

Balasubramaniam, Divya; Chatterjee, Santanu; Mustard, David B.

**Working Paper**

## Got water? Social divisions and access to public goods in rural India

IZA Discussion Papers, No. 5977

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Balasubramaniam, Divya; Chatterjee, Santanu; Mustard, David B. (2011) : Got water? Social divisions and access to public goods in rural India, IZA Discussion Papers, No. 5977, Institute for the Study of Labor (IZA), Bonn, <https://nbn-resolving.de/urn:nbn:de:101:1-201110102341>

This Version is available at:

<https://hdl.handle.net/10419/55044>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

IZA DP No. 5977

**Got Water?  
Social Divisions and Access to Public Goods  
in Rural India**

Divya Balasubramaniam  
Santanu Chatterjee  
David B. Mustard

September 2011

# **Got Water? Social Divisions and Access to Public Goods in Rural India**

**Divya Balasubramaniam**

*St. Joseph's University*

**Santanu Chatterjee**

*University of Georgia*

**David B. Mustard**

*University of Georgia  
and IZA*

Discussion Paper No. 5977  
September 2011

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### Got Water? Social Divisions and Access to Public Goods in Rural India<sup>\*</sup>

We use data for 436 rural districts from the 2001 Census of India to examine whether different aspects of social divisions help explain the wide variation in access to tap water across rural India. Studies linking social fragmentation to public goods usually aggregate different types of fragmentation into one index. In contrast, we use disaggregated measures of social fragmentation to show that different types of social fragmentation are associated with dramatically different outcomes for access to tap water in rural India. Communities that are heterogeneous in terms of caste (within the majority Hindu religion) have lower access to tap water than correspondingly homogeneous communities. Communities that are fragmented across religions have higher access to tap water than correspondingly homogeneous communities. This underscores the importance of heterogeneity both *within and across* religions. Therefore, relying on aggregate measures of social fragmentation may conceal different effects of the component measures and obscure important information regarding the design of policies related to public goods.

JEL Classification: H4, O2

Keywords: public goods, social fragmentation, water, public policy, India

Corresponding author:

David B. Mustard  
Department of Economics  
Terry College of Business  
University of Georgia  
Athens, GA 30602  
USA  
E-mail: [mustard@uga.edu](mailto:mustard@uga.edu)

---

<sup>\*</sup> We are grateful to Rohini Somanathan for providing the Hindu Caste data. This paper has benefited from the helpful comments of two reviewers and presentations at the University of Georgia, St. Joseph's University, the SEA annual meeting in Washington, DC, and the Conference on International Development at the Federal Reserve Bank of Atlanta. Balasubramaniam acknowledges financial assistance provided by the Dean's Award at the University of Georgia, which enabled the purchase of the Indian Census and Meteorological data.

## 1. Introduction

Water is essential for economic activity and an important determinant of health outcomes and living standards. In most developed countries, water services are provided through a pricing mechanism. By contrast, in developing countries, the government usually acts as the sole provider of water services. India is a prime example of a fast-growing developing country where drinking (tap) water is a “public” good, especially in rural areas, where government provision is intended to provide universal and free access to a pre-dominantly poor population. The data, however, reveal wide variation in access to tap water across rural India, with some rural districts having no access at all (see Table 1)

India is characterized by deep and historic social divisions on the lines of caste (within the majority Hindu religion) and religion. In this paper, we test whether different types of social divisions can help explain the observed variation in access to tap water in rural India. The link between social fragmentation and the provision of public goods has received a lot of attention in both the public and development economics literature, starting mainly with Easterly and Levine (1997) and Alesina et al. (1999). The literature typically uses an aggregate index of social fragmentation, which is made up of several socio-economic characteristics such as ethnicity, race, language, religion, caste, etc. Our study differs from existing work in several important dimensions. First, we argue that the aggregate social fragmentation index is not useful for designing public policy, as it obscures information on the magnitude and the direction of the individual characteristics that comprise the index. Different types of social divisions may have disparate effects on the provision of (and access to) public goods, and also call for different policy measures. In this sense, our study is a new approach that uses disaggregated measures of social fragmentation. Specifically, we ask whether measures of diversity based on caste, religion,

and political preferences move in the same or different directions when determining access to tap water in rural India. In essence, we examine how fragmentation *within* a religion (caste system among Hindus) compares with fragmentation *across* religions (Hindus, Muslims, Christians, etc.) in determining access to drinking water.

Second, our paper is related to a recent contribution by Banerjee and Somanathan (2007), who examine how access to public goods across rural India changed between 1971 and 1991. An important contribution of our paper is that disaggregating the standard measure of social fragmentation helps resolve a puzzling result reported by Banerjee and Somanathan (2007). Specifically, while their aggregate social fragmentation index reduces access to public goods in 1971, this effect is statistically insignificant in 1991. They interpret this result as indicating that between 1971 and 1991, social and religious groups in India mobilized themselves politically, which diminished the importance of social fragmentation in rural India over time. However, we show that when employing an aggregate measure of social fragmentation, the effects of the individual components may offset each other, thereby explaining the statistical insignificance. Indeed, even ten years beyond Banerjee and Somanathan's period of study, we find that social fragmentation *does* matter, with its components moving in different directions. This result is only evident when one examines the relative contribution of *each* component of the fragmentation measure. Our results therefore indicate that public policy must be designed to target different aspects of social divisions, rather than the one-size-fits-all approach often adopted.

We use data from the 2001 Census of India for 436 rural districts to test for the effect of caste and religious heterogeneity on tap water access in rural India.<sup>1</sup> We use three measures of tap

---

<sup>1</sup> The analysis is restricted to rural areas because the caste classification, which is one of the primary variables of interest, is available only for rural India.

water access: the share of households in a district having access to (i) total tap water, (ii) tap water within a residence, and (iii) tap water outside a residence. Compared to previous studies that use aggregated measures of fragmentation and aggregate measures of public goods, we use household data and disaggregated measures of both public good access and social fragmentation. In this respect, our approach provides a more granular view. We find that aggregate measures of social heterogeneity are unrelated to access to water, and using these measures conceals important differential effects of the individual components. Specifically, districts that are more heterogeneous in terms of caste have lower access to both total tap water and within-residence tap water.<sup>2</sup> On the other hand, districts that are more heterogeneous in terms of religion have greater access to total tap water, within-residence tap water, and water delivered outside the residence.<sup>3</sup>

The rest of the paper is organized as follows. Section 2 discusses the nature of social divisions in India and the possible mechanisms through which they affect access to drinking water. Section 3 presents a simple stylized model highlighting the link between different measures of social fragmentation and the provision of public goods. Section 4 describes the empirical specification, while Section 5 characterizes the data and discusses some econometric issues such as endogeneity and selective migration. Section 6 discusses the results and some underlying intuition, and Section 7 concludes with a brief discussion of how private provision of water services might overcome the barriers imposed by social divisions.

---

<sup>2</sup> The heterogeneity index reflects the mean within-group affinity for a public good. The higher the share of an individual's own group in the population, the higher is the probability that he or she will have access to public goods.

<sup>3</sup> This result is consistent with Alesina et al. (2003), who find that countries with higher degrees of religious heterogeneity also tend to have a higher quality of infrastructure services.

## 2. Socio-religious Stratification and Access to Drinking Water in Rural India

The provision of drinking water services in India has predominantly been controlled by the government. India follows a top-down approach of water management, where the central government has a monopoly in providing water with some limited degree of decentralization (Saleth, 2005). The data reveal a stark story about access to drinking water for rural India: according to the 2001 Census of India, the average share of households in a rural district that had access to tap water (inside and outside the residence) ranged from 3 percent in the eastern state of Orissa to 83 percent in the northern state of Himachal Pradesh. The numbers for tap water access within a residence are even worse: from 0 percent in Orissa to 27 percent in the western state of Maharashtra. Not only does access to government-provided water services vary widely across the country, no rural district has 100 percent access to tap water (See Table 1).<sup>4</sup>

The majority of India's population lives in rural areas and depends heavily on publicly provided water. As such, a sustained water shortage can amplify economic hardship and intensify competition among social groups for a scarce but essential public commodity. Even though one expects economic factors such as poverty and inequality to play a pivotal role in determining access to public goods, these factors are intricately linked to historically persistent social divisions (Banerjee et al., 2005).<sup>5</sup>

The predominant form of social stratification in India is the caste system, which has deep historic roots in the majority Hindu religion. Historically, caste divisions were designed to facilitate the identification of social groups based on their skill levels. However, over time,

---

<sup>4</sup> Several recent studies indicate a looming water crisis in India, with demand far outstripping supply. See for example, Haarmeyer and Mody (1997), Saleth (2005), Ramachandran (2006), Bajpai (2007), and Brooks (2007).

<sup>5</sup> India has had a long history of invasions (from Mongolia and Persia) and external occupation (French, Dutch, Portuguese, and British), which ended with Independence from British colonial rule in 1947. Consequently, Indian society is deeply fragmented along social and religious lines, and these divisions continue to play a pivotal role in both politics and the allocation of scarce public resources more than six decades after independence.

certain groups gained power over others, forming the basis for social discrimination and caste-based stratification became a determinant of an individual's relative position in the social hierarchy. Even though discrimination based on caste is illegal in India, there is a sharp social difference between high and low castes even today. According to the Census of India, there are more than 180 caste groups within the Hindu religion: *Brahmans* belong to the highest caste, while *Scheduled Castes* and *Scheduled Tribes* are at the bottom of the social hierarchy.<sup>6</sup>

The Indian caste system has been historically pervasive and the source of segregation and intolerance in society from ancient times. Indeed, the practice of “untouchability,” whereby people belonging to lower castes were prohibited from interacting socially or sharing public resources with those from higher castes, provided the foundation for segregation and lack of tolerance. These historically pervasive social divisions create social norms over time, which tend to be very persistent and critical in determining economic outcomes (Ray 1998, chapter 5).

Another dimension of social division in India is religion, even though the Indian Constitution provides for a secular state. More than 80 percent of the population belongs to the Hindu religion, while the principal minorities include Muslims, Christians, Sikhs and Buddhists (Census of India, 2001). Religious conflicts, often based on economic discrimination between Hindus and Muslims, and more recently between Hindus and Christians, have been historically pervasive in India. India's multi-party democracy, with regional and caste-based political parties often playing important roles in national politics, and high levels of poverty and inequality, underscore the nature and depth of its social fragmentation.

---

<sup>6</sup>Brahmans are at the highest end of the caste spectrum and have been historically associated with being priests, teachers, and philosophers. At the other end of the spectrum lie the Scheduled Caste and Scheduled Tribes. The Government of India uses these terms to classify the poorest and most disadvantaged communities in India; Scheduled Caste refers to the *Dalit* community and Scheduled Tribe to the tribal communities or *adivasis*.

Why might social (caste) and religious divisions determine access to public goods such as drinking water? With respect to caste-based divisions, the answer lies in the nature of social stratification and the competition it creates for scarce public goods. Several authors, starting with Easterly and Levine (1997) and Alesina et al. (1999) highlight the need for coordination within communities to gain access to public goods.<sup>7</sup> Social divisions based on caste and associated externalities such as untouchability and rivalry from the use of public goods may impede this coordination mechanism, thereby lowering access to public goods.

With respect to religion-based divisions, however, its link with public goods may be different relative to that generated by caste-based divisions. As Besley and Persson (2009, 2010) argue, a threat of internal conflict might influence an incumbent government's allocation decisions for public goods. Therefore, in the context of India, an increased risk of internal conflict due to religious divisions may cause the state to send more public goods to regions that have higher religious fragmentation. This may be particularly relevant for rural areas, where it may be hard for the state to provide law and order in times of conflict due to a lack of adequate infrastructure. So the provision of public goods may be a mechanism for the state to lower the risk of a religion-based conflict in remote rural areas.<sup>8</sup> Therefore, it is not clear *a priori* how different dimensions of social divisions (within a religion and across religions) affect access to public goods. Our paper attempts to shed some new light on this important public policy issue.

---

<sup>7</sup> Other important papers include Dayton-Johnson (2000), Miguel and Gugerty (2005), and Khwaja (2009).

<sup>8</sup> Indeed, Varshney and Wilkinson (2004) and Jha (2010) document that more than 90 percent of deaths due to religious riots in India were concentrated in urban areas, with rural areas virtually unaffected by religious strife. The point here is that the conventional channel of collective action through which social fragmentation adversely affects public good provision also applies to religion, but there may be other channels intrinsic to religious diversity that may work in the opposite direction.

### 3. A Stylized Model of Social Fragmentation and Public Good Provision

We present a simple stylized model that highlights the link between different aspects of social heterogeneity and the provision of public goods. Consider a median voter in a community who derives utility from the consumption of a private good,  $x$ , and a public good,  $g$ . The voter also derives disutility from various aspects of social heterogeneity in his/her community. Specifically, we assume that there are two forms of social heterogeneity, namely religious heterogeneity, given by the Herfindahl index  $h_r$ , and caste heterogeneity, given by a corresponding index  $h_c$ . These Herfindahl indices are calculated as

$$h_i = 1 - \sum_{j=1}^n s_j^2, \quad i = r, c \quad (1)$$

where  $s_j$  is the population share of the  $j$ -th caste or religious group.<sup>9</sup> The median voter's utility function is given by

$$U = x(1-h_r) + g^\alpha (1-h_c), \quad \alpha \in (0,1), \quad h_i \in [0,1], \quad i = r, c \quad (2)$$

The key assumption here is that preferences for private and public consumption are affected differentially by the two sources of heterogeneity: while religious heterogeneity reduces the utility derived from private consumption, caste heterogeneity has a similar effect on public consumption. This can happen if an increase in either form of heterogeneity raises the risk of internal conflict. It is plausible that the risk of internal conflict will affect the marginal utility from private and public consumption differently, depending on the source of heterogeneity (religion or caste).<sup>10</sup>

---

<sup>9</sup> The index in (1) measures the probability that two distinct individuals picked randomly from the population in a given district belong to distinct social groups (caste or religion).

<sup>10</sup> For example, conflicts between religious groups can lead to the destruction of private property and loss of income. This is essentially a reduced-form argument based on Besley and Persson (2009, 2010). An aversion towards members of lower castes might reduce the utility derived from the consumption of a public good since public goods

Following Alesina et al. (1999), let the median voter have a fixed exogenous source of income,  $y$ , so that consumption of the private good depends on after-tax disposable income

$$x = y - t \quad (3)$$

where  $t$  represents lump-sum taxes paid to the government or planner. The planner, in turn, finances the provision of the public good by the tax revenues raised from voters

$$g = t \quad (4)$$

Substituting (3) and (4) into (2), the utility function of the median voter can be written as

$$U = (y - g)(1 - h_r) + g^\alpha (1 - h_c) \quad (5)$$

The planner chooses the optimal provision of the public good to maximize (5), given the degree of religious and caste heterogeneity in the community. Maximization yields

$$\hat{g} = \left[ \frac{\alpha(1 - h_c)}{1 - h_r} \right]^{\frac{1}{1 - \alpha}} \quad (6)$$

where  $\hat{g}$  represents the optimal provision of the public good. From (6), it is easy to see that

$$\frac{\partial \hat{g}}{\partial h_r} > 0, \quad \frac{\partial \hat{g}}{\partial h_c} < 0 \quad (7)$$

Therefore, an increase in religious heterogeneity increases the provision of the public good by the planner, while an increase in caste heterogeneity has precisely the opposite effect. If an increase in religious heterogeneity increases the risk of internal conflict, the planner might increase the provision of the public good to offset for the disutility from a higher level of religious fragmentation. On the other hand, since an increase in caste-based fragmentation

---

must be “shared” by members of the community. Therefore, caste-based social divisions can generate some form of rivalry associated with the usage of public goods.

reduces the median voter's preference for public consumption, the planner chooses a lower level of public goods in equilibrium.

The issue at hand is essentially empirical. The set-up described in (1)-(7) above represents a specific example of the link between underlying measures of social fragmentation and the provision of public goods. One can construct other scenarios as well, leading to different implications.<sup>11</sup> For example, if we assume that there is only one aggregate measure of social fragmentation that includes both religious and caste-based heterogeneity, as is done in much of the literature, we can show that an increase in social fragmentation can either leave the provision of public goods unaffected, as in Banerjee and Somanathan's (2007) analysis using the 1991 Census of India, or lower, as in Easterly and Levine (1997). To see this, assume in (2) that  $h_r = h_c = h$ , where  $h$  represents an aggregate index of social fragmentation. From (6), it is immediately evident that an increase in this aggregate index will not affect the provision of the public good in equilibrium. Another possibility might be that this aggregate index only impacts public consumption, and not private consumption, as in Alesina et al. (1999). The utility function (5) is then modified to

$$U = y - g + g^\alpha (1 - h)$$

In this case, an increase in  $h$  will reduce the provision of public goods in equilibrium.

#### **4. Empirical Specification**

The central focus of this paper is to identify how different measures of social divisions are associated with access to tap water in rural India. In particular, we will test whether using an

---

<sup>11</sup> Esteban and Ray (1999) and Banerjee and Iyer (2007) construct scenarios in the context of models of collective action where heterogeneity can improve the provision of public goods.

aggregate measure of homogeneity produces different results than disaggregated measures. The specification we test is given by the following cross-section regression:

$$Y_{i,k} = \alpha_1 HI_{i,k} + \alpha_2 C_{i,k} + \alpha_3 R_{i,k} + \alpha_4 E_{i,k} + \alpha_5 P_{i,k} + \alpha_6 X_{i,k} + \alpha_7 M_k + \varepsilon_{i,k} \quad (8)$$

$Y_{i,k}$  is the share of households with access to tap water (within and outside the residence) in district  $i$  of state  $k$ .  $HI_{i,k}$  is the homogeneity index for social divisions in a district  $i$  of state  $k$ , and is defined in a manner similar to a Herfindahl index:

$$HI_j = \sum_{j=1}^n s_j^2$$

where  $s_j$  is the population share of the  $j$ -th caste or religious group. Therefore, a positive coefficient estimate indicates that an increase in social homogeneity increases tap water access, or conversely, an increase in social heterogeneity, decreases tap water access, since the corresponding heterogeneity index is simply equal to  $1 - HI_j$ ; see (1) above. We decompose the aggregate homogeneity index into caste and religious indices to determine whether different types of homogeneity have differential effects. The disaggregated indices measure the probability that two distinct individuals picked randomly from the population in a given district belong to the same (a) Hindu caste, or (b) religion.

$C_{i,k}$  is a set Hindu caste groups (Brahman, Scheduled Caste and Scheduled Tribes).  $R_{i,k}$  contains the shares of Muslims, Christians and other religious minorities in the sample.  $E_{i,k}$  represents a set of economic variables, and includes the land Gini coefficient (to capture inequality), and bank deposits per capita.  $P_{i,k}$  is a set of election outcomes that include the share

of total votes received by the winning party and a political heterogeneity index.<sup>12</sup>  $X_{i,k}$  is a vector of geographical characteristics that include average annual rainfall, average temperatures, terrain, a dummy for coastal areas, and controls for population density including the average village population, average number of villages, and household size.  $M_k$  is a set of state fixed-effects, and  $\varepsilon_{i,k}$  is a district-specific shock.

The empirical specification we adopt is quite standard in the social heterogeneity-public goods literature (Alesina et al., 1999). Though there is conclusive evidence on the effects of ethnic heterogeneity with respect to public goods, the interpretation of these effects is complicated by the nature of the aggregation mechanism that links individual preferences to community outcomes. Vigdor (2004) shows how the theory of altruistic behavior can be used to aggregate individual decisions (to contribute to public goods) to community outcomes. We therefore assume that the effects of group heterogeneity on access to public goods (tap water, in our case) are derived from the altruistic behavior of individuals in each community (district).<sup>13</sup>

Another important point of the model specification is the inclusion of individual group shares, with individual coefficient estimates for each group, and a single coefficient for the homogeneity (or heterogeneity) index. The homogeneity index reflects the mean within-group affinity for a public good. The higher the share of an individual's own group in the population, the higher is the probability that he or she will contribute to a given set of public goods. If the model fails to control for the group shares, it then imposes an implicit behavioral restriction that all groups have the same propensity to contribute to public goods, regardless of the composition (caste or religious) of the community. Consequently, the Hindu caste shares enter the

---

<sup>12</sup> The political heterogeneity index estimates the probability that two individuals randomly drawn from a population will belong to different political parties.

<sup>13</sup> Banerjee and Somanathan (2007) and Rushton (2008) also adopt a similar procedure.

specification in two important ways. First, following Vigdor (2002, 2004), the shares of Hindu caste groups appear as a linear term with a separate coefficient estimate for each caste group.<sup>14</sup> Second, they appear as squared terms in the caste homogeneity index, with a single coefficient estimate. We include individual shares in the specification to capture the effect of the presence of other caste groups on the access to tap water for individuals in a specific group. The homogeneity index captures how the number of groups affects access to water, assuming that the within-group affinity is equal across all the groups in the district.<sup>15</sup> When both the index and the individual shares are included in the same regression, we cannot make a *ceteris paribus* argument, since when the share of a group changes, the homogeneity index also changes simultaneously. We follow the same procedure to analyze and interpret the effects of religious heterogeneity.

We estimate the model with ordinary least squares and calculate Huber-White standard errors. All regressions include state fixed-effects, unless otherwise noted, to account for unobserved characteristics across states, and the standard errors are clustered to account for within-state unobserved variation.<sup>16</sup> We discuss related econometric issues such as endogeneity and selective migration in detail in the next section.

---

<sup>14</sup>We use data on 180 caste groups for the Hindu religion, which makes the inclusion of each caste group separately in the model complex. Therefore, we aggregate these 180 groups into three sub-groups: Brahmans, Scheduled Castes, and Scheduled Tribes. Each of these three sub-groups appears linearly in the specification, which is consistent with the literature.

<sup>15</sup> Ideally, one could include the shares and their squared terms with separate coefficients in the regression. The squared group shares capture the within-group-affinity across groups and a coefficient estimate for each squared group share allows for affinity to vary across caste groups. We do not make this assumption because of the large number of caste groups in our sample and including each of the 180 groups and its squared term is very complex.

<sup>16</sup> Though we use state fixed effects in our main specification, we do not use state fixed effects in some robustness checks in Table 7 that include some state-level control variables like per-capita GDP and the number of police personnel per 100,000.

## 5. Data

We use district-level data from the 2001 Census of India, the latest year for which data are available. The cross-section precludes us from identifying intertemporal trends. Our dataset includes 436 of the 593 rural districts in India, because data for some of the explanatory variables are not available for districts in several states. Table 1 reports the summary statistics for the variables we use and, as mentioned in the introduction, reveals some striking patterns with respect to access to tap water across rural India.

### *5.1. Tap Water Access*

The share of households with access to tap water is the main dependent variable for our analysis, and is obtained from the Houses, Households and Amenities Section of the 2001 Census of India. We use three measures of tap water access: (i) total tap water access,<sup>17</sup> (ii) within-residence tap water access, and (iii) outside-residence tap water access.<sup>18</sup> This measure is calculated over those households that do not have access to water at home rather than all households. Table 2 shows that the mean share of a household's total tap water access ranges from 3 percent in the eastern state of Orissa to 83 percent in the northern state of Himachal Pradesh; the mean share of a household's within-residence tap water access ranges from nearly 0 percent in Orissa to 27 percent in the western state of Maharashtra. The mean share of a household's outside-residence tap water access ranges between 5 percent and 58 percent across the sample. These numbers reflect large variation across Indian states and document that no state has 100 percent access to tap water in their rural districts. The segment of the population not served by tap water uses hand pumps, wells, rivers or other water sources to meet their daily

---

<sup>17</sup> Total access to tap water is the sum of access to tap water within and outside the residence.

<sup>18</sup> Outside-residence tap water access refers to a rural household traveling 100 meters or more to fetch drinking water.

water needs. Outside-residence tap water access implies substantial costs borne by the households, including travel time to the water source and waiting time to get access to water.

## 5.2. *Caste and Religion*

Detailed caste data in India were last enumerated under the British Colonial regime in 1931. After independence, caste-based data collection was discontinued to prevent discrimination and, from 1951 onwards, the Indian government has collected data based on three broad categories: Scheduled Castes, Scheduled Tribes and Others. In calculating the caste homogeneity index, we use the method described in Banerjee and Somanathan (2007).<sup>19</sup> Because the caste data are from the 1931 Census, and a significant portion of the Muslim population immigrated to Pakistan after Independence in 1947, they adjust the increase in the proportion of Hindus after 1931 by scaling up the numbers in each caste group based on the Hindu share in the current census.<sup>20</sup> We similarly scale up the caste figures by the share of Hindu population in 2001 and also adjust for newly created districts between 1991 and 2001. In all, we have 180 caste groups within the Hindu religion in our sample.

One important point of departure from the Banerjee-Somanathan study is in the construction of the caste homogeneity index. Their study combines 185 Hindu caste groups with six non-Hindu religions in the construction of a “socio-religious” heterogeneity index, thus assuming that other religious (non-Hindu) groups are internally homogeneous. We do not make

---

<sup>19</sup> We are grateful to Rohini Somanathan for sharing the caste data used in the Banerjee-Somanathan study. The 1931 Census had a very large list of caste groups for each British province and princely state, by district. Over time, even though state boundaries were redrawn, district boundaries remained intact. After independence, a few districts were created by subdividing old ones. For these new districts, Banerjee and Somanathan weigh the caste data by the area of the new district that was created from the original districts. Since the number of caste groups is very large, they restrict the number to the Hindu caste that constitutes more than one percent of the population of each state or province in 1931. This approach yields 185 Hindu caste groups. Since Hindus are a majority, this restriction is reasonable. We use data on 180 out of the 185 caste groups as some states were not included in the sample due to lack of data on other crucial explanatory variables.

<sup>20</sup> This assumes that over time, all Hindu castes grew at a similar rate. Since this paper focuses on rural areas, the method is a reasonable approximation, since the percentage of rural to urban migration has been relatively slow in India (Haub and Sharma, 2006).

any assumptions about the caste structure in other religions, but instead restrict our analysis to only the Hindu castes. Further, we also do not combine other forms of heterogeneity such as language and religion in constructing the caste index because in India there are many languages with several sub-dialects and it is very difficult to group people or communities by language.<sup>21</sup>

Data on the various religious groups (Hindus, Muslims, Christians, and other religious minorities) are from the religion tables of the 2001 Census of India. The religion index may have potential problems because a person can potentially hide his or her religion to avoid oppression or persecution. Individuals may change religions as well, though conversion *to* the majority Hindu religion is prohibited, since one can only be born into the religion. Given the history of the caste system within Hinduism, mobility across castes is also prohibited (Alesina, et al. 2003). We address these issues in section 5.8.

Table 1 shows that the Hindu caste homogeneity index ranges between 0 and 0.32, with a mean of 0.04, which implies that rural districts are highly heterogeneous in terms of caste. The religion homogeneity index ranges between 0.37 and 0.99 with a mean of 0.80. Since 85 percent of the population is Hindu, the sample mean of 0.80 implies that districts are highly homogeneous with respect to religion.

### *5.3. Economic Controls*

Private wealth, an indicator of economic status, may affect access to water. To this end, we obtained per-capita bank deposits across rural districts in 2001 from the Reserve Bank of India database. The number and area of operated land holdings by different sizes (measured in hectares) are obtained from the 2001 Agricultural Census of India. We calculate the land Gini

---

<sup>21</sup> Easterly and Levine (1997) use measures of ethno-linguistic heterogeneity constructed from the former Soviet Union in 1960. The data, however, were based on linguistic classification rather than on race or color. One problem with this measure is that it may hide other aspects of ethnicity; see Alesina et al. (2003). For example, if two ethnic groups speak the same language but have different customs and beliefs, then classification based purely on language combines these two different ethnic groups in one category, which in turn may generate measurement error.

coefficient using these data to proxy for land inequality across districts. We assign zero land holdings to agricultural laborers. Because there are no data on ownership of land holdings, the use of operated land holdings may be less than a perfect measure for land distribution. However, one defense of this variable is that since Independence most rural land has been owner-cultivated (Banerjee and Somanathan, 2007). Table 1 shows that the land Gini ranges between 0.41 and 0.86 with a sample mean of 0.71, implying that rural districts have a high degree land inequality.

#### 5.4. Political Controls

In India, political power is crucial in determining access to water across districts. Politicians are concerned about the number of votes they will receive in the next election based on the satisfaction of the public on the various public goods provided during their last term in office. More homogeneous political districts may provide more effective services, or when water is not provided, the incumbent may be defeated, which could increase the heterogeneity in the vote share in future elections.

Data on general elections for 1999 are from the Election Commission of India (1999) website.<sup>22</sup> We use two variables, namely (i) the vote share of the winning party, and (ii) a political fragmentation index, which is calculated using:

$$p = 1 - \sum_{i=1}^n v_i^2$$

where  $v_i$  is the vote share of the  $i$ -th party. The index  $p$  lies between 0 and 1, where 1 represents complete political heterogeneity and 0 represents political homogeneity. The political

---

<sup>22</sup> We use 1999 because general elections are conducted once in five years and 1999 is the closest year to the analysis period. The data on general elections are available for the 543 parliamentary constituencies in India. We use a mapping method from Banerjee and Somanathan (2007) that makes the data consistent at the district level. Specifically, the data are mapped by visually comparing the number of districts that go to each parliamentary constituency. We further compare the number of constituencies that go to each district by assigning weights by visual inspection using maps of districts and maps of parliamentary constituencies.

fragmentation index ranges between 0.14 and 0.80, with a mean of 0.59, thereby implying a high degree of political heterogeneity across rural districts (Table 1).

### *5.5. Geography Controls*

Because average rainfall and temperature affects access to water, we also control for these measures. The relationship between average rainfall and piped water is uncertain. In wet regions, the cost of building a piped water system may be lower, but the cost of accessing water at a river or well may also be lower. Data on average annual rainfall and average temperatures come from the Indian Meteorological Department (IMD) and the rainfall and average temperature maps released by the Maps of India website.<sup>23</sup> The share of wastelands in each district also determines the level of water access. The shares of land that are steep, barren, and sandy were obtained from the Wasteland Atlas of India (Ministry of Rural Development) for 2003.<sup>24</sup> Finally, we also include a coastline fixed effect.

### *5.6. Other Population Controls*

Average household size, the number of villages in each district, and the average village population from the 2001 Census of India captures the population density in a district, because population composition may affect access to tap water.

### *5.7. Correlation between Dependent and Control Variables*

Table 3 shows the correlation between the Hindu caste homogeneity index, religion homogeneity index, and each of the three dependent variables. Of the three dependent variables, the within-residence tap water share is the most positively correlated with the Hindu caste

---

<sup>23</sup> We map the available data from 109 weather stations in India on to districts in the following manner. First, we assign average rainfall and average temperature values to each district based on its proximity to each weather station. Second, for a few states in the northeastern region where no data is available, we use the median value calculated using the maps that contain the average annual rainfall and average temperatures.

<sup>24</sup> According to the Ministry of Rural Development of India sandy areas have stabilized accumulation of sand, in coastal, riverine, or inland areas, and steeply lands are steep sloping wasteland areas.

homogeneity index (0.295). The total tap water share and outside residence tap water share have a very high correlation of 0.92, which implies that most of the tap water access for the given sample is from outside the residence. The Brahman share has a positive correlation with tap water share access, because historically Brahmans are considered the upper-most caste group in India and we expect them to have greater access to public goods, including tap water. The religion homogeneity index is not consistently correlated with the dependent variables.

Table 4 shows the correlation coefficients between the dependent variables and economic and political variables. It reveals consistent signs for each of the variables. One expects a negative sign for the land Gini, which implies that more land inequality is associated with lower access to tap water. Similarly, the higher is political heterogeneity, the lower is the access to tap water. Other economic variables such as per-capita bank deposits are also important because wealthier districts may have more public funding and hence better access to tap water. However, the magnitude of these correlations is small.<sup>25</sup>

Finally, we compare the two homogeneity indices with Alesina et al. (2003). They find a small but positive correlation between their measures of ethnic and religion fragmentation (0.142). In our sample, there is also a small positive correlation between the Hindu caste homogeneity index and the religion homogeneity index (0.201).

### *5.8. Endogeneity*

A potential econometric issue is endogeneity, either through omitted variables, simultaneity, or measurement error. Hence, it is important to determine if the effect of Hindu caste and religion heterogeneity on tap water access is robust to correcting for these issues. The caste-based classifications in India created during pre-historic times are determined at the time of

---

<sup>25</sup> There is also a possibility that there may be reverse causality between the economic and political variables in our specification. We tried different specifications with various combinations of control variables as well as interaction terms, and our results remain robust to these sensitivity checks.

birth, and mobility across castes is prohibited by social norms. Therefore, the Hindu caste homogeneity index is not likely to be endogenous. Moreover, the Hindu caste variable is based on 1931 data, and therefore, is unlikely to be influenced by contemporaneous changes in districts.<sup>26</sup>

A more serious issue is the endogeneity of the religion homogeneity index. Data on religious conversion rates over time may solve the issue, but such data are not available for India. Though it is difficult to control fully for endogeneity related to religious heterogeneity, the following considerations perhaps help minimize the problem: (i) though religious conversions can undermine the causal effects of religious heterogeneity, conversion into the majority Hindu religion (more than 80% of the population) is prohibited, since one can only be *born* a Hindu, and (ii) the religion data from various rounds of the Census of India between 1961 and 2001 show that the population proportions across religions have been very stable.<sup>27</sup> In any case, we treat the link between religious fragmentation and access to drinking water cautiously, with the theoretical discussion in Sections 2 and 3 and also in Section 6 providing some suggested underlying mechanisms driving this relationship.<sup>28</sup>

### *5.9. Selective Migration*

There is also the problem of selective migration because people may prefer to migrate to districts that have higher access to tap water. Moreover, inter-district migration (say, due to effective district governance) can also affect the measure of religious diversity. This can

---

<sup>26</sup> Anderson (2011) also provides a similar argument for the exogeneity of caste-based fragmentation.

<sup>27</sup> The 2001 Census of India shows that the proportion of Hindus (including both rural and urban areas) in 1961 was 84.4 percent, while in 2001 it was 81.4 percent; proportion of Muslims in 1961 was 10.2 percent and in 2001 it was 12.4 percent; proportion of Christians in 1961 was 2.4 percent and in 2001 it was 2.3 percent. Other religious groups grew at a similar rate between 1961 and 2001.

<sup>28</sup> We have not been able to identify research in the social fragmentation-public goods literature that controls for endogeneity arising from religious fragmentation. One notable exception is Michalopoulos (2011), who uses geographic variability and the heterogeneous endowments of landholdings from colonial times to identify the sources of ethnic fragmentation, but does not address the issue of religion. Banerjee and Iyer (2007, p. 3138-9) discuss at length the difficulties with addressing the causality problem in this context.

influence the caste and religion group shares, as well as the size of the village. However, Haub and Sharma (2006) show that rural-urban migration has been very low in India, especially when compared to countries in Latin America. Urban migration in India rose from 11 percent in 1901 to only 28 percent in 2001. Further, Dreze et al. (1999) report that less than 1 percent of rural land is sold in India each year, and Munshi and Rosenzweig (2009) document that there has been very little caste-based migration in rural India.<sup>29</sup> Most Indians live their entire lives in rural areas. However, the lack of disaggregated data on net migration hinders a richer specification for our analysis. To test for selective migration, we exclude the population density variables from the model specification to check for the robustness of the results. Further, since we restrict the analysis to only rural areas, this minimizes the effects of urbanization on water access.

## 6. Results and Discussion

Table 5 reports the ordinary least squares coefficient estimates of a regression of three types of water access on the *aggregate* social homogeneity index (that combines both caste-based and religious homogeneity) used by Banerjee and Somanathan (2007) and an extensive set of control variables. Consistent with their results, in each of the three specifications, the coefficient estimates are not statistically different from zero. Banerjee and Somanathan interpret this result to mean that the role of social fragmentation, which they find to have been important in previous decades, is no longer important for determining access to public goods.

However, an alternative explanation of these results is shown in Table 6, which reports the results of the OLS regression of the same three dependent variables. The only difference between Table 5 and Table 6 is that the former table uses the aggregate social homogeneity index while that latter disaggregates the total index into the caste and religious homogeneity indexes.

---

<sup>29</sup> For a detailed discussion of why migration is not an issue for rural India, see Anderson (2011).

After controlling for all the covariates, the Hindu caste homogeneity index is positive and statistically significant at the 5 percent level for both the total tap water and the within home share of tap water. A 10 percentage-point increase in a households' own caste share of the population increases the household's total probability of access to tap water by 3.25 percentage points and within-residence tap water access by 2.73 percentage points. The religion homogeneity measure is also statistically significant for all three dependent variables. A 10 percentage-point increase in a households' own religion share of the population decreases the probability of having any type of tap water by 3.17 percentage points, inside tap water by 2.66 percentage points, and outside tap water by 3.51 percentage points.<sup>30</sup>

The primary conclusion of this paper is shown in the top rows of Tables 5 and 6, which clearly illustrate that the using an aggregate social index, as in Banerjee and Somanathan (2007) and much of the existing literature, may obscure important differences in the way that individual components of that index affect access to public goods. This conclusion is robust to excluding different combinations of the control variables. Our results remain robust to the inclusion of none, some, and all of the groups of control variables listed in Tables 5 and 6.

The striking feature of our results in Table 6 is that religious heterogeneity is associated with higher access to tap water, while the effect of caste-based heterogeneity works in the opposite direction. While the effect of caste heterogeneity is well-understood, the causal link between religious diversity and access to tap water is difficult to establish. However, the theoretical literature on public goods provides some insights into the mechanisms that might be driving this correlation. First, as Besley and Persson (2009, 2010) point out, the threat of internal conflict can influence the incumbent government's choice of public goods provision. In our

---

<sup>30</sup> An alternative way to interpret these coefficient estimates is to think of them as the effects of moving from a perfectly homogeneous to a perfectly heterogeneous community. For example, for the caste index, this would reduce access to tap water by 32.5 percent, and for the religion index it would increase access by 31.7 percent.

context, we use a reduced form of this argument in the stylized model of section 3 to show how this result may depend on the asymmetric way in which different sources of heterogeneity affect the marginal utility from public and private consumption. Specifically, a threat of internal conflict due to religion might cause the government to send more public goods to areas with a higher degree of religious fragmentation (see footnote 8). Second, Esteban and Ray (1999) and Banerjee and Iyer (2007) demonstrate that when smaller groups are less subject to failures of collective action, a community with many small groups does better at providing public goods than a community with few large groups, thereby mitigating the problem of collective action.<sup>31</sup>

Consistent with what would be expected for a disadvantaged group, an increase of 10 percentage points in the share of Scheduled Tribes is associated with a reduction of 2.20 percentage points in the probability of having any type of tap water and of 1.95 percentage points in the probability of having access to outside tap water (Table 5). Table 6, which includes the disaggregated measures, shows that Scheduled Tribes are associated with a reduction of 2.75 and 2.45 percentage points in access to total tap and outside tap water, respectively. This is interesting because given the position of Scheduled Tribes in the socio-economic scale in India their main source of drinking water is likely to be outside the residence.

When the disaggregated homogeneity measures are included in Table 6, the coefficient estimates on the shares of Muslims, Sikhs, and Christians tend to become more negative. The fraction of households that are Sikhs is associated with higher probabilities of having any access (Table 6) and access to inside to tap water (Tables 5 and 6).

Another interesting difference between our results and those of Banerjee and Somanathan (2007) is that they find that areas with larger shares of Brahmans, the most elite caste, have

---

<sup>31</sup>Banerjee and Iyer (2007, p. 3132), while discussing this issue, quote the French philosopher Voltaire, “If there were one religion . . . , its despotism would be terrible; if there were only two, they would destroy each other; but there are 30, and therefore they live in peace and happiness.”

greater access to public goods. In contrast, after controlling for economic, political, geographic, population, and land-type variables, we find that the fraction of Brahmans has no statistically significant relationship on access to tap water in rural India. Our result is consistent with Anderson (2011), whose study of the effects of caste composition across villages in two states of North India also document that villages with a higher share of Brahmans are not associated with more access to public goods relative to villages where lower castes are the dominant group. As Anderson (2011) points out, the difference could stem from the fact that Banerjee and Somanathan use data from parliamentary constituencies from the entire country, which presumably includes both rural and urban areas, while we use household and district-level data only for rural areas.

### *6.1. Robustness Checks*

We perform a number of sensitivity checks to ensure that the results from using the two measures of social divisions are robust. The first row of Table 7 reports the baseline coefficient estimates for the caste and religion heterogeneity indexes for all three measures of water access shown in Table 6. Specifications (2) and (3) exclude the population density variables and political variables, respectively. The population and political variables may be affected by endogeneity and these specifications test whether the main variables of interest are affected by including them. Specification (4) controls for per capita GDP, agricultural yield (to proxy for soil quality), and police officers per 100,000 population (to proxy for law enforcement).<sup>32</sup>

---

<sup>32</sup> The state-level per-capita GDP is from the Directorate of Economics and Statistics of respective state governments and from the Central Statistical organization (CSO). The state-level police personnel per 100,000 residents is from the Bureau of Police Research and Development. The data on agricultural yield per acre are from the Ministry of Agriculture, Government of India. District-level data on soil quality are available for some, but not all districts. In cases where the yield is not available, we impute the state-level yield for all districts in that state.

The central conclusion of Table 7 is that the coefficient estimates on the caste and religion heterogeneity variables are robust to excluding and including different combinations of variables. In the baseline regression, the coefficient estimate of Hindu caste homogeneity index is .325 for the total tap, .273 for the indoor tap, and .187 for the outside tap, the first two of which are statistically significant at the .05 level. In specifications 2-4 the coefficient estimate on caste index varies from .331 to -.344, all of which are statistically significant and larger than the coefficient estimate in the base specification. Similarly, the coefficient estimates for indoor tap are all statistically significant and vary from .265 to .269, which are not statistically different from the base estimate of .273. For access to outside water, the three robustness checks report coefficient estimates that are all larger than the base specification (none of which are statistically significant).

The coefficient estimate for the religion homogeneity index exhibits similar stability for the total, indoor, and outdoor access to water. All 3 of the baseline estimates and all 9 of the robustness estimates are statistically significant at the .05 level. For total access, the baseline estimate of -.317 is not statistically different from the three robustness checks that vary from -.290 to -.316. Similarly, the coefficient estimates for indoor vary from -.256 to -.268. For indoor access, the coefficient estimate varies from -.256 to -.268 with the baseline estimate of -.266. The coefficient estimates for outdoor access vary from -.312 to -.353 with the baseline estimate of -.350. In sum, Table 7 clearly demonstrates that the coefficient estimates on the religion and caste homogeneity indexes are highly robust to the inclusion or exclusion of different sets of variables.

To explore how the relationship between the two homogeneity indexes and public access to water, Table 8 reports the results of two regressions that interact the caste and religion homogeneity indexes with other control variables. The first section of the table reports the results

of a regression that interacts the two indexes with population and the second section reports the results of a regression that interacts these two indexes with per capita bank deposits. In the first half of the table, the coefficient estimates for the caste and religion variables continue to be similar to the baseline estimates reported in Table 6. For all three types of water access, the caste homogeneity index continues to have positive coefficient estimates (.426 for total, .169 for indoor, and .308 for outdoor) and the religion homogeneity variable has negative coefficient estimates (-.422 for total, -.310 for indoor, and -.423 for outdoor). The magnitudes of the caste and religion coefficient estimates with the population interactions are larger in five of the six cases than they are in the baseline regression. The levels of statistical significance are also similar with the exception that the caste homogeneity index is now no longer statistically significant for inside access to water. The interaction between religion and population is statistically significant for total and indoor access. None of the interactions between caste and population are statistically significant.

In the second half of the table, the coefficient estimates on the religion homogeneity index continue to be negative and statistically significant (-.361 for total, -.259 for indoor, and -.383 for outdoor access) and are similar to the base line coefficient estimates from Table 6 of (-.317 for total, -.266 for indoor, and -.351 for outdoor). The coefficient estimates on the caste homogeneity index and the interactions between the two indexes and the per capita bank deposits are not statistically different from zero in any of the three cases. In sum, the results from the regressions with the interaction variables confirm the central conclusion of the paper—that multiple types of heterogeneity (in this case caste and religious), can have different effects on the access to public goods and should be disaggregated when analyzing the relationship.

## 7. Conclusions

We examine whether different aspects of social divisions in India help explain the wide variation in access to tap water across rural India. In contrast to most studies, which use aggregated measures of social fragmentation that are comprised of several socio-economic characteristics such as ethnicity, race, language, religion, and caste, we employ disaggregated measures. Consequently, our approach allows for individual measures of diversity to have heterogeneous effects on outcomes.

Our empirical analysis documents that communities that are heterogeneous in terms of caste within the majority Hindu religion are likely to have lower access to tap water than correspondingly homogeneous communities. Historical divisions created by the caste system in India may generate rivalry associated with the usage and sharing of public goods, and thereby reduce the coordination required to get the state to provide more public goods. By contrast, communities that are fragmented across religions are likely to have more access to tap water than correspondingly homogeneous communities. This may be because the risk of internal conflict due to religious divisions (a common source of violent riots in India) cause the state to send more public goods to areas fragmented by religion. This may be more prevalent in rural areas, where the effectiveness of law and order in times of conflict may be limited due to inadequate infrastructure (roads, communications, etc.).

These interesting results indicate that both heterogeneity *within* and *across* religions matter for access to public goods, but they may work in opposite directions. Consequently, studies that use an aggregated measure of social fragmentation by combining many characteristics of social divisions are unlikely to reveal reliable information regarding its impact.

Our data and results point to the need for policy reform in the water sector in India. Given that certain types of social divisions like the caste system creates barriers for adequate public provision of drinking water, should public policy encourage private participation in the water sector? Two recent examples strengthen this view. Davis et al. (2008) survey 800 poor households in the southern city of Hyderabad. A large majority reported inadequacies in government provided water and sanitation. Interestingly, their regression analysis suggests that even if faced with non-concessional market rates of financing, these households would prefer to pay for private investment in water and sewer connections. These results underscore the vital role micro-financing can play in overcoming social barriers. Another example comes from the region of Tirupur in southern India. A recent public-private partnership has ensured the supply of drinking water for 4-6 hours each day for 80,000 households, compared to getting water every alternate day of the week before the partnership. More interestingly, 100% of the residents (mostly poor) now pay for drinking water (Mulford, 2006).

We end our discussion with a caveat. Because we use district-level data, it is not possible for us to identify the underlying mechanism that drives the opposite signs for the group heterogeneity results (caste and religion). The problem is that, for local public goods like water, the measurement of social heterogeneity and water access would ideally occur at a smaller kilometer grid because each district may contain multiple communities and therefore investments in one community may not benefit others (Jackson, 2007). The lack of data at a more disaggregated level for rural India precludes a solution for this problem. However, in the wake of the current water crisis in India, these results provide insights into the role played by two important sources of social fragmentation in India, namely caste and religion, and will, in turn, direct future research to analyze the underlying mechanism that drives these opposing effects of

heterogeneity. In this context, an interesting extension might be to consider the role of elected women leaders in villages fragmented along the lines of caste and religion in mitigating the effects of diversity in determining access to public goods. The recent contribution of Beaman et al. (2009) is a good starting point for extending this line of research.

## References

- Agricultural Census of India (2001), Department of Agriculture & Cooperation, Agricultural Census Division, Ministry of Agriculture, India.
- Alesina, A., Baqir, R. and Easterly, W. (1999), “Public goods and ethnic divisions.” *Quarterly Journal of Economics* 114 (4): 1243–1284.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003), “Fractionalization.” *Journal of Economic Growth*, 8, 155–194.
- Anderson, S. (2011), “Caste as an impediment to trade.” *American Economic Journal: Applied Economics* 3, 239-263.
- Bajpai, N. (2007), “India’s Growing water crisis.” *Hindu Business Line*, 12, January.
- Banerjee, A., Iyer, L. and Somanathan, R. (2005), “History, Social Divisions, and Public Goods in Rural India.” *Journal of the European Economic Association*, 3(2-3), 639–647.
- Banerjee, A., and Iyer, L. (2007), “Public action for public goods,” in T. Paul Schultz and John Strauss (ed.), *Handbook of Development Economics*, Vol. 4. Amsterdam: Elsevier North-Holland.
- Banerjee, A. and Somanathan, R. (2007), “The Political Economy of Public Goods: Some Evidence from India.” *Journal of Development Economics* 82, 287-314.
- Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., and Topalova, P. (2009), “Powerful women: does exposure reduce bias?” *Quarterly Journal of Economics* 124, 1497-1540.
- Besley, T. and Torsten, P. (2009), “The origins of state capacity: property rights, taxation, and politics.” *American Economic Review* 99, 1218-1244.
- Besley, T. and Torsten, P. (2010), “State capacity, conflict, and development.” *Econometrica* 78, 1-34.

Brooks, N. (2007), “Imminent water Crisis in India.”

<http://www.arlingtoninstitute.org/wbp/global-water-crisis/606#>

Census of India (2001), *Office of the Registrar General & Census Commissioner*, India.

Davis, J., Gary W., Said D., and Rich, T. (2008), “Improving access to water supply and sanitation in urban India: microfinance for water and sanitation infrastructure development.” *Water Science and Technology* 58, 887-891.

Dayton-Johnson, J. (2000), “Determinants of collective action on the local commons: a model with evidence from Mexico.” *Journal of Development Economics* 62, 181-208.

Dreze, J., Lanjouw, P., and Sharma, N. (1999), “Economic development in Palanpur, 1957-93,” in P. Lanjouw and N. Stern (ed.), *Economic Development in Palanpur Over Five Decades*. Oxford, UK: Oxford University Press.

Easterly, W., and Levine, R. (1997), “Africa’s Growth Tragedy: Policies and Ethnic Divisions,” *The Quarterly Journal of Economics* 112, 1203–1250.

Election Commission of India (1999), *Statistical Reports of Lok Sabha Elections*.

<http://www.eci.gov.in/StatisticalReports/ElectionStatistics.asp>

Esteban, J. and Ray, D. (1999), “Conflict and distribution.” *Journal of Economic Theory* 87, 379-415.

Haarmeyer, D. and Mody, A. (1997), “Private capital in water and sanitation.” *Finance and Development* 34, 31-37.

Haub, C. and Sharma, O. (2006), “India’s population reality: reconciling change and tradition,” *Population Bulletin*. 61, 1-20.

Jackson, K. (2007), “Why does diversity affect public good provision? An empirical analysis of water provision in Africa.” Dissertation Chapter, University of British Columbia.

- Jha, S. (2010), "Trade, complementarities and religious tolerance: evidence from India." Unpublished manuscript, Stanford University.
- Khwaja, A. (2009), "Can good projects succeed in bad communities?" *Journal of Public Economics* 93, 899-916.
- Michalopoulos, S. (2011), "The origins of ethnolinguistic diversity." *American Economic Review*, forthcoming.
- Miguel, E., and Gugerty, M. (2005), "Ethnic diversity, social sanctions and public goods in Kenya." *Journal of Public Economics* 89, 2325–2368.
- Mulford, D. (2006), "Public-Private partnership can help solve India's water and sanitation challenges." *The Hindu*, March 22 issue.
- Munshi, K. and Rosenzweig, M. (2009), "Why is mobility in India so low? social insurance, inequality, and growth." Unpublished manuscript, Brown University.
- Ramachandran, A. (2006), "Any hope for India's water woes?" India/Kerala News.  
[http://www.greentreks.org/woodchuckcafe/features/feature\\_indiaswater.asp](http://www.greentreks.org/woodchuckcafe/features/feature_indiaswater.asp)
- Ray, D. (1998), *Development Economics*. Princeton University Press, Princeton, NJ.
- Rushton, M. (2008), "A note on the use and misuse of the racial diversity index." *Policy Studies Journal*, Forthcoming.
- Saleth, M. (2005), "Water institutions in India: structure, performance, and change," in Gopalakrishnan, C., Biswas, A., and Tortajada, C. (ed.), *Water Institutions: Policies, Performance and Prospects*, 47-80. New York, NY.
- Varshney, A., and Wilkinson, S. (2004), "Varshney-Wilkinson dataset on Hindu-Muslim violence in India, version 2."
- Vigdor, J. (2002), "Interpreting ethnic fragmentation effects." *Economics Letters*, 75, 271-276.

Vigdor, J. (2004), "Community composition and collective action: analyzing initial mail response to the 2000 Census." *The Review of Economics and Statistics*, 86, 303–312.

Wasteland Atlas of India (2003), Ministry of Rural Development, Department of Land Resources, Government of India. <http://dolr.nic.in/fwastecatg.htm>

**Table 1: Summary Statistics**

Variables	Number of Observations	Mean	Std. Dev.	Min.	Max.
<b><i>Dependent Variables</i></b>					
Total tap share	436	.27	.24	.01	.91
Total Tap Share (Within)	436	.11	.11	0	.57
Total Tap Share (Outside)	436	.24	.23	0	.88
<b><i>Independent Variables</i></b>					
<b>Caste Variables</b>					
Hindu Caste Index (Homogeneity)	436	.04	.05	0	.32
Brahman	436	.04	.04	0	.27
Scheduled Caste	436	.11	.09	0	.46
Scheduled Tribe	436	.04	.06	0	.53
<b>Religion Variables</b>					
Religion Index (Homogeneity)	436	.80	.16	.37	.99
Hindus	436	.85	.18	.04	.99
Muslims	436	.09	.12	0	.79
Christians	436	.02	.06	0	.47
Sikhs	436	.03	.15	0	.94
Buddhists	436	.01	.04	0	.59
Jains	436	.002	.004	0	.04
Others	436	.003	.016	0	.19
Not Stated	436	.001	.001	0	.01
<b>Economic Variables</b>					
Land Gini	436	.71	.09	.41	.86
Per Capita Bank Deposits	436	.02	.02	.002	.23
<b>Political Variables</b>					
Political Index	436	.59	.09	.14	.80
Vote Share of Winning Party	436	.48	.08	.27	.71
<b>Population Variables</b>					
Household Size	436	5.38	.85	4	8
Number of Villages ('000s)	436	1.13	.821	.041	10.54
Avg. Village Population ('000s)	436	1.98	3.13	.115	26.79
<b>Geography Variables</b>					
Avg. Rainfall (meters)	436	.98	.68	.07	5.88
Avg. Temperature (Celsius)	436	25.44	2.89	14.47	29.88
Coastline	436	.13	.34	0	1
Sandy	436	.04	.11	0	.97
Barren	436	.04	.07	0	.50
Steepy	436	.01	.03	0	.49

**Table 2: Mean Share of Households with Access to Tap Water**

State Name	No. of Districts	Total	Within	Outside
Andhra Pradesh	22	0.40 (0.18)	0.13 (0.08)	0.27 (0.13)
Assam	23	0.06 (0.06)	0.01 (0.01)	0.05 (0.05)
Chattisgarh	16	0.05 (0.02)	0.02 (0.01)	0.03 (0.02)
Gujarat	25	0.47 (0.23)	0.25 (0.15)	0.22 (0.12)
Haryana	19	0.38 (0.14)	0.12 (0.06)	0.26 (0.11)
Himachal Pradesh	12	0.83 (0.06)	0.25 (0.09)	0.58 (0.11)
Karnataka	27	0.48 (0.17)	0.11 (0.04)	0.37 (0.15)
Kerala	14	0.14 (0.08)	0.08 (0.04)	0.06 (0.05)
Madhya Pradesh	45	0.11 (0.09)	0.05 (0.05)	0.06 (0.04)
Maharashtra	30	0.45 (0.16)	0.27 (0.12)	0.18 (0.07)
Orissa	29	0.03 (0.01)	0.01 (0.01)	0.02 (0.01)
Punjab	17	0.17 (0.08)	0.11 (0.04)	0.06 (0.05)
Rajasthan	32	0.21 (0.13)	0.13 (0.09)	0.08 (0.06)
Tamilnadu	28	0.6 (0.13)	0.07 (0.04)	0.53 (0.11)
Uttarkhand	11	0.67 (0.17)	0.19 (0.09)	0.49 (0.20)
Uttar Pradesh	70	0.15 (0.10)	0.09 (0.08)	0.06 (0.04)
West Bengal	16	0.08 (0.07)	0.02 (0.02)	0.06 (0.05)
Total	436	0.27 (0.24)	0.11 (0.10)	0.17 (0.18)

*Source:* Census of India, 2001. Standard deviations in parentheses.

**Table 3: Correlations between Tap Water Access and Hindu Caste Variables**

	Tap share (Total)	Tap share (within)	Tap share (outside)	Hindu Caste Index	Braham	SC	ST	Religious Index
Tap share (total)	1.00							
Tap share (within)	0.71	1.00						
Tap share (outside)	0.95	0.54	1.00					
Hindu Caste Index	0.28	0.29	0.21	1.00				
Brahman	0.13	0.13	0.1	0.42	1.00			
Scheduled Caste	0.01	0.08	-0.07	0.25	0.3	1.00		
Scheduled Tribe	0	0.24	-0.134	0.29	0.06	0.24	1.00	
Religion Index	0.1	-0.03	0.03	0.2	0.28	0.19	0.14	1.00

Note: SC- Scheduled Caste; ST- Scheduled Tribe

**Table 4: Correlations between Tap Water Access and Economic and Political Variables**

	Tap share (total)	Tap share (within)	Tap share (outside)	Land Gini	PCBD	Vote share	Political Index
Tap share (total)	1.00						
Tap share (within)	0.71	1.00					
Tap share (outside)	0.92	0.36	1.00				
Land Gini coefficient	-0.11	-0.1	-0.09	1.00			
PCBD	0.32	0.17	0.39	-0.17	1.00		
Vote share	0.15	0.04	0.16	0.08	0.14	1.00	
Political Index	-0.22	-0.06	-0.23	-0.04	-0.19	-0.91	1.00

Note: PCBD- Per Capita Bank Deposits.

**Table 5: Social Composition and Total Tap Water Access**  
*Aggregated Social Homogeneity Index*

	Tap water Share (Total)	Tap water Share (Within)	Tap water Share (Outside)
<b><u>Index</u></b>			
Social Homogeneity Index (aggregated)	-.129 (0.77)	-.114 (1.15)	-.228 (1.11)
<b><u>Share of Hindu Groups</u></b>			
Share of Brahmans	-.195 (-0.61)	-.125 (0.64)	-.112 (0.42)
Share of Scheduled castes	.139 (0.98)	.102** (1.99)	.119 (0.77)
Share of Scheduled Tribes	-.220** (2.06)	-.061 (0.72)	-.195* (1.79)
<b><u>Religion Variables</u></b>			
Share of Muslims	.112 (0.59)	.157 (1.20)	.098 (0.59)
Share of Christians	-.016 (0.11)	.098 (1.17)	0.053 (0.35)
Share of Sikhs	.358 (1.25)	.384** (2.38)	.100 (0.24)
Constant	.558*** (2.66)	.038 (0.31)	.591*** (2.41)
<i>N</i>	436	436	436
<i>Economic Control Variables</i>	YES	YES	YES
<i>Political Control Variables</i>	YES	YES	YES
<i>Geography Control Variables</i>	YES	YES	YES
<i>Population Control Variables</i>	YES	YES	YES
<i>Wasteland Control Variables</i>	YES	YES	YES
<i>State Fixed Effects</i>	YES	YES	YES
<i>Overall R<sup>2</sup></i>	0.028	0.003	0.085

*Heteroskedasticity-consistent t-statistics* (clustered by state) included in parentheses; \*, \*\* and \*\*\* represent 10, 5 and 1% significance level respectively. Columns (2)-(3) include shares of other religion groups.

**Table 6: Caste and Religious Composition and Total Tap Water Access**  
*Caste and Religious Homogeneity Indexes*

	Tap water Share (Total)	Tap water Share (Within)	Tap water Share (Outside)
<b><u>Indexes</u></b>			
Caste Homogeneity Index	.325** (2.23)	.273** (2.38)	.187 (1.16)
Religion Homogeneity Index	-.317** (2.46)	-.266*** (3.53)	-.351** (2.71)
<b><u>Share of Hindu Groups</u></b>			
Share of Brahmins	-.249 (0.84)	-.172 (0.98)	-.159 (0.65)
Share of Scheduled castes	.063 (0.47)	.038 (0.73)	.050 (0.34)
Share of Scheduled Tribes	-.275** (2.06)	-.109 (1.08)	-.245* (1.79)
<b><u>Religion Variables</u></b>			
Share of Muslims	-.184* (1.92)	-.094 (1.63)	-.286** (2.45)
Share of Christians	-.420* (1.89)	-.244* (1.87)	-.430* (1.89)
Share of Sikhs	.254* (1.83)	.291*** (3.90)	-.104 (0.65)
Constant	.781*** (3.29)	.225* (1.73)	.853*** (3.09)
<i>N</i>	436	436	436
<i>Economic Control Variables</i>	YES	YES	YES
<i>Political Control Variables</i>	YES	YES	YES
<i>Geography Control Variables</i>	YES	YES	YES
<i>Population Control Variables</i>	YES	YES	YES
<i>Wasteland Control Variables</i>	YES	YES	YES
<i>State Fixed Effects</i>	YES	YES	YES
<i>Overall R<sup>2</sup></i>	0.039	0.018	0.151

*Heteroskedasticity-consistent t-statistics (clustered by state) included in parentheses; \*, \*\* and \*\*\* represent 10, 5 and 1% significance level respectively. Columns (2)-(3) include shares of other religion groups.*

**Table 7: Robustness Check for the Hindu Caste and Religion Homogeneity Indexes**

Specification	Tap water share (Total) (1)	Tap water share (Within) (2)	Tap water share (Outside) (3)
(1) Baseline (including all controls)			
Caste Homogeneity Index	.325** (2.23)	.273** (2.38)	.187 (1.16)
Religion Homogeneity Index	-.317** (-2.46)	-.266*** (-3.53)	-.350*** (-2.71)
(2) Excluding population density variables			
Caste Homogeneity Index	.339** (2.21)	.269** (2.44)	.204 (1.19)
Religion Homogeneity Index	-.316** (-2.50)	-.256*** (-3.38)	-.353*** (-2.72)
(3) Excluding both the political controls			
Caste Homogeneity Index	.344** (2.06)	.265** (2.23)	.211 (1.18)
Religion Homogeneity Index	-.310** (-2.37)	-.268** (2.23)	-.343** (-2.55)
(4) Including police, State GDP per capita, and yield per acre			
Caste Homogeneity Index	.331** (2.26)	0.266*** (2.72)	0.189 (1.07)
Religion homogeneity index	-.290*** (-2.69)	-.264*** (-3.90)	-.312*** (-2.99)
<i>N</i>	436	436	436
<i>Economic Control Variables</i>	Yes	Yes	Yes
<i>Political Control Variables</i>	Yes	Yes	Yes
<i>Geography Control Variables</i>	Yes	Yes	Yes
<i>Population Control Variables</i>	Yes	Yes	Yes
<i>Wasteland Control Variables</i>	Yes	Yes	Yes
<i>State Fixed Effects</i>	Yes	Yes	Yes

*Heteroskedasticity-consistent t*- statistics (clustered by state) included in parentheses; \*, \*\* and \*\*\* represent 10, 5 and 1% significance level respectively.

**Table 8: Interactions with Caste and Religion Homogeneity Indexes**

Specification	Tap water Share (Total)	Tap water Share (Within)	Tap water Share (Outside)
<b><u>Caste and Religion Homogeneity Indexes Interacted with Population</u></b>			
Caste Homogeneity Index	0.426** (2.41)	0.169 (1.39)	0.308 (1.40)
Religion Homogeneity Index	-0.422*** (2.69)	-0.310*** (3.60)	-0.423*** (2.88)
Caste Index * Population	-0.084 (0.81)	0.121 (1.27)	-.114 (0.96)
Religion Index * Population	-0.0001* (1.91)	.00003* (1.76)	.00004 (1.28)
<b><u>Caste and Religion Homogeneity Indexes Interacted with per capita bank deposits</u></b>			
Caste Homogeneity Index	.062 (0.29)	0.132 (0.66)	-0.108 (0.64)
Religion Homogeneity Index	-0.361** (2.49)	-0.259*** (3.02)	-0.383** (2.48)
Caste Index * per capita bank deposits	9.036 (1.26)	4.650 (1.22)	10.034 (1.32)
Religion Index * per capita bank deposits	1.647 (1.15)	-3.087 (0.28)	1.216 (0.70)
<i>N</i>	436	436	436
<i>Economic Control Variables</i>	Yes	Yes	Yes
<i>Political Control Variables</i>	Yes	Yes	Yes
<i>Geography Control Variables</i>	Yes	Yes	Yes
<i>Population Control Variables</i>	Yes	Yes	Yes
<i>Wasteland Control Variables</i>	Yes	Yes	Yes
<i>State Fixed Effects</i>	Yes	Yes	Yes

*Heteroskedasticity-consistent t-statistics (clustered by state) included in parentheses; \*, \*\* and \*\*\* represent 10, 5 and 1% significance level respectively.*