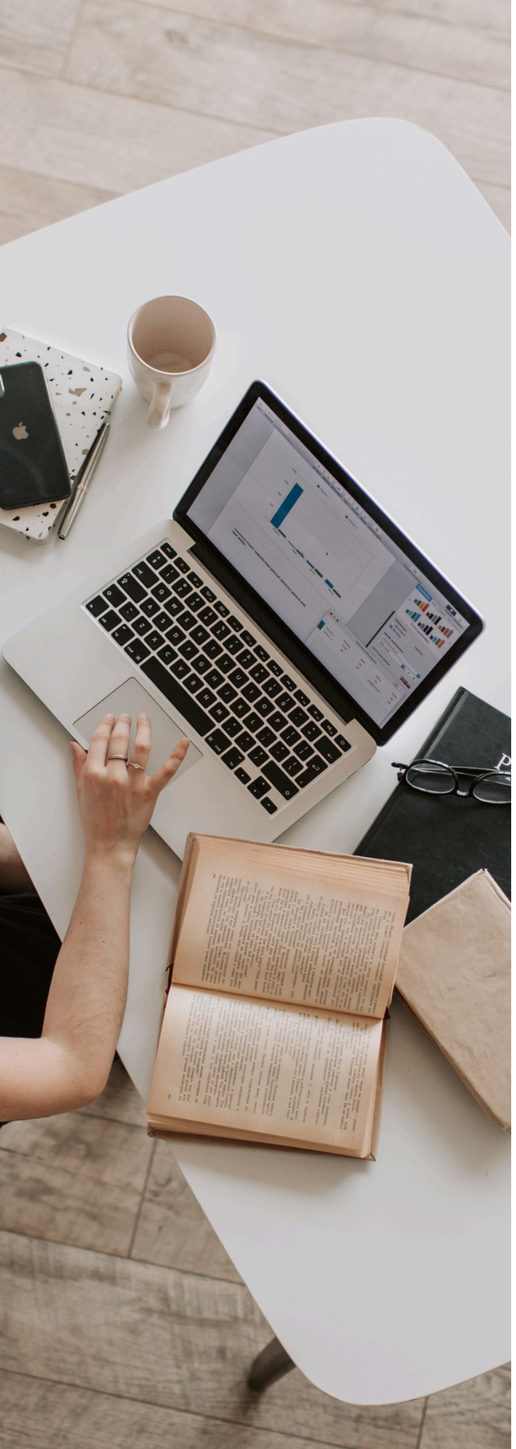
Reproducible Research

The colonial origins of comparative development: an empirical investigation

Alexey Demchuk 447348 Filip Ferens 420872 Dominika Dunda 422722



Main goal of the analysis and subject chosen

- Reproduction and enhancement of a data analysis and regression originally conducted using Python and Pandas.
- Source – analysis of article published on QuantEcon.
- Authors wanted to determine whether differences in institutions can help to explain observed economic outcomes.
    - economic outcomes - proxied by log GDP per capita in 1995, adjusted for exchange rates.
    - institutional differences are proxied by an index of protection against expropriation on average over 1985-95, constructed by the Political Risk Services Group.

How will we proceed?

What tools will be used?

- The intention was to translate the analysis into R and try to replicate and then compare the results to those obtained in the paper.

- To prepare the analysis we have used RStudio and GitHub.

- The final report is prepared in R markdown.

Packages used

- Data Import and Manipulation:
  haven, dplyr
- Data Visualization:
  ggplot2, ggrepel
- Regression Analysis:
  lmtest, AER, stargazer, broom
- Report and Presentation:
  knitr

Data Overview

Source: QuantEcon repository

Key Variables:

- avexpr (Average protection against expropriation risk)
- logpgp95 (Logarithm of GDP per capita in 1995)

Initial dataset inspection and summary statistics to identify missing values.

Data Preparation

Omitted rows with missing values in key variables (avexpr and logpgp95).

Ensured the dataset is ready for regression analysis by confirming the absence of missing values in critical variables.

Simple Linear Regression Model

- Relationship between avexpr and logpgp95.
- Positive correlation between protection against expropriation and GDP per capita.

Coefficient	Results_in_R	Results_in_Python
The intercept	4.626	4.6261
The slope	0.532	0.5319
R-squared	0.611	0.611
F-statistic	171.438	171.4

Multivariate Regression Model

- Extended model to include additional explanatory variables (latitude, continent dummies).
- Consistent findings between R and Python implementations.

Coefficient	Model 1 in R	Model 1 in Python	Model 3 in R	Model 3 in Python	Model 4 in R	Model 4 in Python
The intercept	4.63	4.63	4.87	4.87	5.85	5.85
Average Expropriation Risk	0.53	0.53	0.46	0.46	0.39	0.39
Latitude			0.87	0.87	0.33	0.33
Asia					-0.15	-0.15
Africa					-0.92	-0.92
Other					0.30	0.30
R-squared	0.61	0.61	0.62	0.62	0.72	0.72

Endogeneity and 2SLS Regression

- Addressing endogeneity using Two-Stage Least Squares (2SLS).
- Instrument: Settler mortality rates.

First stage

Coefficient	Results_in_R	Results_in_Python
The intercept	9.34	9.3414
logem4(i)	-0.61	-0.6068
R-squared	0.27	0.270
F-statistic	22.95	22.95

Second stage

Coefficient	Results_in_R	Results_in_Python
The intercept	1.91	1.9097
avexpr(i)	0.94	0.9443
R-squared	0.19	0.1870

Summary of Findings

- OLS regression suggests a positive correlation between institutional quality and economic outcomes.
- 2SLS regression confirms the relationship, addressing endogeneity concerns.
- Institutional differences, rooted in colonial history, significantly impact current income levels across countries.
- Robustness of findings validated through consistent results in both R and Python.

Conclusions

- The pivotal role of institutions in economic development.
- Importance of addressing endogeneity in regression analysis.
- Consistency of results across different software environments (R and Python).
- Validation of Acemoglu, Johnson, and Robinson's findings on the impact of colonial institutions on economic outcomes.

References

- https://python.quantecon.org/ols.html
- Daron Acemoglu, Simon Johnson, and James A Robinson. The colonial origins of comparative development: an empirical investigation. The American Economic Review, 91(5):1369–1401, 2001.

Thank you for attention!