Empirical Economics (2018) Assignment 4

Suggested Solutions – Andrew Proctor

Question 1

An instrumental variable strategy is preferred when there is suspected correlation between the explanatory variable of interest and omitted variables that we cannot reasonably control for in an OLS regression. In practice, absent any other strategies like diff-in-diff etc, we typically suspect that most regression will suffer from omitted confounders.

In the case of property rights (or, more broadly, quality of government enhancing institutions) and economic development, there are a number of reasons to suspect omitted variable bias. Principally, institutions that affect the quality of government do not arise in a vacuum – they are a result of economic development, of culture, and a myriad of historical processes that may also have direct impacts on economic development. Moreover, leaving aside long-term determinants of institutions, variation in institutions in recent history seem likely to be reactions (at least in part) to economic and political pressures on the country, such as worsening economic conditions, extreme inequality, corruption, civil or geopolitical strife, etc.

Not all the confounders above may be expected to specifically affect property rights. Since property rights is very likely to correlate with other measures of institutions/quality of government, however, any endogeneity of other institutional measures will likely also bias estimates using the property rights measure. For this reason, researchers like Acemoglu, Johnson, and Robinson generally look to find an instrument that can predict exogenous variation in the quality of institutions.

Acemoglu, Johnson, and Robinson propose that a good instrument for property rights is the mortality rate that European settlers faced when settling overseas colonies. This proposed strategy is based on the claim that European colonizers had different motives for setting up different colonies. Colonies that were less hospitable environments (and typically had greater natural resources) were not settled for the purpose of creating a new home, but instead were settled for the purposes of extracting wealth from the colony. As a result, the institutional structures created in these environments provided less individual protection of property so that colonizers could more easily extract wealth. Meanwhile, colonies with milder climates were more likely settled as "New Europes," with institutions that were more protective of private interests, including protection of private property. One can hence instrument for private property protection by using the mortality rate for European settlers, a measure of how (in)hospitable the colony was.

Questions 2 and 3

Sample and Key Variables

For this analysis, I choose a sample consisting of countries which were at some point colonized by a European colonial power and that moreover have complete data for GDP, settler mortality, and property rights.

The measure of GDP I use is gross domestic product (GDP) per capita, reported in terms of constant dollars purchasing power parity (PPP). I moreover log-transform GDP per capita, so that marginal effects can be interpreted as percent changes in GDP.

For the property rights measure, I use the Heritage Foundation Property Rights score, which is a measure of the degree of protection of private property from government expropriation. This score is measured on a 0-100 scale, with a higher score representing better protection from expropriation.

The setter mortality measure is the log of the estimated European settler mortality rate per 1000 persons from Acemoglu, Johnson and Robinson (2001).

For all time-varying variables, I take observations reported for the year 2013, because this selection provides the best data coverage for property rights and GDP data among recent years. The resulting dataset corresponds to 71 countries, colonized by the Netherlands, Spain, UK, France, Portugal, or Belgium.

Instrument variables regression without controls

To explore the relationship between property rights and economic development, I perform regression analysis of log GDP per capita on property rights, without additional controls. In the IV regression, I instrument property rights with log settler mortality. The structural equation and first stage are specified as:

Model 1: Instrumental variables model without controls

$$\log(GDP)_i = \alpha + \beta_1 PropertyRights_i + u_i \quad \text{(Structural Equation)}$$

$$PropertyRights_i = \pi_0 + \pi_1 SettlerMortality_i + \epsilon_i \quad \text{(First Stage Equation)}$$

Before producing the IV estimates, I first assess the results of naïve OLS regression of GDP on property rights, treating these rights as exogenous (this is simply OLS regression of the structural equation). Under this specification, reported in column (i) of Table 1, I estimate that a 1-unit change in the property score is estimated to increase GDP per capita by 3%. Of course, given that there is no credible case for exogeneity of property rights in this model, using neither instruments nor controls, the estimate is likely biased.

Using the first stage equation, I attempt to isolate exogenous variation in property rights based on countries' historical exposure to different rates of settler mortality. I find that settler mortality is a relevant predictor of property rights – and that moreover the effect is significant enough to avoid a weak instrument problem (the F statistic for the instrument equaling 33.116, much high than the F=10 rule of thumb).

Instrumenting property rights with settler mortality produces larger estimates of the structural equation coefficient. As reported in column (iv), the estimated effect of a 1-unit increase in the property score is a 6% increase in GDP per capita, twice the effect size of OLS. The standard errors of the IV estimates are also much larger than the OLS estimates. This is expected, because the IV/2SLS regression procedure only uses predicted exogenous variation in property rights (a small share of the overall variation) to estimate the effect of property rights on GDP.

One can easily verify the IV property rights coefficient by dividing the reduced form estimate (-0.60) by the first-stage estimate (-9.85), yielding a coefficient of 0.0609. Alternately, I also compute the estimate by manual performing two-stage least squares. To do this, I first regress GDP on settler mortality and then generate the predicted outcomes. I then regress property rights on settler mortality and generate the predicted outcomes for this regression as well. Finally, I regress the predicted log GDP on the predicted property rights.1 Using this method, I get the same results as the Stata IV estimator, although the standard errors are several orders of magnitude smaller. This is because the manual 2SLS estimates have not been adjusted to account for the fact that GDP per capita and property rights are no

[1] It is important to run these regressions in Stata only for complete observations (no missing data) for GDP, property rights, and settler mortality. One could also use GDP instead of the predicted GDP from the reduced form equation, since the predicted property rights is a linear function of settler mortality and the residual of GDP is by construction uncorrelated with settler mortality.

long fixed (nonstochastic), but are instead estimates in this framework (with corresponding sample variance).

Possible Violations of the Exclusions Restriction

While the IV regression results seem more plausible than plain OLS, an important question is whether the exclusion restriction holds. The exclusion restriction states that the instrument should be uncorrelated with the dependent variable except through its correlation with the endogenous explanatory variable. It seems likely that this result does not hold, since the mortality rates are likely associated with long-term aspects of the disease and climate environment. To the extent that the disease and climate environment is persistent to contemporary times and has a direct effect on economic performance, then failing to control for these direct effects will bias the IV regression estimates.

Moreover, it is possible that mortality rates correlate with economic performance other than through direct contemporary effects of environment, either because of historical chance or via indirect environmental effects. One possible correlate is with measures of civic "fractionalization" – the degree to which society is fragmented into different groups, for instance by language, ethnicity, or religion. Ethnic fractionalization is notably correlated with geographical variables, such as continent and latitude, which also correlate with disease and climate environment. This may be the result of historical happenstance, but one might also expect that historically, the optimal size of communities and degree of contact between communities could be affected by the environment.

One might also think that fractionalization could vary with settler mortality in more direct ways. For instance, it is possible that societies tend to become more internally homogenous (ie less fractionalized) in response to increased contact or pressure from foreign societies. At the same time, settler mortality rates were often principally driven by a lack of acquired immunities to local diseases by European settlers. To the extent that a legacy of contact between Europeans and native communities offered opportunities to develop immunities, then a history of foreign contact may predict both settler mortality rates and fractionalization.2 To account for the possible violations of the exclusions restricted from either direct effects of the environment or correlation with civic fractionalization, in the following analysis I include these variables as two separate sets of controls.

Instrument variables regression with disease and climate controls

In the specification with disease and climate environmental controls, I include the distance from the equator (expressed as the absolute value of latitude for the country's capital), the extent of Malaria in 1994, and the contemporary presence of yellow fever. This regression model is specified as:

Model 2: Instrumental variables model with disease and climate controls

$$\log(GDP)_i = \alpha + \beta_1 PropertyRights_i + |Latitude|_i + Malaria_i + YellowFever_i + u_i \text{ (Structural Equation)}$$

$$PropertyRights_i = \pi_0 + \pi_1 SettlerMortality_i + |Latitude|_i + Malaria_i + YellowFever_i + \epsilon_i \text{ (First Stage Equation)}$$

I report the estimates for the model with disease and climate controls in Table 2. The OLS estimates of this model are similar – but smaller – than the OLS estimates of the relationship between property rights and GDP without controls. A 1-unit change in the property score is estimated to increase GDP per capita

[2] A corollary concern is that fractionalization might change as a response to European settlement, in which case fractionalization might be endogenous to the settler mortality rate. To the extent that fractionalization is an intermediate outcome of mortality rates, then one would be "overcontrolling" for part of the effect of mortality rates on property rights.

by 2%. The IV model predicts that GDP will increase by 3% for a 1-unit change in property rights. This is half of the estimated effect from IV regression without controls – suggesting (though not proving) that the exogeneity assumption does not hold without controlling for these additional correlates of settler mortality that affect GDP.

Looking at the reduced form estimates, one can gain a clearer impression of how omitting the disease and climate controls will lead to bias. From the reduced form, it is clear that malaria (but not distance from equator or yellow fever presence) has a sizeable impact on current GDP. The inclusion of the fractionalization controls moreover has a large impact on the reduced form effect of settler mortality, with the estimated affect about a third of the size in magnitude relative to the no controls estimate (-0.22 compared to -0.60). As a result, the effect of settler mortality on GDP and property rights are both overstated, due to correlation of the instrument with climate and disease controls.

Hence, without controlling for climate and disease characteristics, the exogeneity assumption does not appear to hold. If we control for these variables, then the instrument may satisfy conditional exogeneity (exogeneity conditional on the controls).

In addition to exogeneity, the second criterion for instrument validity is relevance. The F-test statistic for property rights is equal to 9.638 assuming homoskedastic standard errors and 8.384 assuming heteroskedastic standard errors. This is below the rule of thumb value of 10 in each case, although just slightly if we were to (unreasonably) assume homoscedasticity. As a result, the IV strategy suffers from a weak instruments problem.

To review, although the inclusion of climate and disease environment controls has addressed a potential violation of exogeneity induced by their omission, after controlling for these variables we now face the problem that the instrument is weak.

Instrument variables regression with fractionalization controls

To evaluate whether the proposed instrument suffers from still further correlation with other determinants of GDP, I estimate the effects of property rights on GDP using fractionalization controls. I include measures of fractionalization separately for ethnicity, language, and religion. Each measure is defined as the probability "that two randomly selected people from a given country will not belong to the same … group."[3] The regression model is specified as:

Model 2: Instrumental variables model with fractionalization controls

$$\log(GDP)_i = \alpha + \beta_1 PropertyRights_i + EthnicFract_i + LanguageFract_i + ReligFract_i + u_i \quad \text{(Structural Equation)}$$

$$PropertyRights_i = \pi_0 + \pi_1 SettlerMortality_i + ReligFract_i + LanguageFract + ReligFract + \epsilon_i \quad \text{(First Stage Equation)}$$

I report the estimates for the model with fractionalization controls in Table 3. Estimates for the property rights variable in both OLS and IV are very similar to the no controls. In the OLS estimation, a 1-unit change in the property score is estimated to increase GDP per capita by 3%. The corresponding IV estimate is (like the no controls case) 6% (using more precise estimates, the point estimate declines from 0.061 to 0.059). Given that some of the controls are significant in the first stage and reduced form (although not individually in each regression), this suggests that the correlation between the instrument and the omitted determinant of GDP (language fractionalization) is not strong.

[3] The QoG Basic Dataset 2017 Codebook, https://www.qogdata.pol.gu.se/data/qog_bas_jan17.pdf.

Finally, it is important to investigate instrument relevant and strength once again. The F-test statistic for the instrument in the case of fractionalization controls is equal to 12.390 in the case of heteroskedasticity-robust standard errors (15.357 if one assumes homeskedasticity). In each case, this is slightly above the rule of thumb value for weak instrument.

Conclusion

By performing instrumental variables analysis of the effects of property rights on economic development, it appears that OLS regression understate the effects of property rights. A 1-unit increase (on a scale of 0-100) is estimated to increase GDP by between 2-3% using OLS methods, while using IV methods, the estimated effects are instead between 3-6%. Of course, as with any instrumental variables analysis, it is important to consider whether the identifying assumptions hold. By investigating climate/disease and fractionalization controls, I find that climate and disease controls are particularly important to control for in order to assume exogeneity. Fractionalization also appears as another set of relevant controls, but their omission produces minimal differences in model estimates. While the settler mortality is a relevant instrument in all models, in the preferred specification using disease and climate controls, the instrumental variables model suffers from a weak instruments problem.

Table 1: IV Regression of GDP and Property Rights (no controls)

	(i) Log GDP (OLS)	(ii) Log GDP (Reduced Form)	(iii) Property Rights (1st Stage)	(iv) Log GDP (IV)	(v) Log GDP (Manual IV)
Property Rights	0.03^{***}			0.06^{***}	
	(0.00)			(0.01)	
Log Settler Mortality		-0.60^{***}	-9.85^{***}		
		(0.07)	(1.71)		
Predicted Property Rights from first stage					0.06^{***}
					(0.01)
Constant	7.66^{***}	11.54^{***}	83.91^{***}	6.42^{***}	6.42^{***}
	(0.15)	(0.34)	(8.76)	(0.34)	(0.30)
Observations	171	84	84	84	84

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: IV Regression of GDP and Property Rights (disease and climate controls)

	(i) Log GDP (OLS)	(ii) Log GDP (Reduced Form)	(iii) Property Rights (1st Stage)	(iv) Log GDP (IV)	(v) Log GDP (Manual IV)
Property Rights	0.02^{***}			0.03^{***}	
	(0.00)			(0.01)	
Latitude of capital (absolute value)	-0.00	0.00	0.44^{**}	-0.01	-0.01
	(0.00)	(0.01)	(0.21)	(0.01)	(0.01)
Malaria index in 1994	-1.48^{***}	-1.65^{***}	-4.68	-1.49^{***}	-1.49^{***}
	(0.20)	(0.23)	(5.94)	(0.24)	(0.26)
Yellow fever present today	-0.46^{**}	-0.02	0.47	-0.04	-0.04
	(0.19)	(0.20)	(5.82)	(0.17)	(0.20)
Log Settler Mortality		-0.22^{**}	-6.55^{***}		
		(0.09)	(2.26)		
Predicted Property Rights from first stage					0.03^{**}
					(0.01)
Constant	9.00^{***}	10.42^{***}	61.67^{***}	8.36^{***}	8.36^{***}
	(0.22)	(0.51)	(12.09)	(0.38)	(0.43)
Observations	144	79	79	79	79

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: IV Regression of GDP and Property Rights (fractionalization controls)

	(i) Log GDP (OLS)	(ii) Log GDP (Reduced Form)	(iii) Property Rights (1st Stage)	(iv) Log GDP (IV)	(v) Log GDP (Manual IV)
Property Rights	0.03***			0.06***	
	(0.00)			(0.01)	
Ethnic Fractionalization	0.02	-0.47	-24.79*	1.00	1.00
	(0.47)	(0.52)	(12.94)	(0.84)	(0.71)
Language Fractionalization	-1.11***	-1.20***	-1.12	-1.13*	-1.13***
	(0.40)	(0.40)	(8.62)	(0.64)	(0.41)
Religion Fractionalization	-0.46	0.53	25.99***	-1.00	-1.00*
	(0.32)	(0.35)	(8.05)	(0.61)	(0.56)
Log Settler Mortality		-0.41***	-7.03***		
		(0.09)	(2.00)		
Predicted Property Rights from first stage					0.06***
					(0.01)
Constant	8.51***	11.25***	73.07***	6.94***	6.94***
	(0.27)	(0.38)	(8.00)	(0.69)	(0.68)
Observations	160	81	81	81	81

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$