Supplementary Information

There are some small correlations between predictor variables (see Table 4). Distance from the capital is positively correlated with the percentage of households using surface water (0.409), and negatively correlated with the percentage of households with permanent construction materials (-0.324) and average years of study (-0.246). Having a health post is positively correlated with land ownership (0.344) and the presence of non-indigenous community members (0.360). The presence of non-indigenous community members is also positively correlated with average years of study (0.284). However the variance inflation factors for both of the regression models was < 10, thus collinearity is not an issue.

Table 4. Correlations between predictor variables

	Remoteness	Land Ownership	House Type	Crowding	Education	Wage Labor	Health Post	Surface Water Usage	Non-Indigenous Presence
Remoteness	1.000	0.147	-0.324	0.075	-0.246	-0.195	-0.020	0.409	-0.054
Land Ownership	0.147	1.000	-0.200	-0.020	-0.130	-0.002	0.344	-0.068	0.099
House Type	-0.324	-0.200	1.000	-0.197	0.146	-0.007	0.101	-0.233	0.042
Crowding	0.075	-0.020	-0.197	1.000	0.044	0.001	-0.170	0.097	0.111
Education	-0.246	-0.130	0.146	0.044	1.000	0.109	0.317	-0.154	0.284
Wage Labor	-0.195	-0.002	-0.007	0.001	0.109	1.000	-0.052	-0.301	0.171
Health Post	-0.020	0.344	0.101	-0.170	0.317	-0.052	1.000	-0.170	0.360
Surface Water Usage	0.409	-0.068	-0.233	0.097	-0.154	-0.301	-0.170	1.000	-0.009
Non-Indigenous Presence	-0.054	0.099	0.042	0.111	0.284	0.171	0.360	-0.009	1.000

The variance of the predictor variables was reasonably similar across divisions by Linguistic Family and Department. Years of study was the only normally distributed predictor variable (Shapiro-Wilk test, $p = 0.54$), and a Bartlett test of homogeneity of variances revealed no significant difference in variance across linguistic families (Bartlett's $K^2 = 3.50$, $df = 4$, $p = 0.48$) or departments (Bartlett's $K^2 = 0.65$, $df = 2$, $p = 0.72$). The non-parametric Fligner-Killeen test of homogeneity of variances was used to examine the remaining predictor

variables. There were significant differences in variance for Surface Water Usage across linguistic families (Fligner-Killeen chi-squared = 13.45, df = 4, p = 0.009) and departments (Fligner-Killeen chi-squared = 8.43, df = 2, p = 0.01). As the difference in variances occurs between the two linguistic families with the fewest communities (Guaicuru with 4 communities and Guarani with 8 communities) and the three with the most communities (Zamuco with 14 communities, Mataco Mataguayo with 27 communities, and Lengua Maskoy with 52 communities), the difference may be an artefact of population size. The same is true of the difference in variance across departments, where Alto Paraguay, with only 17 communities, differs from Boquerón and Presidente Hayes, which have 39 and 49 communities respectively. The only other significant difference in variance was for the Non-Indigenous Presence predictor across linguistic families (Fligner-Killeen chi-squared = 9.85, df = 4, p = 0.04), however the median Non-Indigenous Presence was not significantly different (Kruskal-Wallis chi-squared = 7.69, df = 4, p = 0.10). All other predictor variables demonstrated homogeneity of variance across linguistic families and departments ($p > 0.05$).

Other predictor variables that may affect the variance of Surface Water Usage and Non-Indigenous Presence are already included in the regression models. Geographic differences between linguistic families and departments that might affect water access are included in the model through the Remoteness predictor variable. The Health Post and Years of Study predictors had small positive correlations with Non-Indigenous Presence (perhaps because non-indigenous persons are working in indigenous health posts and/or schools, or they tend to marry spouses in communities with these features). With this in mind, and in consideration of the small number of communities belonging to the three smallest linguistic families and the smallest department, it is reasonable to analyze the linguistic families and departments in aggregate.