

Supplement

Ethnic diversity, social sanctions, and public goods in Kenya

This serves as a supplement to the published paper *Ethnic diversity, social sanctions, and public goods in Kenya*, by Edward Miguel, Mary Kay Gugerty (Journal of Public Economics, 2006). It outlines minor discrepancies found in data replication which do not, as a whole, affect the reported data analysis published in the paper. See *README.txt* for a detailed update. The tables below come from the output generated with the replication code in *kenya_ethnic_analysis.do* and *water_analysis.do*.

We have replicated the results for the paper, and we show here the replicated tables with any adjustment. There were several typos in the original version, as well as some minor discrepancies between the reproduced and published versions due to changes in STATA over time. Color code: **discrepancy**, **significance discrepancy**,

The data replication generated minor discrepancies in **Table 1: Descriptive statistics**. The discrepancies generated from *kenya_ethnic_analysis.do* are highlighted in Table 1 below, and are especially present in the 1996 data.

Table 1

Ethnic diversity across geographic divisions in Busia and Teso districts in 1962 and 1996

Geographic division	Name in 1962	Proportion of largest residential ethnic group (group in parentheses)	
		1962	1996 (Pupil Questionnaire data)
Budalangi	Bunyala	0.99 (Luhya)	0.94 (Luhya)
Funyula	Samia	0.98 (Luhya)	0.94 (Luhya)
Butula	Marachi	0.92 (Luhya)	0.86 (Luhya)
Amukura/Chakol	South Teso	0.92 (Teso)	0.87 (Teso)
Angurai/Amagoro	North Teso	0.87 (Teso)	0.86 (Teso)
Nambale/Matayos	Bukhayo	0.68 (Luhya)	0.76 (Luhya)

The 1962 data is from the 1962 Kenyan Census (Government of Kenya, 1965). The 1996 data is from the ICS Pupil Questionnaire, which relies on self-described ethnic affiliation.

The data replication generated a minor discrepancy in **Table 2: Pupil descriptive statistics by ethnic group**. This discrepancies generated from *kenya_ethnic.do* are highlighted in Table 2 below, and it does not impact the reported data analysis in the paper.

Table 2
Pupil descriptive statistics by ethnic group

VARIABLES	(1) <u>mean</u>	(2) <u>mean</u>	(3) <u>mean</u>	(4) <u>mean</u>
Age in years	14.51	14.44	14.71	14.51
Father, years of education	7.53	7.54	7.36	7.54
Mother, years of education	5.01	4.97	4.89	5.14
Fathers with formal employment	0.24	0.26	0.17	0.29
Mothers with formal employment	0.05	0.04	0.05	0.04
Proportion latrine ownership	0.85	0.84	0.87	0.81
Proportion iron roof ownership	0.26	0.26	0.23	0.30
Proportion livestock ownership	0.78	0.78	0.78	0.79
Proportion cultivates corn (maize)	0.87	0.87	0.86	0.92
Proportion cultivates cash crop	0.39	0.33	0.56	0.29
Attends school not closest to home	0.18	0.19	0.17	0.17
Resid. & school different geo zones	0.12	0.15	0.06	0.06
Lives with a parent if at least one parent is alive	0.82	0.81	0.85	0.81
Average number of full siblings	4.45	4.39	4.63	4.09
Proportion, Catholic	0.56	0.57	0.56	0.63

1 = Entire Sample, 2 = Luhya pupils, 3 = Teso pupils, 4 = Luo pupils

Table 2
Pupil descriptive statistics by ethnic group

VARIABLES	(1) <u>mean</u>	(2) <u>mean</u>	(3) <u>mean</u>	(4) <u>mean</u>
Lives with a parent if at least one parent is alive	0.85	0.84	0.88	0.84

1 = Entire Sample, 2 = Luhya pupils, 3 = Teso pupils, 4 = Luo pupils

The data replication generated minor discrepancies in **Table 3: Primary school descriptive statistics**. The discrepancies generated from *kenya_ethnic.do* are highlighted in Table 3 below. These discrepancies do not impact the reported data analysis in the paper.

Table 3
Primary school descriptive statistics

VARIABLES	(1) N	(2) mean	(3) sd
Zonal residential ELF across tribes,'96 PQD	84.00	0.23	0.14
Proportion of largest ethnic group in zone,'96 PQD	84.00	0.86	0.11
School ELF across tribes,'96 PQD	84.00	0.20	0.17
School ELF across tribes,'96 END	84.00	0.21	0.15
ELF across tribes for ...5km,'96 END	84.00	0.24	0.14
Proportion of largest ethnic group in school,'96 PQD	84.00	0.79	0.18
Total local school funds collected per pupil,'95 KS	84.00	152.58	99.43
Harambee donations collected per pupil,'95 KS	84.00	44.80	88.18
School fees collected per pupil,'95 KS	84.00	107.78	48.58
Desks per pupil,1995	80.00	0.22	0.12
Pupil latrines per pupil,1995	84.00	0.02	0.01
Classrooms per pupil, 1995	84.00	0.03	0.01
School-owned texts per pupil, 1995	84.00	0.34	0.21
Private texts per pupil, 1995	84.00	0.07	0.10
Pupil enrollment per primary school, 1996	84.00	296.31	146.36
School record sanctions/verbal pressure,1997	84.00	3.15	2.99
School record administrative activities,1997	84.00	18.92	11.43
Parent school meetings, 1997	83.00	3.45	1.90
Parent cooperation from 0 to 1 (RFO),1998	84.00	0.49	0.33
Teacher motivation from 0 to 1 (RFO),1998	84.00	0.54	0.30
Pupil-teacher ratio, 1996	84.00	29.12	9.82
Proportion teachers with high school education,1996	83.00	0.79	0.16
Years of teaching experience, 1996	83.00	14.03	3.01
Proportion of male teachers, 1996	83.00	0.74	0.19
Latitude (degrees north), GPS data	84.00	0.43	0.19
Longitude (degrees east), GPS data	84.00	34.23	0.13
Number of other primary schools w/in 5 km,GPS data	84.00	14.49	3.70

The file replicates 'Proportion of male teachers', while the paper published had
'Proportion of female teachers' = (1 - result replicated)

The data replication generated minor discrepancies in **Table 4: Ethnic diversity and local characteristics**. The discrepancies are highlighted below. These discrepancies do not impact the reported data analysis in the paper.

Table 4
Ethnic diversity and local characteristics

Dependent variable	Coefficient estimate on zonal residential ELF across tribes (OLS)	Coefficient estimate on ELF across tribes among schools within 5 km (spatial OLS)	Number of schools	Mean dependent variable
<i>(A) Pupil characteristics (1996 Pupil Questionnaire)</i>				
Father, years of education	0.5 (1.0)	−0.4 (1.2)	84	7.3
Mother, years of education	1.2 (1.3)	0.2 (1.4)	84	4.9
Fathers with formal employment	−0.09 (0.07)	−0.24*** (0.07)	84	0.23
Mothers with formal employment	−0.01 (0.02)	0.01 (0.02)	84	0.04
Proportion of latrine ownership	0.13 (0.09)	0.06 (0.09)	84	0.84
Proportion of iron roof ownership	0.04 (0.11)	0.02 (0.10)	84	0.25
Proportion of livestock ownership	0.16* (0.08)	0.12 (0.11)	84	0.78
Proportion, cultivates corn (maize)	−0.03 (0.06)	−0.16** (0.08)	84	0.87
Proportion, cultivates cash crop	0.26 (0.31)	0.67*** (0.20)	84	0.40
Average number of full siblings	1.7 (1.5)	2.5 (1.6)	84	7.4
Proportion, Catholic	−0.03 (0.19)	0.07 (0.17)	84	0.57
<i>(B) School and teacher characteristic</i>				
Pupil enrollment per primary school, 1996	72.2 (103.0)	−13.1 (104.4)	84	296.3
Pupil-teacher ratio, 1996	−4.2 (10.0)	−8.8 (6.4)	84	29.1
Proportion, teachers with HS education, 1996	0.10 (0.08)	−0.10 (0.14)	83	0.79
Years of teaching experience, 1996	0.3 (3.0)	2.7 (2.4)	83	14.0
Proportion of female teachers, 1996	−0.12 (0.15)	−0.12 (0.16)	83	0.26

Huber robust standard errors in parentheses. Significantly different than 0 at 90% (*), 95% (**), 99% (***) confidence. Regression disturbance terms are clustered at the zonal level. Ethnolinguistic fractionalization is defined as $ELF = 1 - \sum_i (\text{proportion of ethnolinguistic group } i \text{ in the population})^2$. School ELF considers Luhya as a single group. The coefficient estimate on zonal residential ELF across tribes uses data from the 1996 Pupil Questionnaire. In these specifications, observations are assumed to have independent error terms across geographic zones but not necessarily within zones. The coefficient estimate on ELF across tribes among schools within 5 km uses 1996 Exam Namelist data. In these specifications, regression disturbance terms are allowed to be correlated across schools as a general function of their physical distance, using the estimation strategy developed in Conley (1999).

The data replication generated minor discrepancies in **Table 5: Ethnic diversity and local primary school funding**. The discrepancies generated from *kenya_ethnic.do* are highlighted in Table 5 below. These discrepancies do not impact the reported data analysis in the paper.

Table 5.A
Ethnic diversity and local primary school funding
Dependant variable: total local primary school funds collected per pupil '95 KS

VARIABLES	(1) School ELF across tribes OLS 1st stage	(2) Total local OLS	(3) primary school OLS	(4) funds collected IV-2sls
Zonal ELF across tribes	0.87*** (0.07)		-185.7** (77.9)	
School ELF across tribes		-32.9 (64.0)		-216.4** (88.4)
Observations	100	84	84	84
R-squared	0.4	0.0	0.1	
Root MSE	0.138	99.84	96.73	105.5

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5.B
Ethnic diversity and local primary school funding
Dependant variable: total local primary school funds collected per pupil '95 KS

VARIABLES	(1) OLS	(2) OLS	(3) OLS
Zonal ELF across tribes	-145.2*** (49.6)	-143.6* (82.1)	
1 - (Prop... in zone)			-162.9** (66.6)
Proportion of fathers with formal employment	189.5 (165.1)	-220.6* (120.5)	184.6 (170.9)
Proportion of pupils with a latrine at home	-431.6*** (139.9)	-286.3 (228.0)	-429.8*** (150.3)
Proportion livestock ownership	120.1 (136.9)	186.2 (130.4)	110.6 (144.3)
Proportion cultivates cash crop	35.7 (61.4)	22.2 (106.9)	27.8 (62.4)
Proportion Teso		67.9 (181.4)	
Observations	84	84	84
R-squared	0.1	0.3	0.1
Geographic division indicators	No	Yes	No
Root MSE	94.96	92.97	95.36

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5.C		
Ethnic diversity and local primary school funding		
Dependant variable: total local primary school funds collected per pupil '95 KS		
VARIABLES	(1) Spatial OLS	(2) Spatial OLS
ELF across tribes for all schools within 5 km	- 173.0** (76.2)	- 171.9** (80.6)
Proportion of fathers with formal employment		143.2 (167.0)
Proportion of pupils with a latrine at home		- 466.9 (250.2)
Proportion livestock ownership		116.6 (117.5)
Proportion cultivates cash crop		84.2 (78.5)
Observations	84	84
R-squared	0.06	0.09
Geographic division indicators	No	Yes
Root MSE	97.1	95.1
Robust standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

The data replication generated minor discrepancies in **Table 6: Ethnic diversity impacts, controlling for ethnic population shares**. The discrepancies generated from *kenya_ethnic.do* are highlighted in Table 6 below. These discrepancies do not impact the reported data analysis in the paper.

Table 6.A				
Ethnic diversity impacts controlling for ethnic population shares				
Dependant variable: total local primary school funds collected per pupil '95 KS				
VARIABLES	(1) OLS	(2) OLS	(3) IV-2sls	(4) IV-2sls
Zonal ELF across tribes	-185.7** (77.9)	-189.1** (77.5)		
School ELF across tribes			-216.4** (88.4)	-208.4** (96.1)
Proportion Luhya		-196.5 (393.8)		-136.4 (415.7)
Proportion Teso		-247.3 (366.6)		-184.8 (386.2)
Observations	84	84	84	84
R-squared	0.1	0.1		
Root MSE	96.73	94.75	105.5	104.0
Robust standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				

The data replication generated minor discrepancies in **Table 7: Other primary school outcomes**. The discrepancies generated from *kenya_ethnic.do* are highlighted in Table 7 below. These discrepancies do not impact the reported data analysis in the paper.

Table 7
Other primary school outcomes

Dependent variable	Coefficient estimate on zonal residential ELF across tribes (OLS)	Coefficient estimate on ELF across tribes among schools within 5 km (spatial OLS)	Number of schools	Mean dependent variable
<i>(A) Local school funding</i>				
Harambee donations collected per pupil, 1995 (Kenyan Shillings)	-157.1** (61.6)	-182.1** (68.5)	84	44.8
School fees collected per pupil, 1995 (Kenyan Shillings)	11.9 (35.2)	8.1 (64.6)	84	107.8
<i>(B) School facilities, inputs</i>				
Desks per pupil, 1996	-0.20** (0.08)	-0.31*** (0.08)	84	0.21
Pupil latrines per pupil, 1996	-0.007 (0.009)	-0.007 (0.013)	84	0.016
Classrooms per pupil, 1996	-0.016 (0.016)	-0.023* (0.013)	84	0.030
School-owned textbooks per pupil, 1996	-0.17 (0.13)	-0.27 (0.17)	84	0.34
Private texts (at home) per pupil, 1996	-0.03 (0.07)	-0.10 (0.09)	84	0.07
Number of other primary schools within 5km	-10.2*** (3.5)	-12.2*** (3.7)	84	14.5
<i>(C) Test scores</i>				
Average school score on 1996 NGO exams, Grades 3–8 (in standard deviations)	0.10 (0.52)	0.11 (0.52)	84	0.05
Socioeconomic controls (zonal averages)	Yes	Yes		

The data replication generated minor discrepancies in **Table 8: School committee records and field officer observations**. The discrepancies generated from *kenya_ethnic.do* are highlighted in Table 8 below. These discrepancies do not impact the reported data analysis in the paper.

Table 8
School committee records and field officer observations

Dependent variable	Coefficient estimate on zonal residential ELF across tribes (OLS)	Coefficient estimate on ELF across tribes among schools within 5 km (spatial OLS)	Number of schools	Mean dependent variable
<i>(A) School Committee Records</i>				
School committee record items regarding sanctions or verbal pressure, 1997	-3.7** (1.6)	-4.2* (2.3)	84	3.2
School committee record items regarding administrative activities, 1997	5.7 (6.1)	6.2 (10.3)	84	18.9
Parent school meetings, 1997	-1.6 (1.1)	-1.3 (1.6)	84	3.4
<i>(B) Field Officer observations</i>				
Parent cooperation from 0 to 1 (reported by field officers), 1998	-0.77*** (0.26)	-0.84** (0.35)	84	0.49
Teacher motivation from 0 to 1 (reported by field officers), 1998	-0.39** (0.17)	-0.49* (0.29)	84	0.54
Socioeconomic controls (zonal averages)	Yes	Yes		

The data replication from *water_analysis.do* did not generate any discrepancies for **Table 9: Well descriptive statistics**.

Table 9
Well descriptive statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) N	(5) mean	(6) sd
ELF...5 km from well	667.00	0.23	0.14			
Indic...“normal” water flow	667.00	0.57	0.49			
Indic.... no broken/missing	667.00	0.66	0.48			
Indic... another well				196.00	0.32	0.47
Year well stopped				196.00	1,997.46	3.09
Latitude (north)	667.00	0.36	0.17			
Longitude (east)	667.00	34.20	0.12			

The data replication generated minor discrepancies in **Table 10: Ethnic diversity and well maintenance**. The discrepancies generated from *water_analysis.do* are highlighted in Table 10 below. These discrepancies do not impact the reported data analysis in the paper.

Here are the Probit regressions:

Table 10 (PROBIT) Ethnic diversity and well maintenance			
VARIABLES	(1) Normal water flow Probit	(2) No broken or missing well parts Probit	(3) Get water from another local well Probit
ELF...5 km from well	-0.26* (0.14)	-0.24* (0.13)	-0.73** (0.31)
Observations	667	667	196
Root MSE	.	.	.
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			

Right number of wells off by one unit in table output

Here are the Spatial OLS regressions (see log-file)

Table 10 (SPATIAL OLS) Ethnic diversity and well maintenance			
VARIABLES	(1) Normal water flow Spatial OLS	(2) No broken or missing well parts Spatial OLS	(3) Get water from another local well Spatial OLS
ELF...5 km from well	-0.26 (0.17)	-0.25 (0.22)	-0.71* (0.36)
Observations	667	667	196
Root MSE	.	.	.
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			

Copied from STATA's log-file – reported s.e. corrected for spatial dependence.

The data replication from *kenya_ethnic_analysis*.do did not generate any discrepancies for **Table A1: Selection into NGO assistance program in 1995**.

Table A1 Selection into NGO assistance program in 1995				
	(1)	(2)	(3)	(4)
	1995 Pupils enrollment (District Educational Office records)		1995 Average government exam result Grades 6–8	
VARIABLES	OLS	OLS	OLS	OLS
SAP	-99.6*** (17.1)	-115.6** (43.2)	-72.0*** (10.7)	-53.2** (22.4)
ELF		73.3 (110.2)		116.4 (89.1)
SAP*ELF		65.0 (128.5)		-89.5 (73.2)
Observations	300	300	300	300
R-squared	0.06	0.07	0.10	0.11
Root MSE	182.3	182.5	102.5	102.0

Robust standard errors in parentheses

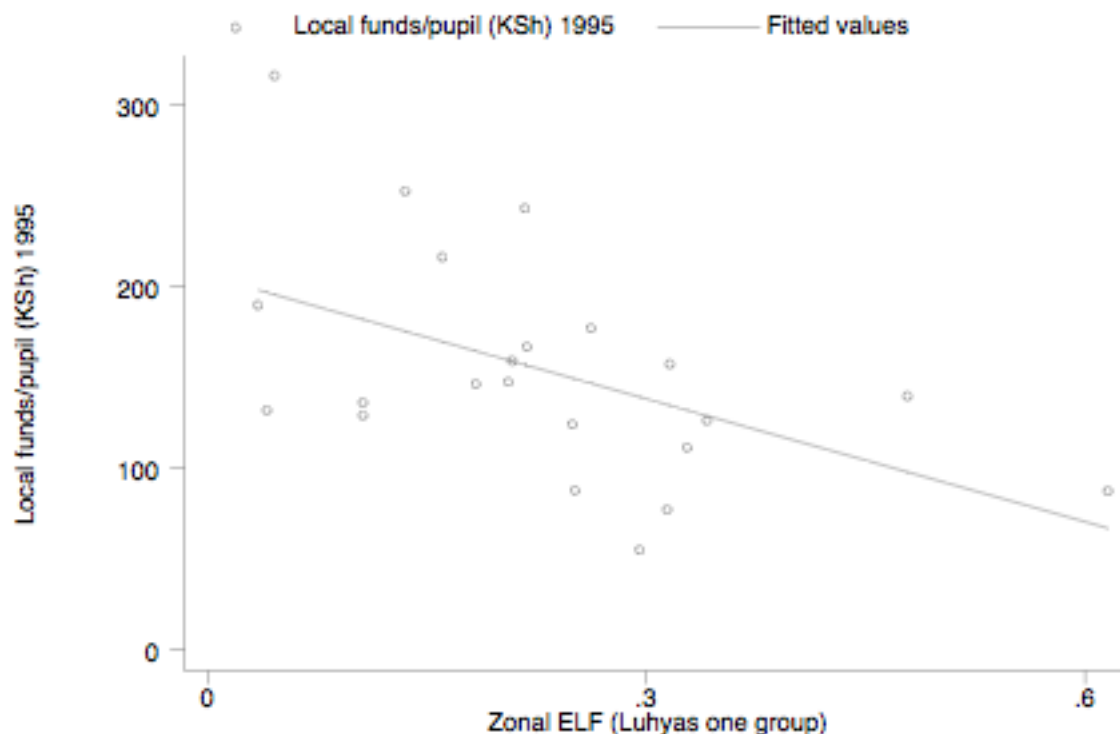
*** p<0.01, ** p<0.05, * p<0.1

SAP: Indicator for selection into NGO assistance program

ELF: Zonal residential ELF across tribes in 1996

SAP*ELF: (Indicator for selection into NGO assistance program)*(zonal residential ELF across tribes in 1996

This is the replicated version of **Figure 3**, without discrepancies with respect to the published version:



This is the replicated version of **Figure 4**, without discrepancies with respect to the published version:

