Supplementary Material

MASS RESETTLEMENT and POLITICAL VIOLENCE: Evidence from Rwanda

By LACHLAN McNAMEE

World Politics

doi: 10.1017/S0043887118000138

Replication data are available at:

McNamee, Lachlan 2018. "Replication Data for Mass Resettlement and Political Violence: Evidence from Rwanda." Harvard University Dataverse, V1. doi: 10.7910/DVN/VYBJRM.

Linear probability model:
The determinants of post-independence *paysannat* settlement

	(1)	(2)	(3)	(4)	(5)	(6)
Cattle per 100 persons (1949)	0.06	0.06	0.05	0.06	0.05	0.07
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Distance to border, log	−0.19	−0.22	−0.19	−0.19	−0.198	−0.20
	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Log pop. density (1949)			−0.08	0.007	−0.06	−0.005
			(0.07)	(0.06)	(0.07)	(0.06)
Tutsi percentage (1956)					−0.007	−0.008
					(0.003)	(0.004)
Observations	1,046	1,046	1,046	1,046	1,046	1,046
Territory FE	No	Yes	No	Yes	No	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes

Table 1: Linear probability model with the dependent variable a binary indicator of whether a village was settled by a *paysannat* in the post-independence period. As pop. density and cattle density are measured at the chieftaincy-level, standard errors are clustered at the chieftaincy level.

Logit: The determinants of post-independence *paysannat* settlement

	(1)	(2)	(3)	(4)	(5)	(6)
Hutu parties (1960) vote %	−5.19	−4.94	−4.91	−5.09	−4.82	−4.93
	(1.51)	(1.71)	(1.68)	(1.92)	(1.66)	(1.89)
Distance to border, log	−1.03	−1.22	−1.04	−1.22	−1.06	−1.10
	(0.25)	(0.26)	(0.27)	(0.34)	(0.27)	(0.35)
Log pop. density (1949)			−0.16	−0.09	−0.16	−0.07
			(0.44)	(0.45)	(0.42)	(0.43)
Tutsi percentage (1956)					−0.03	−1.68
					(0.03)	(0.14)
Observations	1,046	1,046	1,046	1,046	1,046	1,046
Territory FE	No	Yes	No	Yes	No	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes

Table 2: Logistic regression with the dependent variable as whether a village was settled by a *paysannat* in the post-independence period. As population density and Hutu party vote share are measured at the chieftaincy level, standard errors are clustered at the chieftaincy level.

Logit: The determinants of post-independence *paysannat* settlement

	(1)	(2)	(3)	(4)	(5)	(6)
Cattle per 100 persons (1949)	0.39	0.40	0.32	0.41	0.33	0.43
	(0.12)	(0.13)	(0.16)	(0.13)	(0.17)	(0.13)
Distance to border, log	−1.19	−1.47	−1.18	−1.47	−1.22	−1.31
	(0.27)	(0.37)	(0.26)	(0.36)	(0.28)	(0.37)
Log pop. density (1949)			−0.38	0.05	−0.32	−0.09
			(0.42)	(0.37)	(0.43)	(0.34)
Tutsi percentage (1956)					−0.04	−1.78
					(0.03)	(0.11)
Observations	1,046	1,046	1,046	1,046	1,046	1,046
Territory FE	No	Yes	No	Yes	No	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes	Yes

Table 3: Logistic regression with the dependent variable as whether a village was settled by a *paysannat* in the post-independence period. As population density and cattle density are measured at the chieftaincy level, standard errors are clustered at the chieftaincy level.

Reporting all coefficients: The effect of resettlement on number of prosecutions for participation in the Rwandan genocide

	(1)	(2)	(3)	(4)	(5)
paysannat %	1.75	0.65	0.72	0.23	0.49
	(0.86)	(0.33)	(0.26)	(0.08)	(0.19)
mean distance to town log		37.42	12.24	−10.20	22.44
		(25.51)	(31.06)	(7.40)	(25.18)
distance to border log		5.57	19.83	2.41	17.42
		(27.86)	(26.54)	(4.71)	(23.68)
mean distance to road log		−45.88	−44.12	−11.88	−32.24
		(18.13)	(17.95)	(4.03)	(16.16)
population density (1949) log		2.15	5.48	1.54	−1.54
		(35.82)	(34.58)	(7.42)	(28.79)
Hutu parties % vote (1960)		212.46	−171.04	−26.82	−144.22
		(166.41)	(151.73)	(37.34)	(124.16)
Cattle per 100 residents (1949)		0.85	-8.81	1.41	−10.22
		(18.83)	(19.91)	(4.42)	(16.69)
Tutsi % (1956)		13.42	12.29	4.17	8.11
		(3.67)	(4.08)	(0.48)	(4.37)
population log			249.52	27.15	216.12
			(54.62)	(9.64)	(47.61)
population density log			−33.66	−0.22	−35.20
			(25.27)	(5.92)	(20.34)
MRND control			−30.44	−4.58	−25.87
			(31.58)	(5.99)	(28.86)
RTLM radio reception			−1.64	−7.90	−9.53
			(64.55)	(13.25)	(60.73)
Tutsi % (1991)			109.40	10.04	99.36
			(121.63)	(17.84)	(114.20)
Hutu income			−415.79	−30.91	−384.87
			(420.29)	(77.83)	(377.14)
Hutu education			−1.99	−-0.49	−1.50
			(3.76)	(1.15)	(3.17)
Hutu literacy			7.89	1.47	6.44
			(5.78)	(1.65)	(4.74)
Observations	1,065	1,046	1,031	1,031	1,031
District FE	No	Yes	Yes	Yes	Yes
Clustered SE	Yes	Yes	Yes	Yes	Yes
Gacaca Category	1 & 2	1& 2	1& 2	1	2

Table 4: OLS regression with dependent variable as the total number of Category 1 and 2 village-level prosecutions with standard errors clustered at the district level. Replication of Main Table 4 reporting covariate coefficients.

Placebo test: The effect of resettlement on violence during the revolution (1959-62)

	(1)	(2)
paysannat %	−0.0003	−0.0003
	(0.0001)	(0.0001)
Constant	0.028	0.014
	(0.005)	(0.005)
Observations	1,065	1,065
R^2	0.004	0.057
District FE	No	Yes

Table 5: OLS with heteroskedasticity robust standard errors. The dependent variable is a binary indicator for whether a village is reported as in 'area of conflict' between 1959-62 (*lieu d'affrontement*) in Lugan (1975). OLS is employed because a logistic model cannot be fitted - no village with a *paysannat* was reported as an area of conflict

	Prosecutions RDD dropping Hutu SES controls					
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient	160.62	186.84	157.66	163.09	26.33	132.49
	(49.66)	(46.17)	(57.38)	(65.74)	(16.38)	(47.87)
Observations	1031	1031	1031	1031	1031	1031
Kernel Type	Triangular	Uniform	Triangular	Triangular	Triangular	Triangular
Polynomial	Linear	Linear	Linear	Quadratic	Linear	Linear
Bandwidth Type	MSE	MSE	CER	MSE	MSE	MSE
Eff. # of treated obs	55	48	48	56	54	56
Eff. # of untreated obs	278	254	255	286	270	279
Cluster Std. Error	Yes	Yes	Yes	Yes	Yes	Yes
Gacaca Category	1 & 2	1& 2	1& 2	1& 2	1	2

Table 6: Replication of Main Table 2 dropping controls for Hutu income, literacy and education. Geographic regression discontinuity of total number of village-level prosecutions for participation in the Rwandan genocide where the forcing variable is geodesic distance of each village to the nearest *paysannat* border. Standard errors clustered at the district level.

	Prosecutions RDD Controlling for Number of Survivors					
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient	141.66	168.50	141.16	145.82	19.65	120.66
	(43.16)	(40.90)	(48.48)	(55.42)	(14.99)	(41.76)
Observations	1030	1030	1030	1030	1030	1030
Kernel Type	Triangular	Uniform	Triangular	Triangular	Triangular	Triangular
Polynomial	Linear	Linear	Linear	Quadratic	Linear	Linear
Bandwidth Type	MSE	MSE	CER	MSE	MSE	MSE
Eff. # of treated obs	55	48	48	56	55	56
Eff. # of untreated obs	277	252	254	282	274	277
Cluster Std. Error	Yes	Yes	Yes	Yes	Yes	Yes
Gacaca Category	1 & 2	1& 2	1& 2	1& 2	1	2

Table 7: Replication of Main Table 2 controlling for the number of genocide survivors in each village. Geographic regression discontinuity of total number of village-level prosecutions for participation in the Rwandan genocide where the forcing variable is the distance of each village to the nearest *paysannat* border. Standard errors clustered at the district level.

OLS: Interaction between resettlement and population growth % on village-level prosecutions for participation in the Rwandan genocide

	(1)	(2)	(3)	(4)	(5)
paysannat, percentage	−3.15	−2.07	−1.27	0.06	−1.33
	(1.72)	(0.93)	(0.92)	(0.22)	(0.76)
pop. growth 1978-91 %	−268.77	−501.34	630.07	−369.94	1000.01
	(2273.34)	(1647.97)	(2116.39)	(343.16)	(1982.46)
paysannat % *					
pop. growth 1978-91 %	145.50	80.25	58.53	5.58	52.96
	(48.09)	(24.89)	(24.62)	(5.40)	(20.49)
Observations	1,048	1,030	1,030	1,030	1,030
Controls	No	Yes	Yes	Yes	Yes
District FE	No	Yes	Yes	Yes	Yes
Clustered FE	Yes	Yes	Yes	Yes	Yes
Gacaca Category	1 & 2	1& 2	1& 2	1	2

Table 8: Replication of Main Table 4 including a reported interaction between *paysannat* percentage and commune population growth between 1978-1991 taken from Verpoorten (2012). As in Table 4, the controls in specification (2) are pre-treatment variables and those hypothesized to shape *paysannat* settlement: log distance to border, town and road, log colonial population density (1949), vote share for the revolutionary Hutu parties (1960), Tutsi percentage (1956) and cattle per 1000 residents (1949). Specifications (3) - (5) add a battery of post-independence controls using the 1991 census: proportion Tutsi, log population, log population density, MNRD control, RTLM reception, and measures of Hutu income, literacy and education. Standard errors clustered at the district level.

	Prosecutions RDD Controlling for Distance to the Congo					
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient	158.87	185.22	159.30	160.03	23.03	134.50
	(48.58)	(44.94)	(56.42)	(65.74)	(15.35)	(45.70)
Observations	1031	1031	1031	1031	1031	1031
Kernel Type	Triangular	Uniform	Triangular	Triangular	Triangular	Triangular
Polynomial	Linear	Linear	Linear	Quadratic	Linear	Linear
Bandwidth Type	MSE	MSE	CER	MSE	MSE	MSE
Eff. # of treated obs	55	48	48	56	51	56
Eff. # of untreated obs	278	255	255	283	268	281
Cluster Std. Error	Yes	Yes	Yes	Yes	Yes	Yes
Gacaca Category	1 & 2	1& 2	1& 2	1& 2	1	2

Table 9: Replication of Main Table 2 including a control for log distance of a village to the Democratic Republic of the Congo. Geographic regression discontinuity of total number of village-level prosecutions for participation in the Rwandan genocide where the forcing variable is geodesic distance of each village to the nearest *paysannat* border. Standard errors clustered at the district level.

	RDD Estimated Tutsi Death Toll					
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient	-0.089	-0.072	-0.069	-0.085	-0.081	-0.107
	(0.04)	(0.049)	(0.053)	(0.051)	(0.051)	(0.054)
Observations	1029	1011	1011	1011	1011	1011
Kernel Type	Triangular	Uniform	Triangular	Triangular	Triangular	Triangular
Polynomial	Linear	Linear	Linear	Linear	Quadratic	Third-order
Bandwidth Type	MSE	MSE	CER	MSE	MSE	MSE
Eff. # of treated obs	64	53	46	47	65	60
Eff. # of untreated obs	341	267	246	246	347	300
Controls	No	Yes	Yes	Yes	Yes	Yes
Cluster Std. Error	Yes	Yes	Yes	Yes	Yes	Yes

Table 10: Geographic regression discontinuity of total number of estimated Tutsi death toll (proportion 0-1 killed) during the Rwandan genocide where the forcing variable is geodesic distance of each village to the nearest *paysannat* border and standard errors are clustered at the district level.

Ratio correction

In this section, I discuss why, when seeking to measure the effect of resettlement, using the total number of prosecutions controlling for the confounder of population is an improvement on using the rate of prosecutions per population as a dependent variable. Discussion of the inferential pitfalls faced when controlling for a confounder by using a ratio has a long intellectual heritage in statistics (e.g. Pearson 1897; Neyman 1952; Firebaugh and Gibbs 1985; Kronmal 1993) and this short summary is based on the comprehensive discussion in Kronmal (1993).

Suppose the true relationship between total prosecutions *Y*, total population *X* and a binary indicator of whether a village was settled by a *paysannat D* can be represented as:

$$Y = \alpha + \beta_0 X + \beta_1 D + e \tag{1}$$

Suppose that we wish to control for the effect of population by dividing total prosecutions by total population. This may be because the rate of prosecutions is more easily interpretable, more clearly controls for population, or is the proper quantity of interest. The consequent estimated model can be written as:

$$\frac{Y}{X} = \sigma + \beta_2 D + e \tag{2}$$

Which, by expanding out, can be re-written as:

$$Y = \sigma X + \beta_2 DX + e \tag{3}$$

Comparing the equations, we can see that model (3) differs from model (1) in that it drops the intercept α and first-order term $\beta_1 D$ and contains a new interaction term $\beta_2 DX$.

Thus, ratio correction in model (2) introduces two potential statistical issues. First, if α or $\beta_1 \neq 0$, then the estimates from model (2) will be biased. Second, if we use a ratio as a dependent variable without also dividing the predictor of interest D by the denominator X, then we would have included an interaction without including all the variables that comprise the interaction as predictors, which is not best statistical practice. On this point, as Firebaugh and Gibbs (1985) suggest, "If Z is controlled by division rather than by residualization, all of the other variables should be divided by Z. Should only some of the variables be divided by Z, the effect of Z is "controlled" for some variables and not for others, and a defensible interpretation of the results is difficult" (p. 721). We could implement this recommendation by replacing D in model (2) with $\frac{D}{X}$ - a binary indicator of whether a village is covered by a *paysannat* divided by total population. But, unlike for a population variable such as proportion female or proportion Tutsi, the resulting parameter has no clear interpretation and moreover cannot be estimated via a regression discontinuity design, which relies on a binary treatment.

Given these issues, Kronmal (1993) concludes that "the common practice of using ratios for either the dependent or the independent variable in regression analyses can lead to misleading inferences and rarely results in any gain" (p. 391). Hence, in general, I estimate the effect of *paysannat* using total prosecutions as the dependent variable whilst controlling for the confounder of total population. Nevertheless, using a rate seems an entirely reasonable and more easily interpretable alternative specification and so I have included RDD and OLS models using the rate of prosecutions as a dependent variable which, aside from the shift in statistical significance from the 5 to the 10 percent level in some OLS specifications, does not change the results (Tables 11 and 12 below). In the ratio RDD results, resettled villages are estimated to have an approximately 3.45 percentage point increase in the number of prosecutions per total population which is approximately identical in magnitude to the RDD estimates using total prosecutions controlling for total population.[1]

[1] The average population of a village in 1991 was 4852, and the effect of *paysannat* was previously estimated at approximately 160 prosecutions which constitutes 3.3% of the average total population of a village

	Rate of Prosecutions Regression Discontinuity at Paysannat Border					
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient	3.45	3.66	3.64	4.24	0.51	2.92
	(1.29)	(1.28)	(1.59)	(1.62)	(0.39)	(1.11)
Observations	1031	1031	1031	1031	1031	1031
Kernel Type	Triangular	Uniform	Triangular	Triangular	Triangular	Triangular
Polynomial	Linear	Linear	Linear	Quadratic	Linear	Linear
Bandwidth Type	MSE	MSE	CER	MSE	MSE	MSE
Eff. # of treated obs	55	47	47	63	51	55
Eff. # of untreated obs	272	250	253	308	265	273
Cluster Std. Error	Yes	Yes	Yes	Yes	Yes	Yes
Gacaca Category	1 & 2	1& 2	1& 2	1& 2	1	2

Table 11: Replication of Main Table 2 where the dependent variable is the population rate (0-100) of village prosecutions for participation in the Rwandan genocide and where the forcing variable is geodesic distance of each village to the nearest *paysannat* border. Standard errors clustered at the district level.

OLS: The effect of resettlement on rate of village-level prosecutions for participation in the Rwandan genocide

	(1)	(2)	(3)	(4)	(5)
paysannat, percentage	0.023	0.018	0.015	0.006	0.009
	(0.015)	(0.008)	(0.007)	(0.002)	(0.005)
Observations	1,049	1,030	1,030	1,030	1,030
Controls	No	Yes	Yes	Yes	Yes
District FE	No	Yes	Yes	Yes	Yes
Clustered FE	Yes	Yes	Yes	Yes	Yes
Gacaca Category	1 & 2	1& 2	1& 2	1	2

Table 12: Replication of Main Table 4 where the rate of gacaca prosecutions per total population (0-100) of a village is the dependent variable. Standard errors clustered at the district level.

References

Firebaugh, Glenn and Jack P Gibbs. 1985. "User's Guide to Ratio Variables." *American Sociological Review* pp. 713–722.

Kronmal, Richard A. 1993. "Spurious Correlation and the Fallacy of the Ratio Standard Revisited." *Journal of the Royal Statistical Society. Series A* pp. 379–392.

Lugan, Bernard. 1975. *Vers L'Indépendance du Rwanda: Evénements Politiques de 1959-1962*. St Germain en Laye: Edition M.D.I.

Neyman, Jerzy. 1952. *Lectures and Conferences on Mathematical Statistics*. 2nd ed. Graduate School, U.S. Department of Agriculture.

Pearson, Karl. 1897. "Mathematical Contributions to the Theory of Evolution - On a form of Spurious Correlation which may Arise when Indices are used in the Measurement of Organs." *Proceedings of the Royal Society* 60(359-367):489–498.

Verpoorten, Marijke. 2012. "Leave None to Claim the Land: A Malthusian Catastrophe in Rwanda?" *Journal of Peace Research* 49(4):547–563.