of V-Dem

(beyond the usual quantitative advice)

Avoiding pitfalls, leveraging strengths, being involved in the project

Opportunities and challenges of these data

1. The large number of indicators
2. They include estimates of measurement uncertainty
3. The large number of years in the data set
4. The potential for spatial dependence

The main challenges

- The large number of variables make it easier to capitalize on chance
- It is possible that scores contain some bias associated with the characteristics of the experts who assigned them
- The long time series increases the need to model or rule out causal heterogeneity across historical periods.

The main advantages

- Having many disaggregated variables that make it possible to precisely test hypotheses and specific causal mechanisms
- And to perform robustness checks
- Being able to account for measurement uncertainty in any kind of analysis
- Leveraging large samples to obtain greater statistical power
- Reducing omitted variable bias
- Having less risk of Nickell bias in models with fixed effects and a lagged dependent variable.

Also,

Greater possibilities for designing

- Designing differences-in-differences tests
- Designing synthetic control type analyses
- Using the Generalized Method of Moments
- Testing complex models that can differentiate between true spatial dependence and other processes that generate similar patterns in the data.

The large number of variables

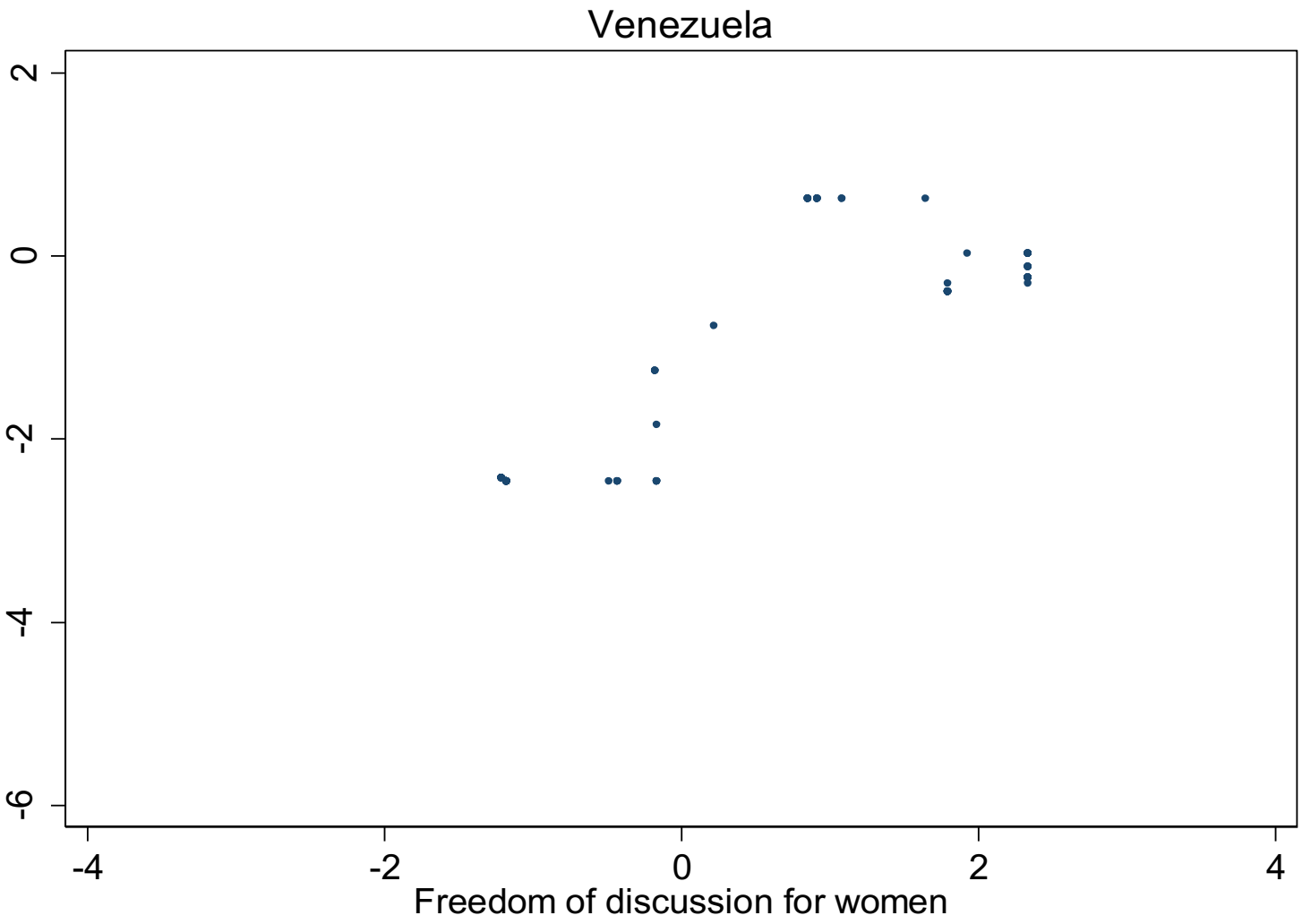
- With more possible independent variables, there are more opportunities to explore and test different causal mechanisms
  - E.g., are parties among the many institutions that increase economic growth? (Bizzarro et al. 2018).
- With more possible dependent variables, there are more opportunities for placebo tests: If X should cause Y but not Z, if X does not cause Z that increases our confidence in X→Y.
- However, it also increases opportunities for fishing expeditions, p-hacking, capitalizing on chance, etc.

Estimates of measurement uncertainty

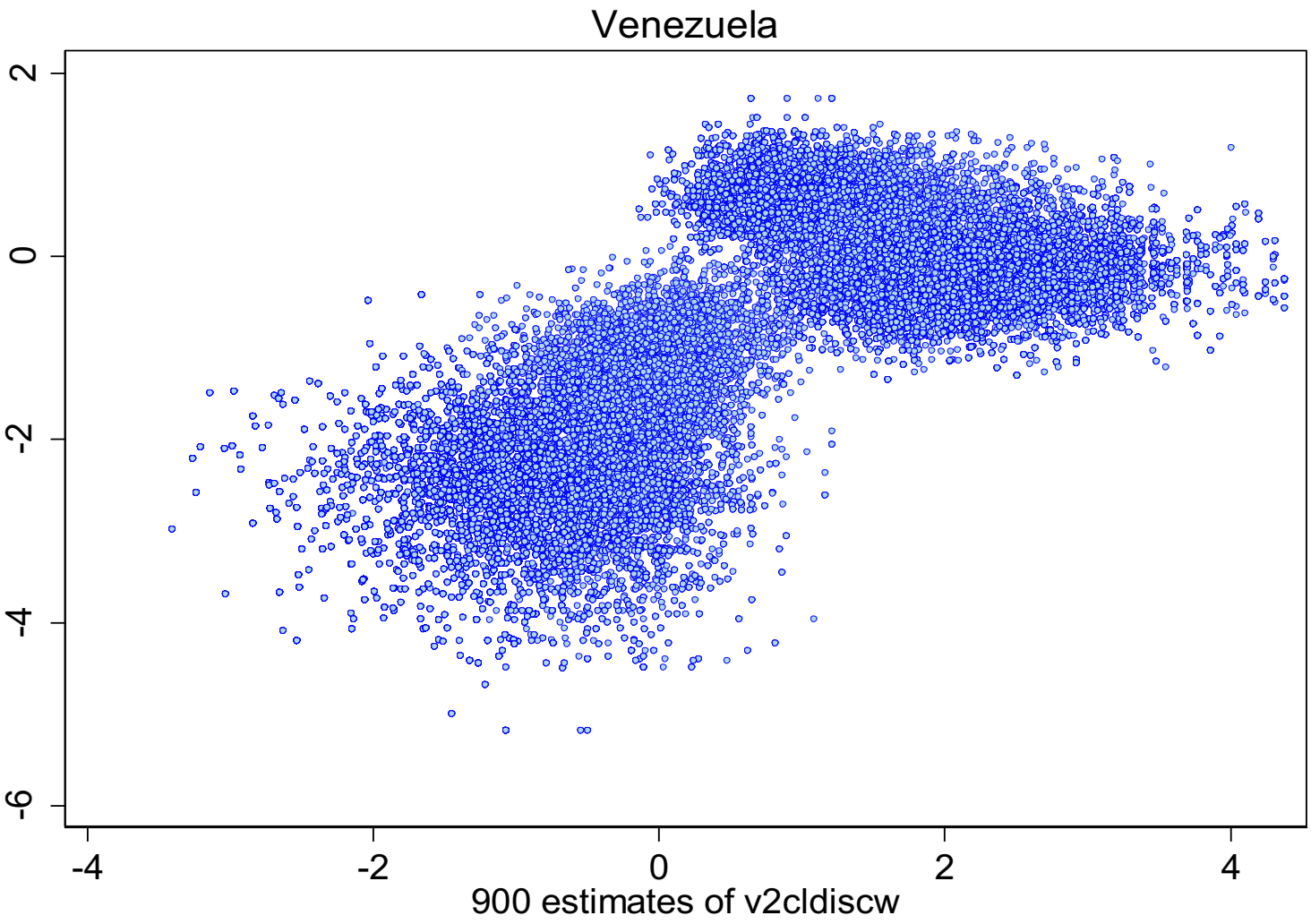
- Almost everyone has ignored how measurement uncertainty affects regression estimates.
- It's not as simple as "when it's in the independent variable, it biases coefficients toward zero, and therefore leads to conservative estimates."
- With V-Dem (C) data, we provide the whole distribution of estimates, so it's possible to do a proper test for the impact of measurement uncertainty using the Method of Composition.

Method of Composition

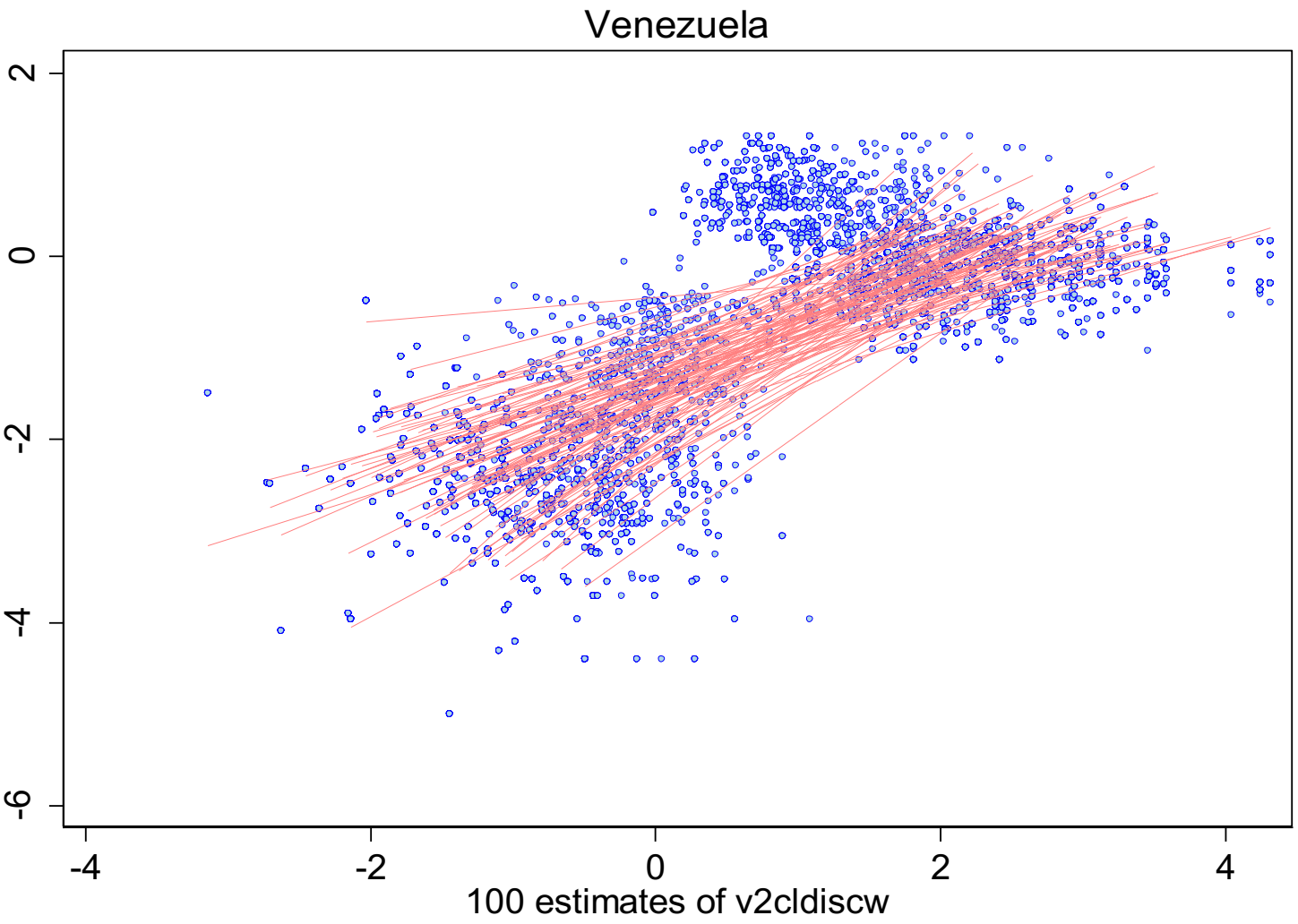
- Run your model 900 times, each time using a random draw from the different estimates
- Save your estimated coefficients
- The median coefficient becomes the new estimate
- The standard deviation of the 900 estimated coefficients becomes the new standard error.
- Better instructions at https://kellogg.nd.edu/content/workshop-varieties-democracy-data-incorporating-measurement-error



Here are point estimates for Venezuela, 1900-2012, on two variables. This relationship implies that there is no uncertainty about what these values are.



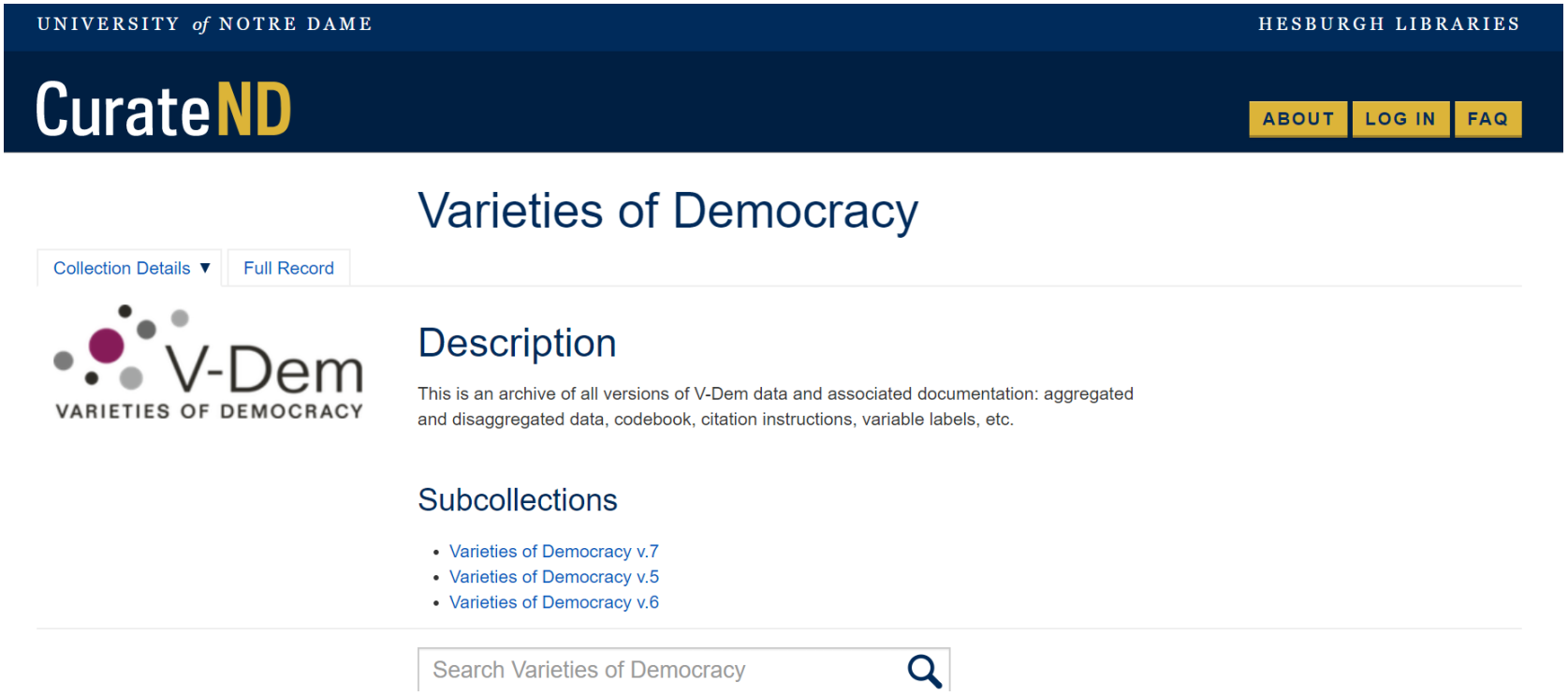
But if we look at 900 estimates of the true values, we get a more complete sense of what the relationship is.



The workshop materials explain how to estimate relationships many times for different draws from these distributions and then combine those estimates into summary estimates.

Posteriors are archived at CurateND

- (and supposedly at the Swedish National Data Archive)



However

- Exercise caution when using V-Dem variables on both sides of the equation, because
- “To the extent that coders exhibit biases that are not consistent with DIF, relationships between V-Dem variables may represent those biases rather than actual relationships between measured concepts.”

Long Panel Structure

- Helps with established temporal priority
- However, opens the door to causal heterogeneity: The causal process is likely to evolve over different historical periods
- This can be modeled. It's an opportunity, too!
- Increased opportunities for diff-in-diff
- We recommend using lagged dependent variables
  - With panels this long, the risk of Nickell bias is reduced

Spatial Relationships

- The potential for confounding due to interference among units has always been present
- Having a (nearly) full set of countries, and long time series, makes it possible to model these relationships
- E.g., Michael Coppedge, Benjamin Denison, Paul Friesen, Lucia Tiscornia, and Yang Xu, Chapter 4, "International Influence: The Hidden Dimension," in Michael Coppedge, Amanda Edgell, Carl Henrik Knutsen, and Staffan I. Lindberg, eds. *Why Democracies Develop and Decline* (Cambridge University Press, June 2022).