UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

MMEF Econometrics - Mid-term Submission

Sukanya Mukherjee, Tom Konigkramer and Xiao Jiang

April 2022

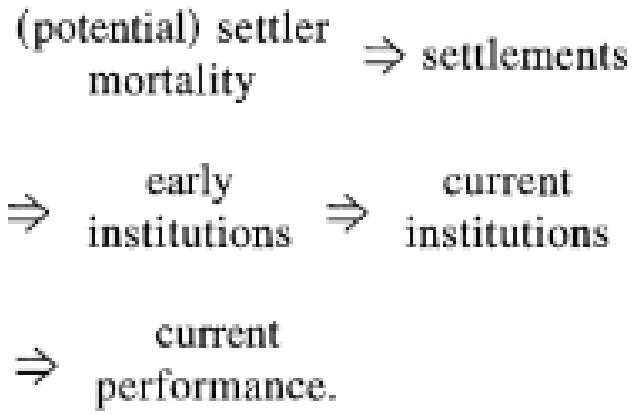
1 Introduction

With this paper, we aim to reproduce in part "The Colonial Origins of Comparative Development: An Empirical Investigation" by Daron Acemoglu, Simon Johnson and James A. Robinson (AJR 2001). The original paper explores the effects of Colonial Europe's different colonising style via the institutions it set up in the colonised countries on the colonised countries' present day GDP per capita.

It further decomposes the colonising style and in effect the institution type to the instrumental effect of settler mortality. It argues that Europeans adopted very different colonisation policies in different colonies, with different associated institutions. In places where Europeans faced high mortality rates, they

could not settle and were more likely to set up institutions that were primarily extracting resources, resulting in higher expropriation risk. These institutions persisted to the present, and continue to negatively impact growth per capita.

The theory can be summarized as below -



In our attempt, we reproduce the OLS regression on the variables of concern, and conduct various tests to check the validity of the model. We also briefly comment on the instrumentality of the settler mortality on GDP per capita.

2 Descriptive Statistics

```
. sum
```

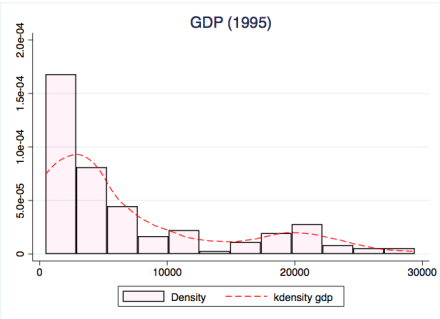
Variable	Obs	Mean	Std. dev.	Min	Max
shortnam	0				
gdp	148	7064.866	7331.979	450	29399.99
loggdp	148	8.302509	1.105342	6.109248	10.28875
mort	87	220.9264	411.4982	2.55	2940
logmort	87	4.595984	1.303333	.9360933	7.986165
alat	162	.2955502	.1903789	0	.7222222
logalat	161	-1.526488	.951462	-4.49981	-.3254224
avexpr	121	7.066491	1.804287	1.636364	10
logavexpr	121	1.917398	.2929078	.4924765	2.302585
africa	163	.3067485	.4625652	0	1
asia	163	.2576687	.4386984	0	1
neoeuro	163	.0245399	.1551948	0	1

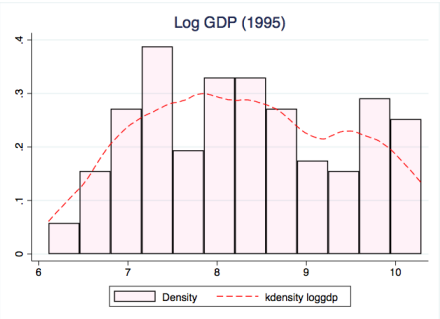
We have 148 data points, which corresponds to 148 countries. We note that the instrumental variable of interest has fewer data points, but the sample size is still sizeable even if not large.

3 Model Methodology

3.1 Variables of Interest

Our dependent variable of concern is GDP per capita.

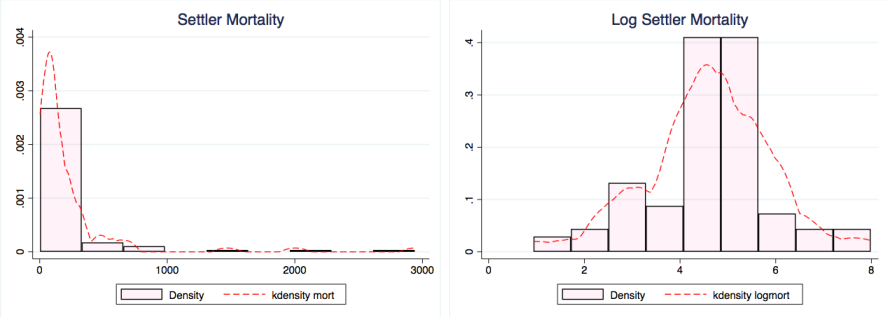




We note that the distribution is skewed, and transforming it using logarithm yields a normally distributed variable. Thus, we use log GDP per capita as our dependent variable.

Our explanatory variables are -

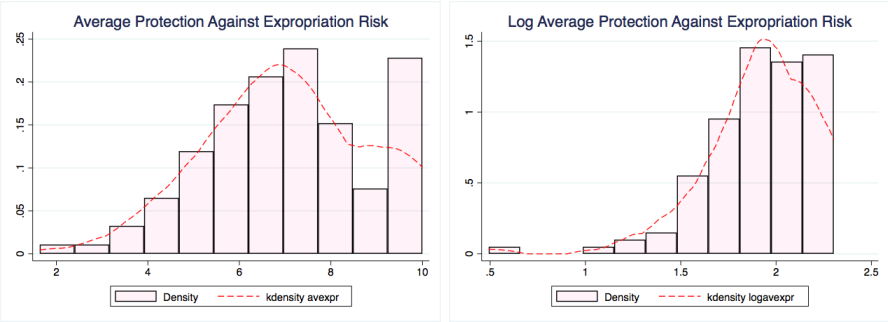
1. Settler Mortality



We note that the distribution is skewed, and transforming it using logarithm yields a normally distributed variable. Thus, we use log settler mortality as our factor variable.

We further note that this is an instrumental variable. It does not directly impact the GDP but has an indirect influence, through the institution type / protection against expropriation risk.

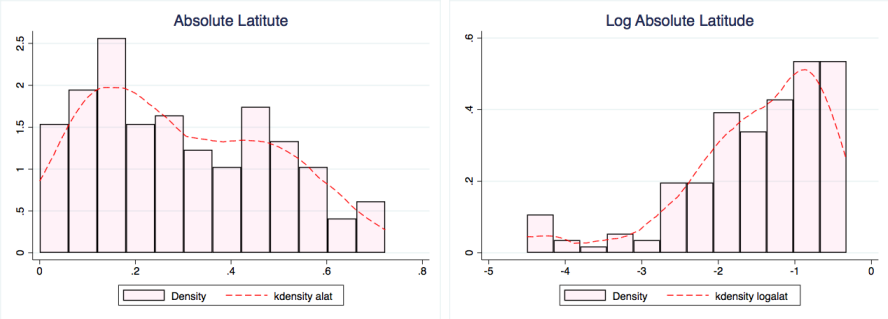
2. Average Protection against expropriation risk



We note that the distribution is slightly skewed, and transforming it using logarithm just skews it more. Thus, we use the original variable as our factor variable.

This is our main factor of interest in the OLS regression.

3. Absolute value of the latitude of the capital of the country



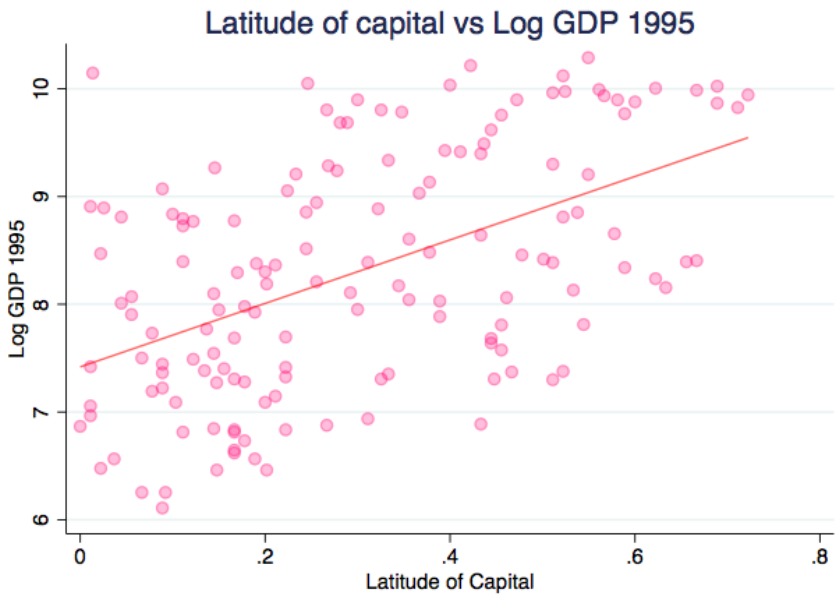
This is a measure of colonial influence on the country. This fits with our wider view of "colonial experience" as not just countries that have been direct colonies, but also countries that have had "colonial experience", that is have been strongly affected by colonialism (for instance slave trade).

We further introduce cross-sectional variables to study the geographical effects. We employ the method of dummy variables to do so. Our dummy variables include -

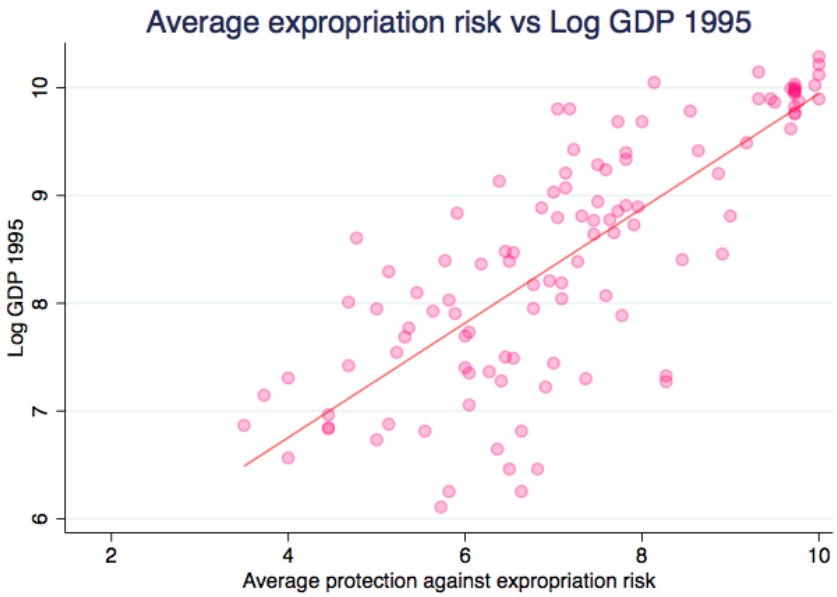
1. Africa
2. Asia
3. Neo-Euro - This categorises a group of countries - namely Australia, Canada, New Zealand and the USA - where settlers attempted to replicate European institutions. The intuition behind this grouping is that these countries would face particularly low expropriation risk, and hence have higher present-day GDP's per capita.

3.2 Relationship with the dependent variable

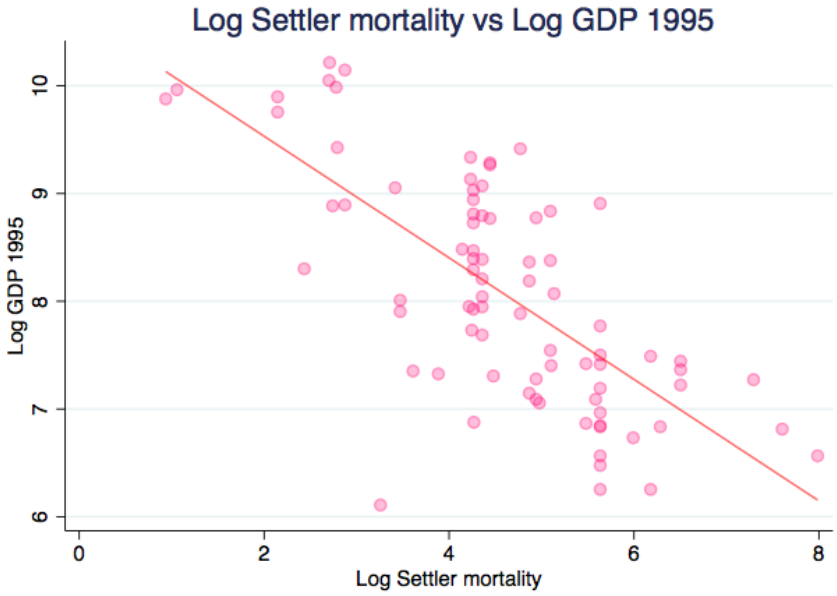
We look at the relationship between the factor variables and the dependent variable.



While there is a positive relationship, there seems to be high variance. We note that the GDP has multiple causal factors, but even in our context of colonized geographies, the colonisation type is spread across the latitude and hence, this may not be a significant factor. We check this in the regression model.



There is a positive relationship as expected in the theory proposed earlier.



There is a negative relationship as expected in the theory proposed earlier.

4 Model Estimation

4.1 First Estimates

Our linear regression model is

$$Y_i = \alpha_i * X_i + \beta_i * Z_i + \epsilon_i$$

where Y_i is the log of GDP per capita of a country i, X_i is the protection against expropriation measure, Z_i is a vector of other covariates - we group them since they are not the focus, and ϵ_i is a random error term.

The coefficient α encapsulates the effect of institutions on GDP per capita respectively.

4.2 Regression Tables

Regression1: With all variables

```
reg loggdp avexpr alat africa asia neoeuro
```

Source	SS	df	MS	Number of obs	=	111
				F(5, 105)	=	52.49
Model	103.238691	5	20.6477381	Prob > F	=	0.0000
Residual	41.3045388	105	.39337656	R-squared	=	0.7142
				Adj R-squared	=	0.7006
Total	144.543229	110	1.31402936	Root MSE	=	.6272

loggdp	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
avexpr	.3879406	.0516789	7.51	0.000	.2854709	.4904102
alat	.3231819	.4459846	0.72	0.470	-.5611233	1.207487
africa	-.9249072	.1659692	-5.57	0.000	-1.253993	-.595821
asia	-.1585018	.1550343	-1.02	0.309	-.4659061	.1489025
neoeuro	.1839098	.3364556	0.55	0.586	-.4832195	.8510391
_cons	5.870766	.3445111	17.04	0.000	5.187664	6.553868

We infer the following -

1. F-test : Our model is globally significant.
2. R-square : Our model explains 71% of the variance.
3. Coefficients : As expected, latitude is positively correlated and a higher average protection against expropriation risk leads to higher GDP per capita.
4. P-test : All factor variables are significant in this sample except Absolute latitude, Neo-Euro and Asia.

We conduct the Fisher test to check contingent insignificance.

```
. test alat asia neoeuro

 ( 1)  alat = 0
 ( 2)  asia = 0
 ( 3)  neoeuro = 0

       F(  3,   105) =    0.78
            Prob > F =    0.5052
```

We note that the probability is higher than 5 % and hence we fail to reject the hypothesis. That is these variables are indeed insignificant. We conclude that we can remove all the insignificant variables.

Regression 2 : With Significant Variables

```
. reg loggdp avexpr africa
```

Source	SS	df	MS	Number of obs	=	111
				F(2, 108)	=	130.83
Model	102.312981	2	51.1564903	Prob > F	=	0.0000
Residual	42.2302489	108	.391020823	R-squared	=	0.7078
				Adj R-squared	=	0.7024
Total	144.543229	110	1.31402936	Root MSE	=	.62532

loggdp	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
avexpr	.4241635	.0397114	10.68	0.000	.3454487	.5028784
africa	-.8784368	.1470671	-5.97	0.000	-1.169949	-.5869243
_cons	5.655561	.3134418	18.04	0.000	5.034265	6.276857

We infer the following -

1. F-test : Our model is globally significant.
2. R-square : Our model explains 71% of the variance.
3. Coefficients : As expected, a higher average protection against expropriation risk leads to higher GDP per capita.
4. P-test : All factor variables are significant.

This is our final OLS regression model with a good-fit and significant variables.

4.3 Tests and Corrections

Testing for Heteroskedasticity

We run the Breusch Pagan test.

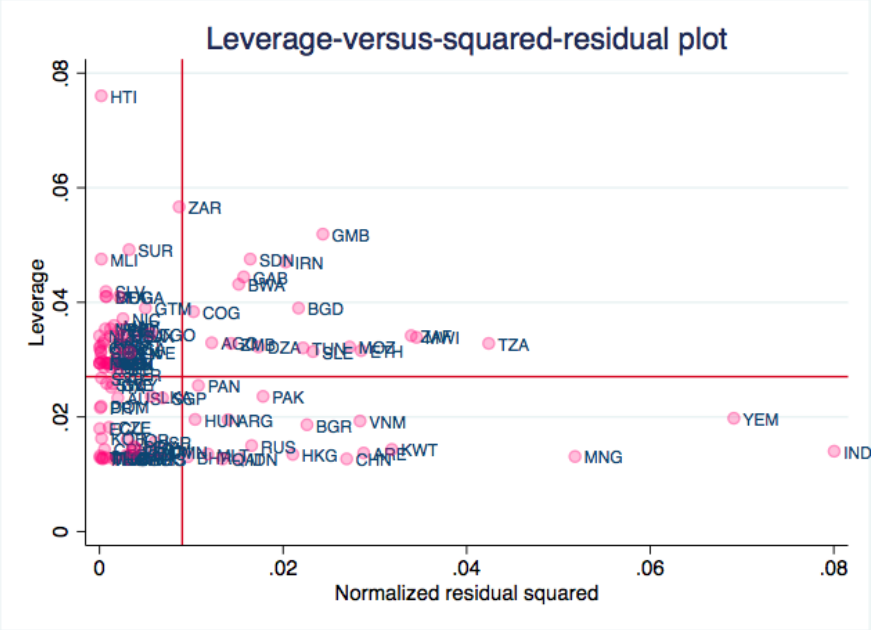
```
H0: Constant variance

    chi2(2) =    2.17
Prob > chi2 = 0.3381
```

We note that the probability is greater than 5% so we fail to reject the null hypothesis, and note that our model is homoskedastic.

Outliers and Influential points

We look at the leverage - normalized residuals chart.



We note that there are a few outliers like Mongolia and India but we note they do not have high leverage, and hence we do not remove it. Otherwise, we find a fairly tight grouping of our observations in the bottom left quadrant of our graph, which coincides with low leverage, and low normalised squared residuals.

Endogeneity

We note that our model does not include many factor variables that explain GDP growth. We are looking fundamentally at causality. However, there is an issue of endogeneity here. The higher the GDP, the better the country's economy and investment into the economy in the form of institutions amid other resources. This strengthening of the institutions increases the protection against expropriation risk.

We thus introduce the instrumental variable of Settler mortality to explain the protection against expropriation risk.

4.4 Final Model

IV Model Our final linear regression model is a 2-step linear regression model which accounts for the log of settler mortality as an instrumental variable.

Step 1 :

$$Y_i = \alpha_i * X_i + \beta_i * Z_i + \epsilon_i$$

Step 2 :

$$X_i = \gamma_i * S_i + \hat{\epsilon_i}$$

where Y_i is the log of GDP per capita of a country i, X_i is the protection against expropriation measure, S_i is the log of settler mortality, Z_i is a vector of other covariates. ϵ_i and $\hat{\epsilon_i}$ are random error terms.

The coefficients α enumerates the effects of institutions on income per capita and incorporates the β effect of the setter mortality within it.

We run the IV-regress on our final OLS regression with our settler mortality instrument of protection against expropriation risk.

IV Regress table with Africa

```
. ivregress 2sls loggdp (avexpr = logmort) africa

Instrumental variables 2SLS regression          Number of obs   =         70
                                                Wald chi2(2)    =      75.59
                                                Prob > chi2     =     0.0000
                                                R-squared       =     0.4896
                                                Root MSE        =     .75133

------------------------------------------------------------------------------
     loggdp | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
------------+-----------------------------------------------------------------
     avexpr |   .7438437   .1474711     5.04   0.000     .4548056    1.032882
     africa |  -.3732258   .2612792    -1.43   0.153    -.8853236     .138872
      _cons |   3.341892   1.057696     3.16   0.002     1.268845    5.414939
------------------------------------------------------------------------------
Instrumented: avexpr
 Instruments: africa logmort
```

We note that Africa is not significant but we do not drop it yet. We posit that there might be a bias given that we have not incorporated other regions such as Asia and Neo-europe. We run it again with these variable.

IV Regress table with all dummies

```
. ivregress 2sls loggdp (avexpr = logmort) africa asia neoeuro

Instrumental variables 2SLS regression          Number of obs   =         70
                                                Wald chi2(4)    =      77.17
                                                Prob > chi2     =     0.0000
                                                R-squared       =     0.4636
                                                Root MSE        =     .77026

------------------------------------------------------------------------------
     loggdp | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
------------+-----------------------------------------------------------------
     avexpr |   .8662774   .2204763     3.93   0.000     .4341519    1.298403
     africa |  -.6080968    .266889    -2.28   0.023    -1.13119    -.085004
       asia |  -.9619332   .3187829    -3.02   0.003    -1.586736   -.3371302
    neoeuro |  -1.290321   .7924798    -1.63   0.103    -2.843552    .2629112
      _cons |   2.857413   1.468066     1.95   0.052    -.0199424    5.734769
------------------------------------------------------------------------------
Instrumented: avexpr
 Instruments: africa asia neoeuro logmort
```

Neo-europe is insignificant now, so we drop it.

```
. ivregress 2sls loggdp (avexpr = logmort) africa asia

Instrumental variables 2SLS regression          Number of obs   =         70
                                                Wald chi2(3)    =      95.44
                                                Prob > chi2     =     0.0000
                                                R-squared       =     0.5851
                                                Root MSE        =     .67747

------------------------------------------------------------------------------
      loggdp | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      avexpr |   .7126764   .1304441     5.46   0.000     .4570107    .9683421
      africa |  -.6174934   .2326284    -2.65   0.008    -1.073437   -.1615501
        asia |  -.7351747   .2359481    -3.12   0.002    -1.197624    -.272725
       _cons |   3.769281   .9238343     4.08   0.000     1.958599    5.579963
------------------------------------------------------------------------------
Instrumented: avexpr
 Instruments: africa asia logmort
```

Diagnostics of the IV Procedure

We now run the Wu Hausman test -

```
. estat endogenous

  Tests of endogeneity
  H0: Variables are exogenous

  Durbin (score) chi2(1)          =  8.60938  (p = 0.0033)
  Wu-Hausman F(1,65)              =  9.11555  (p = 0.0036)
```

Our p-value is lower than 5 % so we reject the null hypothesis that the variables are exogenous. Our IV regress is a better fit than the OLS.

5 Conclusion

Our models have shown that the theoretical hypothesis presented at the beginning of the paper holds true. Indeed the colonised countries' GDP per capita continue to be affected via the colonisation styles of the European settlers decades, if not centuries in some cases, ago. The detrimental effects are more pronounced in some geographies than others. The depression of GDP per capita is seen via the channels of institutions and their role in furthering expropriation risks. Many economists and social scientists believe that differences in institutions and state policies are at the root of large differences in income per capita across countries.

Our argument rests on the following premises:
(1) Europeans adopted varied colonisation strategies, with different associated institutions. On one hand,as in the case of Neo-Europe,they went and settled in

the colonies and set up institutions that enforced the rule of law and encouraged investment.
On the other hand, as in Africa, they set up extractive states with the intention of transferring resources rapidly to the metropole. These institutions were detrimental to investment and economic progress.
(2) The colonisation strategy was in part determined by the feasibility of European settlement.In places where Europeans faced very high mortality rates, they could not go and settle, and they were more likely to set up extractive states.
(3) Finally, we argue that these early institutions persisted to the present. Determinants of whether Europeans could go and settle in the colonies, therefore, have an important effect on institutions today. We exploit these differences as a source of exogenous variation to estimate the impact of institutions on economic performance.

It is useful to point out that the findings do not imply that institutions today are predetermined by colonial policies and cannot be changed. An immediate policy implication to propel GDP growth is thus to invest in strengthening the institutions of the country, and one way to do that is particularly by empowering them against extractive practices.

6 Appendix

Stata Code

For complete access to the Stata code and output, see our GitHub Repository. The supplementary zipped file contains the same file structure and contents.