Settler Mortality Rate as an Instrument for Institutional Quality

Nazrul Islam
Research Professor and Head of the Quantitative Analysis Section,
ICSEAD

Working Paper Series Vol. 2004-28
October 2004

The International Centre for the Study of East Asian Development, Kitakyushu

Settler Mortality Rate as an Instrument for Institutional Quality

Nazrul Islam[1]

Research Professor and Head of the Quantitative Analysis Section, ICSEAD

Abstract

This short note demonstrates that mortality rates among early European settlers in colonies can work as an instrument for institutional quality only when entered in log form. It fails to succeed as an instrument if instead is entered in level form. This calls into question the theory regarding the relationship between settler mortality rate and institutional quality that underpins the use of the former as instrument for the latter. (*JEL Classification:* O1, O4, N9; *Keywords:* Economic Growth, Development, Colonial heritage.)

In their recent *AER* paper, "Colonial Origin of Comparative Development: An Empirical Investigation," Acemoglu, Johnson, and Robinson (2001) (henceforth AJR) present 'Early European settler mortality rate' as an instrument for 'institutional quality' (IQ). They measure institutional quality by a variable named 'RISK,' which represents 'the average protection against expropriation risk during 1985-1995.' RISK is therefore endogenous in a regression with per capita income of 1995 as the dependent variable, and hence it is necessary to find an instrument. Using European settler mortality rate as the instrument, AJR show that institutional quality is the most important determinant of per capita income level across countries.

The purpose of this Comment is to draw attention to the fact that European settler mortality rate works as an instrument only if it is entered in *log* form. It does not work if it is instead entered in *level* form. While this does not damage AJR's regression results, it calls into question some aspects of the theory that underlie the instrument. For AJR, the settler mortality rate is not a casually picked instrument. A whole theory rides on its back.[2] AJR correctly note that there is nothing in this theory that suggests that the postulated relationship between institutional quality and settler mortality rate should hold only if the latter is taken in log form.[3] The fact that the relationship does not hold when the settler mortality rate is taken in level form may therefore pose some question for this theory.

---

[1] I would like to thank Joy Mazumdar for helpful discussions. All remaining errors are mine. Send your comments to nislam@icsead.or.jp

[2] As AJR put it, "In this paper, we propose a theory of institutional differences among countries colonized by Europeans, and exploit this theory to derive a possible source of exogenous variation." (p. 1370)

[3] As AJR note, "We use the log of the settler mortality rates, since there are no theoretical reasons to prefer the level as a determinant of institutions rather than the log…." (p. 1383) Clearly, the same reasoning may be applied to argue for the level of the mortality rate to be a determinant of institutions.

Table-1 presents a summary of the results that form the basis of this Comment. To save space, it shows only the estimated values of the RISK coefficient. However, column (1) provides the list of other right hand side variables so that the specification of the regression is clear in each case.[4] Column (2) provides information regarding the sample. Column (3) reproduces the results *reported* in AJR. Column (4) shows the results obtained when AJR estimation is replicated as it is (i.e., using *log* of settler mortality rate as instrument). This is done only to check whether we are using the same data and sample. The results in column (3) and (4) should ideally be the same, though in some cases they differ, as we note later. Column (5) shows the results when the specifications are re-estimated with mortality rate *level* as the instrument.

The first few specifications are from AJR Table-4, which presents their basic regressions. Specification 1 (shown in row 1)[5] is the bivariate regression and specification 2 includes *Latitude* as the only other right hand side variable. Comparing the results in column (5) with those in either (3) or (4),[6] we see that re-estimation of these two specifications with mortality rate level as the instrument leads to considerable changes in the coefficient values and their standard errors. However, the RISK variable remains significant, though at weaker significance levels. The situation changes with slightly more extended specifications. Row 3 shows the regression that includes the continent dummies. As we see from column (5), the RISK coefficient becomes insignificant with a *p*-value of .31, when estimated with mortality rate level as instrument. Row 4 shows that when *Latitude* is added to this regression, the *p*-value of the RISK coefficient further deteriorates to .40.

AJR's Table-5 presents regressions with additional controls. The results from re-estimation of these specifications are in the next few rows of Table-1. Row 5 shows the specification with *Latitude* and *Colonizer* dummies as additional right hand side variables.[7] As we can see from column (5), the *p*-value for the RISK coefficient equals .17 when estimated with mortality rate level as the instrument. This may be compared with the *p*-value of .01 (see column (4)) when the RISK coefficient is estimated with log as the instrument. With *Latitude* and *Predominant Religion* dummies as regressors, the magnitude of the RISK coefficient proves similar from estimation using alternative forms of the instrument. However,

[4] Detailed results on other coefficients of the second-stage regressions and on the first stage regressions are available from the author upon request.

[5] 'Specification' and 'Row' have the same number, and we use them interchangeably.

[6] For these specifications the results in column (3) and (4) are almost identical.

[7] The results in column (3) and (4) for this specification differ from each other somewhat, perhaps because of some difference in variable construction.

as column (5) of row 6 shows, the RISK coefficient again proves insignificant when estimated using mortality rate level as the instrument (with a *p*-value of .16). Row 7 shows the results when *Ethno-linguistic Fractionalization* variable is included in the regression instead of *Religion*. Despite the closeness of magnitude, the RISK coefficient again proves insignificant, judged by conventional levels, when estimated with mortality rate level as the instrument.

In the regressions of their Table-7, AJR introduce another set of controls. These include *Life expectancy*, *Infant mortality*, and exposure to *Malaria*. Re-estimation results of these specifications with mortality rate level as the instrument are reported in the next few rows of Table-1. Rows 8 and 9 show that with *Malaria* included in the specifications, the RISK coefficient again proves insignificant when estimated with mortality level as the instrument. Things prove worse when specifications include *Life expectancy*, as can be seen in rows 10 and 11. In addition to being insignificant, the risk coefficient estimates in column (5) now prove negative.

While there are no *theoretical* reasons suggesting that the mortality rate variable should be taken in level or log form, an argument may be made for the latter on the *statistical* ground of undue influence of outliers in estimation.[8] AJR indeed mention this as an argument for using log. In order to check this argument, we next re-estimate the basic AJR specifications leaving out the outliers. As can be seen from AJR's Appendix Table A2, Gambia, Nigeria, and Mali are the three African countries having extremely high mortality rates of 1470, 2004, and 2940, respectively. Row 12 shows what happens when these countries are excluded in estimating AJR's basic specification that includes just the *Continent* dummies. The RISK coefficient again proves insignificant when mortality rate level is used as the instrument, even though it proves highly significant when estimated with log of mortality rate as the instrument. When in addition *Latitude* is added to the specification, the RISK coefficient becomes numerically implausible in addition to being insignificant. Madagascar, Ivory Coast, Ghana, and Togo are the next set of African countries having high settler mortality rates of 536, 668, 668, and 668, respectively. When the specification with *Latitude* and *Continent* dummies is re-estimated after leaving these four countries (in addition to the previous three countries), the RISK regression coefficient obtained with mortality rate level as the instrument still remains insignificant. This shows that the outlier-argument cannot resolve the issue entirely. While exclusion of the (African) outliers from the sample may help

[8] As AJR note, "… using the log ensures that the extreme African mortality rates do not play a disproportionate role." (p. 1383)

the mortality rate level perform better as an instrument in some specifications, it does not work in most of the other specifications.

Overall, therefore it remains the case that the 'early European mortality rate' does not generally succeed as an instrument for the institutional quality variable, RISK, if it is entered in level form.[9] This failure does not damage AJR's regression results. However, it raises a question for the theory of institutional quality that AJR propounds. In particular, it poses to this theory the additional task of explaining why the relationship between institutional quality and early settler mortality rate is of such non-linear nature as to make this rate work as an instrument only when it is taken in log form.

In pondering about this issue one may of course point out that the institutional quality variable, RISK, is arbitrary too. First of all, it is subjective. Second, it does not have any natural origin or scale. Hence one may be tempted to fix the absence of a robust relationship between mortality rate level and institutional quality by opting for an alternative transformation of the RISK variable. However, that will throw the question of the relationship between institutional quality and settler mortality rate further up in the air. It therefore seems that the issue of a relationship between institutional quality and settler mortality rate is yet to be fully settled.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001), "The Colonial Origins of Comparative Development," *American Economic Review*, Vol. 91, No. 5 (December), pp. 1369-1401

Gallup, John L., Andrew D. Mellinger, and Jeffrey D. Sachs (1998), "Geography and Economic Development," National Bureau of Economic Research (Cambridge, MA)Working Paper No. 6849

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny (1999), "The Quality of Government," *Journal of Law, Economics, and Organization*, 15(1), April, pp. 222-79

McArthur, John W. and Jeffrey D. Sachs (2001), "Comment on Acemoglu, Johnson, and Robinson," NBER (Cambridge, MA) Working Paper No. 8114

[9] This is not entirely surprising given that the correlation between RISK and the level of mortality rate is much lower (equaling to -.25) compared with that between RISK and log of mortality rate (equaling to -.52).

Table-1

RISK Coefficients Obtained from 2SLS Estimation with *Log* and *Level* of 'European Settler Mortality Rate' as Instrument

No.	*Specification:* Other right hand side variables included in the second stage regression.	Sample	RISK Coefficient reported in AJR with *log* of mortality rates as IV	RISK Coefficient from re-estimation with *log* of mortality rates as IV	RSIK Coefficient from re-estimation with *level* of mortality rates as IV
	(1)	(2)	(3)	(4)	(5)
1.	None	Base (64)	.94 (.16) Table- 4(1)	.95 (.17) [.00]	1.17 (.40) [.01]
2.	*Latitude*	Base (64)	1.00 (.22) Table- 4(2)	1.00 (.22) [.00]	1.27 (.58) [.03]
3.	*Asia, Africa, 'Other Continents'*	Base (64)	.98 (.30) Table- 4(7)	.98 (.30) [.00]	1.33 (1.29) [.31]
4.	*Latitude, Asia, Africa, Other*	Base (64)	1.10 (.46) Table- 4(8)	1.11 (.46) [.02]	1.45 (1.69) [.40]
5.	*Latitude, Britain, France*	Base (64)	1.16 (.34) Table- 5(2)	1.33 (.49) [.01]	1.53 (1.10) [.17]
6.	*Latitude, Catholics, Muslim, 'Other Religions'*	Base (64)	1.00 (.25) Table- 5(8)	.99 (.25) [.00]	1.01 (.71) [.16]
7.	*Latitude, Ethnolinguistic fractionalization*	Base (64)	.79 (.17) Table- 6(8)	.79 (.18) [.00]	.83 (.50) [.10]
8.	*Malaria*	Base (62/61)	.69 (.25) Table- 7(1)	.61 (.18) [.00]	1.14 (1.70) [.50]
9.	*Latitude, Malaria*	Base (62/61)	.72 (.30) Table- 7(2)	.64 (.20) [.00]	1.03 (1.29) [.43]
10.	*Life expectancy*	Base (60/61)	.63 (.28) Table- 7(3)	.61 (.44) [.17]	-.76 (5.18) [.88]
11.	*Latitude, Life expectancy*	Base (60/61)	.68 (.34) Table- 7(4)	.63 (.60) [.29]	-1.06 (8.64) [.90]

12.	*Asia, Africa, Other Continents*	Dropping *GMB, MLI,* and *NGA* (61)		.87 (.22) [.00]	1.98 (2.62) [.45]
13.	*Latitude, Asia, Africa, Other Continents*	Dropping *GMB, MLI,* and *NGA* (61)		.96 (.34) [.01]	75.05 (7125) [.99]
14.	*Latitude, Asia, Africa, Other Continents*	Dropping *GMB, MLI, NGA,CIV, MDG, GHA,* and *TGO* (57)		.81 (.23) [.00]	1.11 (.79) [.17]

Notes to Table-1:

1. The variable definition, construction, and source are as follows: (i) *Risk* stands for institutional quality and is defined as the 'Average protection against expropriation risk 1985-1995' measured on a scale from 0 to 10, where a higher score means more protection against risk of appropriation of investment by the government. (Source: AJR, Appendix, Table A2. (ii) *Latitude* denotes how far a country is away from the equator. (Source: La Porta et al., 1999.) (iii) *Asia*, *Africa*, and *Other* are continent dummies with obvious construction. (iv) Three *Colonizers* are distinguished: *Britain*, *France*, and *Other*. (v) Religions distinguished are: *Catholics*, *Muslims*, *Protestants*, and *Other*. (Source: La Porta et al., (1999).) (vi) *Ethnolinguistic Fractionalization* denotes probability of two randomly chosen person of the country belonging to same ethnic group or speaking the same language. (Source: La Porta et al. (1999).) (vii) *Malaria* denotes percentage of the population living where falciporum malaria is endemic. (Source: Gallup and Sachs, 1998) (viii) *Life expectancy at birth* in 1995 is from McArthur and Sachs (2001). (ix) *Mortality* is the 'European settler mortality rate,' reported in AJR, Appendix Table A2. *Lmortality* is log of *Mortality*.

2. *Risk* is the dependent variable instrumented by *Lmortality* in column (3) and (4) and by *Mortality* in column (5). Such variables as *Continent* dummies, *Latitude*, *Colonizer* dummies, *Religion* dummies, *Malaria* and *Ethnolingusitic fractionalization* are exogenous by their nature or are deemed exogenous. *Life expectancy* is instrumented by its lagged values obtained from Barro and Lee (1994).

3. Estimation is conducted using 2SLS. The numbers in parentheses are standard errors and the numbers in square brackets are the corresponding *p*-values.

4. Column (3) also indicates the source from where the AJR results are taken. For example in row 1, Table-4(1) indicates that these results are from column (1) of AJR's Table-4.

5. *Sample size:* Most of the specifications are estimated using AJR's base sample of 64 countries. For specifications 1 to 7, the samples in AJR's and the current Comments' estimation are identical. For specifications 8 to 11, the sample sizes differ, but only very slightly. The number in front of "/" denotes the AJR sample size, while the one behind denotes the sample size obtained by this author. The specifications 12 to 15 again correspond exactly to AJR's base sample minus the outliers mentioned.