A Template for Reproducible Empirical Research*

Joachim Gassen

TRR 266 "Accounting for Transparency"

Humboldt-Universität zu Berlin

May 10, 2021

Abstract

This is a paper stub generated out of a template repository. It repliactes the association of national income per capita with life expectancy that has become known as the 'Preston Curve.'

*I acknowledge financial support from Deutsche Forschungsgemeinschaft-Project-ID 403041268 - TRR 266.

1 Introduction

The scientific community widely agrees that reproduction, defined as reproducing findings using the same sample and method, and replication, defined as retesting findings using a new sample while applying the same method, of empirical results are a central building block of sound empirical evidence. Yet, although data and code repositories are on the rise, the additional materials only rarely lead to studies being directly reproducible. While several reasons for this observation have been discussed by prior literature, most commentators agree that the lack of established reproducible workflows makes reproductions and replications inherently costly to do (Gertler, Galiani, and Romero (2018)).

On the GitHub repository of the TRR 266 "Accounting for Transparency," we host and maintain a repository (https://github.com/trer) that implements a bare-bones reproducible workflow. This paper stub is included in the repo and can be reproduced in its entirely by running the code included in the repository.

2 The Preston Curve as a Case Study

To demonstrate the workflow, I replicate a well-known association in global health and development economics: the positive association of national income per capita as assessed by country-level real Gross Domestic Product (GDP) per capita with country-average life expectancy at birth. This association has become known as the "Preston Curve" (Preston (1975)), as reproduced in Figure 1.

[Figure 1 about here.]

While it is unclear whether the association is indicative for a causal link of national income on life expectancy, it is often referred to in public debates as one of the main arguments why economic growth is socially desirable. The analysis presented here is only meant to demonstrate the workflow of the repository and not to contribute to the underlying question of causality.

The data for the analysis has been collected from the Open Data repository of the World Bank (https://data.worldbank.org). The sample is limited to country-year observations that have non-missing values for all variables and I use the natural logarithm of GDP per capita as my

measure for national income per capita to account for the underlying growth process that shapes national income. The descriptive statistics of the data are presented in Table 1.

Table 2 reports the correlations between dependent and independent variables. As expected, it features a strong positive correlation between national income and life expectancy. Figure 2 visualizes this association.

To verify that this association is robust to including controls and general cross-sectional as well as time variance, I estimate the respective coefficient for GDP per capita in a multiple regression framework. Using a level-log multiple regression setup, Table 3 yields a set of model estimates. As you can see, the association between national income and life expectancy is significant at conventional levels across all specifications. The magnitude of the association, however, is significantly reduced once one controls for unobserved cross-sectional variation by including country fixed effects.

3 Conclusion

The Preston curve is alive and well. While being simplistic, the presented analysis contains all main building blocks of a typical empirical study based on observational data: data collection, sample specification, variable construction, exploratory as well as confirmatory analysis and result presentation. Have fun using this template for your own projects and consider contributing to its development, e.g., by providing code for Software packages besides R.

[Figure 2 about here.]

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

References

Gertler, Paul, Sebastian Galiani, and Mauricio Romero. 2018. "How to Make Replication the Norm." *Nature* 554: 417–19. https://doi.org/10.1038/d41586-018-02108-9.

Preston, Samuel H. 1975. "The Changing Relation Between Mortality and Level of Economic Development." *Population Studies* 29 (2): 231–48. https://doi.org/10.1080/00324728.1975.10410201.

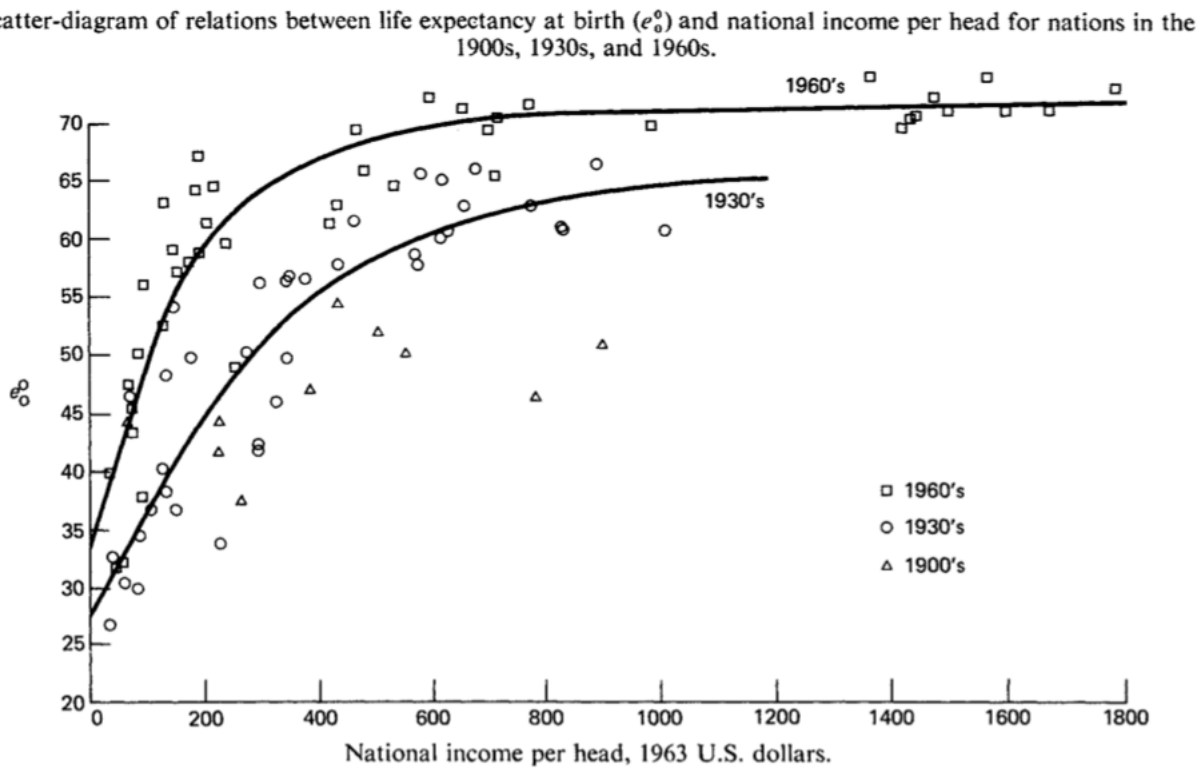

Figure 1: Preston Curve (Preston (1975): 235)

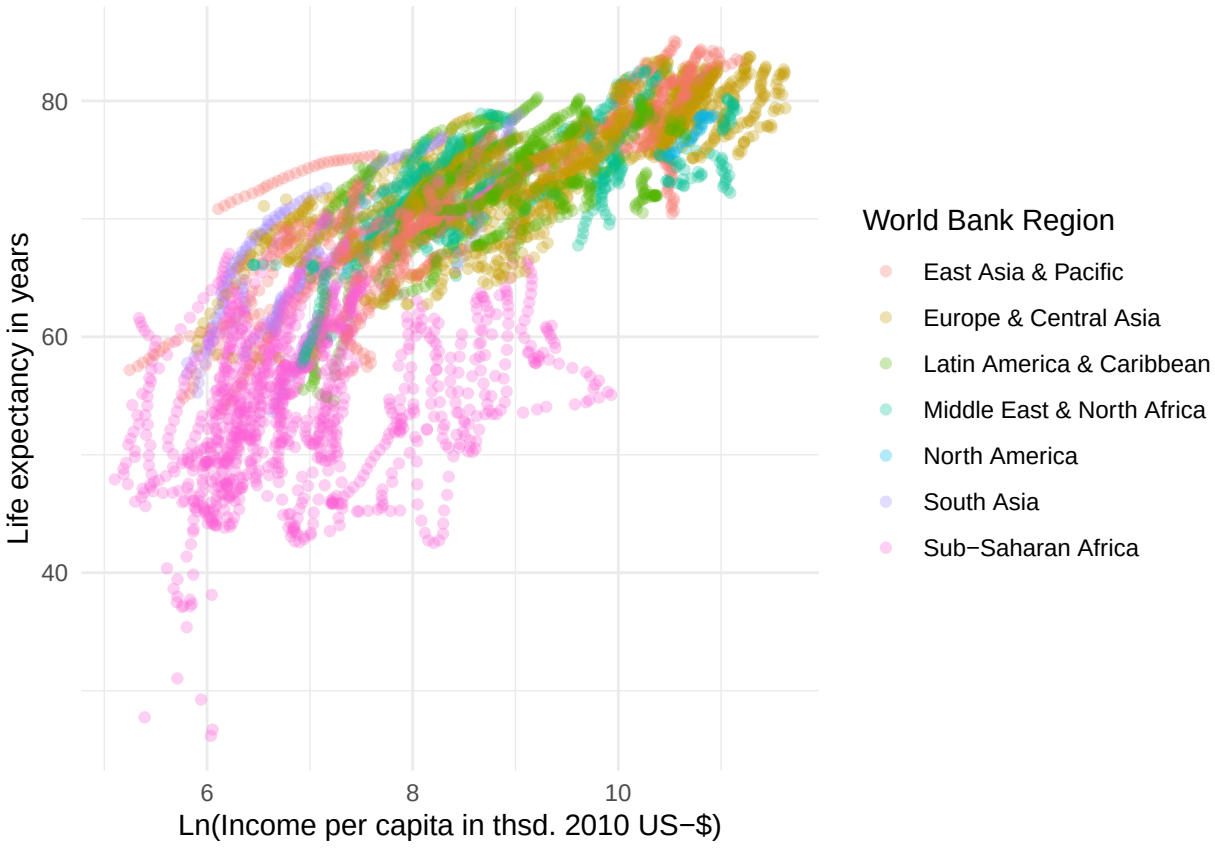

Figure 2: National Income and Life Expectancy

Table 1: Descriptive Statistics

	N	Mean	Std. dev.	Min.	25 %	Median	75 %	Max.
National Income per Capita	5,054	12,243.967	17,525.796	164.337	1,306.658	4,217.818	15,157.499	111,968.349
Unemployment (in %)	5,054	7.967	6.217	0.110	3.600	6.260	10.428	37.970
Life Expectancy (in years)	5,054	68.633	9.744	26.172	62.505	71.006	75.836	85.078

Note: The data is obtained from the World Bank. The sample covers 181 countries and the period 1991 to 2019. *National Income per captia* values are in constant 2010 thousand U.S. dollars.

Table 2: Correlations

	A	B	C
A: Unemployment (in %)		**0.04**	**0.09**
B: Life Expectancy (in years)	**0.12**		**0.80**
C: ln(National Income per Capita)	**0.20**	**0.85**	

This table reports Pearson correlations above and Spearman correlations below the diagonal. Number of observations: 5054. Correlations with significance levels below 5% appear in bold print.

Table 3: Regressions

	Dependent variable:			
	Life Expectancy (in years)			
	(1)	(2)	(3)	(4)
ln(National Income per Capita)	5.151***	5.171***	5.034***	1.400***
	(0.055)	(0.055)	(0.070)	(0.484)
Unemployment (in %)		−0.052***	−0.040**	0.042
		(0.013)	(0.017)	(0.042)
Constant	25.313***	25.559***		
	(0.472)	(0.475)		
Estimator	ols	ols	ols	ols
Fixed effects	None	None	year	country, year
Std. errors clustered	No	No	year	country, year
Observations	5,054	5,054	5,054	5,054
R^2	0.633	0.634	0.635	0.018
Adjusted R^2	0.633	0.634	0.633	-0.024
Note:				*p<0.1; **p<0.05; ***p<0.01