Topics in Middle Eastern and African Economies
Vol. 13, September 2011

Tribal Heterogeneity and the Allocation of Development Resources: Evidence from Yemen

Daniel Egel[*] University of California, Institute on Global Conflict and Cooperation

February 8, 2011

Abstract

This paper examines how tribes, the dominant political structure in rural areas of many developing countries, affect the allocation of development resources. I create a dataset of Yemeni tribes and demonstrate that areas with more tribes per capita i.e more inter-tribal heterogeneity -receive larger allocations of development resources. My empirical analysis, which focuses on development resources in the education sector, demonstrates that this effect reflects the enhanced ability of smaller tribes to both provide public goods and extract patronage from the central government. The intuition for this result is developed in a model that links inter-tribal heterogeneity and access to development resources; as tribes are endogenous political structures that formed to solve public good problems during an agrarian period, smaller tribes are less internally heterogeneous and more effective in engaging in political action.

[*]Post-Doctoral Scholar. E-mail: degel@ucsd.edu. I thank Eli Berman, Barry Eichengreen, Bryan Graham and Ted Miguel for their encouragement and support throughout this project. Robert Anderson, Fred Finan, Silke Forbes, Avner Greif, Jonathan Rose, and Philippe Wingender provided valuable advice and helpful comments. I am deeply indebted to Lamis Aliryani, Betty Sadoulet and Alain de Janvry and the Social Fund for Development for supporting my fieldwork in Yemen. Nada Alhajjri, Mohammed Almawri, Mervat Al-Naimee, Ahmed Barakani, Ahmed Basharahi, Marta Colburn, Roberta Contin, Khalid Moheyddeen, Kunera Moore and Tareq Yeslam provided very helpful discussions during my stay in Yemen. I also thank Garret Christensen, Luke Condra, Lindsay Heger, Juan Sebastian Lleras, Guillermo Noguera, Anna Spurlock, Chloe Thurston, and Lanwei Yang and participants in the Berkeley Development Lunch and the Stanford Labor and Development Reading Group for helpful discussions and comments. The data collection would not have been possible without Fayadh Thawabah's excellent research assistance and a team of hardworking Yemeni researchers.

1 Introduction

The relationship between social heterogeneity and local access to development resources has been studied extensively in recent years. Following the early work of Alesina, Baqir, and Easterly (1999) [henceforth ABE], most of these studies have found a negative relationship between heterogeneity and local resource provision.[1] However, several empirical studies have found a positive relationship between heterogeneity and the distribution of government resources. These latter studies have typically implicated ethnic or caste patronage as the key reason for these findings.[2]

This paper, which is the first economic study to my knowledge that examines the impact of tribes on the allocation of development resources, has two central contributions. First, I extend the heterogeneity and public good model developed by ABE to demonstrate that areas with more tribes per capita will have higher per capita levels of local resources. Second, I create a unique dataset of tribal structure in over 1,000 administrative units in rural northern Yemen to test this model; I demonstrate that areas with more tribes per capita - i.e. areas with greater 'tribal heterogeneity' - do indeed have higher levels of resources.

The ABE model predicts a negative relationship between inter-group heterogeneity and the level of public good provision (i.e. municipalities with more ethnic groups provide less public goods). My theoretical contribution extends this model to demonstrate that inter-group heterogeneity has a *positive* impact on the per capita level of resources if individuals select both the type and level of the public good *within* their group. This revised structure is consistent with the *de facto* decentralized structure of rural Yemen, as well as many other rural areas in developing country contexts.

My model relies on an assumption motivated by the endogenous structure of tribes. Tribes formed during an agrarian period in order to provide public goods (e.g. agricultural risk diversification, defense, maintenance of markets) that were important for the success of agrarian rural societies (Dresch 1989, Weir 2007). The key implication of this endogenous structure is that the individuals' preferences for public goods (e.g. location of a school or well) will be more heterogeneous across tribes than within.

[1] While many of these studies use relatively stylized political models to understand the observed negative correlation between ethnic heterogeneity and public good provision in developed countries (e.g. Alesina and Ferrara 2005), more recent studies have examined the mircoeconomic mechanisms driving a similar negative relationship found in several developing countries (e.g. Miguel and Gugerty 2005; Habyarimana, Humphreys, Posner, and Weinstein 2007).

[2] Both Banerjee and Somanathan (2007) and Bros (2010) provide evidence that the caste structure may strengthen the ability of some Indian communities to extract resources from the state. However, both Alesina, Baqir, and Easterly (1999) and Alesina and Ferrara (2005) suggest that a positive correlation could be found between diversity and public good provision in the presence of ethnic patronage, and Alesina, Baqir, and Easterly (2000) provide evidence of its existence in U.S. cities.

My empirical analysis tests the relationship between inter-tribal heterogeneity and per capita development resource provision. While the model studies public good provision at the level of the tribe, I show how the predicted impact of tribe size - the specific form of heterogeneity implied by the model - can be tested using more aggregated data. This allows me to use data from rural subdistricts, the most disaggregated administrative structure in Yemen, to study the impact of average tribe size on log per-capita resources.

Tribal heterogeneity is calculated using local tribal identities. While these local tribes are embedded in a complex hierarchical structure - with every family maintaing local, regional, and national tribal identities - local tribes are typically believed to play the most important role in local politics and public good provision. Indeed, these local polities seem to have developed to provide public goods that were essential to the survival of communities during an agrarian period (Weir 2007). My subdistrict-level tribal dataset, which contains data on over 4,000 unique local tribal identities, allows me to calculate the number of tribes per capita for over 1,000 subdistricts.[3]

Development resources in this analysis focus on the education sector. Several school censuses and a nationwide educational facility survey provide a tool for measuring the local availability of educational resources. The school censuses provide an estimate of the number of school teachers for each subdistrict and the educational facility survey reports the number of school classrooms for each subdistrict.

The main empirical specifications includes geographic variables and regional fixed-effects to control for, respectively, the cost of providing educational resources and the potential impact of regional bureaucrats (e.g. influence of local Ministry of Education officials).[4] A variety of agricultural, population, and access to services controls allow me to examine the robustness of these results to factors that likely affected both the development of the tribes as well as the ability of the subdistrict to engage in political action independent of the tribe.

I first test the relationship between tribe size and educational resources as measured by the number of teachers and the number of classrooms in a given subdistrict. I find a negative and significant relationship between tribe size and both the number of teachers and classrooms which is consistent with the predictions of my theoretical model. The estimated relationship between tribe size and the number of teachers is robust to the inclusion of a rich set of controls - i.e. controls for population density, total population, number of villages, government services, agricultural conditions, and ruggedness of the terrain. However, the relationship between tribe size and school infrastructure is more tenuous; the significant

[3]The average subdistrict in my data contains 13 villages, 6,500 residents and 5 different tribes.

[4]Specifically, district fixed-effects are included as district-level offices are the most disaggregated Ministry of Education structures. Note that these offices are typically staffed by individuals from outside the area as an effort to prevent corruption.

relationship between tribal size and the number of classrooms disappears when additional controls are included.

A second empirical section, which studies the difference between the effect on educational infrastructure and educational employment, concludes that tribes are engaged in both patronage extraction and public good provision. The strong and robust impact of tribe size on the number of school teachers reflects tribes' role in the educational patronage system as suggested by previous authors (e.g. Posusney and Angrist 2005; DRI 2008, Phillips 2008). Indeed, it provides evidence that tribes are able to affect the allocation of the nearly 40% of school teachers that are "ghost teachers" - i.e. teachers that do not teach (ARD 2006). However, I also demonstrate tribes do have a significant and robust effect on school infrastructure built by Western donors. As Western-built infrastructure is less likely to be patronage construction, as Western donors insist on significant monitoring and oversight, this provides compelling evidence that tribes also provide modern public goods.

Throughout the analysis, identification rests on the assumption that the number of tribes is exogenous to the development of the education system. This is motivated directly by the theoretical model; as tribes formed during a historical period to provide a public good essential to that earlier period, the size of tribes today is not affected by modern public goods. There is significant empirical evidence to support this result. Phillips (2008) has argued that tribal structure has been stable since at least the time that the borders of subdistricts were demarcated in the 1930s-1960s, and potentially much longer. Similarly, both Dresch (1989) and Weir (2007) suggest that the tribes have maintained a very similar structure for hundreds of years through an Ottoman occupation, the rule of several Imams and the arrival of the Yemeni Republic. Given the recent arrival of widespread education in Yemen - the first rural schools did not appear until the 1960, and very few rural areas had schools before the mid-1970s - it seems unlikely that the education system has had a significant impact on tribal structure.

My analysis offers three other insights about the role of social institutions in development. First, most econometric analyses treat ethnic groups, caste structures, religious groups, etc. as well defined and internally homogenous structures. Here I provide a theoretical motivation, based on well characterized previous results, for why tribes are expected to be internally heterogeneous and provide empirical evidence consistent with this internal heterogeneity. As the development of Yemeni tribal structures mirrors the development of ethnic and other social identities in other rural developing country contexts, i.e. they emerged to provide public goods required of rural agrarian societies, it is likely that my results are applicable in other developing country contexts.

Second, my analysis suggests that engaging indigenous social structures, such as tribes,

may influence the effectiveness of development efforts. In particular, if these indigenous polities are indeed the primary actor in the distribution of development resources in rural areas, as compared to more aggregated state constructed polities, engaging these groups directly may enhance the effectiveness of these efforts.

Third, this analysis shows the important role that a non-Islamic institution can play in the modern development of a country of the Muslim Middle East.[5] Most discussions of the role of institutions in development in the Middle East have focused on Islam (cf. Kuran 2004, Chaney 2008). Additionally, while non-Islamic institutions may have a direct impact on development, as I have shown here, they may also have indirect impacts. As an example, there is significant evidence that tribal institutions have had an important impact on the development of legal systems across the Middle East (e.g. Charrad 2001, Weir 2007). Thus, a more careful examination of non-Islamic institutions is important in understanding development in the Middle East.

In the following section I extend two existing models of public good provision to provide intuition for why heterogeneity in the size of tribes should be expected and why smaller tribes should be more effective public good providers. Section 3 discusses the empirical approach including a motivation and description of the empirical estimation, the identification strategy, the measure of heterogeneity employed, and the other data that is used for the analysis. Section 4 examines the relationship between average tribe size and per capita educational resources. Section 5 then examines whether these tribes are engaged in public good provision or patronage extraction. Section 6 concludes.

2 Model

The model is developed in three subsections. The first describes the setup of the model. The second adapts the approach of Alesina, Baqir, and Easterly (1999) in describing the public good provision of tribes. The third section uses the endogenous structure of the tribes to demonstrate the positive expected relationship between tribal heterogeneity and the level of public good provision.

[5]Tribes are particularly interesting as they are a pre-Islamic institution in Yemen and many other Middle Eastern countries (Khoury and Kostiner 1990). Indeed, in the apocryphal tale describing the arrival of Islam to Yemen, the Islamic cleric that eventually started Islam in Yemen came originally at the behest of two powerful tribes in order to settle a dispute.

2.1 Setup

Consider a fixed geographical region of population P. I will refer to this geographical region as a subdistrict, as this will be the geographical unit that is the focus of the empirical work. The population P is composed of individuals with ideal points for public goods distributed uniformly on the segment $[0, P]$.[6] Individuals' utility is decreasing with the distance of their tribe from their ideal point. These ideal points should be interpreted as ideological, geographic, taste-based, or income-based preferences for the types of public goods that tribes can provide.[7] Each individual has ideal point for public goods important during an agrarian period (e.g. wells, conflict mediation) and a modern public good (e.g. school teachers, lobbying).

Public goods are provided by an individual's tribe and each individual's utility is given by

$$U_i = g^{\alpha}(1 - D_i) + y - \tau_i, \qquad 0 < \alpha < 1 \tag{1}$$

where g is his per-capita share of the public good provided by the tribe, D_i is the preference distance from individual i to the public good, y is private income, τ_i is the tax paid by the individual to the tribe, and α is a parameter that reflects individuals' distance-adjusted utility from consuming the public good (assumed to be constant across individuals). $y - \tau_i$ is private consumption.

2.2 Equilibrium Public Good Provision within a Tribe

Take a single tribe, n, among the N tribes in this subdistrict. The members of this tribe have to decide, by majority rule, on a modern public good, both on its size and type. After the public good is selected, a lump-sum tax is levied on all tribes members in order to fund the good. The lump-sum tax is assumed to be identical for everyone.

Using the implied budget constraint – i.e. $g = t$ – individual utility becomes

$$U_i = g^{\alpha}(1 - D_i) + y - g. \tag{2}$$

Solving for the equilibrium type and level of public good provision requires the following

[6] This assumption makes the model tractable in closed form. I discuss the implications of relaxing this uniformity assumption at the end of this section.

[7] Individuals may actually differ on multiple dimensions simultaneously, but here I map all types of heterogeneity into a single-dimension to makes the predictions of the model clear. Relaxing this single-dimensionality will not change the key comparative statics of Proposition **??** (see Section II.B of Alesina, Baqir, and Hoxby (2004)).

assumption about the voting process:

Assumption 1. *Individuals vote first on the amount of taxation, and then on the type of the public good.*

This assumption is made for tractability, in order to avoid issues of multidimensional voting. However, this assumption approximates political behavior within the tribes as tribal members are first taxed and then the leadership decide on how resources would be spent.

I use the median voter theorem to determine both the type and level of the public good provided. Though all members of a tribe are not typically consulted when these decisions are made, tribal elders *are* involved in these processes. As each tribal member is represented by an elder, this application of the median voter theorem provides an approximation for the median elder. This difference is important as it may affect the type of "public" good provided (i.e. it may be a public good provided only to the tribal elders involved in the political process).

I now solve the model backward, starting with the following result, which derives from a straightforward application of the median voter theorem.

Proposition 1. *For any positive amount of public good g, the type chosen is the one most preferred by the median member of the tribe.*

Let me now consider the choice of the size of the public good g. Individual i's preferred size is given by the solution to the following problem:

$$\max_g U_i = g^\alpha(1 - \widehat{D_i}) + y - g \tag{3}$$

where $\widehat{D_i}$ is the distance of the individual i from the ideal type of median voter. This formulation incorporates the fact that the voters know that, after a decision is reached on the size of g, the type chosen is the one most preferred by the median voter. The solution of (3) is

$$g_i^* = \left[\alpha(1 - \widehat{D_i})\right]^{1/(1-\alpha)} \tag{4}$$

Define $\widehat{D^m}$ as the median distance from the type most preferred by the median voter - in short the "median distance from the median". A straightforward application of the median voter theorem leads to Proposition 2:

Proposition 2. *The amount of public good provided in equilibrium is given by*

$$g_i^* = \left[\alpha(1 - \widehat{D^m})\right]^{1/(1-\alpha)} \tag{5}$$

2.3 Tribe Size and Public Good Provision

My key result derives from the endogenous structure of tribes. Tribes formed during an agrarian period in order to provide public goods that were important for the success of agrarian rural societies (Dresch 1989, Weir 2007). In particular, tribes organized a variety of goods with economies of scale (e.g. agricultural risk diversification, defense, and maintenance of markets) for members of the tribe.

The key implication of this endogenous structure is that there will be more inter-tribal *(across)* than intra-tribal *(within)* heterogeneity with respect to the distribution of ideal points for the agrarian public good. Thus, increasing the number of tribes in a subdistrict will decrease the internal heterogeneity of each tribe, increase the per capita public good provision of each tribe, and consequently increase the per capita public good provision of the entire subdistrict.[8]

The following assumption, which makes my result tractable in closed form, is motivated by this endogenous structure. Note that this is a limiting case as it is analogous to assuming that (1) tribal structure achieves the social welfare maximum with respect to providing an agrarian public good with economies of scale, as characterized by Alesina and Spolaore (1997), and (2) individuals' ideal points are perfectly correlated across the agrarian and modern public good.[9]

Assumption 2. *All tribes in a subdistrict are of equal size and each tribe represents a contiguous segment of the distribution of ideal types for the public good.*

Given Assumption 2 and the assumption that the population, P, is uniformly distributed on $[0, P]$, the median distance from the median, $\widehat{D^m}$, becomes

$$\widehat{D^m} = \frac{1}{4} \cdot \frac{P}{N} \tag{6}$$

so that Equation 5 becomes

$$g_i^* = \left[\alpha \left(1 - \frac{1}{4} \cdot \frac{P}{N}\right)\right]^{1/(1-\alpha)} \tag{7}$$

[8]This analysis takes the number of tribes formed during this agrarian period as given. Footnote 9 provides some intuition for why the number of tribes is likely to vary across subdistricts.

[9]The equivalence of Assumption 2 to these two assumptions requires one further assumption - i.e. that there is a strong positive correlation between geographical location and preferences for a public good. Under these three assumptions, Proposition 1 of Alesina and Spolaore (1997) follows directly and each tribe in a subdistrict will be (1) equally sized and (2) cover a continuous range of ideal points for that subdistrict. This equivalence provides intuition for why we would expect variation in the size of tribes across districts. In particular, Proposition 1 of Alesina and Spolaore (1997) implies that tribe size should be inversely related to the cost of providing public goods in a subdistrict .

Equation 7 demonstrates that public good provision should be *increasing* with the number of different tribes in a given subdistrict. A proof of this result is presented in Appendix A.

Three assumptions drive the closed form result presented in Equation 7. The first assumption, Assumption 2, is equivalent to assuming that tribes were formed in order to minimize heterogeneity within the tribes. However, the qualitative interpretation of Equation 7 is maintained for any assumption that results in less intra-tribal than inter-tribal heterogeneity. As an example, the same negative relationship is obtained if (1) individuals migrated across tribes in response to their access to the agrarian public good, which is suggested by the gradual migration rate of individuals between tribes,[10] and (2) ideal types for the agrarian and modern public goods are positively correlated. The migratory process would equalize the maximal distance from the median across tribes so that the median distance from the median would decrease as the number of tribes increased.

The second assumption, that the distribution of ideal types is uniform, is not restrictive; the qualitative result will not change for more general distributions. As long as tribes are of equal size (i.e. Assumption 2 is maintained) the range covered by each tribe, as well as the median distance from the median, will decrease monotonically with the number of tribes regardless of the distribution of types.

The third assumption is that α is constant across individuals, and thus across subdistricts. If α, which measures individuals' distance-adjusted utility from consuming the public good, is allowed to vary across subdistricts, then it is not possible to identify the independent impact of tribe size on public good provision. While α may differ across regions for certain classes of public goods (e.g. girls' education), I believe that this is a reasonable assumption for the problem studied here (i.e. the value of boys' education is widely appreciated throughout rural Yemen).

3 Empirical Approach

In this section I elaborate my empirical approach in three subsections. The first uses results from my theoretical discussion to motivate my basic estimating equation, the second explains my identification strategy, and the third describes the data used in the empirical analysis.

[10] The migration of individuals between adjacent tribes is a common, if infrequent, occurrence in rural Yemen. Approximately 1% of the households of any given tribe will migrate to another tribe every decade (author's calculations using a qualitative village survey instrument).

3.1 Estimating the Impact of Diversity on Patronage

My empirical strategy is based on Equation 7 though my unit of analysis is the subdistrict and not the individual. Since all tribes in a subdistrict are of equal size, per-capita public good will be equalized across tribes. Consequently, per-capita public good provision will be identical for every individual i in a given subdistrict and it follows that

$$g_s = \left[\alpha\left(1 - \frac{1}{4}\cdot\frac{P_s}{N_s}\right)\right]^{1/(1-\alpha)} \tag{8}$$

where s indexes subdistricts, g is per-capita public good provision, P is total population, N is the number of tribes, and α is a parameter that reflects individuals' distance-adjusted utility from consuming the public good (assumed to be constant across subdistricts).[11] I can take logs of both sides and re-write this equation as

$$\log(G_s) - \log(P_s) = \frac{\log(\alpha)}{1-\alpha} + \frac{1}{1-\alpha}\log\left(1 - \frac{1}{4}\frac{P_s}{N_s}\right). \tag{9}$$

$$\approx \frac{\log(\alpha)}{1-\alpha} - \frac{1}{1-\alpha}\left(\frac{1}{4}\frac{P_s}{N_s}\right) \tag{10}$$

where G_S is total public good provision in subdistrict s.[12] The second (approximate) equality follows from the empirical definition of P_s used here. In particular, P_s is calculated as the total quantity of men in a subdistrict divided by 10,000 so that $\left(1 - \frac{1}{4}\frac{P_s}{N_s}\right)$ is close to one.

Equation 10 motivates my basic estimating equation

$$\log(G_s) = \beta_0 + \beta_1\left(\frac{P_s}{N_s}\right) + \beta_2\log(P_s) + \epsilon_s \tag{11}$$

where the subscripts s denotes subdistricts and ϵ_s is an error disturbance which is assumed to be i.i.d. As the total quantity of public good provision should be decreasing in the average tribe size, $\frac{P_s}{N_s}$, my theoretical predicts that the sign on β_1 should be negative.

3.2 Identification Strategy

Identification here rests on the central assumption that the observed number of tribes within an administrative subdistrict is exogenous to the modern public goods that are the focus of this analysis. This assumption requires that these public goods have not affected the number of tribal units within a subdistrict.

[11] See Section 2 for a further discussion of the α parameter.
[12] Note that $\log(g_s) \equiv \log(G_s) - \log(P_s)$.

This assumption is motivated directly by the theoretical model; as tribes formed during a historical period to provide a public good essential to that earlier period, the size of tribes today is not affected by modern public goods. This assumption is supported by the very slow rate of adaptation of the tribes. Dresch (1986, 1989) uses historical texts to show that many tribes have maintained the same borders for over a millennium. Weir (2007), who uses legal documents maintained by the tribe themselves in her analysis, similarly concludes that the tribes of her study, which do not overlap with those of Dresch, were stable for over four hundred years.[13]

The public goods that are the focus of this analysis are a very recent phenomenon. Access to education, health, and modern roads was extremely limited throughout the country until the 1960s and most rural areas did not have access to these facilities until the 1970s at the earliest (Appendix B provides more details on the emergence of the education system that is the focus of this analysis). Thus, given the historical stability of the tribes, and the slow rate of evolution, it is unlikely that these modern public goods have affected the structure of the tribes.

3.3 Data

In this subsection I describe the tribal and educational data that are at the core of this analysis. A description of the administrative structure and the variety of population, area, terrain, economic and agricultural controls used throughout the analysis is deferred to Appendix C though Table 2 provides key summary statistics for these variables.

3.3.1 Tribal Data

For this analysis I collected data on the tribal structure in 1,073 rural Yemeni administrative subdistricts from eight Yemeni governorates.[14] For each governorate studied, I identified a Yemeni research assistant from that governorate to assist me in collecting the data. These research assistants then collated information on the tribal structure in each subdistrict of their respective governorates through a combination of field visits to district capitals and conversations with friends and family members.

The variable that is the focus of this analysis is the number of local tribes per subdis-

[13]Though tribes do evolve over time, qualitative research done in conjunction with this project suggests that only one percent of tribal families will switch tribes in any given decade (author's estimates using qualitative village survey data).

[14]There are a total of 19 rural governorates with a total of 2,031 rural administrative subdistricts. These eight governorates are not fully representative of Yemen as they are all drawn from the tribal regions of northern Yemen. However, they still provide a useful sample for testing the plausible impact of tribes, as is the focus of this analysis.

trict. Local tribes is defined here as the most disaggregated tribal identity identifiable by my research assistants.[15]

Figure 1 illustrates the significant local variation in tribal structure within one administrative district. As most of my empirical specifications include district-level fixed effects, in order to control for unobserved local factors that might affect tribal size and access to educational public goods, my results are identified off of this within-district variation. Figure 2 then demonstrate the variation of tribal diversity across one governorate (this is the same governorate from which Figure 1 was drawn). Finally, Figure 3 describes the full coverage of my data; areas colored in blue are places for which there is no tribal data available.

Table 2 reports summary statistics for the average size of tribes at the subdistrict level, which is the tribal measure motivated by the theoretical analysis.

3.3.2 Education Data

The education variables that are the primary outcome of interest in this analysis are drawn from two key sources: (1) three national school surveys conducted during the 1999-2000, 2000-2001 and 2005-2006 school years and (2) an educational facility census from 2007. Table 1 provides summary statistics for the variety of educational variables that are used as dependent variables throughout this analysis.

Two key types of variables are used for the analysis. The first type of variable is the number of male school teachers per subdistrict. This analysis excludes female teachers as they tend not to function as substitutes for male teachers - as female teachers typically will not teach male students in rural areas - and the politics of female education are not the focus of this analysis. Data for this variable is available for all rural schools in each of the three school surveys. These surveys contain data for 11,916 *(1999-2000)*, 12,748 *(2000-2001)*, and 14,367 *(2005-2006)* rural schools.

The second type of data is the number of classrooms per subdistrict. This data is drawn from the 2007 educational facility census which provides detailed data on school infrastructure. While this facility census provides more detailed information on school infrastructure, the available data is somewhat incomplete and only contains data on 11,375 schools.

These data allow me to identify the agent who funded the construction of each classroom. I aggregate these agents into seven different groups: (1) Arab donors - Egyptian, Emirati, Iraqi, Kuwaiti, and Saudi donors, (2) Western donors - American, Asian, and European donors, (3) local donors - Yemeni individuals, private corporations, and religious

[15] In practice, this was predominantly tribes identified of the "fourth branch" - i.e. four times removed from national-level tribal confederations - though in some cases they were identified as "fifth branch". Though I also have data on all the higher order affiliations of each of these local tribes, I use the disaggregated tribal identities in calculating my measures of tribal diversity following the advice of several Yemeni colleagues.

charities, (4) local councils - district parliaments designed to interface between government ministries and citizens of the district, (5) government ministry - the President's Office, the Office of Public Works, and the Ministries of Agriculture, Education, and Electricity, (6) Social Fund for Development (SFD) - Yemeni parastatal development organization that receives funding from foreign and Yemeni governments, and (7) community - built by the local community only.

There are two important differences between the first and second measure. First, while the second measures the total *stock* of physical educational infrastructure built locally by communities, central government ministries and development donors, the first measures the *flow* of the educational public good to a subdistrict. Second, the number of teachers per subdistrict is only a public good if those teachers are actually teaching. A discussion of this latter point is the focus of Section 5 below.

4 Educational Resources and Tribe Size

This section focuses on estimation of Equation 11. However, I will estimate

$$\log(G_{d,s}) = \beta_0 + \beta_1 \left(\frac{P_{d,s}}{N_{d,s}} \right) + \beta_2 \log(P_{d,s}) + \gamma' X_{d,s} + \eta_d + \epsilon_{d,s} \tag{12}$$

where d denotes subdistricts, s denotes subdistricts, $\log(G_{d,s})$ is the log of total public good availability, $\log(P_{d,s})$ is the log of the total population, and $\epsilon_{d,s}$ is an error disturbance which is assumed to be i.i.d. Equation includes two types of controls not predicted directly by the theory: (1) district fixed-effects (η_d) which adjusts for the fact that educational resources are typically allocated through district-level representatives and (2) a vector of geographic and population variables (X_s) that control for the cost of of providing educational resources.

The following subsection examines the relationship between tribal size and the amount of public goods available. A second subsection examines the robustness of these results to a broader range of control variables.

4.1 Tribe Size and the Availability of Educational Resources

Table 3 examines the impact of tribe size on the distribution of educational resources. Columns (1) and (4) report the most basic regression which compares the impact of tribe size on local access to educational resources across all subdistricts. Columns (2) and (5) include district fixed effects so that the point estimate on tribe size reports the impact *within* districts. Columns (3) and (6) look at the impact of tribe size controlling for factors that

potentially affect the cost of providing those resources.

The key result of Table 3 is found in columns (3) and (6) which demonstrate the significant negative relationship between tribe size and both the flow of educational resources (i.e. the number of teachers) and the stock of educational resources (i.e. the number of classrooms) at the subdistrict level. These results provide empirical evidence for the key theoretical result presented in Equation 7. Importantly, the significance of the relationship differs across the two differ measures; understanding this difference is the focus of Section 5.

There are two other important differences between the impact of tribe size on the flow of educational resources and the stock of educational resources. The first important difference is that the impact of tribe size on the stock of educational resources seems is only *within* districts while the impact of tribe size on the flow operates both *across* and *within* districts. This difference can be seen by comparing the difference in the estimated relationship of tribe size across columns (1) and (2) to the difference across columns (4) and (5). This difference reflects the difference in the nature of the two educational resources. While the construction and placement of new educational infrastructure necessarily involves district level bureaucrats, as permission for new construction typically requires obtaining permission from local councils, the distribution of teachers is more fungible. Indeed, as I find that tribe size affects the number of teachers with and without fixed effects, this indicates that tribes are able to petition both district and higher-level bureaucrats to obtain additional positions.

The second difference between the results for these two different variables is that the estimated relationship for variables that proxy for the cost of providing educational resources differ. In particular, though the impact of most of these control variables is similar, which can be seen by comparing columns (3) and (6), the estimated impact of population density differs substantially across the two variables. The estimated negative impact of population density on the number of classrooms, even if insignificant, is unsurprising as classroom sizes are necessarily smaller in low density areas as children are limited in the distance that they can travel to attend school. However, the estimated *positive* and significant relationship of population density on the number of teachers is more unexpected.

4.2 Robustness of Results

The focus of the empirical analysis presented directly above (Section 4.1) was to test the predictions of the theoretical model. Thus, only a limited number of control variables, i.e. those factors likely to affect the cost of providing educational resources, were included in the analysis. In this Section I examine the robustness of these results to the inclusion of a variety of agriculture and access to services control. The agricultural variables provide proxies for

agricultural wealth and local political structure (see discussion below) and the access to service variables provide additional controls for local wealth and government penetration into the region.

The key result to emerge from Table 4 is that the estimated relationship between tribe size and educational resource provision is robust to a diversity of agricultural and access to service controls. However, while the significance of the estimated impact of tribe size on school teachers is not affected by the inclusion of these variables, the impact on school infrastructure seems somewhat more fragile.[16] This fragility is the focus of the subsequent section.

Several other important results emerge from this robustness analysis. First, Table 4 provides further evidence of the importance that local political structure can have on access to development resources. This is demonstrated most strongly in columns (2)-(3), though also in columns (5)-(6), by the point estimates on the two variables that measure the structure of land ownership in these rural areas. The positive coefficients indicate that areas with more small landholders provide more of these development resources.[17]

The importance of local non-agricultural wealth in the provision of these educational resources is also demonstrated in Table 4. Indeed, the share of households without sanitation and the share of households using primitive cooking fuels exhibit a robust negative relationship for both teachers and for school infrastructure. These two variables are the best proxies for local non-agricultural wealth available in these data.

The results for agricultural wealth are more mixed. The share of cultivable land and the number of sheep and goats both exhibit a positive relationship with educational resource provision, suggesting a positive relationship for agricultural wealth. However, the production of both qat and grains show a weakly negative relationship with educational resource provision. The difference between these results likely reflects the political implications of each: areas with more cultivable land are more sheep and goats tend to be more equal while areas with high production of the the two most profitable crops are more unequal.

5 Educational Goods: Public Goods or Patronage?

The results from Section 4 provided mixed results on the relationship between the size of tribes and educational public goods. While the number of teachers in a subdistrict exhibited a strong, significant, and robust negative relationship with the size of tribes as predicted by

[16]The loss of significance between columns (4) and (5) is partially because of the reduced number of observations available for the specification in column (5). See Appendix C.3 for more details.

[17]The omitted categories are households with more than 20,000 square meters and households with no land.

the model, the relationship between the size of tribes and school infrastructure was more fragile.

This difference suggests that tribal heterogeneity affects both access to real educational resources, i.e. classrooms and teachers that actively teach, as well as the ability of tribes to extract patronage from the central government. This result is consistent with the analysis of ARD (2006) who conclude that "ghost teacher" positions, which account for an estimated 40% of all teacher positions, play a key role in Yemen's patronage system. Thus, this result is consistent with previous analyses who have identified the importance of the tribes in Yemen's patronage system (e.g. Posusney and Angrist 2005; DRI 2008, Phillips 2008).

However, while all of these previous analyses have argued forcefully for the role of Yemen's patronage system, none of them have suggested a role for tribes in the provision of modern public goods. Thus, these previous analyses raise an important question: are tribes actually involved in the provision of modern public goods or only in the extraction of patronage?

Though the results for the number of classrooms in column (6) of Table 3 provide some evidence that the tribes are indeed engaged in the provision of modern public goods, it is possible that this result also reflects patronage extraction as tribes secure funding for classroom construction in order to improve their prestige or benefit (either directly or indirectly) from the funds that are used to construct the school. Thus, in Table 5 I re-examine the impact of tribe size on the availability of educational infrastructure by decomposing the number of classrooms by the type of donor funding its construction. This table reproduces the specification in column (6) of Table 3 for each of seven different classes of development donors.

The key result from Table 5, which is presented in column (2), is that the impact of tribe size on local infrastructure happens almost entirely through classrooms built by Western donors. Indeed, while the predicted negative relationship is observed for other types of donors (i.e. local donors, government ministries, SFD), the impact is only significant for Western funded classrooms. Further, as demonstrated in Table 6, this result is robust to the inclusion of geographic, population, access to services, and agricultural controls.

As Western donors typically require significant oversight of how their resources are distributed, it is unlikely that the strong and significant result observed in Table 6 reflects patronage extraction by tribes. Thus, while the comparably strong impact of tribe size on the number of teachers suggest that tribe size does affect patronage extraction, this result provides compelling evidence that tribes also play a role in the provision of public goods.

6 Conclusion

This paper has made several contributions to our understanding of social structures in the developing world. The first one is to demonstrate, using a new dataset of unique tribes covering nearly one-half of rural Yemen, that tribes play an active role in the provision of development resources in rural Yemen. This has important implications for decentralization and other types of development reforms in Yemen as it suggests that tribes will play an important role in these reforms' success.

Second, my findings demonstrate how increased heterogeneity may enhance the ability of groups to extract patronage from the state. Though this result is consistent with previous analysis of patronage in other contexts, I demonstrate how this result is predicted by existing models of heterogeneity and public provision. As many indigenous social structures, like tribes, form in areas with significant underlying heterogeneity, it is likely that these results are generalizable to other similar contexts.

Third, these findings demonstrate that non-Islamic institutions can have an important impact on development outcomes in a country of the Muslim Middle East. Thus, the focus on Islam in current discussions of Middle East development draws attention away from other social institutions that also play an important role in development.

This paper also provides insights into the role that tribes plays in Yemen's patronage system. Though a variety of other other observers have commented on the importance of the tribes in this patronage system (e.g. Posusney and Angrist 2005; DRI 2008; Phillips 2008), mine is the first to provide clear empirical evidence of how tribal structure affects the ability of tribes to extract resources from the central government.

References

Alagbari, B. R. (1992). نظام التعليم في الجمهورية اليمنية *(The Educational System in the Republic of Yemen)*. صنعاء: دار اقراء.

Alesina, A., R. Baqir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*.

Alesina, A., R. Baqir, and W. Easterly (2000). Redistributive Public Employment. *Journal*

of Urban Economics 48(2), 219–241.

Alesina, A., R. Baqir, and C. Hoxby (2004). Political jurisdictions in heterogeneous communities. *Journal of Political Economy 112*(2), 348–396.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature 43*(3), 762–800.

Alesina, A. and E. Spolaore (1997). On the number and size of nations. *The Quarterly Journal of Economics*.

Alwashli, Y. (2007). دراسة في اليمن: سياسية بناء قوة الدولة، دراسة جيوستراتيجية *(A Study in Yemen: Builiding the Political Power of the State, A Geostrategic Study)*. صنعاء.

ARD (2006). Yemen Corruption Assessment. Technical report, US AID.

Banerjee, A. and R. Somanathan (2007, March). The political economy of public goods: Some evidence from India. *Journal of Development Economics 82*(2), 287–314.

Bros, C. (2010). Social Fragmentation and Public Goods: Polarization, Inequality and Patronage in Uttar Pradesh and Bihar. Documents de Travail 26, Centre d'Economie de la Sorbonne.

Burchfield, M., H. G. Overman, D. Puga, and M. A. Turner (2006). Causes of sprawl: A portrait from space. *Quarterly Journal of Economics 121*(2), 587–633.

Chaney, E. (2008, November). Tolerance, religious competition and the rise and fall of muslim science.

Charrad, M. (2001). *States and Women's Rights: The Making of Postcolonial Tunisia, Algeria, and Morocco*. Berkeley: University of California Press.

Contin, R., D. Egel, K. Moore, and H. Ogleh (2009). Chf yemen access-plus baseline report: A report on child labor and education in aden, hajja, hudeida and taiz governorates in yemen. Technical report, CHF International.

Dresch, P. (1986). Significance of the course events take in segmentary systems. *American Ethnologist 13*(2), 309–324.

Dresch, P. (1989). *Tribes, Government and History in Yemen*. Oxford University Press.

DRI (2008, November 2008). Yemen: Comprehensive assessment of eleciton framework. Final report, Democracy Reporting International.

Habyarimana, J., M. Humphreys, D. Posner, and J. Weinstein (2007). Why Does Ethnic Diversity Undermine Public Goods Provision? An Experimental Approach. *American Political Science Review 101*(4), 709–725.

Khoury, P. and J. Kostiner (Eds.) (1990). *Tribes and State Formation in the Middle East*. London: Tauris.

Kuran, T. (2004). Why the middle east is economically underdeveloped: Historical mechanisms of institutional stagnation. *The Journal of Economic Perspectives 18*(3), 71–90.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions and public goods in Kenya. *Journal of Public Economics 89*, 2325–2368.

Nunn, N. and D. Puga (2009). Ruggedness: The Blessing of Bad Geography in Africa. *NBER Working Paper 1498*.

Phillips, S. (2008). *Yemen's Democracy Experiment in Regional Perspective*. Palgrave Macmillan.

Posusney, M. P. and M. P. Angrist (Eds.) (2005). *Authoritarianism in the Middle East: regimes and resistance*. Boulder, Colorado: Lynne Rienner Publishers.

Riley, S. J., S. D. DeGloria, and R. Elliot (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences 5*(1-4), 23–27.

Sappington, J. M., K. M. Longshore, and D. B. Thompson (2007). Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study Using Bighorn Sheep in the Mojave Desert. *Journal of Wildlife Management 71*(5), 1419–1426.

Weir, S. (2007). *A Tribal Order: Politics and Law in the Mountains of Yemen*. University of Texas Press.

World Bank (2006). Tracking basic education expenditures in yemen: Analyses of public resources management and teacher absenteeism. Technical report, Social and Economic Development Group, Middle East and North Africa Region.

A Proof of Equation 6

Equation 6 follows directly from Assumption 2 and the uniform distribution assumption. From Assumption 2 it follows that the size of each tribe is $\frac{P}{N}$. Without loss of generality, consider the tribe on the leftmost end of the range of ideal types for the public good - i.e. the tribe that covers $\left[0, \frac{P}{N}\right]$. As the population is uniformly distributed, the median voter will locate at the point $\frac{1}{2}\frac{P}{N}$. The median distance from this point is $\frac{1}{4}\frac{P}{N}$.

B The Recent Expansion of Yemen's Education System

The education system that exists today in the north of the Republic of Yemen is a very recent phenomenon. The first public schools did not open their doors until the early 1960s, and while the educational system did expand rapidly from the late 1970s until the present, the overall penetration of the education system into rural areas is still quite limited. Importantly, given its recent arrival and the stability and longevity of the tribes discussed above, it is unlikely that the education system has affected the structure of the tribe.

In northern Yemen in 1962, the year of the revolution which saw the removal of a religious monarchy, there were only 23 schools serving a population of over 4 million.[18] And while the education system did expand in the wake of the 1962 revolution, the real expansion of the education system did not happen until the late 1970s.[19]

The rapid expansion of the education system in the late 1970s, and its continued expansion through the 1980s and 1990s, is illustrated in Figure 4 which shows the construction of classrooms during the 20th century.[20] Today, with over 16,000 schools in the education system, there is a school for every third village. However, despite the apparent widespread availability of schools, and the significant resources that are spent on education[21], enrollment rates are still quite low with 68% of eligible boys and 45% of eligible girls currently enrolled.[22]

C Data Appendix

C.1 Yemen's Administrative Structure

The Republic of Yemen currently has three main levels in its administrative hierarchy.[23] The largest administrative structure is the governorate of which there are a total of 21, six of these are from the former People's Democratic Republic of Yemen and the remaining 15 are from the former Yemen Arab Republic (North Yemen). These 21 governorates are then sub-divided into a total of 333 districts which are further divided into nearly 2,200 subdistricts, the smallest official administrative structure. The nearly 40,000 villages, which are themselves composed of approximately 200,000 subvillages, do not have any official status and are typically a locally defined concept.[24]

[18] Under the rule of the Imam, education was done almost entirely in mosques, focused on religious study and the study of Islamic jurisprudence and was restricted to only the elite.

[19] Though Arab and other foreign countries did play an important role in the provision of teachers and the development of curriculum in the early years of this expansion, as is often discussed, local communities also played a central role (Alagbari 1992). Indeed, nearly 60% of the school capacity built in Yemen was built by local communities without the financial support of either the central government or foreign donors (author's calculations using the 2007 Education Census).

[20] The initial expansion of the education system was driven by the first oil boom as the remittances earned by Yemeni workers employed in Saudi Arabia and the Gulf surged. Alwashli (2007) provides a more expansive discussion of the expansion of the education system after the Revolution.

[21] Education is the largest component of current government expenditures and accounts for around 20% of the total budget (though the actual share varies substantially from year to year though it has remained between 15-25% of the budget in recent years).

[22] In addition to lack of access to educational facilities, two central explanations are usually offered for these low enrollment rates. The first is the high cost of education. Though school is purportedly freely provided, a variety of fees are typically charged to students which are often prohibitive (see Contin, Egel, Moore, and Ogleh (2009) for a discussion of this). Note that the Ministry of Education in cooperation with the World Bank and the European Council are currently experimenting with several conditional cash transfer programs to help alleviate this difficulty. The second is the particular severity of teacher absenteeism in Yemen with estimates of absenteeism ranging from 16% (World Bank 2006) to around 50% (Contin, Egel, Moore, and Ogleh 2009).

[23] A variety of other structures, such as subgovernorates, were used in the past but do not find much practical use today.

[24] It is thus impossible to calculate the actual number of villages in Yemen as villages are defined differently in different data sources. Indeed, while a particular hamlet may be reported in a census as part of another

For the rural populations that are the focus of this study, the most recent population census from 2004 reported a total population of just over 14 million individuals residing in 38,736 villages.[25] The median district had a population of approximately 35,000, though there is significant variation in the population of these governorates as the smallest governorate had a population of under 2,000 while the largest had a population of nearly 200,000. And the median subdistrict had 4,000 inhabitants though again there was relatively significant variation of over 8000 with subdistrict populations ranging from only one hundred individuals to approximately 80,000.[26]

C.2 Population and Economic Controls

This analysis uses two types of data from the 1994 and 2004 Population Censuses: (1) population counts and (2) measures of access to government services. The summary statistics for these variables are included in Table 2.

The key independent variable in my analysis is average tribe size per subdistrict. Importantly, the numerator for this independent variable is calculated differently for the analysis of teachers and of school infrastructure. As both the total population and the number of teachers per subdistrict are flow variables - i.e. they change from year to year - I calculate the average tribe size variable separately for each year of teacher data. In particular, I use the 1994 and 2004 Population Censuses to interpolate/extrapolate the population counts for 1999-2000, 2000-2001, and 2005-2006. The analysis of the infrastructure data uses the population estimates from the 2004 Population Census directly.

The second type of data are variables that measure access to public services available for each subdistrict. These variables, which are drawn from the 2004 Population Census only, include (1) percent of citizens without access to sanitation, (2) percent of citizens without access to electricity, (3) percentage of households using wood, coal and kerosene for cooking, and (4) percentage of households without water from a pipe. While these data are available at the village level, subdistrict average/aggregates were calculated for the analysis considered here using the total population in a village as weights.

C.3 Agricultural Controls

In order to control for the potential impact of agricultural factors, this analysis include a variety of agricultural variables drawn from the 2001 Agricultural Census. These variables include: the total amount of land owned by private individuals, the share of the land that is cultivable, the type of water access that is available, the size of land holdings, the amount of animal assets held and the amount of land devoted to grain, *qat* and cash crop production. This last variable is of particular importance as *qat*, which requires significant amounts of

village by a local informant, the informant that is interviewed in another census may indicate that the local hamlet is actually another village. In general, the census officers and field workers defer to the judgement of the local informant which is typically a local elder or leader. In many cases this leads to settlements with only one household and 8-10 members being identified as a separate village in the census.

[25] The total population in 2004 was just over 20 million.

[26] In the southern govenorates, i.e. those of former South Yemen, subdistricts are much more rare and the district is often the smallest administrative structure above the village.

water and grows only in specific climates, is the only true cash crop in Yemen. Though these data are available at the village level, as the primary unit of analysis here is the subdistrict, I calculate subdistrict averages for each of the variables using the total population in each village as the weight for that village. The full list of variables as well as their means and standard deviations are included in Table 2.

Importantly, the agricultural census does not include all of the governorates for which I have tribal data. The governorate of Al-Jawf was not included in the agricultural census as the Ministry of Agriculture and Central Statistical Office judged that the data from this governorate were systematically biased - it was suspected that residents of these areas were instructed to lie about their assets and land by their governor.

C.4 Area and Terrain Controls

Three area and terrain controls are included in this analysis: (1) population density, (2) distance from district capital, and (3) terrain ruggedness. Table 2 provides summary statistics for these variables.

Population density is calculated as the number of individuals per 1,000 square meters. This measure divides the population estimate for a given subdistrict (from the population census) by the area of the subdistrict. The area of each subdistrict was extracted from ArcGIS maps that accompanied the 2004 Population Census.

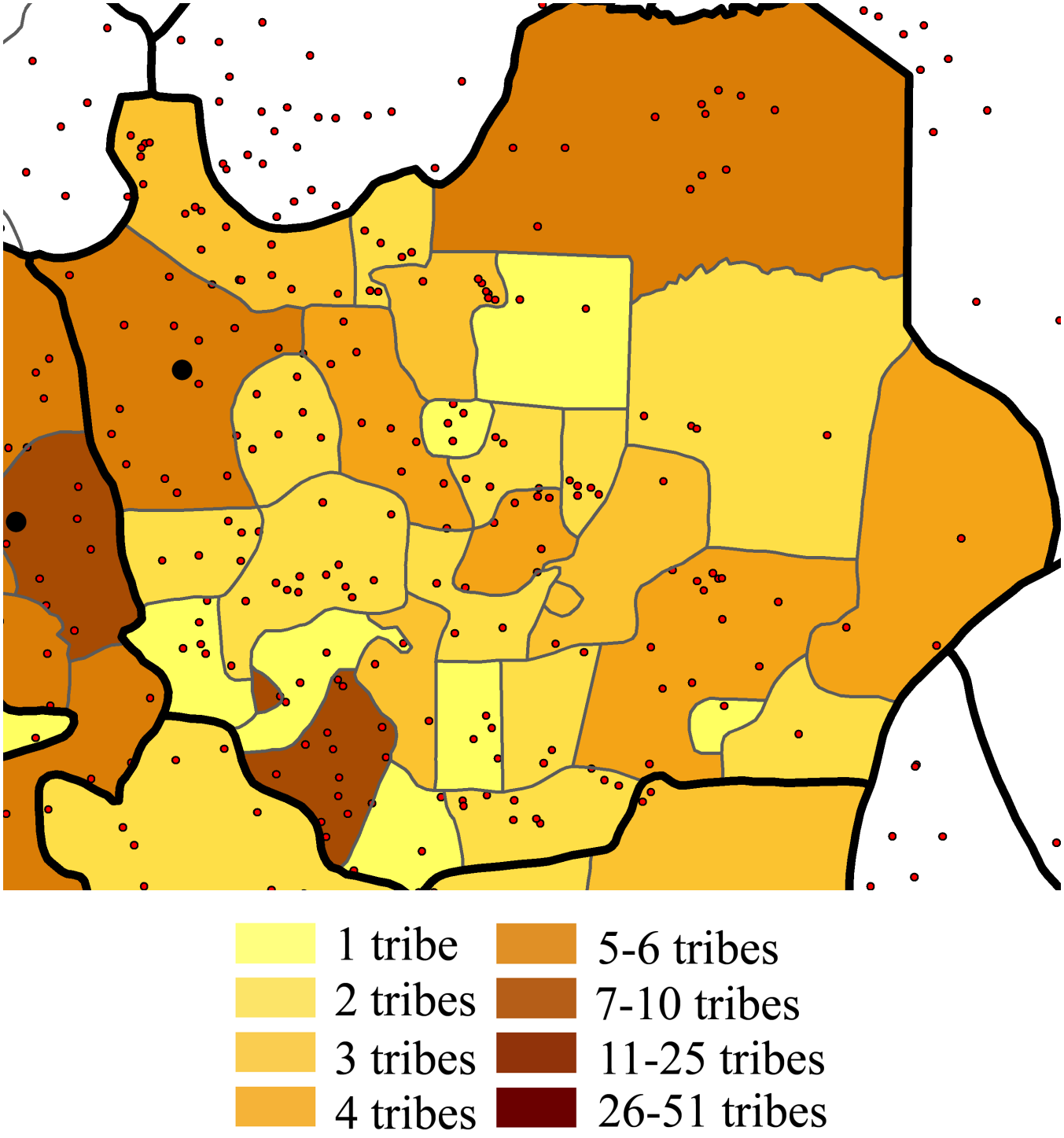
The distance of a subdistrict from the district capital is calculated using the geographical data available for each data. These data allow me to estimate the population centroid for each subdistrict - i.e the population-weighted average of the latitude and longitude location of each village. The distance of a subdistrict to the capital is then calculated as the log distance of this population centroid to the municipal center.

For my measure of terrain ruggedness I use the vector ruggedness measure (VRM) of Sappington, Longshore, and Thompson (2007).[27] Though other studies in economics have focused on the terrain ruggedness index (TRI) of Riley, DeGloria, and Elliot (1999) (cf. Burchfield, Overman, Puga, and Turner 2006; Nunn and Puga 2009), there are two reasons that I have opted to use the VRM. The first is that the VRM is more appropriate for the current analysis as it quantifies ruggedness independently of slope. This is important because even steep terrain is relatively easy to traverse if it is not uneven and broken. Second, calculation of the VRM is facilitated by the availability of a publicly available toolbox for ArcGIS that is designed to calculate this measure.[28]

[27] The USGS GTOPO30 file 'e020n40', which reports the elevation at approximately one kilometer intervals for the Arabian peninsula and eastern Africa, was used for these calculations. It was downloaded from http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info.

[28] This toolbox is available from http://arcscripts.esri.com/details.asp?dbid=15423. It is important to note that the TRI measure also has a script available to facilitate calculation (http://arcscripts.esri.com/details.asp?dbid=12435). However, this script is in practice quite difficult to implement with ArcGIS desktop as it was developed for ArcInfo Workstation.

Figure 1: Tribal Diversity in One District



Note: This figure shows the number of tribes per subdistrict in one district of the governorate of Dhamar. The small red dots indicate settlements and the larger dark dot displays the capital of the district.

Figure 2: Tribal Diversity in One Governorate

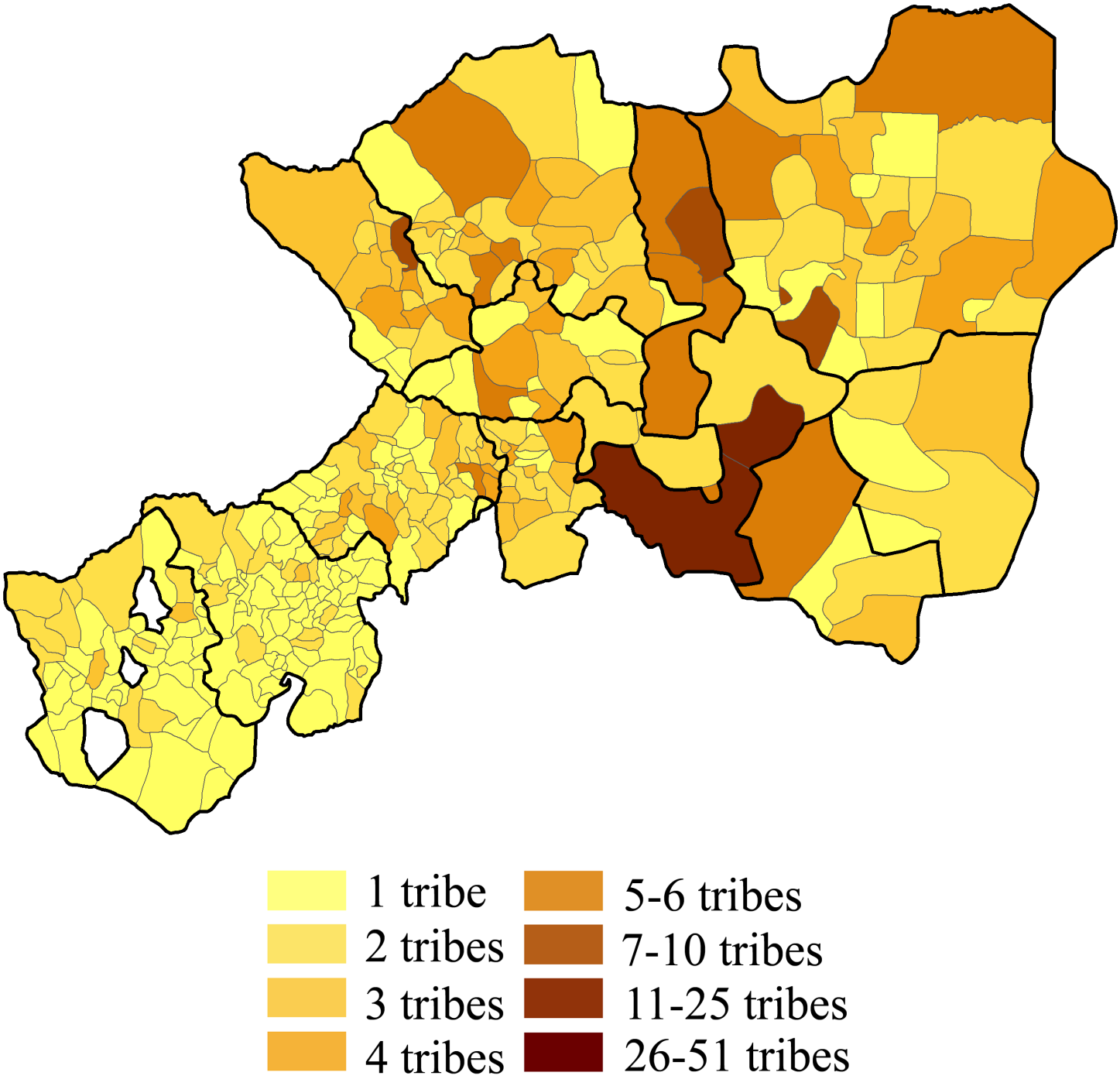


Figure 3: Tribal diversity in Eight Governorates of Northern Yemen

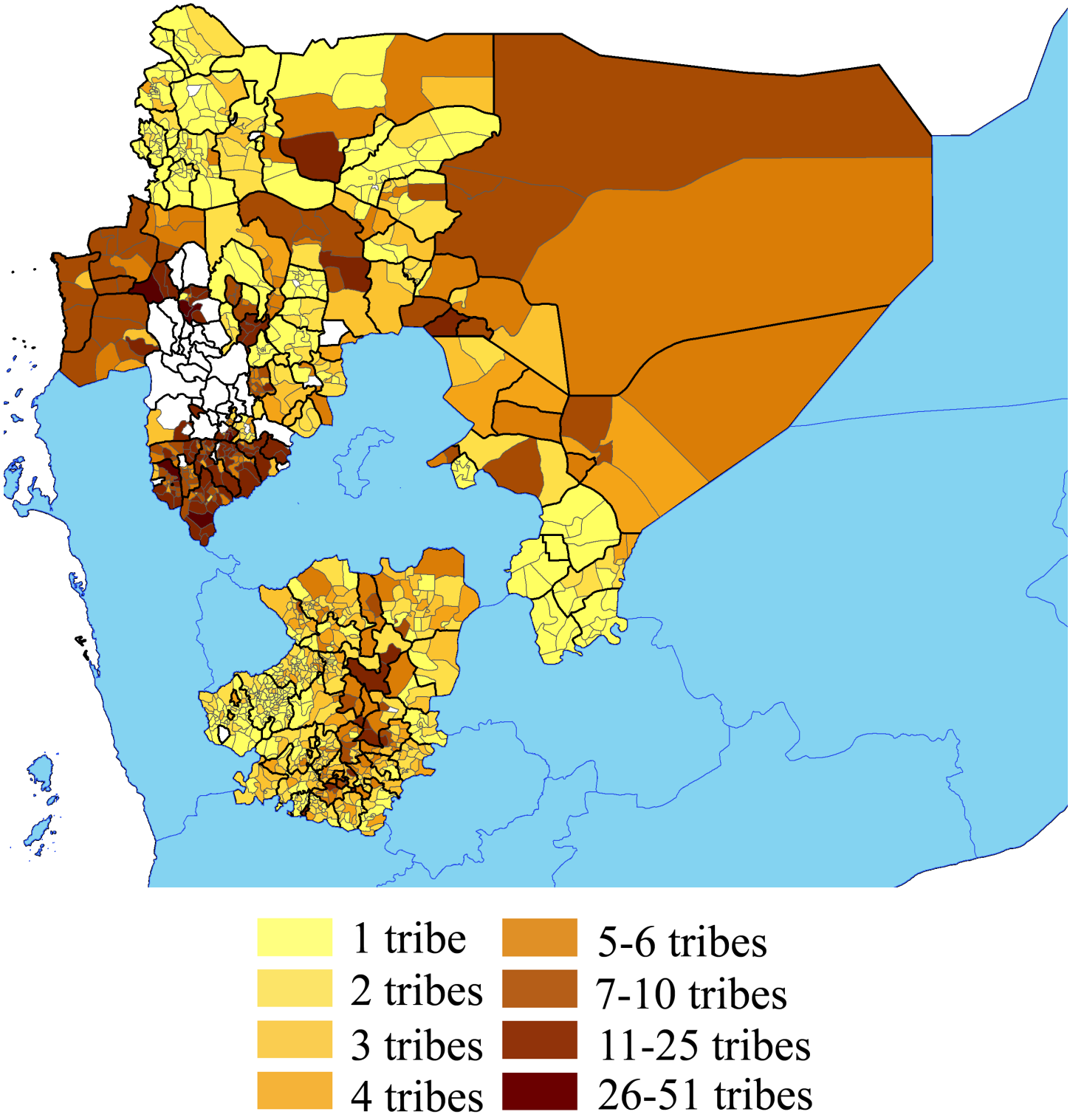


Figure 4: School Room Construction in the Northern Governorates

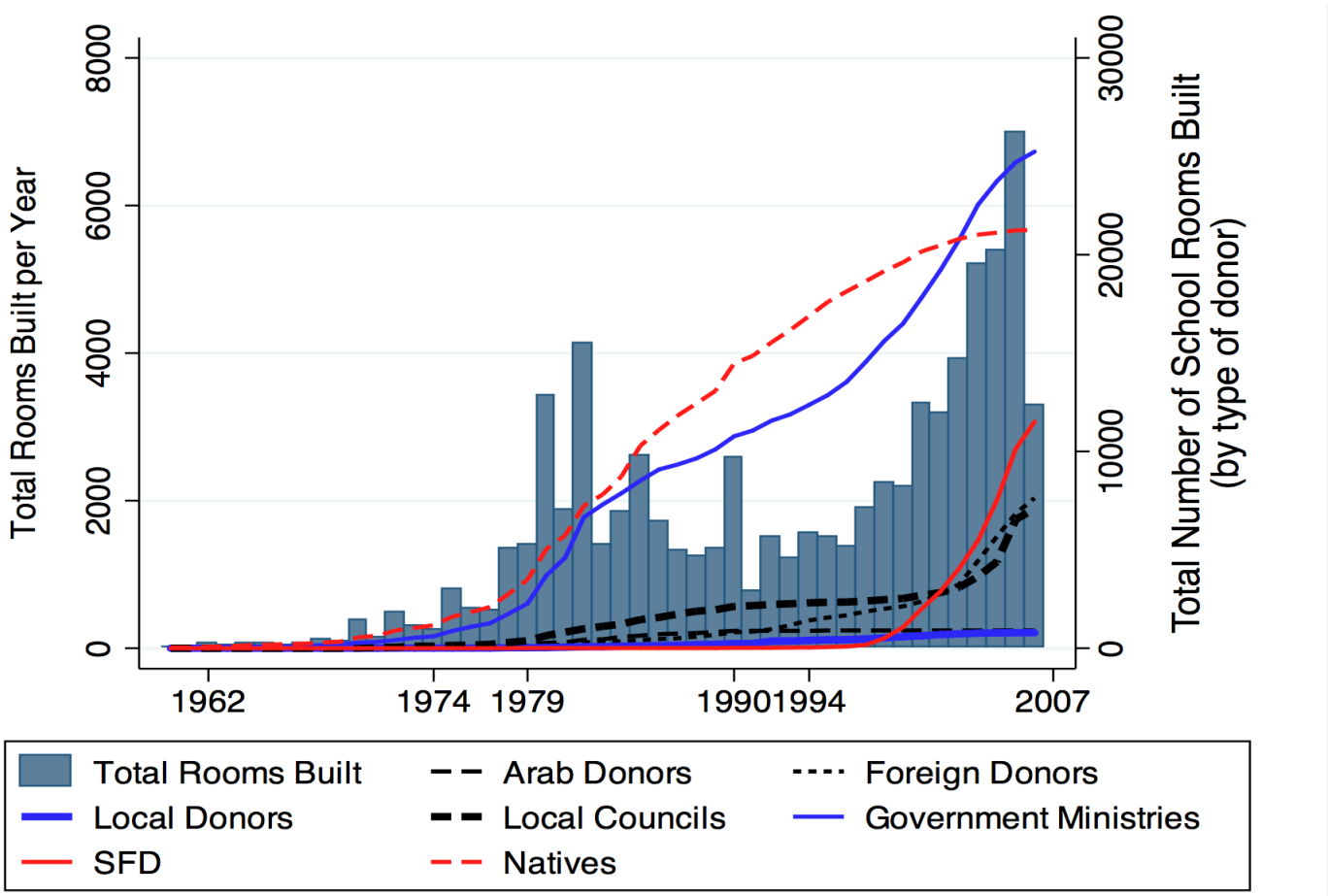


Table 1: Summary Statistics for Dependent Variables

Variable	Mean	Standard Deviation	N =
log(number of male teachers)	3.11	1.18	3572
log(number of classrooms)	3.44	0.94	1000
log(number of classrooms built by Arab Donors)	0.11	0.53	1000
log(number of classrooms built by Western Donors)	0.93	1.18	1000
log(number of classrooms built by Local Donors)	0.10	0.46	1000
log(number of classrooms built by Local Council)	0.89	1.14	1000
log(number of classrooms built by Government Ministry)	2.12	1.35	1000
log(number of classrooms built by Social Fund for Development)	1.33	1.25	1000
log(number of classrooms built by Local Community)	1.92	1.27	1000

Table 2: Summary Statistics for Explanatory Variables

	Variable	Mean	Standard Deviation	N =
Tribal Variable	Number of people (10,000s) per tribe	0.25	0.27	1031
Population Controls	log(total population)	8.23	0.83	1031
	Number of male children (100s)	0.29	0.28	1031
	Population density (number of men per 1,000 square meters)	0.59	0.60	1031
	Number of villages (100s)	0.11	0.10	1031
	Ruggedness (VRM)	0.38	0.12	1031
	log(meters from subdistrict population centroid to district capital)	9.21	1.27	1005
Access to Service Controls	Share of Households using Wood, Coal or Kerosense for Cooking	0.41	0.33	1030
	Share of Households without Sanitation	0.84	0.19	1030
	Share of Households without Piped Water	0.82	0.27	1030
	Share of Households without Electricity	0.73	0.34	1030
Agricultural Controls	log(total area owned by villagers in sub-district)	12.4	1.29	981
	Share of Land Cultivable	0.87	0.12	981
	Share of Cultivable Land that is Rainfed	0.66	0.34	981
	Share of Cultivable Land that is Fed from Wells	0.03	0.11	981
	Share of households owning plots less than 5,000 square meters	0.57	0.27	981
	Share of households owning plots 5,000-20,000 square meters	0.23	0.17	981
	Number of goats & sheep per household (100s)	0.07	0.10	981
	Number of cows per household (100s)	0.01	0.01	981
	Share of land cultivated with grains	0.57	0.30	981
	Share of land cultivated with qat	0.13	0.17	981
	Share of land cultivated with cash crops	0.03	0.07	981

Table 3: Tribe Size and the Distribution of Educational Resources

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	log(number of teachers)			log(number of classrooms)		
Number of men (10,000s) per tribe	-0.30***	-0.26***	-0.33***	-0.00	-0.17*	-0.21**
	(0.07)	(0.07)	(0.07)	(0.08)	(0.10)	(0.10)
log(total population)	1.08***	1.12***	0.98***	0.89***	0.89***	0.77***
	(0.02)	(0.02)	(0.04)	(0.03)	(0.04)	(0.06)
Population density			0.34***			-0.13
			(0.07)			(0.08)
Number of male school age children (1,000s)			0.23***			0.18**
			(0.05)			(0.09)
Number of villages in subdistrict (100s)			0.16			0.36
			(0.16)			(0.24)
Ruggedness (VRM) of subdistrict			-0.04			-0.21
			(0.14)			(0.17)
log(distance of subdistrict centroid from district capital)			-0.05***			-0.06***
			(0.01)			(0.02)
District Fixed Effects?	No	Yes	Yes	No	Yes	Yes
Year Fixed Effects?	Yes	Yes	Yes	No	No	No
R^2	0.57	0.69	0.69	0.60	0.69	0.70
N =	3081	3081	3006	998	998	976

Note: * indicates significance at the 10% level, ** indicates significance at the 5% level, *** indicates significance at the 1% level.

Table 4: Robustness to Agricultural Controls and Access to Service Controls

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	log(number of teachers)			log(number of classrooms)		
Number of men (10,000s) per tribe	-0.33***	-0.34***	-0.21***	-0.21**	-0.17	-0.09
	(0.07)	(0.07)	(0.06)	(0.10)	(0.10)	(0.09)
log(total population)	0.98***	0.96***	0.98***	0.77***	0.74***	0.75***
	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)
Population density	0.34***	0.25***	-0.05	-0.13	-0.13	-0.29***
	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.09)
Number of male school age children (1,000s)	0.23***	0.29***	0.18***	0.18**	0.19***	0.11*
	(0.05)	(0.05)	(0.04)	(0.09)	(0.07)	(0.06)
Number of villages in subdistrict (100s)	0.16	0.28	0.23	0.36	0.82***	0.78***
	(0.16)	(0.19)	(0.18)	(0.24)	(0.28)	(0.27)
Ruggedness (VRM) of subdistrict	-0.04	-0.07	-0.06	-0.21	-0.19	-0.20
	(0.14)	(0.13)	(0.13)	(0.17)	(0.17)	(0.17)
log(distance of subdistrict centroid from district capital)	-0.05***	-0.02	-0.00	-0.06***	-0.04**	-0.03
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
log(total area owned by villagers in sub-district)		0.01	-0.01		0.05	0.04
		(0.02)	(0.02)		(0.04)	(0.04)
Share of Land Cultivable		0.72***	0.50***		0.38*	0.20
		(0.14)	(0.15)		(0.23)	(0.23)
Share of Cultivable Land that is Rainfed		0.05	0.03		0.10	0.10
		(0.06)	(0.06)		(0.10)	(0.10)
Share of Cultivable Land that is Fed from Wells		-0.05	-0.09		0.25	0.21
		(0.17)	(0.18)		(0.20)	(0.20)
Share of households owning plots less than 5,000 square meters		0.60***	0.47***		0.39*	0.36*
		(0.12)	(0.11)		(0.22)	(0.21)
Share of households owning plots 5,000-20,000 square meters		0.77***	0.61***		0.27	0.19
		(0.14)	(0.13)		(0.20)	(0.20)
Number of goats & sheep per household (100s)		0.14	0.54**		0.01	0.25
		(0.25)	(0.26)		(0.39)	(0.36)
Number of cows per household (100s)		-21.97***	-13.66***		-11.20	-5.11
		(4.66)	(4.71)		(6.90)	(7.06)
Share of land cultivated with grains		-0.20**	-0.11		-0.26*	-0.21
		(0.08)	(0.08)		(0.15)	(0.16)
Share of land cultivated with qat		-0.05	-0.07		-0.34*	-0.35*
		(0.12)	(0.12)		(0.20)	(0.19)
Share of land cultivated with cash crops		-0.05	0.05		0.15	0.20
		(0.24)	(0.24)		(0.35)	(0.35)
Share of Households using Wood, Coal or Kerosense for Cooking			-0.75***			-0.32***
			(0.07)			(0.09)
Share of Households without Sanitation			-0.26***			-0.27**
			(0.09)			(0.12)
Share of Households without Piped Water			0.03			-0.07
			(0.06)			(0.09)
Share of Households without Electricity			-0.04			-0.15**
			(0.06)			(0.08)
District Fixed Effects?	No	Yes	Yes	Yes	Yes	Yes
R^2	0.69	0.71	0.73	0.70	0.73	0.74
N =	3006	2856	2856	976	926	926

Table 5: Classroom Construction by Donor

Dependent Variable: log(number of classrooms in subdistrict) [by donor]								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All	Arab Donors	Western Donors	Local Donor	Local Council	Government Ministry	SFD	Community
Number of men (10,000s) per tribe	-0.21**	0.06	-0.54***	-0.14	0.19	-0.27	-0.38	-0.08
	(0.10)	(0.11)	(0.20)	(0.10)	(0.21)	(0.16)	(0.24)	(0.17)
log(number of men)	0.79***	0.07*	0.42***	0.01	0.33***	0.84***	0.40***	0.63***
	(0.06)	(0.04)	(0.10)	(0.04)	(0.08)	(0.11)	(0.10)	(0.09)
Population density	-0.29*	-0.00	-0.23	0.62***	-0.59**	-0.24	-0.78*	-0.31
	(0.17)	(0.11)	(0.36)	(0.22)	(0.29)	(0.36)	(0.40)	(0.32)
Number of male school age children (1,000s)	0.18**	0.01	0.45***	0.05	0.23*	0.20	0.45***	0.42***
	(0.09)	(0.06)	(0.15)	(0.07)	(0.12)	(0.13)	(0.12)	(0.13)
Number of villages in subdistrict (100s)	0.30	-0.25	0.75	0.18	0.18	-0.00	-0.02	0.09
	(0.24)	(0.16)	(0.57)	(0.17)	(0.45)	(0.46)	(0.46)	(0.46)
Ruggedness (VRM) of subdistrict	-0.19	0.23*	0.38	-0.10	0.36	-0.36	-0.13	-0.21
	(0.17)	(0.12)	(0.29)	(0.13)	(0.24)	(0.33)	(0.35)	(0.27)
log(distance of subdistrict centroid from district capital)	-0.06***	0.03	-0.03	-0.00	-0.11***	-0.08**	0.00	-0.08***
	(0.02)	(0.02)	(0.04)	(0.01)	(0.03)	(0.04)	(0.04)	(0.03)
District Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.71	0.22	0.39	0.19	0.34	0.48	0.21	0.56
N =	977	977	977	977	977	977	977	977

Note: "Local Donor" includes mosques, Yemeni corporations, and Yemeni private individuals. "Local Council" are the local councils that were elected during the 2004 local elections. "SFD" is the Social Fund for Development.

Table 6: Tribe Size and the Number of Classrooms Built by Western Donors

Dependent Variable: log(number of classrooms built by Western Donors)					
	(1)	(2)	(3)	(4)	(5)
Number of men (10,000s) per tribe	-0.40**	-0.39*	-0.54***	-0.60***	-0.57***
	(0.19)	(0.21)	(0.20)	(0.21)	(0.21)
log(total population)	0.63***	0.70***	0.42***	0.37***	0.38***
	(0.06)	(0.07)	(0.10)	(0.10)	(0.10)
Population density			-0.23	0.03	-0.13
			(0.36)	(0.39)	(0.41)
Number of male school age children (1,000s)			0.45***	0.60***	0.56***
			(0.15)	(0.12)	(0.12)
Number of villages in subdistrict (100s)			0.75	0.97	1.01
			(0.57)	(0.79)	(0.79)
Ruggedness (VRM) of subdistrict			0.38	0.29	0.27
			(0.29)	(0.30)	(0.30)
log(distance of subdistrict centroid from district capital)			-0.03	-0.01	-0.00
			(0.04)	(0.04)	(0.04)
log(total area owned by villagers in sub-district)				0.02	0.01
				(0.05)	(0.05)
Share of Land Cultivable				-0.03	-0.16
				(0.33)	(0.34)
Share of Cultivable Land that is Rainfed				-0.05	-0.05
				(0.17)	(0.17)
Share of Cultivable Land that is Fed from Wells				0.86**	0.83**
				(0.38)	(0.37)
Share of households owning plots less than 5,000 square meters				-0.27	-0.30
				(0.32)	(0.32)
Share of households owning plots 5,000-20,000 square meters				-0.38	-0.42
				(0.37)	(0.37)
Number of goats & sheep per household (100s)				1.14*	1.24**
				(0.62)	(0.62)
Number of cows per household (100s)				-14.34	-11.14
				(9.85)	(9.84)
Share of land cultivated with grains				-0.10	-0.07
				(0.21)	(0.22)
Share of land cultivated with qat				-0.52	-0.49
				(0.34)	(0.34)
Share of land cultivated with cash crops				-0.33	-0.31
				(0.68)	(0.68)
Share of Households using Wood, Coal or Kerosense for Cooking					-0.14
					(0.16)
Share of Households without Sanitation					-0.42*
					(0.23)
Share of Households without Piped Water					-0.06
					(0.17)
Share of Households without Electricity					0.05
					(0.15)
District Fixed Effects?	No	Yes	Yes	Yes	Yes
R^2	0.15	0.37	0.39	0.40	0.40
N =	1000	1000	977	927	927