Measuring Latent Concepts with Machine Learning

Daniel Weitzel
Department of Political Science
Colorado State University

October 8th, 2024

Are we currently in a period of global democratic backsliding?

Little & Meng: there is no democratic backsliding

Anne Meng
@annemeng_

🚨NEW PAPER ALERT!! @anthlittle and I are excited to share our new working paper "Subjective and Objective Measurement of Democratic Backsliding." 🔥Our hot take🔥: Contrary to the current narrative, we DON'T find evidence that we are in a period of global democratic decline 🧵1/

Subjective and Objective Measurement of Democratic Backsliding*

Andrew T. Little† Anne Meng‡

January 17, 2023

Abstract

Despite the general narrative that we are in a period of global democratic decline, there have been surprisingly few empirical studies to assess whether this is systematically true. Most existing studies of backsliding rely heavily, if not entirely, on subjective indicators which rely on expert coder judgement. We survey other more objective indicators of democracy (such as incumbent performance in elections), and find little evidence of global democratic decline over the last decade. To explain the discrepancy between trends in subjective and objective indicators, we develop formal models that consider the role of coder bias and leaders strategically using more subtle undemocratic action. The simplest explanation is that recent declines in average democracy scores are driven by changes in coder bias. While we cannot rule out the possibility that the world is experiencing major democratic backsliding almost exclusively in ways which require subjective judgement to detect, this claim not justified by existing evidence.

9:21 AM · Jan 18, 2023 · **702.8K** Views

Little & Meng: there is no democratic backsliding

"Despite the general narrative that we are in a period of global democratic decline, there have been surprisingly few empirical studies which assess whether this is systematically true. We [...] find little evidence of global democratic decline over the last decade."

Their argument in a nutshell: We only think there is democratic backsliding because experts are more pessimistic and biased today than in the past.

Little & Meng: there is no democratic backsliding

> *"Despite the general narrative that we are in a period of global democratic decline, there have been surprisingly few empirical studies which assess whether this is systematically true.*
> *We [...] find little evidence of global democratic decline over the last decade."*

Their argument in a nutshell: We only think there is democratic backsliding because experts are more pessimistic and biased today than in the past.

Bias or real backsliding?

How are democracy indicators made?

- All indicators consist of a mix of **objective data** and **subjective coder assessment** evaluating the status of democracy
- V-Dem's Polyarchy:
  - "To what extent is the ideal of electoral democracy in its fullest sense achieved?"
  - Uses a complex aggregation of Bayesian Factor Analysis models to combine numerous features of politics:

Objective	**Subjective**
Are parties banned?	Harassment of journalists?
Are the elections multi-party?	Freedom of discussion for women?
Share of population with suffrage?	Is the media biased?

Weitzel et al.: there is democratic backsliding

In two forthcoming publications, Weitzel, Gerring, Skaaning, and Pemstein use objective data to predict established (and potentially biased) democracy indicators:

- "Electoral Democracy: Global, Historical Measures Based on Observables" (*American Journal of Political Science*)
- "Measuring Backsliding with Observables: Observable-to-Subjective Score Mapping" (*PS: Political Science & Politics*)

How do we do this?

We need the following:

1. **Targets:** democracy indicators
2. **Features:** objectively measured predictors of electoral democracy
3. **Method:** machine learning

Targets: democracy indicators

We predict three established and widely used democracy indicators:

	Range	**Scale**	**Polities**	**Start**
Polyarchy	0 to 1	Interval	202	1789
Polity2	-10 to 10	Ordinal	182	1800
Freedom House	1 to 7	Ordinal	210	1972

All combine objective measures of democracy with subjective expert assessment.

Targets: democracy indicators

We predict three established and widely used democracy indicators:

	Range	**Scale**	**Polities**	**Start**
Polyarchy	0 to 1	Interval	202	1789
Polity2	-10 to 10	Ordinal	182	1800
Freedom House	1 to 7	Ordinal	210	1972

All combine objective measures of democracy with subjective expert assessment.

Features

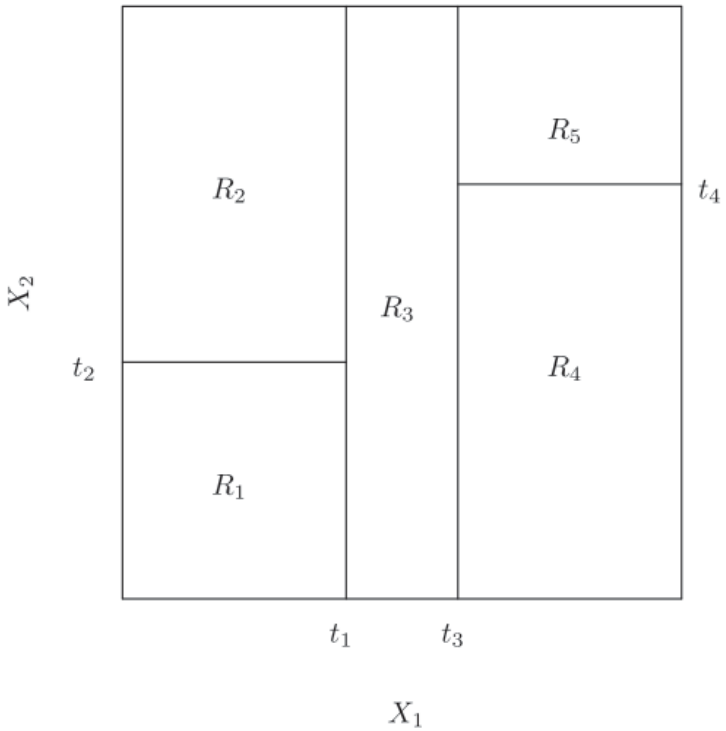
- Our 26 features are observable indicators about (electoral) democracy:
  - Electoral performance (vote and seat shares, $\Delta's$, etc)
  - Turnover
  - Number of consecutive elections
  - Suffrage rates (de jure)
  - Independence
- All these features are:
  - Observable for a broad set of cases and over time
  - Related to democracy
  - Not dependent on interpretation, case knowledge, or expert assessment

Features

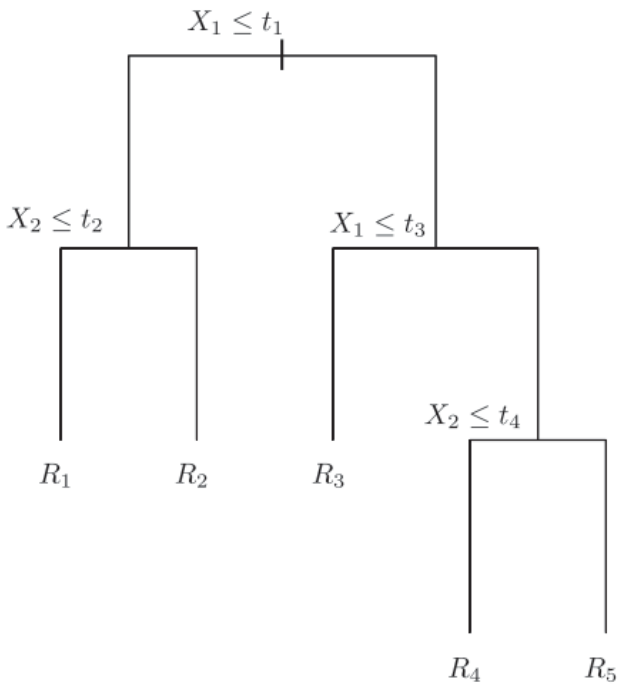
- Our 26 features are observable indicators about (electoral) democracy:
  - Electoral performance (vote and seat shares, $\Delta's$, etc)
  - Turnover
  - Number of consecutive elections
  - Suffrage rates (de jure)
  - Independence
- All these features are:
  - Observable for a broad set of cases and over time
  - Related to democracy
  - Not dependent on interpretation, case knowledge, or expert assessment

Method: random forest

- Supervised algorithm for classification and regression
- Ensemble learner, based on decision trees
- Splits feature space into regions, with each region corresponding to a unique feature combination



(a) Feature Space



(b) Decision Tree

Method: random forest

- Random forests require a training and test set
- We train the algorithm on all years before 2000 (training set) and predict the post-2000 period (test set)
  - The random forest learns the relationship between the target (democracy score) and the features (objective measures) in the training set
  - Based on what it learned, it predicts democracy scores for the unseen test data
- Additional information:
  - 6 fold country-stratified cross-validation
  - Missing values are treated as their own class

Method: random forest

- Random forests require a training and test set
- We train the algorithm on all years before 2000 (training set) and predict the post-2000 period (test set)
  - The random forest learns the relationship between the target (democracy score) and the features (objective measures) in the training set
  - Based on what it learned, it predicts democracy scores for the unseen test data
- Additional information:
  - 6 fold country-stratified cross-validation
  - Missing values are treated as their own class

Assessing backsliding

- **Recap:** Little & Meng (2023) argue that recent values for democracy indicators are biased
- We train a random forest using data from the period before backsliding occurred, then predict the so-called backsliding era
- The model doesn't know that the observed data experienced backsliding
- **A fair test:** Based on the objective features we collected, what are the predicted democracy scores in the test data?

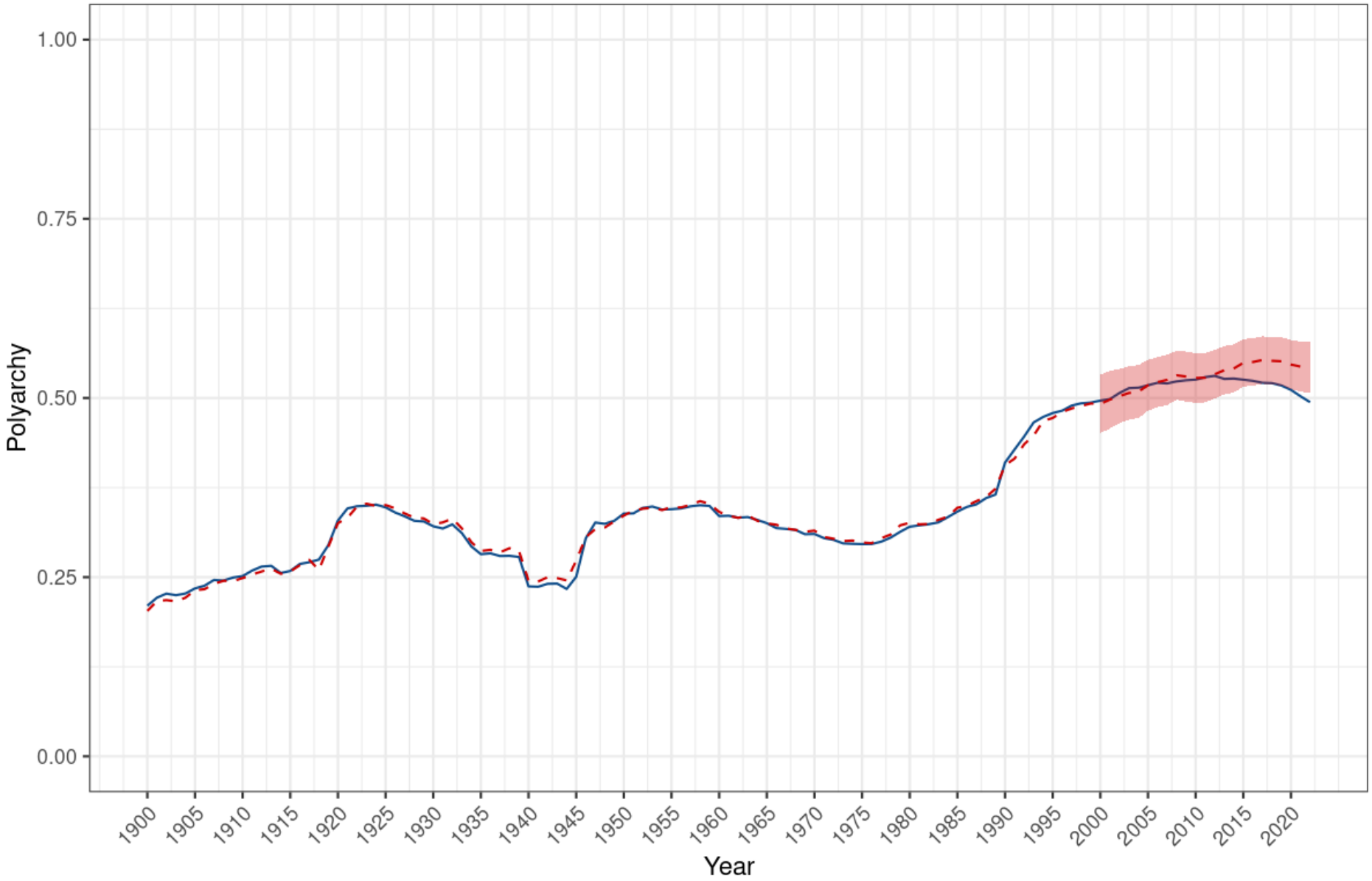
Assessing backsliding

- **Recap:** Little & Meng (2023) argue that recent values for democracy indicators are biased
- We train a random forest using data from the period before backsliding occurred, then predict the so-called backsliding era
- The model doesn't know that the observed data experienced backsliding
- **A fair test:** Based on the objective features we collected, what are the predicted democracy scores in the test data?

Assessing backsliding

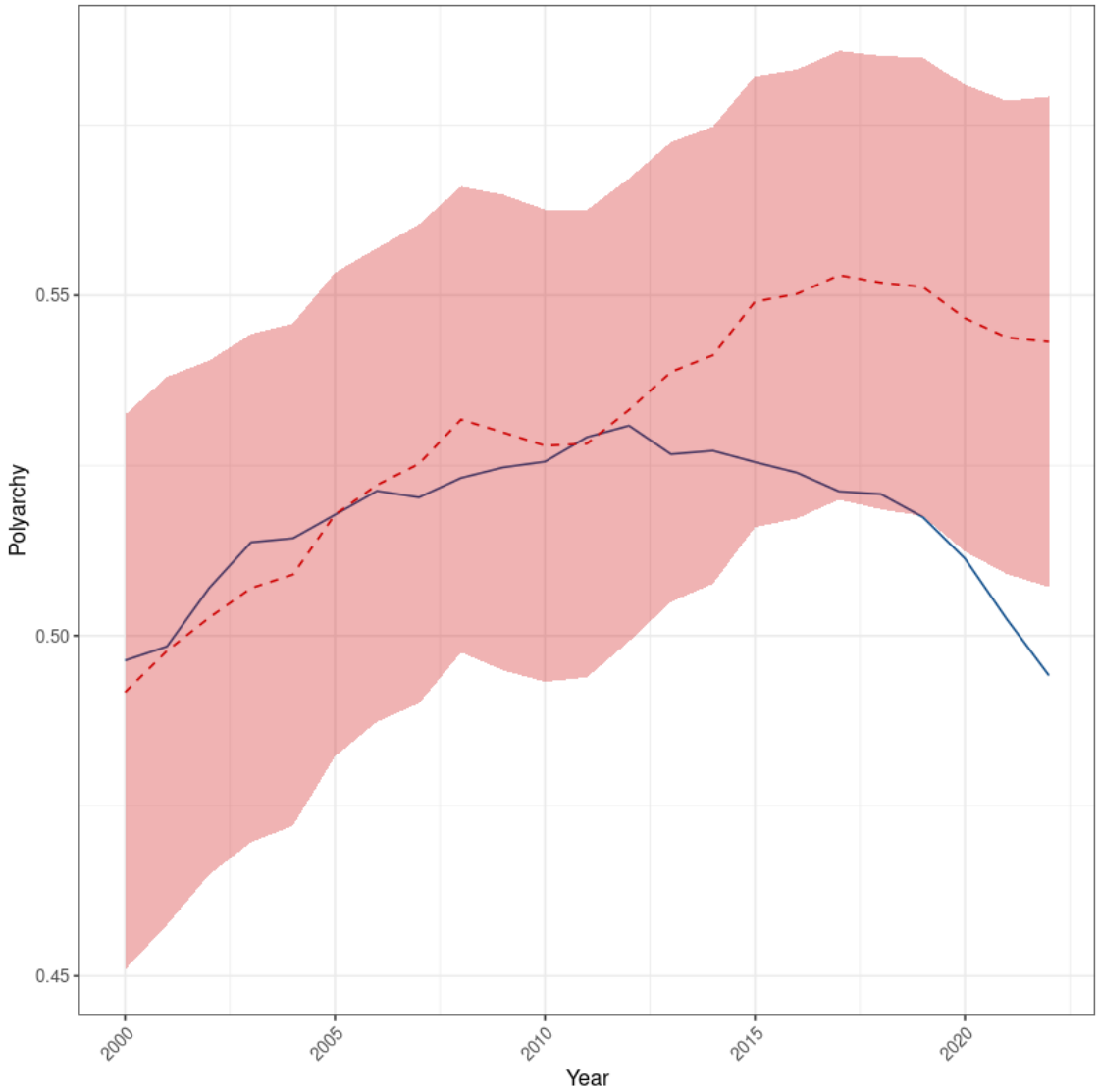
- **Recap:** Little & Meng (2023) argue that recent values for democracy indicators are biased
- We train a random forest using data from the period before backsliding occurred, then predict the so-called backsliding era
- The model doesn't know that the observed data experienced backsliding
- **A fair test:** Based on the objective features we collected, what are the predicted democracy scores in the test data?

Global mean for observed and predicted Polyarchy

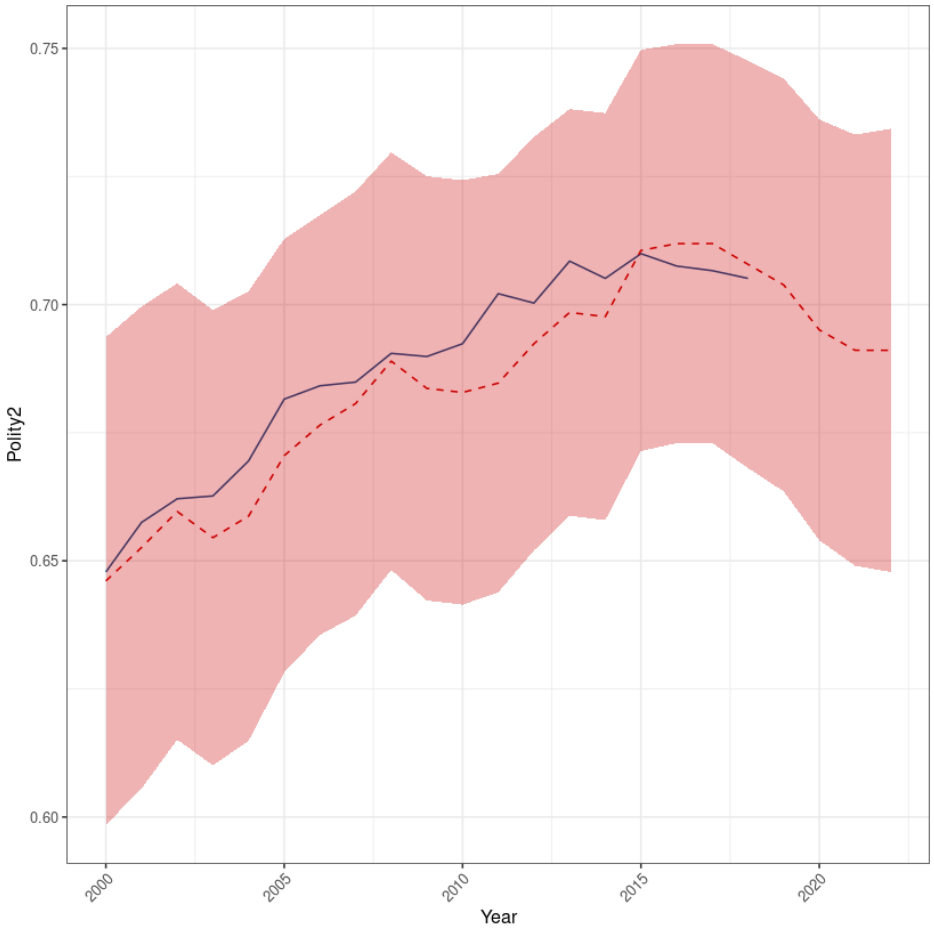


Note: Observed (dark blue) and predicted (red) Polyarchy values. Shaded area indicates out-of-sample predictions.

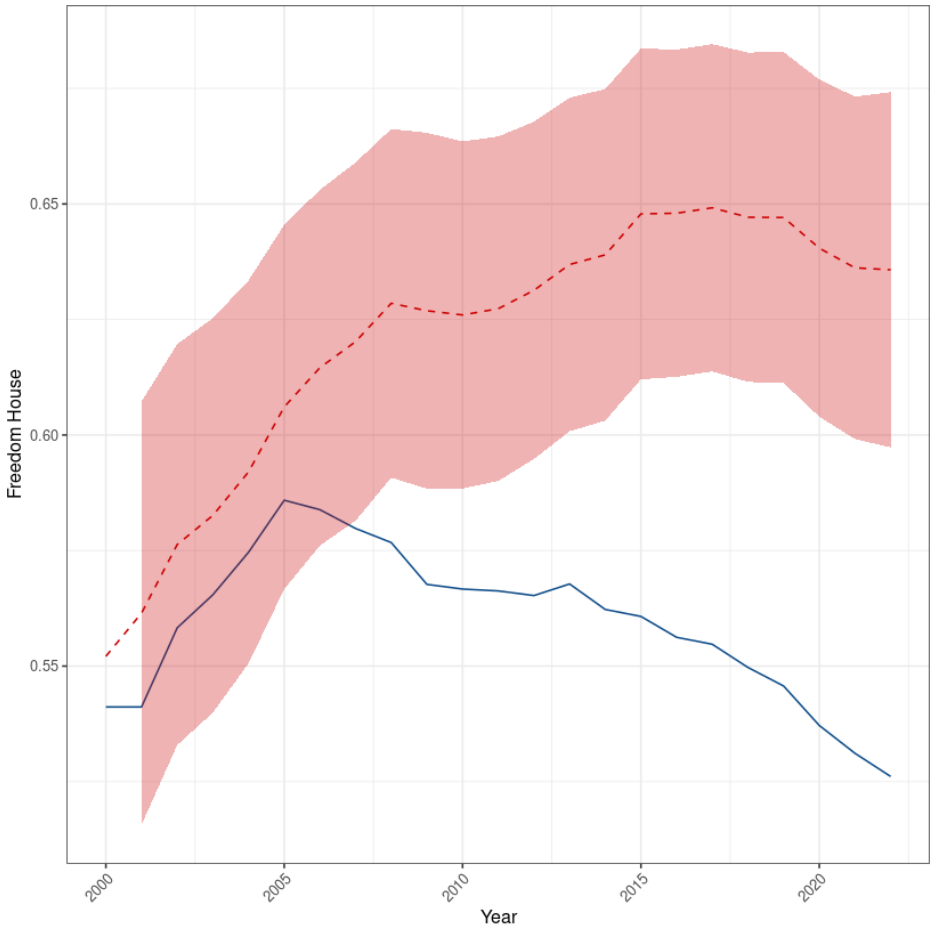
Global means for Polyarchy since 2000



Polity2 and Freedom House since 2000



(a) Polity2



(b) Freedom House

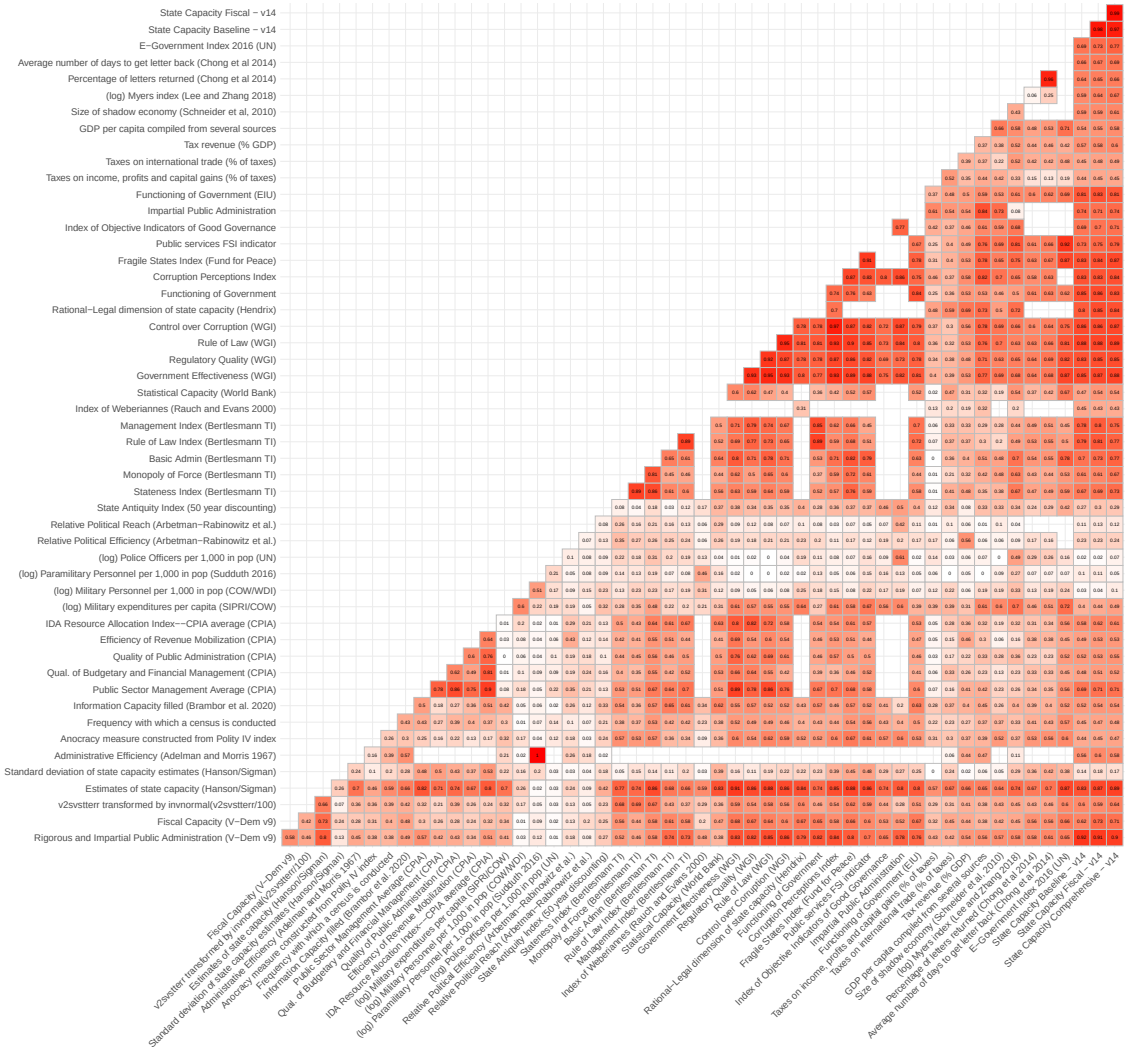
There is global democratic backsliding

- Using objective features, we found evidence of democratic backsliding in Polyarchy, Polity2, and Freedom House predictions
- Coder bias is not causing the current decline!
- Our findings are in line with others in the special edition (Gorokhovskaia 2023; Knutsen et al. 2023)
- Predicted and observed values do not always overlap perfectly, as expected
- Most conservative take based on our predictions: **democratic stasis**

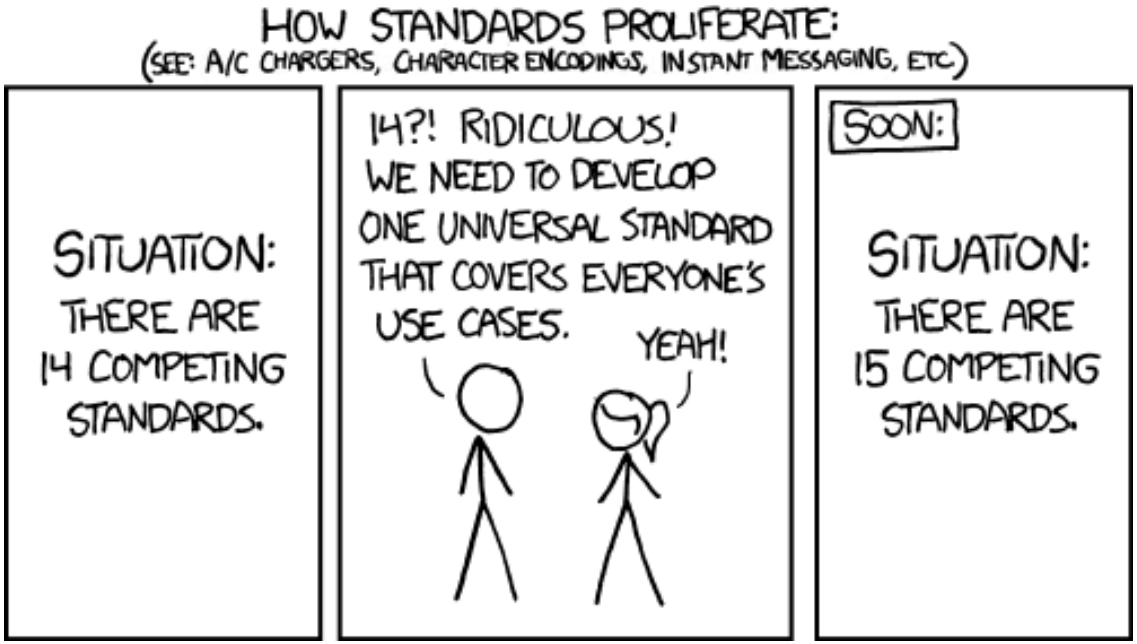
Extending this approach

State Capacity

Existing state capacity measures



Existing state capacity measures



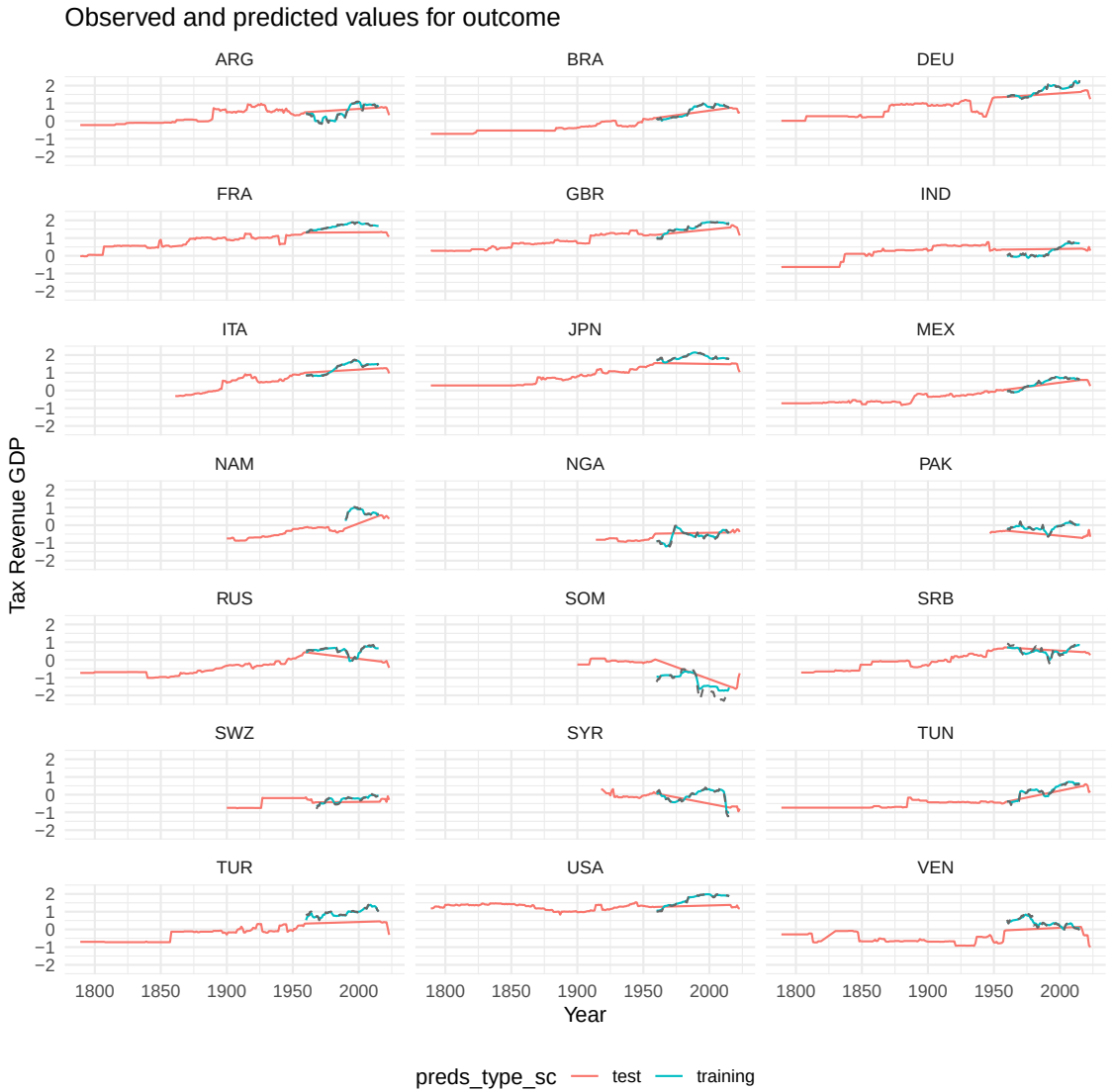
Extending this approach

1. Application of a modified approach to other (more difficult) latent concepts.
   - More difficult: no de facto benchmark measures exist and objective features are harder to collect.
   - Modified: only extreme cases at the end of the spectrum are clear.
2. Currently: **state capacity**
3. What to do? Extend existing measures or build a new one?
4. Trick: Country-year list of either extremely high or low state capacity observations.
5. No need to know the functional form and to describe the functional relationship between target and features.

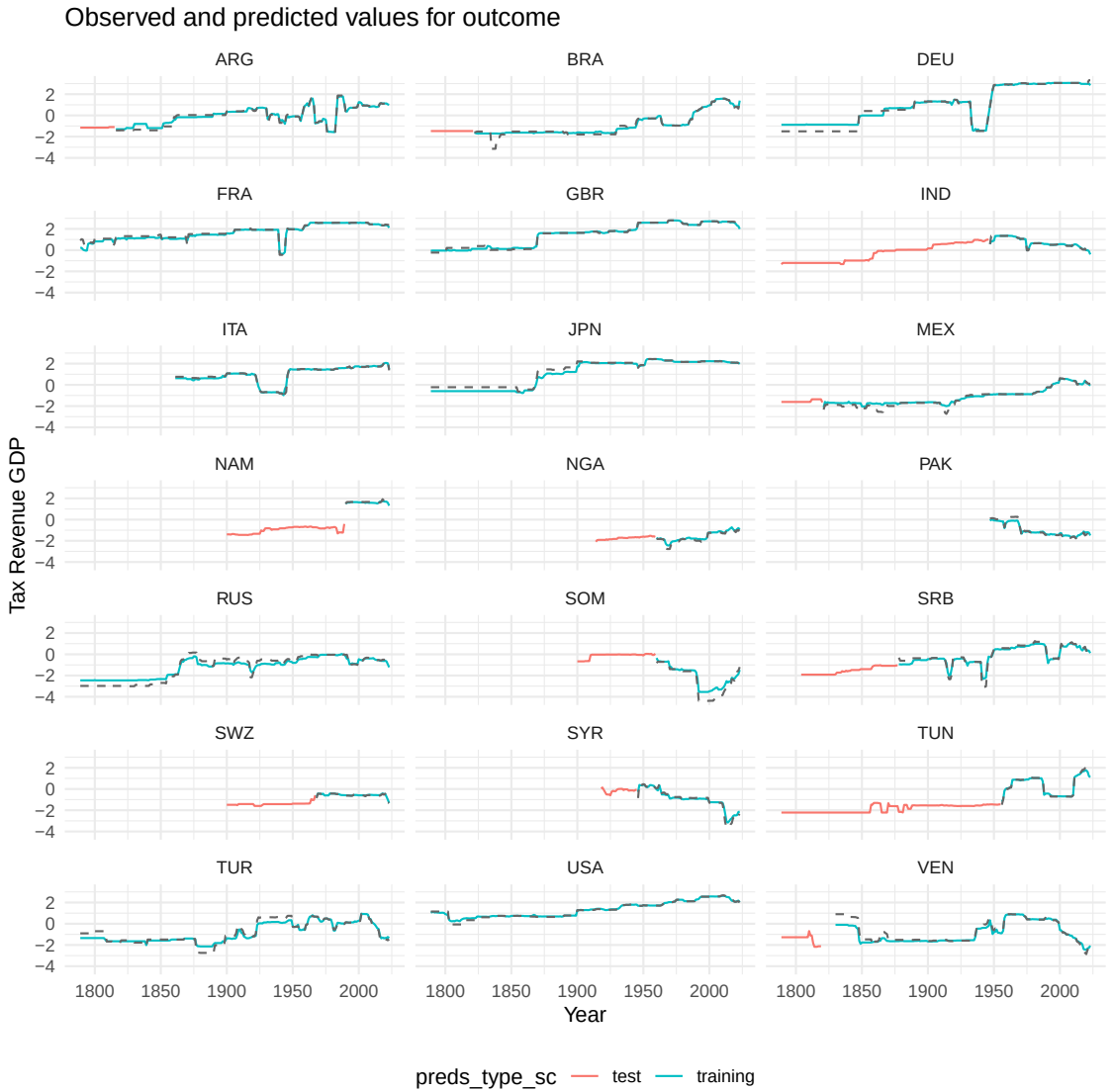
Extending this approach

1. Application of a modified approach to other (more difficult) latent concepts.
   - More difficult: no de facto benchmark measures exist and objective features are harder to collect.
   - Modified: only extreme cases at the end of the spectrum are clear.
2. Currently: **state capacity**
3. What to do? Extend existing measures or build a new one?
4. Trick: Country-year list of either extremely high or low state capacity observations.
5. No need to know the functional form and to describe the functional relationship between target and features.

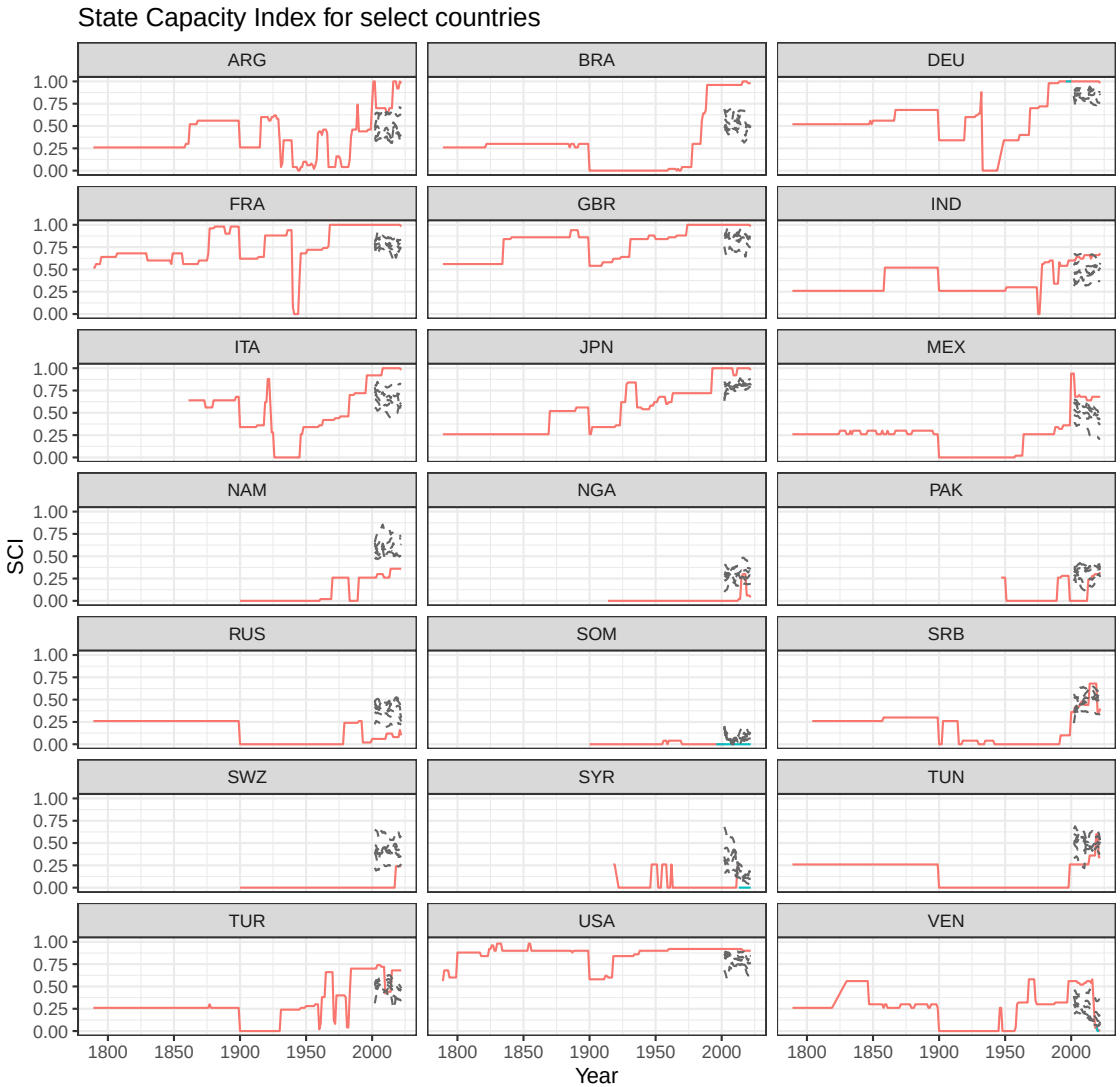
Predicting the Hanson measure of state capacity



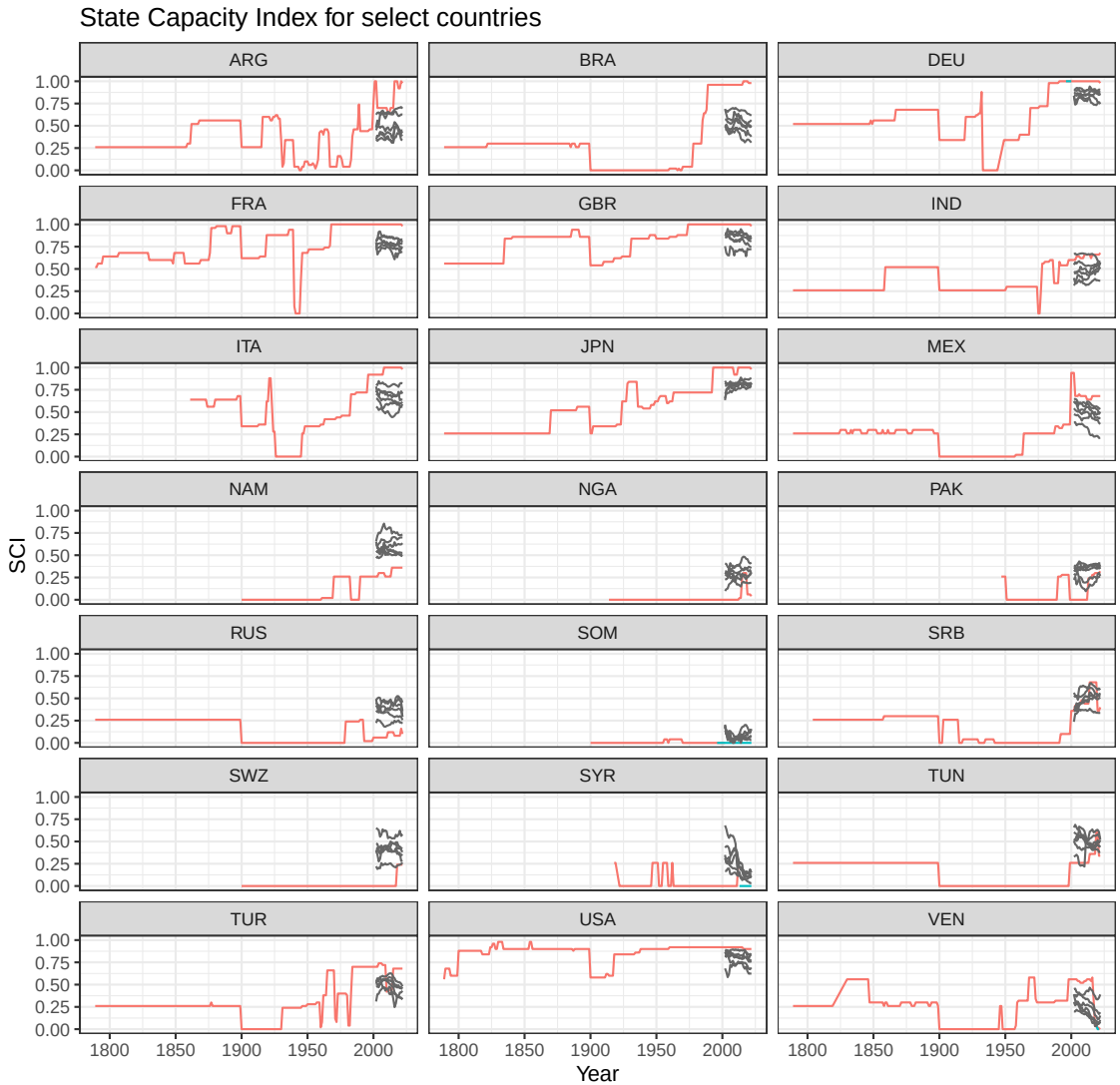
Predicting the LSE measure of state capacity



State Capacity Index



State Capacity Index since 1789



Things that keep me awake at night

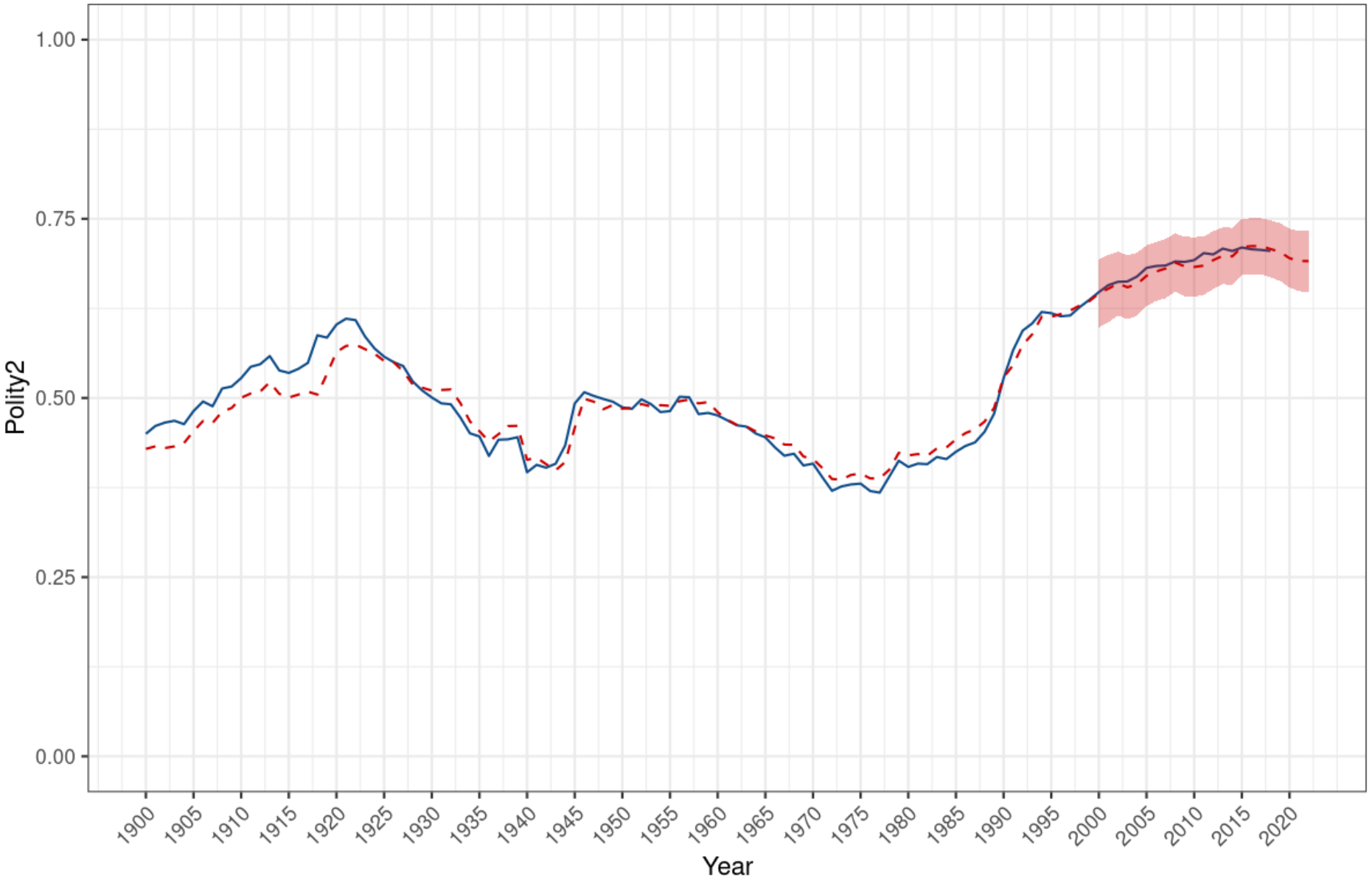
- State capacity is hard!
- Enormous variation in how the concept is understood (monopoly of force, ability to count, taxation).
- Target and features are either closely related or far removed.
- State capacity has changed over time.
- What to do about subnational variation?

Model Performance

	MSE	RMSE	MAE	R^2
Polyarchy	0.0026	0.0513	0.0305	0.9641
Polity2	0.0109	0.1042	0.0596	0.9165
Freedom House	0.0108	0.1038	0.0592	0.9172

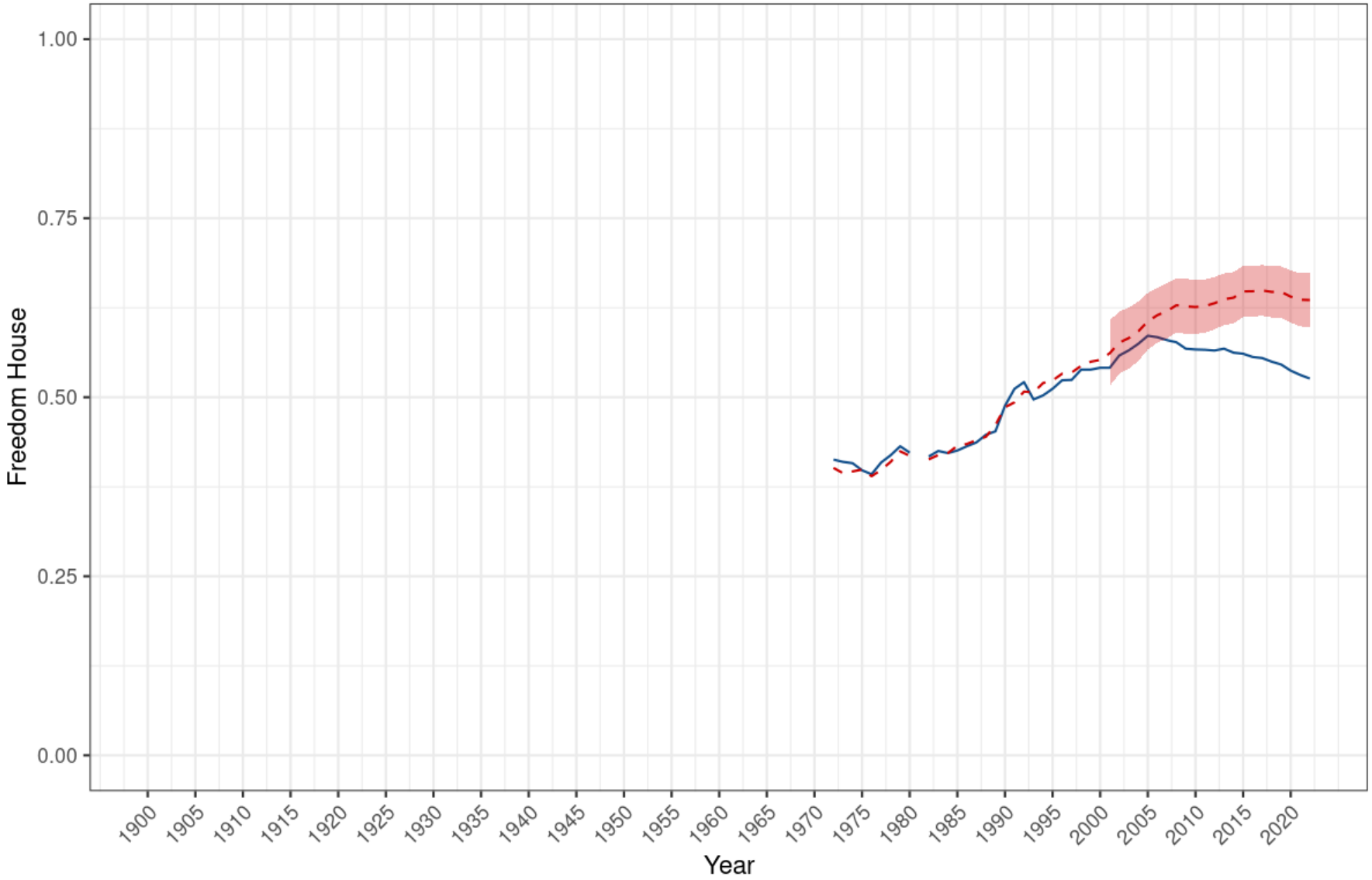
Note: Metrics reported on Out-Of-Bag training samples in the stratified cross-validation set.

Global mean for observed and predicted Polity2



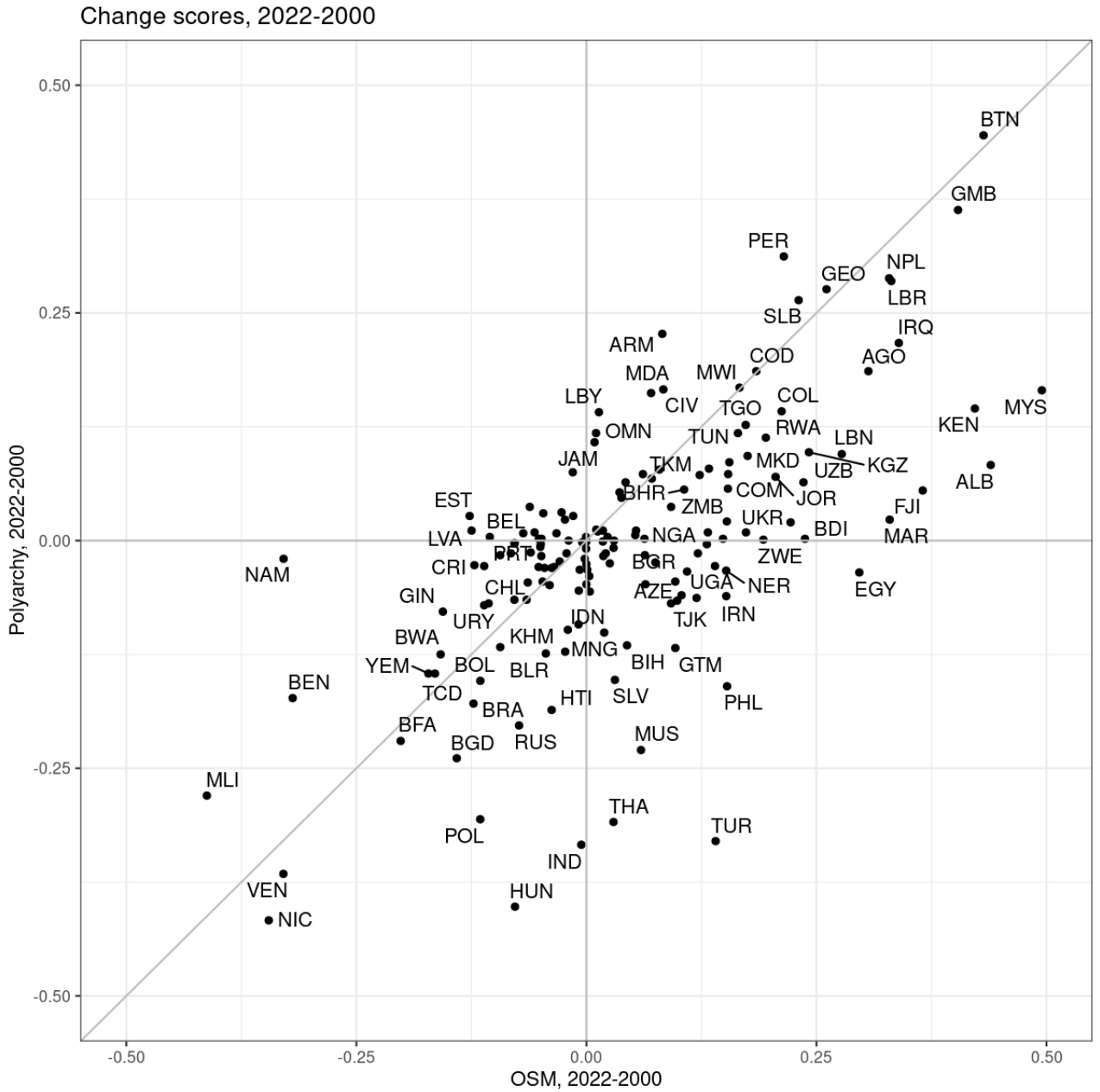
Note: Observed (dark blue) and predicted (red) values for Polity2. Shaded area indicates out-of-sample predictions.

Global mean for observed and predicted FH

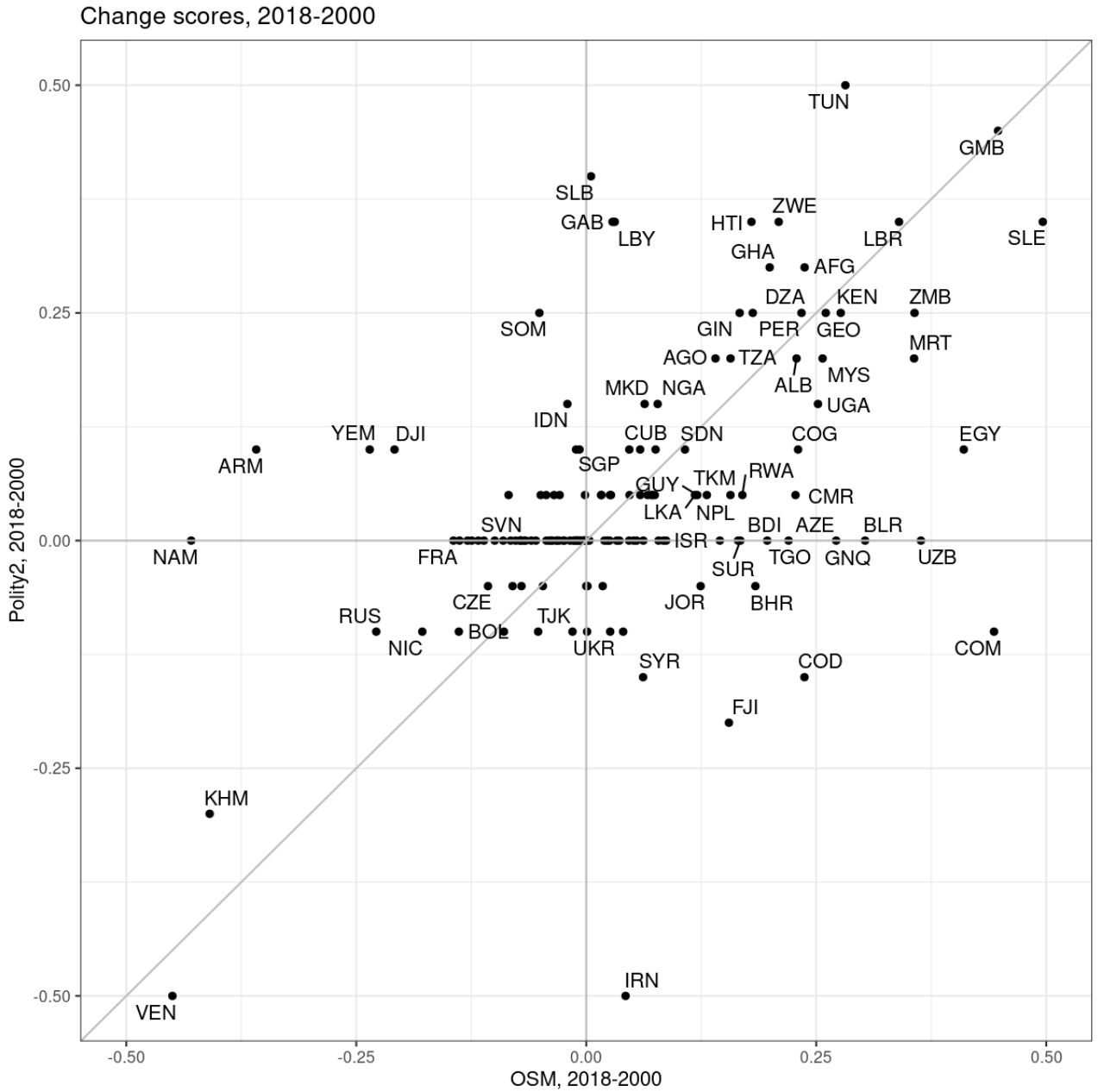


Note: Observed (dark blue) and predicted (red) Freedom House values. The gap from 1981 to 1982 is a gap in Freedom House data collection. Shaded area indicates out-of-sample predictions.

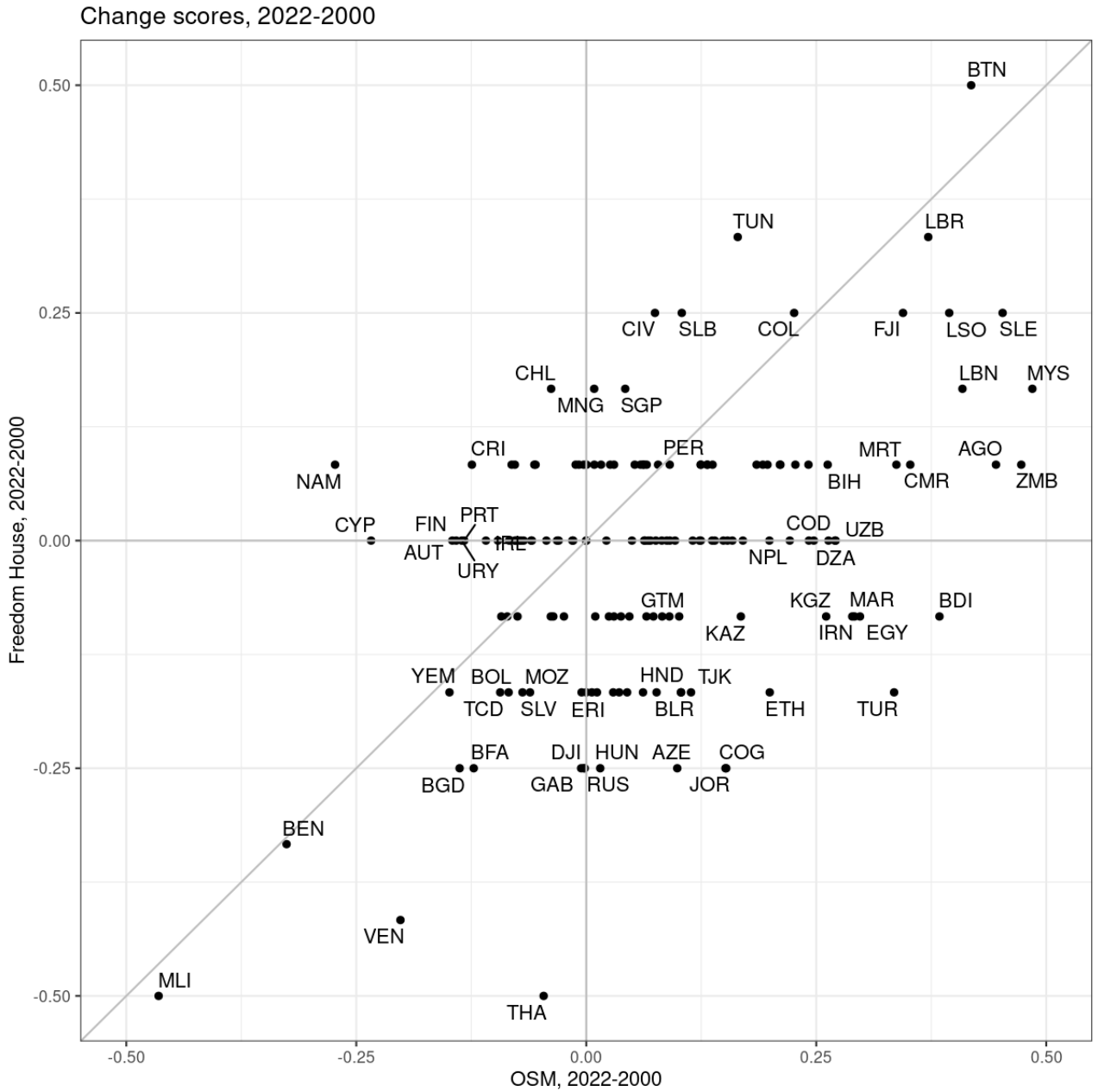
Polyarchy Δ-Plot



Polity2 Δ-Plot



Freedom House Δ-Plot



Robustness Checks

- Reverse the assignment to training and test set.
  - Predictions are very reasonable, even for far out periods.
- Backsliding cut off at 2005 and 2010.
  - Our results are robust to this specification.
- Train on different coder periods in Freedom House.
  - We can observe a divergence between predicted and observed values.
- Predict the interwar period and the 1970s.
  - Predictions are extremely close to observed values.
- Use imputation instead of NA class
  - Results are very close to the initial specification.
- Why not use traditional forecasting?
  - How to model democracy over such a vast set of polity-years?

Robustness Checks

- Reverse the assignment to training and test set.
  - Predictions are very reasonable, even for far out periods.
- Backsliding cut off at 2005 and 2010.
  - Our results are robust to this specification.
- Train on different coder periods in Freedom House.
  - We can observe a divergence between predicted and observed values.
- Predict the interwar period and the 1970s.
  - Predictions are extremely close to observed values.
- Use imputation instead of NA class
  - Results are very close to the initial specification.
- Why not use traditional forecasting?
  - How to model democracy over such a vast set of polity-years?

Robustness Checks

- Reverse the assignment to training and test set.
  - Predictions are very reasonable, even for far out periods.
- Backsliding cut off at 2005 and 2010.
  - Our results are robust to this specification.
- Train on different coder periods in Freedom House.
  - We can observe a divergence between predicted and observed values.
- Predict the interwar period and the 1970s.
  - Predictions are extremely close to observed values.
- Use imputation instead of NA class
  - Results are very close to the initial specification.
- Why not use traditional forecasting?
  - How to model democracy over such a vast set of polity-years?

Robustness Checks

- Reverse the assignment to training and test set.
  - Predictions are very reasonable, even for far out periods.
- Backsliding cut off at 2005 and 2010.
  - Our results are robust to this specification.
- Train on different coder periods in Freedom House.
  - We can observe a divergence between predicted and observed values.
- Predict the interwar period and the 1970s.
  - Predictions are extremely close to observed values.
- Use imputation instead of NA class
  - Results are very close to the initial specification.
- Why not use traditional forecasting?
  - How to model democracy over such a vast set of polity-years?

Robustness Checks

- Reverse the assignment to training and test set.
  - Predictions are very reasonable, even for far out periods.
- Backsliding cut off at 2005 and 2010.
  - Our results are robust to this specification.
- Train on different coder periods in Freedom House.
  - We can observe a divergence between predicted and observed values.
- Predict the interwar period and the 1970s.
  - Predictions are extremely close to observed values.
- Use imputation instead of NA class
  - Results are very close to the initial specification.
- Why not use traditional forecasting?
  - How to model democracy over such a vast set of polity-years?

Robustness Checks

- Reverse the assignment to training and test set.
  - Predictions are very reasonable, even for far out periods.
- Backsliding cut off at 2005 and 2010.
  - Our results are robust to this specification.
- Train on different coder periods in Freedom House.
  - We can observe a divergence between predicted and observed values.
- Predict the interwar period and the 1970s.
  - Predictions are extremely close to observed values.
- Use imputation instead of NA class
  - Results are very close to the initial specification.
- Why not use traditional forecasting?
  - How to model democracy over such a vast set of polity-years?

Variable Importance Plot

List of Features and Variable Names

Democracy Indicators	
Electoral democracy index (D)	*v2x_polyarchy*
Freedom House, combined	*e_fh_combined*
Polity revised combined score (E)	*e_polity2*
Predictors	
Lower chamber election vote share of largest vote-getter (A)	*v2ellovtlg*
Lower chamber election vote share of second-largest vote-getter (A)	*v2ellovtsm*
Lower chamber election vote share of third-largest vote-getter (A)	*v2ellovttm*
Lower chamber election seat share won by largest party (A)	*v2ellostsl*
Lower chamber election seat share won by second largest party (A)	*v2ellostss*
Lower chamber election seat share won by third largest party (A)	*v2ellostts*
Presidential election vote share of largest vote-getter (A)	*v2elvotlrg*
Presidential election vote share of second-largest vote-getter (A)	*v2elvotsml*
Executive electoral regime index (A)	*v2xex_elecreg*
Legislative electoral regime index (A)	*v2xlg_elecreg*
Elections multiparty (LIED)	*multi_party_elections*
Share of population with suffrage (D)	*v2x_suffr*
Dummy for legislative elections	*v2eltype_legislative*
Dummy for presidential elections	*v2eltype_presidential*
Difference in vote share of top two parties	*top2_difference*
Combined vote share of top two parties	*top2_combined*
Top two parties have vote share larger than 59.99	*top2_monopoly*
Legislative elections, consecutive	*v2ellocons*
Legislative elections, cumulative	*v2ellocumul*
Presidential elections, consecutive	*v2ellocons*
Presidential elections, cumulative	*v2elprescumul*
Head of government turnover	*v2elturnhog*
Head of state turnover	*v2elturnhos*
Executive turnover	*v2eltvrexo*
Turnover period (LIED)	*turnover_period*
Turnover event (LIED)	*turnover_event*
Two turnover period (LIED)	*two_turnover_period*

V-Dem's Polyarchy Index

$$
\begin{aligned}
v2x\_polyarchy &= .5 * MPI + .5 * API \\
&= .5 * (v2x\_elecoff * v2xel\_frefair * v2x\_frassoc\_thick * \\
&\quad v2x\_suffr * v2x\_freexp\_altinf) \\
&\quad + .5 * ((1/8) * v2x\_elecoff + (1/4) * v2xel\_frefair \\
&\quad + (1/4) * v2x\_frassoc\_thick + (1/8) * v2x\_suffr \\
&\quad + (1/4) * v2x\_freexp\_altinf)
\end{aligned}
$$