

Table 1: Ethnic Diversity across Geographic Divisions in Busia and Teso districts, in 1962 and 1996<sup>41</sup>

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)

Table 2: Pupil Descriptive Statistics, by Ethnic Group<sup>42</sup>

	Entire sample	Luhya pupils	Teso pupils	Luo pupils
Number of pupils interviewed	5832	3867	1516	301
Proportion of pupil sample	1	0.66	0.26	0.05
Age in years	14.5	14.4	14.7	14.5
Father years of education	7.5	7.5	7.4	7.5
Mother years of education	5.0	5.0	4.9	5.2
Fathers with formal employment <sup>43</sup>	0.24	0.26	0.17	0.29
Mothers with formal employment <sup>30</sup>	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 <sup>44</sup>	0.78	0.78	0.78	0.80
Proportion cultivates corn (maize)	0.87	0.87	0.86	0.92
Proportion cultivates cash crop <sup>45</sup>	0.39	0.33	0.56	0.29
Attends primary school that is not the closest to home	0.18	0.19	0.17	0.17
Residence and school in different geographic zones	0.12	0.15	0.06	0.06
Lives with a parent, if at least one parent is alive	0.85	0.84	0.88	0.84
Average number of full siblings	4.5	4.4	4.6	4.1
Proportion Catholic	0.56	0.57	0.56	0.64

<sup>41</sup> 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.

<sup>42</sup> Data are from the 1996 ICS Pupil Questionnaire administered to pupils in grades 6 to 8. Other ethnic groups, including Kalenjin, Kikuyu, Masaai, Somali, Tachoni, and Taita, comprise 0.020 of the sample.

<sup>43</sup> Formal employment includes government, factory, and corporate jobs, but not petty trading (in contrast to the definition of formal employment in Miguel 2001).

<sup>44</sup> Livestock ownership includes cattle, goats, sheep, and pigs (but not chicken).

<sup>45</sup> Cash crops include tobacco, sugar cane, and cotton.

Table 5: Ethnic Diversity and Local Primary School Funding<sup>50</sup>

Explanatory variable	Dependent variable								
	School ELF across tribes	Total local primary school funds collected per pupil in 1995 (Kenyan Shillings)							
1st stage	(1) OLS	(2) OLS	(3) OLS	(4) IV-2sls	(5) OLS	(6) OLS	(7) OLS	(8) Spatial OLS	(9) Spatial OLS
<i>Ethnic diversity measures</i>									
School ELF across tribes		-32.9 (64.0)		-216.4** (88.4)					
Zonal ELF across tribes	0.86*** (0.07)		-185.7** (77.9)		-145.2*** (49.6)	-143.6* (82.1)			
1 – (Proportion largest ethnic group in zone)							-162.9** (66.6)		
ELF across tribes for all schools within 5 km								-174.0** (76.3)	-174.0** (80.8)
<i>Zonal controls</i>									
Proportion fathers with formal employment					189.5 (165.1)	-220.6* (120.5)	184.6 (170.9)		142.8 (167.3)
Proportion of pupils with a latrine at home					-431.6*** (139.9)	-286.3 (228.0)	-429.8*** (150.3)		-466.9 (250.2)
Proportion livestock ownership					120.1 (136.9)	186.2 (130.4)	110.6 (148.3)		116.9 (117.7)
Proportion cultivates cash crop					35.7 (61.4)	22.2 (106.9)	27.8 (62.4)		85.2 (78.4)
Proportion Teso pupils						67.9 (181.4)			
Geographic division indicators	No	No	No	No	No	Yes	No	No	No
Root MSE	0.14	99.8	96.7	105.5	95.0	93.0	95.4	97.1	95.0
R <sup>2</sup>	0.40	0.00	0.06	-	0.14	0.25	0.12	0.06	0.09
Number of schools	84	84	84	84	84	84	84	84	84
Mean dep. variable	0.20	152.6	152.6	152.6	152.6	152.6	152.6	152.6	152.6

<sup>50</sup> Huber robust standard errors in parentheses. Significantly different than zero at 90% (\*), 95% (\*\*), 99% (\*\*\*) confidence. Observations are assumed to have independent error terms across geographic zones, but not necessarily within zones for Regressions 1 to 7. Ethno-linguistic fractionalization is defined as  $1 - \sum_i (\text{Proportion of Ethno-linguistic group } i \text{ in the population})^2$ . School ELF across tribes and the proportion of the largest ethnic group in the school consider Luhyas a single group. Regression disturbance terms are allowed to be correlated across schools as a general function of physical distance in regressions 8 and 9 (Conley 1999). Geographic indicators are indicators for six (of the seven) geographic divisions.

Table 6: Other Primary School Outcomes<sup>51</sup>

Dependent variable	Coefficient estimate on zonal residential ELF across tribes (OLS) <sup>52</sup>	Coefficient estimate on ELF across tribes among schools within 5 km (Spatial OLS) <sup>53</sup>	Number of schools	Mean dependent variable
<i>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>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>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) <sup>54</sup>	Yes	Yes		

<sup>51</sup> Huber robust standard errors in parentheses. Significantly different than zero at 90% (\*), 95% (\*\*), 99% (\*\*\*) confidence. Regression disturbance terms are clustered at the zonal level. Ethno-linguistic fractionalization is defined as  $ELF = 1 - \sum_i (\text{Proportion of Ethno-linguistic group } i \text{ in the population})^2$ . School ELF considers Luhyas a single group.

<sup>52</sup> 1996 Pupil Questionnaire Data. Observations are assumed to have independent error terms across geographic zones, but not necessarily within zones.

<sup>53</sup> 1996 Exam Namelist data. 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).

<sup>54</sup> Socioeconomic controls include the proportion of fathers in the geographic zone with formal sector employment, the proportion of pupils residing in the geographic zone with a latrine at home, the proportion of pupils whose households own livestock, and the proportion of pupils whose households cultivate a cash crop. The test score results also an additional explanatory variable, an indicator for having received financial assistance through another NGO program.

Table 8: Well Descriptive Statistics<sup>59</sup>

	Mean	Standard deviation	Obs.
<i>Well Characteristic</i>			
ELF across tribes for all primary schools within 5 km of the well, 1996 Exam Namelist data	0.23	0.14	667
Indicator variable “normal” water flow from well, 2000-2001 survey	0.57	0.49	667
Indicator variable no broken or missing well parts, 2000-2001 survey <sup>60</sup>	0.66	0.48	667
Indicator variable people in the area get water from another local well (if not normal water flow), 2000-2001 survey	0.32	0.47	196
Year well stopped functioning (if not normal water flow), 2000-2001 survey	1997.5	3.1	196
Latitude (degrees North), GPS data from 2000-2001 survey	0.36	0.17	667
Longitude (degrees East), GPS data from 2000-2001 survey	34.20	0.12	667

Table 9: Ethnic Diversity and Well Maintenance<sup>61</sup>

Explanatory variables	Dependent variable:					
	Indicator variable for “normal” water flow from well		Indicator variable for no broken or missing well parts		Indicator variable people in the area get water from another local well (if the KEFINCO well does not have normal water flow)	
	(1) Probit	(2) Spatial OLS	(3) Probit	(4) Spatial OLS	(5) Probit	(6) Spatial OLS
ELF across tribes among schools within 5 km	-0.26* (0.14)	-0.26 (0.17)	-0.25* (0.13)	-0.25 (0.22)	-0.73** (0.30)	-0.72* (0.36)
Number of wells	667	667	667	667	196	196
Root MSE	-	0.49	-	0.47	-	0.46
Mean dependent variable	0.57	0.57	0.66	0.66	0.32	0.32

<sup>59</sup> Data are from the 1996 ICS School and Pupil Questionnaires, 1996 Government Examination Namelists, and Global Positioning Systems (GPS) readings taken by NGO field workers. Ethno-linguistic fractionalization is defined as  $1 - \sum_i (\text{Proportion of Ethno-linguistic group}_i \text{ in the population})^2$ . School ELF across tribes and the proportion of the largest ethnic group in the school consider Luhyas a single group.

<sup>60</sup> Well parts include the pump handle, the cover and base, and the external and internal pipes and seals.

<sup>61</sup> Notes: Huber robust standard errors in parentheses. Observations are assumed to have independent error terms across geographic zones, but not necessarily within zones in regressions 1, 3, and 5, and 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), in regressions 2, 4, and 6. Significant at 90(\*), 95(\*\*), 99(\*\*\*) % confidence. Geographic indicators are indicators for six (of the seven) geographic divisions.

Table 7: School committee records and field officer observations<sup>55</sup>

Dependent variable	Coefficient estimate on zonal residential ELF across tribes (OLS) <sup>56</sup>	Coefficient estimate on ELF across tribes among schools within 5 km (Spatial OLS) <sup>57</sup>	Number of schools	Mean dependent variable
<i>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>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) <sup>58</sup>	Yes	Yes		

<sup>55</sup> Huber robust standard errors in parentheses. Significantly different than zero at 90% (\*), 95% (\*\*), 99% (\*\*\*) confidence. Regression disturbance terms are clustered at the zonal level. Ethno-linguistic fractionalization is defined as  $ELF = 1 - \sum_i (\text{Proportion of Ethno-linguistic group, in the population})^2$ . School ELF considers Luhyas a single group.

<sup>56</sup> 1996 Pupil Questionnaire Data. Observations are assumed to have independent error terms across geographic zones, but not necessarily within zones.

<sup>57</sup> 1996 Exam Namelist data. 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).

<sup>58</sup> Socioeconomic controls include the proportion of fathers in the geographic zone with formal sector employment, the proportion of pupils residing in the geographic zone with a latrine at home, the proportion of pupils whose households own livestock, and the proportion of pupils whose households cultivate a cash crop.

Table 3: Group Summary Statistics  
Means and Standard Deviations

	Method of Arrival to Group		Invited to Group Mean & Std Dev & Std Error # of Obs	Uninvited to Group Mean & Std Dev & Std Error # of Obs
	Mean & Std Dev & Std Error	# of Obs		
<b>GEOGRAPHIC CONCENTRATION</b>				
*Average distance to original members from current members (minutes)	12.858 (0.785)	5.088 n=42	12.722 (0.694)	4.495 n=42
*Average percent of original members who live within 10 minutes of current member	0.228 (0.028)	0.182 n=42	0.229 (0.027)	0.178 n=42
GD: Geographic concentration	0.147 (0.016)	0.104 n=42	0.252 (0.027)	0.174 n=42
E(GD): Expected geographic concentration	0.127 (0.014)	0.090 n=42	0.203 (0.026)	0.168 n=42
<b>CULTURAL CONCENTRATION</b>				
*Average percent of original members of same culture as current member	0.200 (0.014)	0.092 n=42	0.204 (0.013)	0.085 n=42
CD: Cultural concentration	0.119 (0.021)	0.136 n=42	0.184 (0.024)	0.155 n=42
E(CD): Expected cultural concentration	0.106 (0.012)	0.078 n=42	0.167 (0.021)	0.136 n=42

\*Variables with asterisks are the key dependent variables used in the specifications in Tables 3, 4, and 5.  
All results calculated on original group members only.

$$GD_{group} = \sum_{neighborhoods} (s_i - x_i)^2,$$

where  $s_i$  is the share of the group from neighborhood  $i$  and  $x_i$  is the share of the general population from neighborhood  $i$ .

$$E(GD)_{group} = [1 - \sum_{neighborhoods} (x_i)^2] / n$$

CD and E(CD) are constructed identically to GD and E(GD), except by cultural group rather than neighborhood.  
The Alesina index for cultural concentration is equal to the sum of squared shares of each cultural group.

Table 4: Individual Default  
OLS, Tobit, and Probit

	Dependent variable: Percent of loan in default at end of cycle					
	1st Loan Only			All Loans		
	OLS (1)	Tobit (2)	Probit (3)	OLS (4)	Tobit (5)	Probit (6)
Distance from individual's home to original members of group	0.019 (0.077) n=616	0.343 (0.342) n=616	0.019 (0.019) n=616	0.049 (0.068) n=1801	0.297 (0.024) n=1801	0.040 (0.027) n=1801
Percent of original members within 10-minute walk of individual's home	-1.536 *** (0.391) n=616	-6.077 *** (1.795) n=616	-0.284 *** (0.079) n=616	-1.556 *** (0.370) n=1801	-3.754 *** (1.078) n=1801	-0.367 *** (0.134) n=1801
Percent of original members with same culture as individual	-0.534 * (0.301) n=616	-4.230 ** (1.791) n=616	-0.200 *** (0.069) n=616	-0.396 (0.308) n=1801	-1.458 (1.116) n=1801	-0.177 (0.111) n=1801

\*\*\* 99% significance; \*\* 95% significance; \* 90% significance

Each cell is a separate specification.

Standard errors corrected for clustering at the group level in all specifications.

Individuals weighted evenly "all loans" specifications.

Individual level specifications include the following control variables (See Appendix Table 2 for results on control variables):

Distance to FINCA (town center), town dummy, neighborhood dummies, age, education, marital status, siblings, children,

# in household, year, and age of group when individual joined.

Loan size estimated using approved loan amount, which is savings balance at end of prior cycle.

Table 6: Dropout  
Probit

	Dependent Variable = 1 if Member Dropped Out after 1st Loan					
	(1)	(2)	(3)	(4)	(5)	(6)
Default	0.115 *** (0.037)	0.112 *** (0.037)	0.113 *** (0.036)	-0.023 (0.056)	0.153 *** (0.043)	0.197 *** (0.041)
Total Accumulated Savings	-0.013 (0.014)	-0.016 (0.014)	-0.014 (0.014)	-0.014 (0.015)	-0.015 (0.014)	-0.013 (0.013)
Distance from individual's home to original members of group	0.037 (0.031)			0.023 (0.032)		
Distance Interacted with default				0.074 *** (0.027)		
Percent of original members within 10-minute walk of individual's home		-0.007 (0.006)			-0.004 0.006	
Percent within 5-minute walk Interacted with default					-0.132 *** (0.047)	
Percent of original members with same culture as individual			-0.192 (0.157)			0.000 (0.144)
Culture Interacted with default						-0.332 *** (0.094)
Observations	616	616	616	616	616	616
# of dropouts	148	148	148	148	148	148
Log-likelihood	-173.47	-173.76	-173.28	-167.39	-171.58	-166.78
Groups	42	42	42	42	42	42

\*\*\* 99% significance, \*\* 95% significance, \* 90% significance

Marginal effects of probit reported.

Standard errors corrected for clustering at the group level.

Individual-level specifications control variables for distance to FINCA (town center), town dummy, neighborhood dummies, age, education, marital status, siblings, children, # in household, year, and age of group.



Figure 3: Map of Baltistan with sampled communities indicated

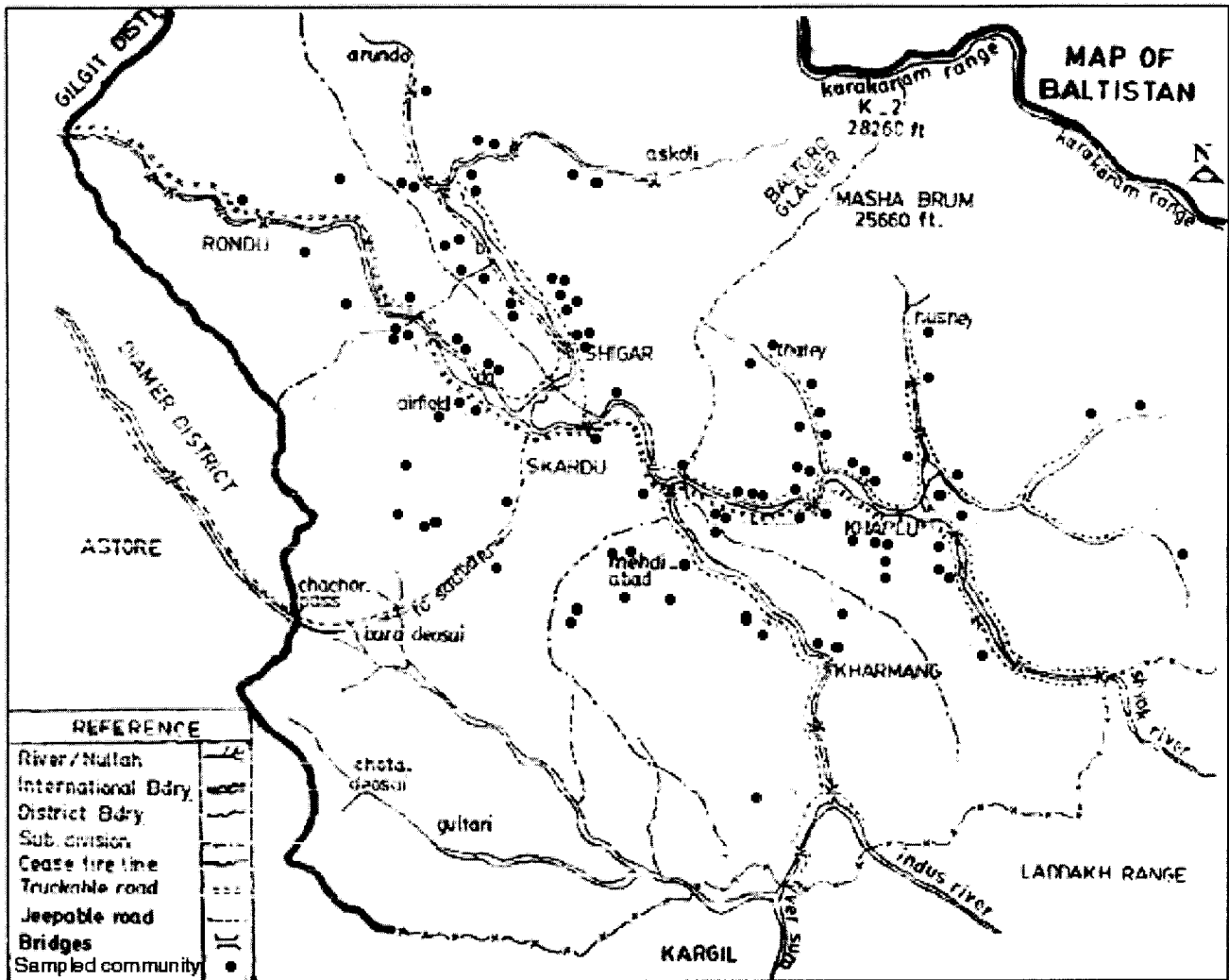
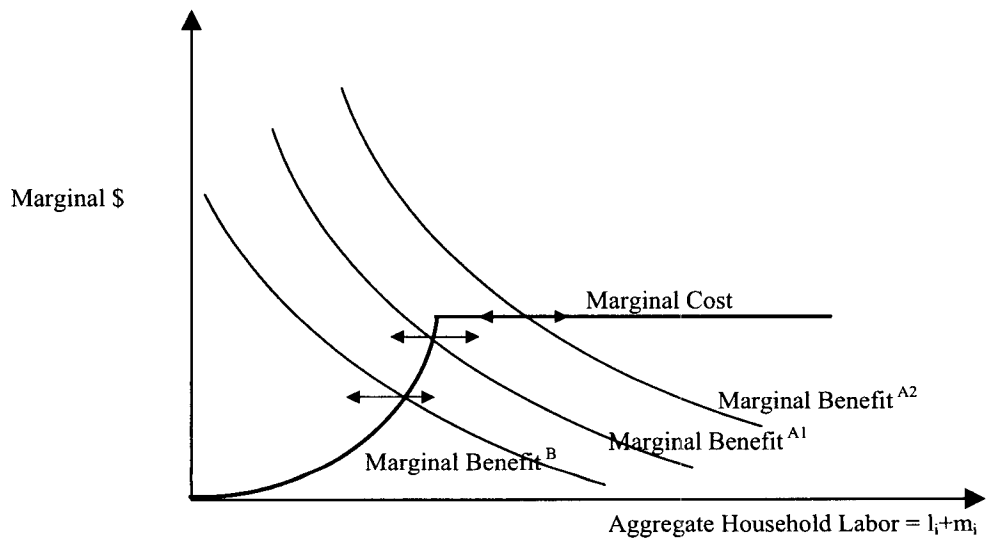


Figure 4: Graphical Illustration of Claim 1



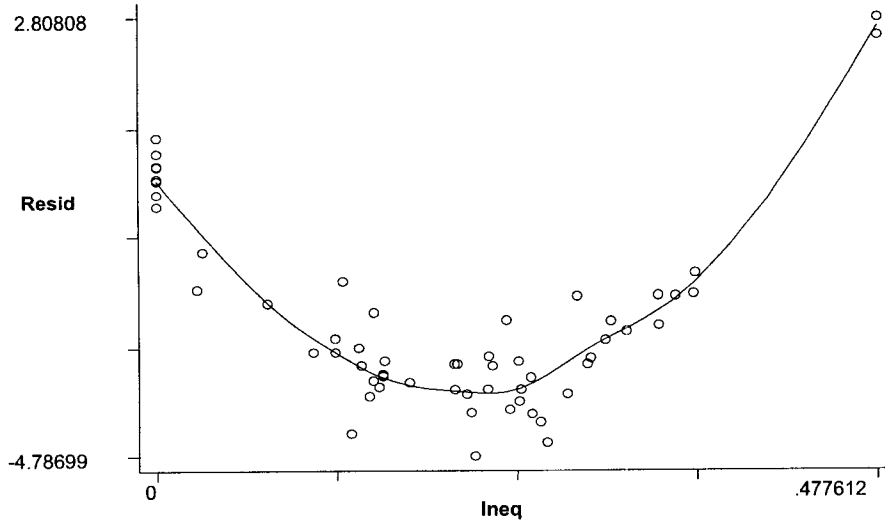
**Table 5. Determinants of Maintenance**

Variables	(1) OLS	Variables	(1-cont) OLS
<b>Community factors:</b>		<i>Physical capital variables:</i>	
Land Inequality	-275.4** (117)	Mean off-farm hh Income	-.0012 (.0021)
Land Inequality Squared	395.7** (194)	Mean real estate value (000)	-.044** (.022)
Social Heterogeneity	-55.1*** (18)	Community wage	.127 (.160)
Community size	-.020 (.062)	No. of Mechanical asset households	2.05*** (0.49)
Total Cultivable land	2e-04 (18e-04)	Access to Electricity?	5.13 (4.38)
Single cropping zone?	-8.79 (5.58)	Access to Health facility?	-1.21 (4.50)
Walk Time	-.046 (.075)	Access to Potable Water?	3.22 (4.37)
Travel Time	-.022 (.040)		
<i>Human capital variables:</i>		<b>Project factors:<sup>s</sup></b>	
Shopkeepers fraction	-71.9*** (27.2)	Project New?	-20.5*** (6.05)
Skilled workers fraction	32.5** (15.0)	Government project?	-8.96 (7.58)
Basic Education fraction	-3.37 (3.72)	Project Leader exists?	11.3** (5.7)
Tertiary Education fraction	36.0 (26.3)	External Funds (000,000)	6.96 (11.3)
Religious Education fraction	-36.8 (24.1)	Project Complexity	-4.20 (2.77)
High school?	20.6*** (6.4)	Non-technical decisions participation	35.5*** (15.4)
		Technical decisions participation	-18 (12.6)
		Controls	Pj age, type
		Adj R <sup>2</sup>	.35
		Prob>F	.00
		N	132

Huber-White robust standard errors in parentheses  
 Disturbance terms clustered at the village level  
 \*\*\*Significantly different from zero at 1%  
 \*\*Significantly different from zero at 5%  
 \* Significantly different from zero at 10%

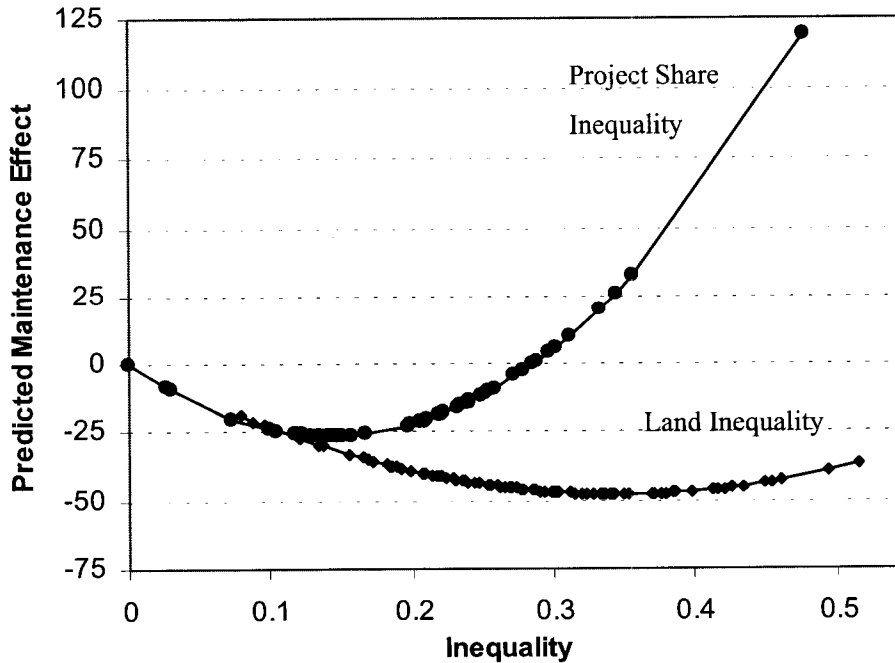
<sup>s</sup> *Project Share Inequality* was not included in these regressions since it is highly correlated with *Land Inequality*.

**Figure 5: total score residuals (Resid) against Project Share Inequality (Ineq)**



The residuals are calculated from a regression of project maintenance on project age, type and community dummy variables only. Excluding the two outliers (highest inequality levels) does change the shape or fit of the plot.

**Figure 6: Estimated magnitude comparisons – Land and Project Share Inequality**



Predicted effects are calculated by estimating the marginal effect of the inequality measure after controlling for all other factors. For the land inequality the estimates used are from the regression in Table 5, and for project share inequality the estimates are from the regression in Column 1, Table 7. Excluding the high project share inequality outlier does not significantly change the point estimates in the regression.

**Table 6. Effect of Project Leadership on Maintenance**  
**OLS and Instrumental Variable (IV-2SLS)**

Variables	(1) 1 <sup>st</sup> stage OLS	(2) 2 <sup>nd</sup> stage IV- 2SLS	(3) OLS	(4) 1 <sup>st</sup> stage OLS	(5) 2 <sup>nd</sup> stage IV- 2SLS
Project Leader exists?	Dependent variable	32.64** (15.70)	36.45*** (13.64)		25.11* (15.04)
Leader Quality				Dependent variable	42.03** (16.94)
Hereditary family 25-50 healthy male?	0.30* (0.18)				
Hereditary family absence (1-3)	-.32*** (.12)				
Hereditary family average age	-.015** (.006)				
Hereditary family 25-50 educated, present male				0.10*** (0.04)	
Hereditary family non-farm?				-0.20*** (0.08)	
Ideal leaders in community? (1-4)				0.21*** (0.08)	
<i>Project Leader attributes:</i>					
Educated?			-3.15 (6.19)		
Age			-.52** (.21)		
Non-farm occupation?			-10.4** (5.95)		
Land holding			-.001 (.056)		
Present throughout year?			7.63 (9.22)		
Trained?			5.05 (5.6)		
From hereditary leader group?			-10.4* (5.94)		
Controls		Community Characteristics, Physical & Human Capital and Project variables	Community Characteristics, Physical & Human Capital and Project variables		Community Characteristics, Physical & Human Capital and Project variables
Adj R <sup>2</sup>	.09	.21	.34	.07	.43
Prob>F	.00	.00	.00	.00	.00
N	132	132	130	132	132

Huber-White robust standard errors in parentheses

Disturbance terms clustered at the village level

\*\*\*Significantly different from zero at 1%

\*\*Significantly different from zero at 5%

\* Significantly different from zero at 10%

**Table 7. Project-specific Determinants of Maintenance  
Community Fixed Effects**

Variables	(1) FE	(2) FE	(3) FE
Project Complexity	-12.76*** (3.85)	-15.19*** (3.04)	-15.44*** (3.92)
Project Share Inequality	-373.3*** (67.7)	-402.7*** (86)	-422*** (69.2)
Project Share Inequality squared	1304*** (225)	1391*** (267)	1381*** (211)
Non-technical decisions participation	55.43* (28.29)		50.87* (24.24)
Technical decisions participation	-38.49* (18.56)		-34* (16.68)
Government project?	-23.63*** (7.95)	-18*** (6.07)	-18.18** (8.03)
Project New?	-41.92*** (13.67)	-40.55*** (11)	-46.77*** (15.06)
Project Leader?			13.36 (8.42)
Controls	Community Fixed Effects, Project Age and type	Community Fixed Effects, Project Age and type	Community Fixed Effects, Project Age and type
Adj R <sup>2</sup>	.71	.63	.73
Prob>F	.00	.00	.00
N	64	64	64

Column 1 presents the primary regression. Column 2 checks to see whether the results remain similar once the potentially endogenous (Halo effects) participation measure is excluded. Column 3 checks to see whether the external agency effect remains once leadership presence (endogenous) and participation are both controlled for.

Huber-White robust standard errors in parentheses

\*\*\*Significantly different from zero at 1%

\*\*Significantly different from zero at 5%

\* Significantly different from zero at 10%