University of Oxford: MPhil in Politics

Causal Inference: Problem Set 2

1090063

Contents

1 Problem 1: Institutions and Economic Development [50 points]

The assignment is based on the famous Acemoglu, Johnson & Robinson (AJR) 2001 study on the importance of inclusive institutions for economic development. AJR argue that institutions leave a long imprint on countries' economic activity. They distinguish between inclusive and extractive institutions. The former diffuses economic returns across different strata of society, whereas the latter facilitates the appropriation of wealth by elites.

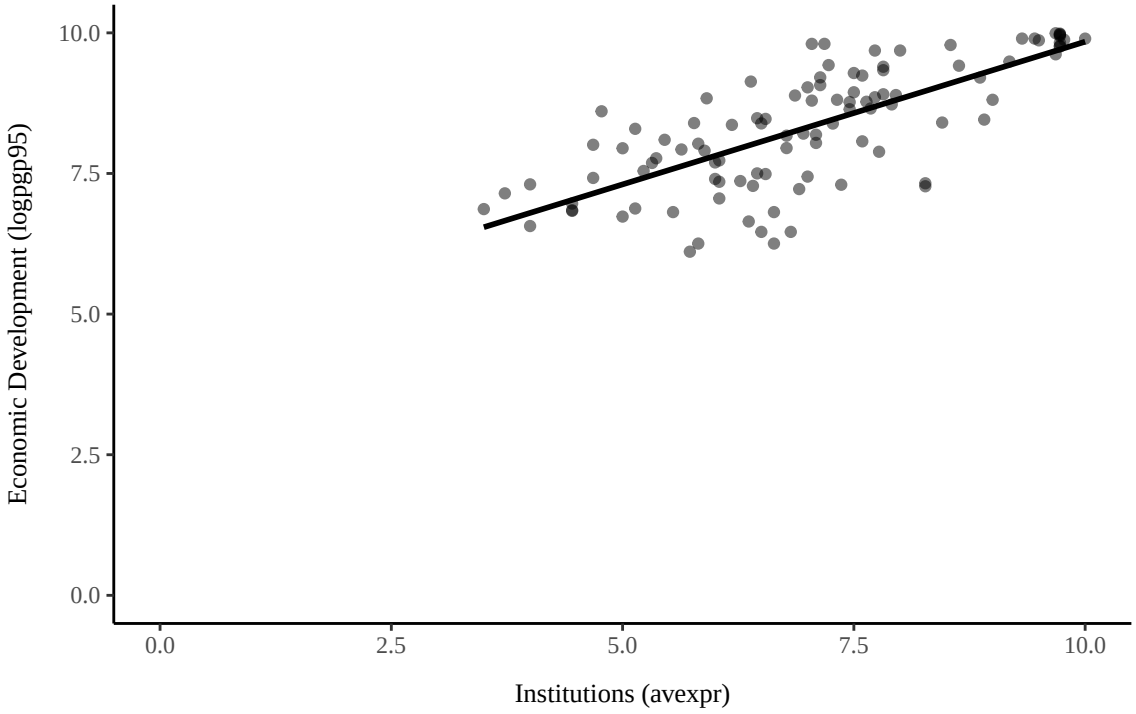
To provide evidence for the importance of institutions, they turn to the colonial structures of the 19th and early 20th centuries. Their identification strategy comes from variation in geography and climate, which determined whether colonizers would establish inclusive or extractive institutions. In those areas in which settlers encountered high mortality rates, they built extractive institutions without long-term planning. In areas with low mortality rates, they built inclusive institutions. Assuming that settler mortality rates satisfy the assumptions of an instrumental variable, this allows the authors to identify the causal effect of institutions on economic development over the long term.

1.1 Correlation

Is there a link between institutions and economic development? This is not a causal question; we are asking if there is any association between the two. Provide a scatterplot to show this is the case.

By using the AJR dataset, we can see that there is a correlation between institutions and economic development of `0.782`. The scatterplot in `Figure 1` shows the relationship between the average protection against expropriation risk between 1985-1995 `(avexpr)` and logged GDP per capita measured in 1995 `(logpgp95)`. This positive, and relatively high correlation factor shows that we would expect there to be a strong, and possibly causal relationship, between increased protection against expropriation risk (strong institutions) and economic development.

Figure 1: Correlation Scatterplot: Institutions and Economic Development



Notes: The correlation factor is calculated with the 'pairwise complete.obs' argument to handle missing values.

1.2 Causal relationship

Is this relationship causal? How do mortality rates help in answering this question?

We are interested in the relationship between institutions and economic development. This can be modelled by the regression specification shown in (1).

$$\texttt{logpgp95}_i = \beta_0 + \beta_1 \texttt{avexpr}_i + \varepsilon_i \tag{1}$$

where:

- $\texttt{logpgp95}_i$ is the log GDP per capita in 1995 for country i,
- $\texttt{avexpr}_i$ is the average protection against expropriation risk between 1985–1995 for country i,
- β_0 is the intercept,
- β_1 captures the effect of institutions on economic development,
- ε_i is the error term.

For the relationship shown in `1.1` to be causal, a number of conditions must hold. Most importantly, the independent variable $\texttt{avexpr}_i$ should not be correlated with the error term ε_i. $\texttt{avexpr}_i$ should therefore

be exogenous and isolated from any unobserved confounding variables to ensure conditional independence. However, this is unlikely to be the case. For example, cultural norms of trust and co-operation can influence the quality and strength of institutions, as may colonial and legal legacies. Moreover, the dataset does not include possible confounders such as education which should be included and controlled for in the model as education can be a determinant of both institutions and economic development. Consequently, ε_i is likely to be correlated with $\texttt{avexpr}_i$ and the model is likely to suffer from omitted variable bias, as well as from the reverse causality of richer countries affording to build better institutions. This means that the estimated coefficient $\hat{\beta}_1$ will be biased and inconsistent if an OLS regression were used, and therefore the relationship shown in `1.1` is not causal.:

$$\mathbb{E}[\hat{\beta}_{OLS}] \neq \beta_{\text{true}} \tag{2}$$

To address this issue, AJR use settler mortality rates as an instrument for institutions. The idea is that settler mortality rates are correlated with the quality of institutions, but not with the error term ε_i. This means that settler mortality rates can be used to isolate the effect of institutions on economic development, ensuring the exogeneity of the independent variable. The authors argue that settler mortality rates are a valid instrument because they are determined by geographical and climatic factors, which are not correlated with the error term. This means that settler mortality rates can be used to identify the causal effect of institutions on economic development, with the expectation being that high mortality rates are correlated with extractive institutions, and lower mortality rates result in inclusive institutions.

1.3 ITT Estimation

Estimate the ITT and interpret it.

The ITT estimates the causal effect of the treatment assignment of our instrument, `logem4` on the outcome of logged GDP per capita, `logpgp95`. The ITT estimate is `-0.564` and is statistically significant. As we have estimated a log-log model, this means that a 1% increase in settler mortality is associated with a 0.5% decrease in GDP per capita. Once we scale the ITT estimate, we find that the ITT estimate is `-0.510`. This means that a one standard deviation increase in settler mortality rates is associated with a `-0.510` standard deviation decrease in log GDP per capita in 1995 which is a moderate effect size. These two ITT estimates both suggest that higher settler mortality rates are associated with lower economic development, which is consistent with the idea that high mortality rates lead to extractive institutions and lower economic development.

1.4 LATE Estimation

Estimate the LATE, using both a Wald estimator and a 2SLS estimator. Interpret your findings.

- Interpret Wald estimator

Table 1: Randomisation Check

	Covariates
Institutions (avexpr)	0.141
	(0.098)
Asia	−0.902***
	(0.306)
Log Population Density	−0.893***
	(0.182)
Rich Countries	−0.904*
	(0.508)
Latitude (absolute)	−0.873
	(0.865)
Africa	0.259
	(0.272)
Num. Obvs	61
R-squared	0.726
Adj. R-squared	0.696

* p <0.1, ** p <0.05, *** p <0.01
Note: Standard errors are in parentheses.

- Interpret 2SLS estimator
- Compare the two estimates
- Discuss the implications of the estimates
- Review Week 4 lab to see if need to show summary table for results

1.5 IV Assumptions

Assess the plausibility of the IV assumptions in this setting. For each of the assumptions relevant first stage, monotonocity, independence, and eclusion restriction below, explain (in words) what it means substantively in the context of this study and provide a statistical test or verbal argument assessing its plausibility.