Elite individuals, insitutions, and economic growth accounting

James Xu

ECON 442, Duke University

April 11, 2024

Table of Contents

Research Question

What is the relationship between elite students, academic institutions, and economic growth?

- Are there outsized returns to economic growth when there are more academic elites?

Method

Through this paper, I investigate the relationship between the number of top-ranked universities, share of top math students, IMO scores, and GDP per capita growth to analyze these questions.

Research Question

Importance and Relevance

Does elite performance matters in education and human capital?

- Rough measures of human capital such as school enrollment rate, are often used (e.g. Mankiw-Romer-Weil)
- Elite students and top university research may give countries a technology advantage

Implications for economic convergence

- Richer, larger countries will likely have better research output and universities
- Technology and trade conflict between China-US have shown that "hoarding" technology may widen gap between nations

Data Sources

- World Bank: World Development Indicators
- International Math Olympiad
- ARWU University Rankings
- PISA Math Scores
- Economist Democracy Scores

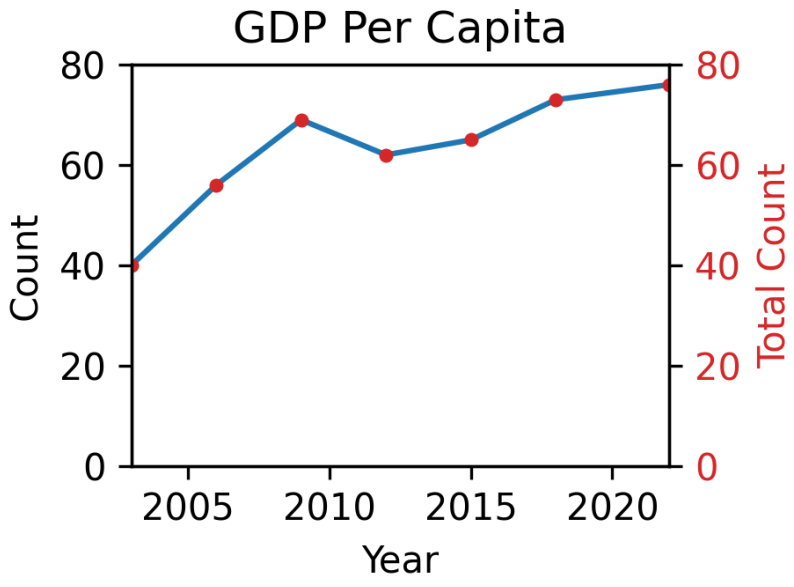
World Bank Data

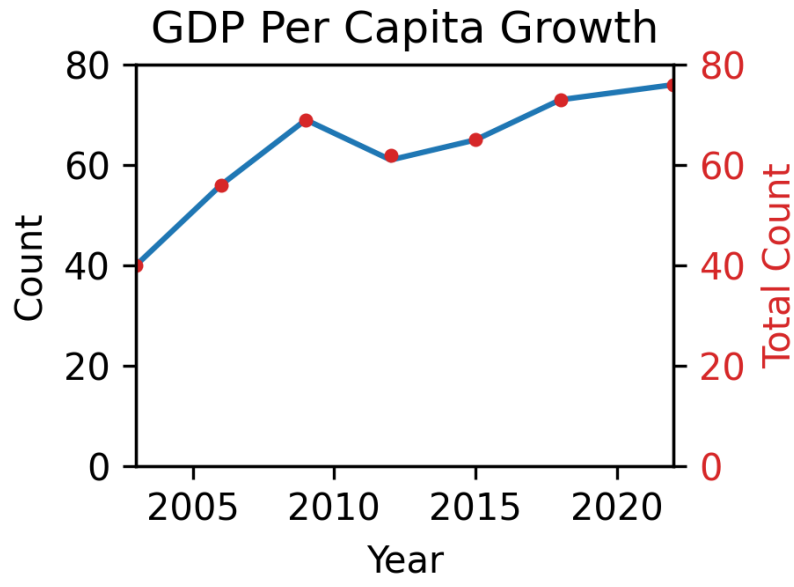
The World Bank collects and publishes development indicators for most countries/economies in the world. In this paper, I use the following variables:

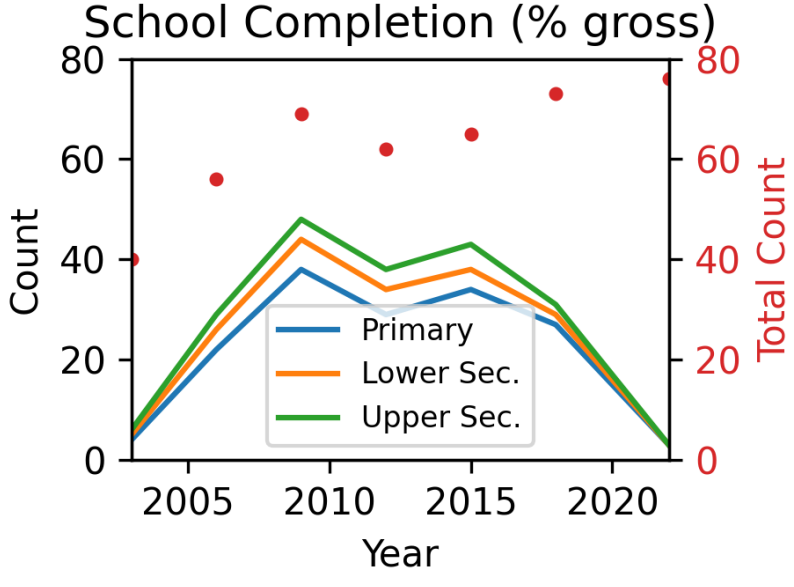
- GDP Per Capita Growth: % growth in constant 2015 $USD
- GDP Per Capita: current $USD
- School Completion Rates: % gross of relevant age group
  - "What % of primary-school-aged population is enrolled in primary school?"
  - This number can be greater than 100%
  - Not available for all years and all countries $\rightarrow$ missing data problem
- Population: all residents of a country/territory

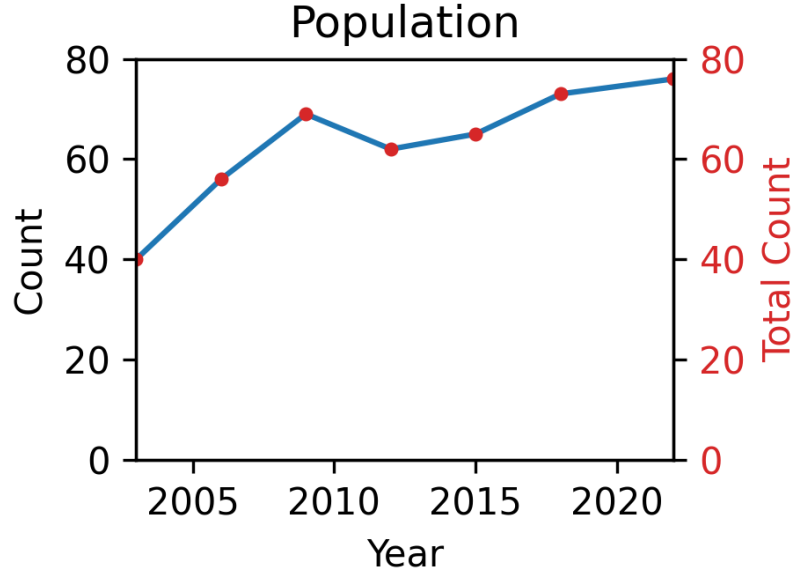
World Bank Data

Missing data









Imputing nulls

Using XGBoost for better predictions

Data is not missing at random

School completion data is only available for a maximum of 80 countries per year and has high variance in this availablility. This is the most limiting factor in the analysis.

Predicting missing values

XGBoost is a tree-based model that has built-in null handling. I use the remaining variables to predict school completion rates and EIU democracy scores. Achieves significantly higher accuracy than linear regression.

IMO Scores

The IMO is an international mathematics contest for high school students

Table: Top 10 countries by IMO score.

Region	IMO Score	GDPpc	GDPpc Growth
KOR	10.2026	26060.9203	293.8552
CHN	9.8296	6424.4972	786.8171
USA	9.6196	54346.5078	125.7613
RUS	9.0800	10432.8366	277.0704
SGP	9.0642	51652.4651	348.0543
BGR	8.8821	7556.3933	417.7564
ROU	8.7367	9432.6698	439.7522
HUN	8.7153	13979.7009	262.9650
VNM	8.5029	2159.2583	527.6404
UKR	8.3941	3079.8378	131.9303

- Scores collected from 2003 to 2022
- Scores are transformed by $t(s) = \frac{s}{\log P}$ where s is the country's raw score and P is population.
  - Team size of 6 means that larger countries have an advantage due to "genius odds"
  - Score is capped, so dividing by $\log P$ will correct for theoretical ceiling of performance

ARWU Rankings

Description and Usage

ARWU (Academic Ranking of World Universities) is a set of university rankings based primarily on research output.

- Rankings are produced annually and are available from 2003
- From 2003 to 2016, 500 top universities were ranked; after 2017, 1000 were ranked.

Per-Capita Scaling

Larger countries naturally have an advantage, so a more fair metric is

$$ARWU_{i,t} = \frac{arwuCount_{i,t}}{P_{i,t}} \cdot 10^6$$

Instead, looking at ARWU insitutions per million population indicates the relative quantity of elite insitutions in a country/region.

PISA Math Scores

World-wide study to evaluate 15-year-old students' performance on math, science, and reading. Testing is done usually every 3 years, starting from 2000. 2021 study was delayed to 2022 due to COVID.

- This paper uses PISA mathematics data from 2003 to 2022
- Countries included are mostly OECD (wealthier), with some other regions/countries participating
- Must use weighted means because of survey methodology
  $math_{c,t} = \frac{\sum_{i \in c \cap t} stu\_math_i \cdot stu\_wgt_i}{\sum_{i \in c \cap t} stu\_wgt_i}$

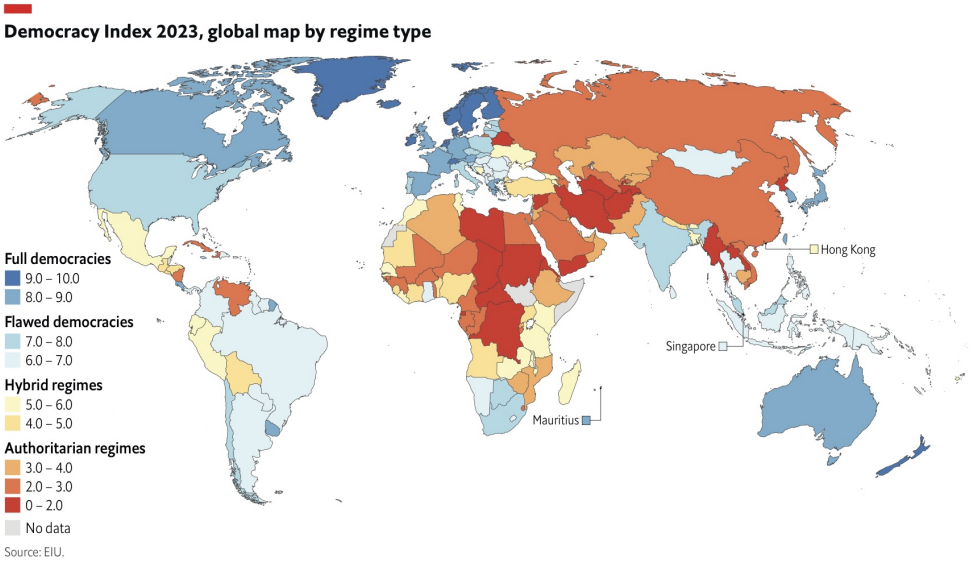
As an indicator of "elite" academic performance, also compute the share of students in a region/country in global 1% of test takers (benchmark score B_t):

$$math99_{c,t} = \frac{\sum_{i \in c \cap t} (stu\_math_i \geq B_t) \cdot stu\_wgt_i}{\sum_{i \in c \cap t} stu\_wgt_i}$$

EIU Democracy Index

Index measuring the quality of democracy around the world published by the Economist Intelligence Unit.

- Published from 2006 every 2 years until 2010, annually afterwards (use XGBoost to impute missing)
- 0-10 scale, where 10 is democracy and 0 is autocracy.



Rationale for variables

IMO Scores

Indicator for a country's (and region's) ability to develop/identify pinncale STEM talent at high school level.

ARWU Rankings

Indicator for a country's (and region's) ability to produce exellence in research output.

Percent in PISA 99th percentile

When controlling for average PISA math scores, this is a partial indicator for whether general excellence in academics is encouraged/necessary.

Model specification

Let elite indicators be: $math99, ARWU, IMO$. For the sake of concision, let $E_{i,t}$ be a 1 by 4 matrix defined as:

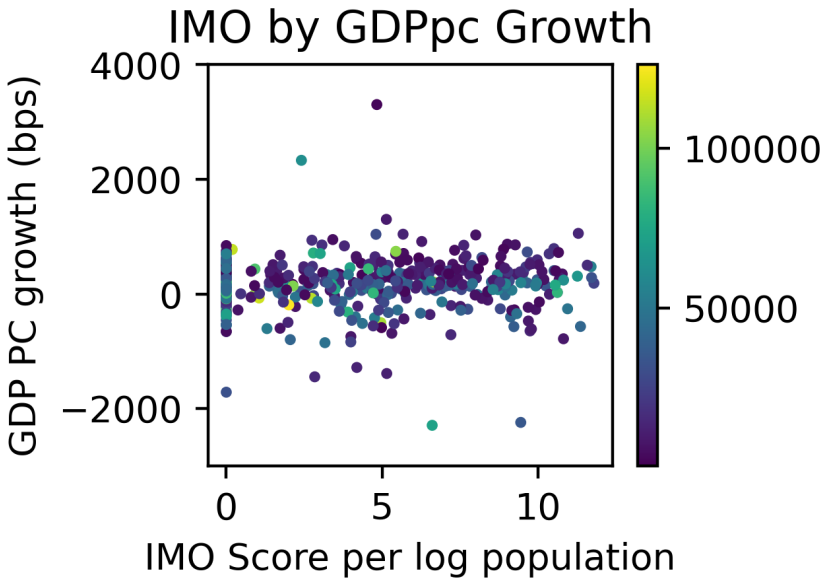
$$E_{i,t} = \begin{bmatrix} math99_{i,t} & ARWU_{i,t} & IMO_{i,t} \end{bmatrix}$$

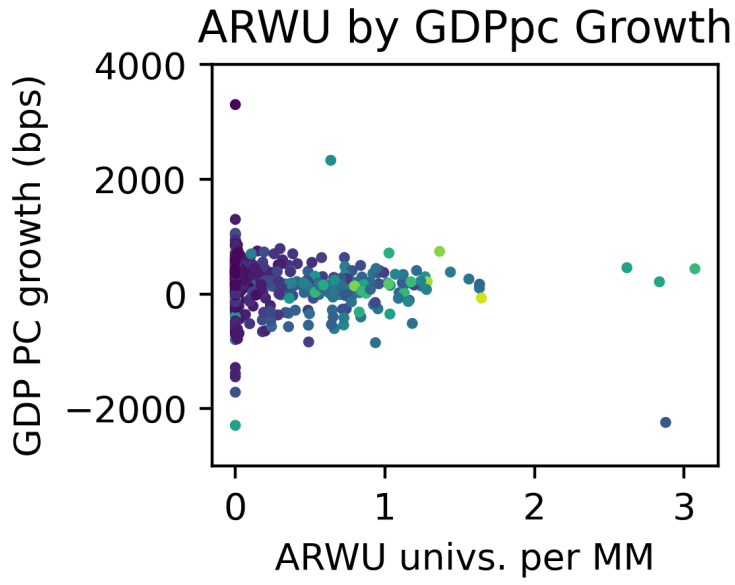
for country/region i in year t. For control variables, $I_{i,t}$ is the matrix of variables and α is coefficients.

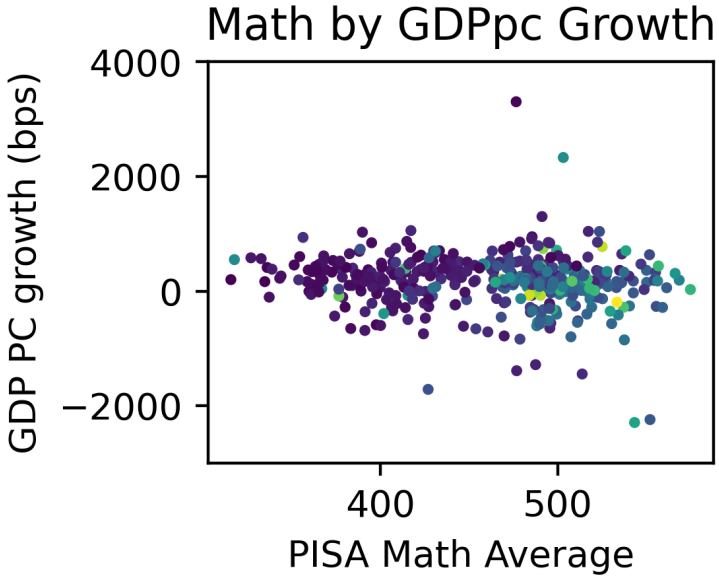
$$Y_{i,t} = \beta_0 + \lambda E_{i,t} + \delta GDPpc \times E_{i,t} + \alpha I_{i,t} + T_t + C_i + \epsilon_{i,t} \tag{1}$$

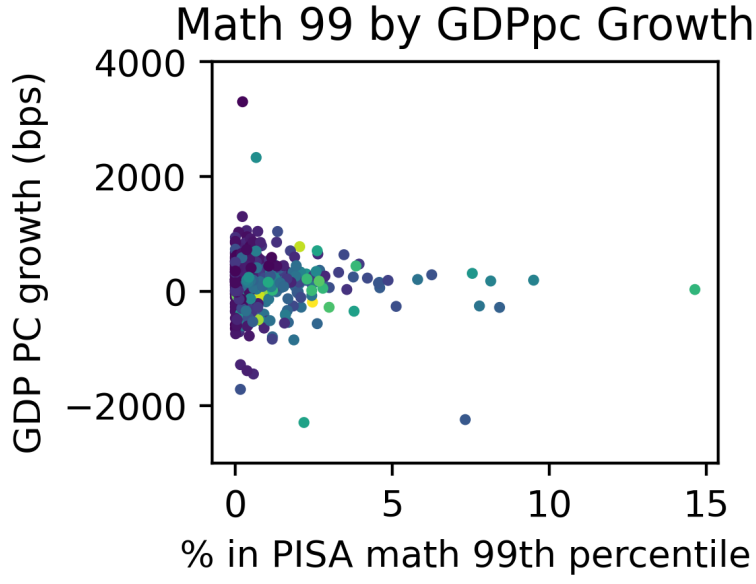
where T, C represent time and entity dummies respectively and $Y_{i,t}$ be the GDP per capita growth in basis points for a country/region i and year t.
Controls: school completion rates, GDP per capita, PISA average math score, democracy index

Relationships









PISA panel regression

Highly dependent on model specification due to rich country bias and limited variation

	Dependent variable: GDP Per Capita Growth (bps)			
	Model 1 (base)	Model 2	Model 3 (Time FE)	Model 4 (Time + Entity FE)
PISA Math in global P99	-31.228*	-38.928*	-34.569*	-115.492***
	(16.314)	(20.802)	(17.850)	(36.599)
IMO score per log population	10.674*	4.924	3.122	-11.158
	(6.363)	(7.470)	(6.441)	(15.728)
ARWU insitutions	-166.641**	-162.705*	-176.049**	-159.960
	(68.800)	(86.871)	(72.281)	(112.884)
Time Effects	No	No	Yes	Yes
Fixed Effects	No	No	No	Yes
Controls	No	Yes	Yes	Yes
Entities	89	89	89	89
Observations	440	440	440	440
R^2	0.037	0.077	0.374	0.553
Adjusted R^2	0.031	0.056	0.351	0.414
Residual Std. Error	445.700 (df=436)	439.888 (df=429)	364.790 (df=423)	346.511 (df=335)
F Statistic	5.616*** (df=3; 436)	3.589*** (df=10; 429)	15.813*** (df=16; 423)	3.983*** (df=104; 335)
Note:				*p<0.1; **p<0.05; ***p<0.01

- Introducing controls and time effects do not have a major effect on coefficients; elite indicators are not highly correlated with controls
- Country/region fixed effects have small effect on ARWU variable, large effects on PISA math 99 and IMO variables

PISA yearly regression

	Dependent variable: GDP Per Capita Growth (bps)							
	2003	2006	2009	2012	2015	2018	2022	Panel FE
PISA Math in global P99	-52.490	-193.475**	154.062**	16.575	-34.309	3.769	-69.681**	-115.492***
	(48.524)	(84.559)	(62.680)	(36.365)	(76.140)	(27.106)	(27.702)	(36.599)
IMO score per log population	4.478	-9.112	14.038	7.368	-16.622	2.193	-0.174	-11.158
	(17.327)	(20.741)	(17.012)	(13.983)	(27.567)	(7.722)	(13.474)	(15.728)
ARWU insitutions	-300.038**	-454.401***	136.623	-344.990**	289.270	61.220	-829.105***	-159.960
	(111.479)	(167.497)	(188.401)	(133.886)	(270.034)	(125.503)	(206.280)	(112.884)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40	56	69	61	65	73	76	440
R^2	0.675	0.499	0.264	0.302	0.100	0.294	0.417	0.553
Adjusted R^2	0.562	0.387	0.137	0.162	-0.067	0.180	0.327	0.414
Residual Std. Error	179.479 (df=29)	365.156 (df=45)	402.134 (df=58)	254.369 (df=50)	490.971 (df=54)	180.728 (df=62)	335.236 (df=65)	346.511 (df=335)
F Statistic	6.011*** (df=10; 29)	4.479*** (df=10; 45)	2.075** (df=10; 58)	2.158** (df=10; 50)	0.598 (df=10; 54)	2.581** (df=10; 62)	4.649*** (df=10; 65)	3.983*** (df=104; 335)

Note: *p<0.1; **p<0.05; ***p<0.01

Relationship changes from year to-year

- 2009 is a significant outlier (coefficients now positive than negative, probably due to recession)
- Panel methods may "average out" the variations in relationship
- High variance in R^2 suggest that importance of controls, variables of interest may also not be constant

PISA panel regression

Without imputation

	Dependent variable: GDP Per Capita Growth (bps)			
	Model 1 (base)	Model 2	Model 3 (Time FE)	Model 4 (Time + Entity FE)
PISA Math in global P99	-31.228*	5.985	8.659	25.421
	(16.314)	(22.468)	(21.493)	(84.291)
IMO score per log population	10.674*	6.824	12.529	50.506
	(6.363)	(9.440)	(8.631)	(32.187)
ARWU insitutions	-166.641**	-190.750	-249.208**	-454.384
	(68.800)	(127.229)	(116.476)	(351.559)
Time Effects	No	No	Yes	Yes
Fixed Effects	No	No	No	Yes
Controls	No	Yes	Yes	Yes
Observations	440	112	112	112
R^2	0.037	0.204	0.383	0.715
Adjusted R^2	0.031	0.125	0.294	0.355
Residual Std. Error	445.700 (df=436)	236.761 (df=101)	212.750 (df=97)	203.207 (df=49)
F Statistic	5.616*** (df=3; 436)	2.587*** (df=10; 101)	4.294*** (df=14; 97)	1.987*** (df=62; 49)
Note:				*p<0.1; **p<0.05; ***p<0.01

- Relationships are very different from imputed version
- Some countries are missing in this regression and some years as well
- School completion rates are only available for certain years and we have seen the relationship to be volatile year-to-year

Interpretation

IMO Scores

- Mostly indicates positive main relationship with GDP per capita growth (large standard errors)
- Negative relationship when fixed effects are included, volatile between years

PISA math 99 percentile share

- Generally negative relationship with GDP per capita growth
- Negative relationship when fixed effects are included, volatile between years

ARWU insitutions per million

- Consistent negative relationship found in regression
- To some degree volatile between years, but is most consistent

Limitations: data problems

Regression results do not present a clear picture of the existence of a statistical relationship between elite indicators and economic growth.

Little variation in data

- Countries and regions with PISA test scores tend to be wealthier, more developed economies
- Low variance within countries over time and between countries as a result

Model specification matters

- Including entity fixed effects a significant difference versus time effects (likely due to above)
- Large variation in $IMO, math99$ between time effects, fixed effects, panel models
- Missing data (only post-2006 EIU Democracy index scores, missing school completion rates)
  - XGBoost is used to fill in this data as controls are somewhat stable year-to-year, but this is not perfect

Limitations: Yearly vs. Panel

Significant year-to-year variation

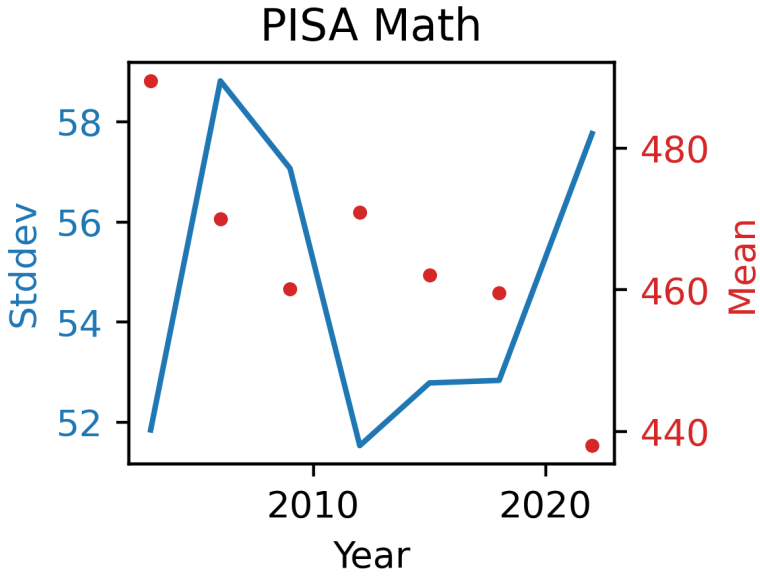
Unstable panel vs. yearly results

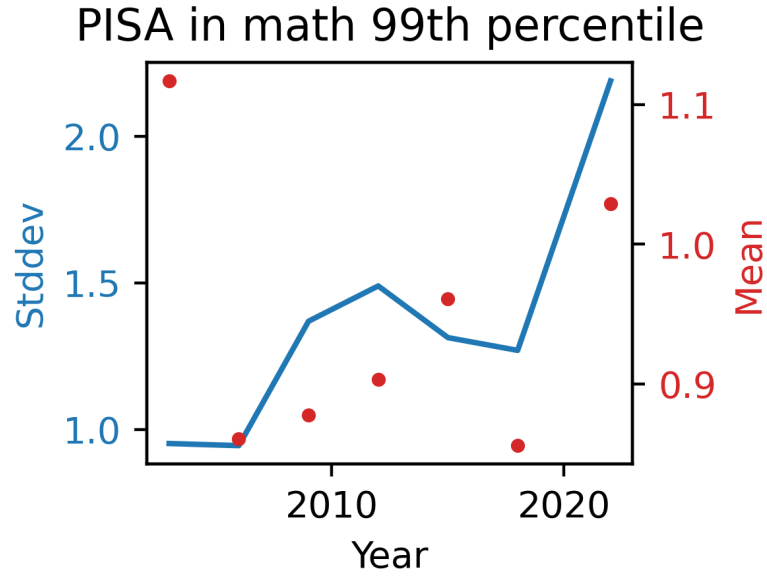
The panel methods mask some of the year-to-year changes in relationships. Some indicators are more stable than others, but $math99$ in particular alternates from positive to negative coefficient with large swings.

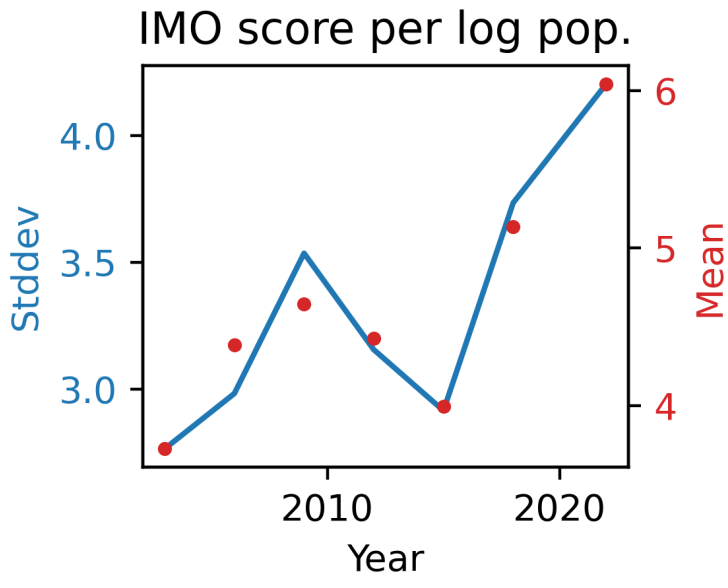
Omitted Variable Bias

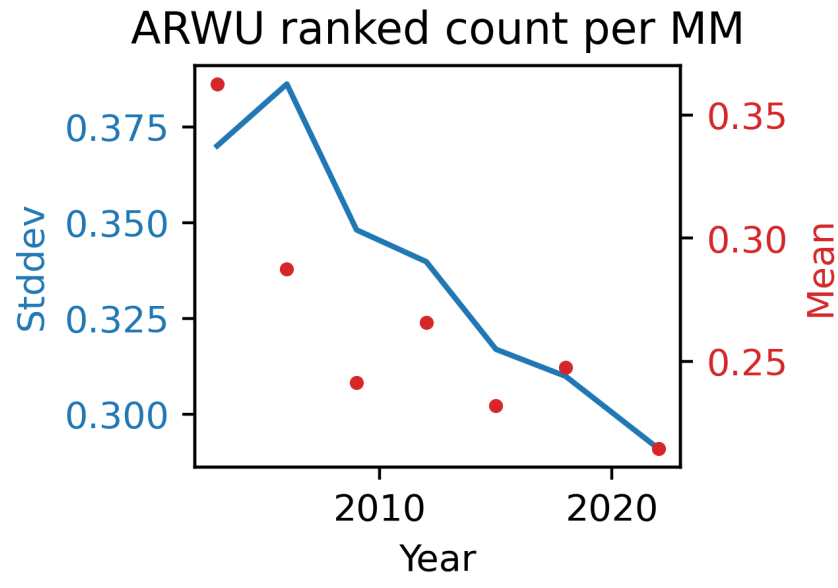
Large changes in relationship magnitude and direction suggests that omitted variable bias is a big problem.

Appendix: convergence and divergence in elite indicators

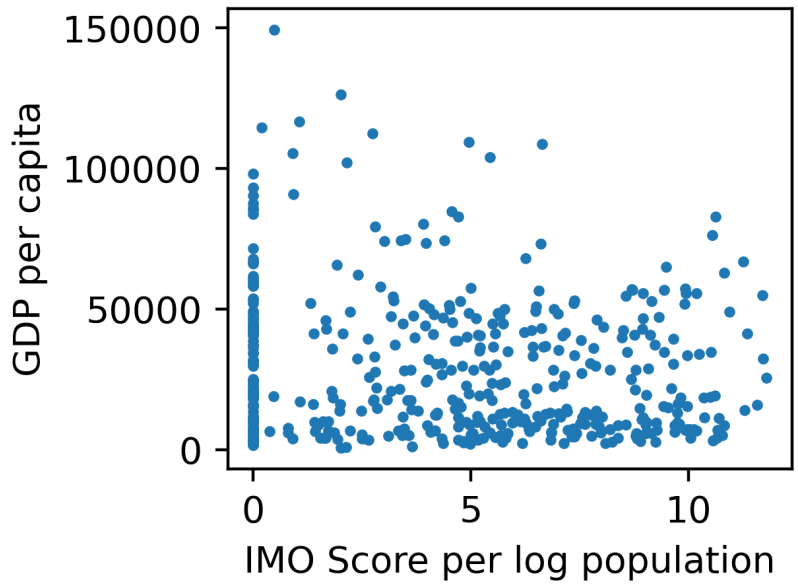


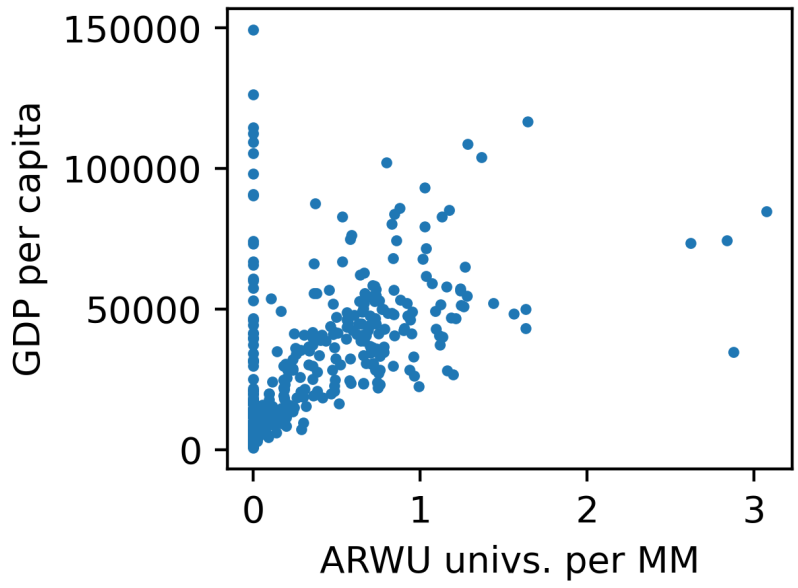


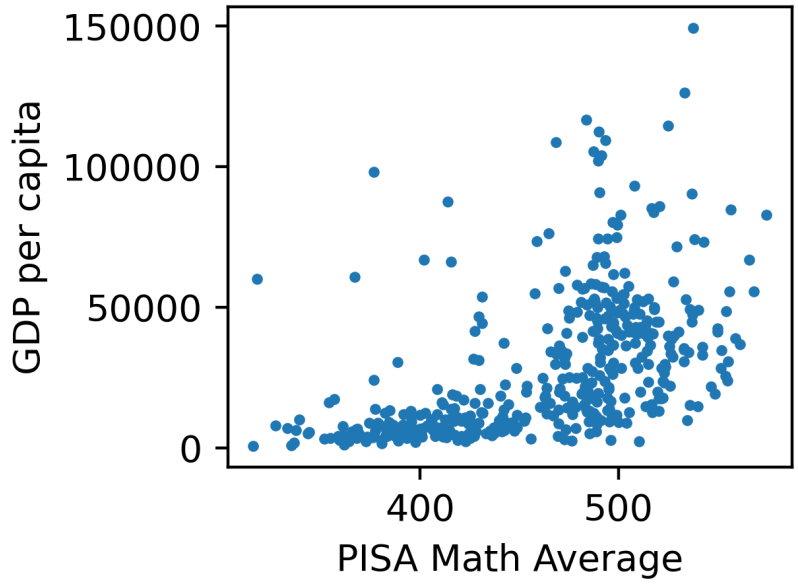


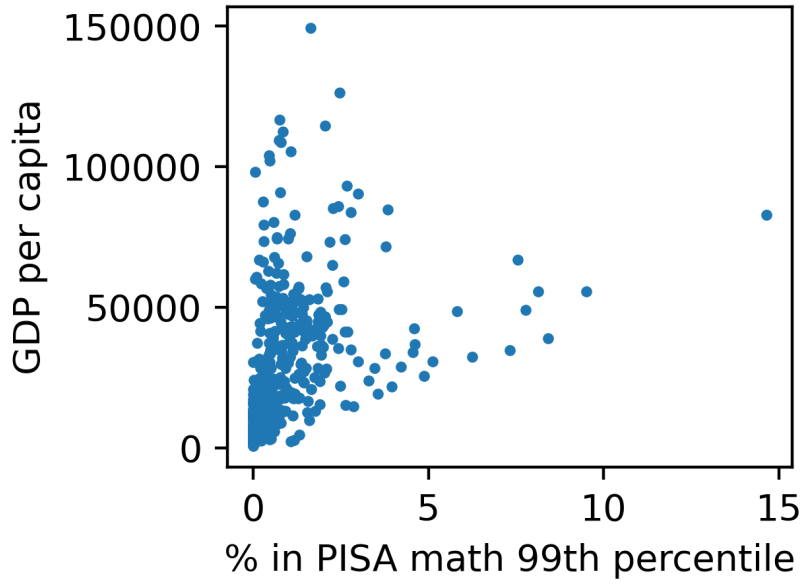


Appendix: elite indicators and GDP per capita









Appendix: Summary Statistics

Table: Summary Statistics

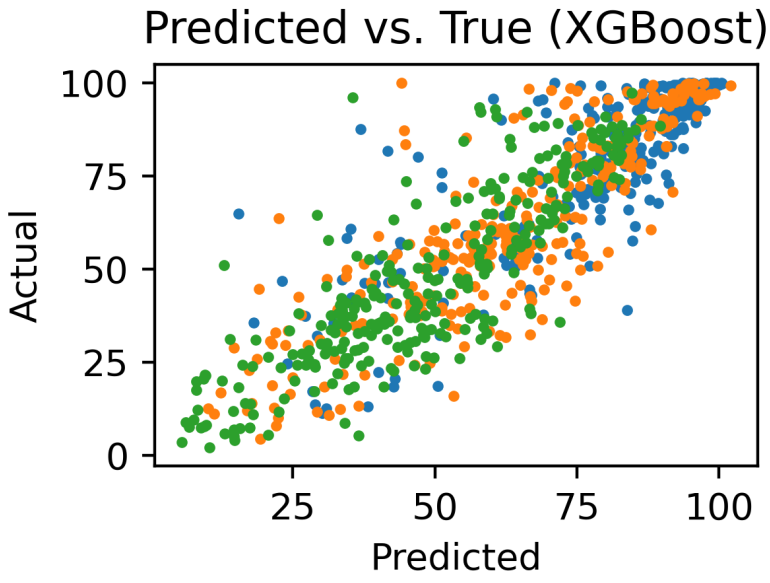
	Count	Mean	Std	Min	Max
GDP per capita	441.00	28463.99	25255.10	543.11	149461.79
GDP per capita growth (bps)	440.00	169.06	452.68	-2292.68	3303.05
EIU Democracy Index	441.00	7.25	1.77	1.93	9.93
Primary Completion	441.00	90.00	10.25	51.35	101.95
Lower Sec. Completion	441.00	77.89	17.14	29.21	101.97
Upper Sec. Completion	441.00	62.20	18.48	18.44	97.40
Population	441.00	35318600.19	59445620.16	34000.00	333287557.00
ARWU per million pop ranked	441.00	0.26	0.33	0.00	1.54
IMO score per log pop	441.00	4.72	3.49	0.00	11.79
PISA math in global 1%	441.00	0.94	1.46	0.00	14.64
PISA math	441.00	461.93	56.23	315.96	574.66

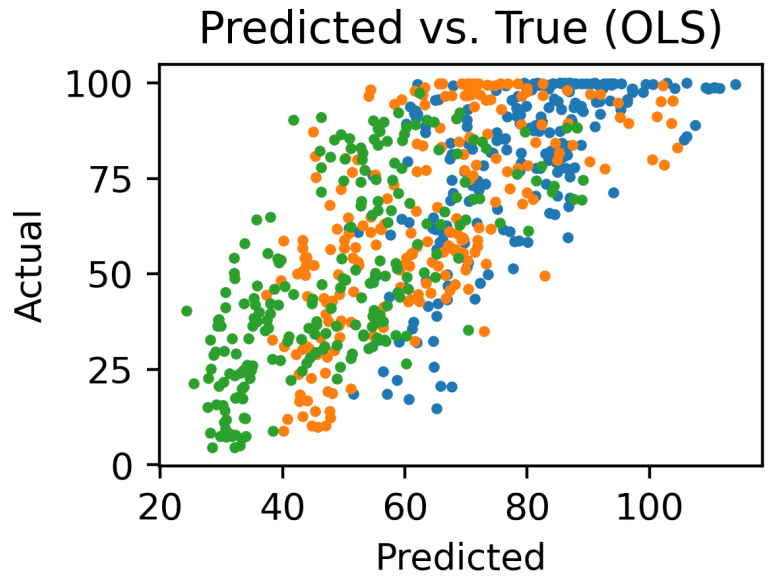
Table: Pre-imputation Statistics

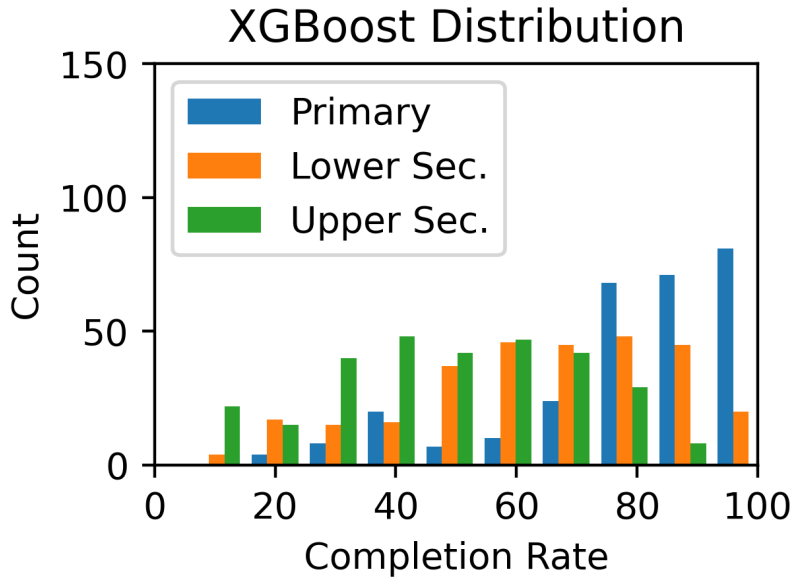
	Count	Mean	Std	Min	Max
EIU Democracy Index	323.00	7.09	1.78	1.93	9.93
Primary Completion	157.00	90.29	10.31	51.35	100.00
Lower Sec. Completion	179.00	76.63	20.03	29.21	99.98
Upper Sec. Completion	198.00	62.32	20.58	18.44	97.40

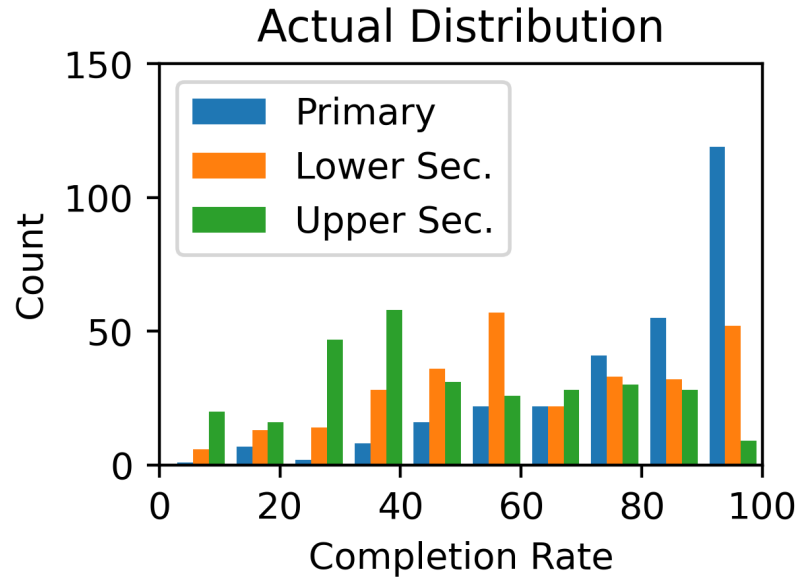
Appendix: Imputation results

School completion rates









Appendix: Imputation results

EIU Democracy Score