ACCESS TO DRINKING WATER IN INDIA: THREE ESSAYS ON CURRENT PROBLEMS AND FUTURE PROSPECTS

by

DIVYA BALASUBRAMANIAM

(Under the direction of Santanu Chatterjee and David B. Mustard)

ABSTRACT

The current work consists of three empirical analyzes on access to drinking water in India. The first study examines the effect that access to safe water and sanitation has on child health outcomes in India. In particular, the paper focuses on "indirect" effects such as diarrhea on the nutritional distribution for children. The findings indicate that the effect of access to safe water and sanitation depends critically on a child's position in the conditional (quintile) nutrition distribution. The analysis is also conducted for urban and rural sub-samples to check for "location" effects. The evidence suggests that the geographical location of children can significantly impact their health status. From a policy perspective, these results indicate that rather than designing one-size-fits-all policies, improving health outcomes require that policies be targeted to the conditional distribution of nutrition as well as geography.

The second study examines whether different social divisions help explain the variation in tap water access across India. We find that communities that are heterogeneous in terms of Hindu caste have less access to tap water than correspondingly homogeneous communities. By contrast, religiously fragmented communities have more access to tap water than correspondingly homogeneous communities. Therefore, heterogeneity within and across religions may work in opposite directions for access to public goods. Consequently, the many studies that use aggregate measures of social fragmentation may obscure important information regarding the design of public policy related to public goods.

There is a large consensus that two broad mechanisms namely, institutions and preferences influence the effects of ethnic diversity on public goods provision. Although a large body of work has analyzed the effect of ethnic heterogeneity and social fragmentation on public goods access and provision, but the channel or mechanism through which these effects occur have not been examined. The final study examines which of these two channels influence social divisions to as a determinant of access to tap water in rural India. The analysis indicates that institutions are important in understanding why social divisions affect access to tap water in rural India.

INDEX WORDS: Public goods, Social fragmentation, Preferences, Institutions Water, Public policy, Child malnutrition, Child height, Child weight, Quantile regressions, India

ACCESS TO DRINKING WATER IN INDIA: THREE ESSAYS ON CURRENT PROBLEMS AND FUTURE PROSPECTS

by

DIVYA BALASUBRAMANIAM

B.COM., The University of Madras, India 2002

M.SC., Anna University, India 2004

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

ACCESS TO DRINKING WATER IN INDIA: THREE ESSAYS ON

CURRENT PROBLEMS AND FUTURE PROSPECTS

by

DIVYA BALASUBRAMANIAM

Major Professors: Santanu Chatterjee
David B. Mustard

Committee: Scott E. Atkinson
Ronald Warren Jr.,

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
July 2010

DEDICATION

I dedicate my Ph.D. to my mother, who passed away a few years back due to Leukemia. She has always been an inspiration and a source of motivation throughout my life. I am also very fortunate to have a husband (Ram) who has been constantly supportive, patient, and understanding throughout my PhD studies. My father has also played an important part in the successful completion of my PhD. He helped me obtain the data needed to undertake my research and I really thank him for his love and support.

ACKNOWLEDGEMENTS

I would like to first thank Professors Santanu Chatterjee and David Mustard for their continued guidance, support and patience throughout my research. I am also grateful to my dissertation committee members Professors Scott Atkinson and Ronald Warren Jr., for their encouragement and support. I would also like to thank Professor Angela Fertig for her helpful comments in my first chapter. I would also like to thank Dr. David Robinson for helping me improve my writing style.

I thank Mr. Madhusudan for his help in collecting data for my research. I am grateful to the Graduate School Dean's award that helped me buy the data for my research.

I would like to thank all professors in my graduate program. Last but not least I would like to thank the Department of Economics for the financial support and encouragement provided to me throughout my graduate studies.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS..v

LIST OF TABLES viii

CHAPTER

1 INTRODUCTION..1

2 DISTRIBUTIONAL CONSEQUENCES OF ACCESS TO SAFE WATER AND SANITATION ON CHILD NUTRITION IN INDIA: A QUANTILE REGRESSIONS APPROACH 4

2.1 INTRODUCTION 4

2.2 A STYLIZED MODEL OF CHILD HEALTH 7

2.3 METHODS..9

2.4 DATA AND DESCRIPTIVE STATISTICS 12

2.5 RESULTS 14

2.6 CONCLUSIONS 22

2.7 REFERENCES..24

3 GOT WATER? SOCIAL DIVISIONS AND ACCESS TO PUBLIC GOODS IN RURAL INDIA..36

3.1 INTRODUCTION 36

3.2 INDIA'S INSTITUTIONAL SETUP AND WATER CRISIS 39

3.3 SOCIO-RELIGIOUS STRATIFICATION AND ITS EFFECTS ON ACCESS TO DRINKING WATER 41
3.4 EMPIRICAL SPECIFICATION 44
3.5 DATA 47
3.6 RESULTS AND DISCUSSION 53
3.7 CONCLUSIONS 59
3.8 REFERENCES 61
4 INSTITUTIONS VERSUS PREFERENCES APPROACH: WHICH UNDERLYING MECHANISM DETERMINES TAP WATER ACCESS IN RURAL AND URBAN INDIA? EVIDENCE FROM THE NATIONAL FAMILY HEALTH SURVEY 74
4.1 INTRODUCTION 74
4.2 THEORETICAL FRAMEWORK 76
4.3 METHODS 78
4.4 DATA, DESCRIPTION, AND ECONOMETRIC ISSUES 79
4.5 RESULTS 81
4.6 CONCLUSIONS 82
4.7 REFERENCES 82
5 CONCLUSIONS 87

LIST OF TABLES

2.1 Summary Statistics 28

2.2 Classification of Malnutrition 29

2.3 OLS and Quantile Regressions (Dependent Variable: Weight-for-Height Z- Scores 30

2.4 OLS and Quantile Regressions (Dependent Variable: Height-for-Age Z-Scores 31

2.5 OLS and Quantile Regressions (Dependent Variable: Weight-for-Age Z- Scores 32

2.6 Urban Vs. Rural-OLS and Quantile Regressions (Dependent Variable: Weight-for-Height Z-Scores) 33

2.7 Urban Vs. Rural-OLS and Quantile Regressions (Dependent Variable: Height-for-Age Z-Scores) 34

2.8 Urban Vs. Rural-OLS and Quantile Regressions (Dependent Variable: Weight-for-Age Z-Scores) 35

3.1 Summary Statistics 65

3.2 Mean Share of Households with Access to Tap Water 66

3.3 Correlation between Tap Water Access and Hindu Caste Variables 67

3.4 Correlation between Tap Water Access and Economic and Political Variables 67

3.5 Caste Composition and Total Tap Water Access...68

3.6 Caste Composition and Within-Residence Tap Water Access 69

3.7 Caste Composition and Outside-Residence Tap Water Access..70

3.8 Religious Composition and Tap Water Access 71

3.9 Robustness Check for the Hindu Caste Homogeneity Index...72

3.10 Robustness Check for The Religion Homogeneity Index 73

4.1 Summary Statistics- Rural 84

4.2 Summary Statistics - Urban 84

4.3 OLS and IV Regressions- Rural..85

4.4 OLS and IV Regressions – Urban 86

CHAPTER 1

INTRODUCTION

Water is a necessity for human existence and sustenance. It is also critical for economic progress, since most activities in agriculture, industry and services depend on water. For developing countries, which are predominantly poor and heavily dependent on agriculture, the efficient provision and management of water resources is essential for economic development. This study seeks to focus on India and its challenges with adequate water provision to a large and growing population.

India is a federal democracy where the government provides most public services like water, healthcare, sanitation, education and transport, to name a few. For many years now, India has faced a growing water crisis. A rapidly growing population, fast-paced urbanization, industrialization and competition between industry, services, and agriculture have generated a huge demand for water services. However, in spite of being endowed with a large river network (indeed, the Ganges forms one of the world's largest river basins) and its tropical geo-position, the supply of water in India has been dramatically inadequate, resulting in severe shortages in both urban and rural areas. India's water crisis threatens the ability of the government to make cities habitable and healthy and thus raises doubts about sustainable economic growth and development.

Given this background, the study seeks answers to the following three questions: (1) What are the implications for access to safe water and sanitation on child health outcomes in India? (2) What are the sociological, economic, and political factors that contribute to the variation in tap water access across India? (3) Why do social factors play an important role in influencing access to these goods?

The first paper examines the effect that access to safe water and sanitation has on child health outcomes in India. In particular, the paper focuses on "indirect" effects such as diarrhea on the nutritional distribution for children. This is the first study to test for indirect effects using the quantile regressions approach. This represents a significant departure from the traditional ordinary least squares estimation, where partial effects may not reveal important information regarding health outcomes across the nutritional distribution. I find that the effect of access to safe water and sanitation depends critically on a child's position in the conditional (quintile) nutrition distribution. Additionally, this is the first study to separately analyze the distributional impacts on child nutrition in urban and rural areas in India. The evidence suggests that the geographical location of children can significantly impact their health status. From a policy perspective, these results indicate that rather than designing one-size-fits-all policies, improving health outcomes require that policies be targeted to the conditional distribution of nutrition as well as geography.

The second chapter examines whether different social divisions help explain the variation in tap water access across India. Using data for 436 rural districts from the 2001 Census of India, the paper finds that communities that are heterogeneous in terms of Hindu caste have less access to tap water than correspondingly homogeneous communities. By contrast, religiously fragmented communities have more access to tap water than correspondingly homogeneous communities. Therefore, heterogeneity *within* and *across* religions may work in opposite directions for access to public goods. Consequently, the many studies that use aggregate measures of social fragmentation may obscure important information regarding the design of public policy related to public goods.

There is a large consensus that two broad mechanisms namely, *institutions* and *preferences* influence the effects of ethnic diversity on public goods provision. Although a large body of work has analyzed the effect of ethnic heterogeneity and social fragmentation on public goods access and provision, but the channel or mechanism through which these effects occur have not been examined. The third paper examines which of these two channels influence social divisions to as a determinant of access to tap water in rural India. It uses the household-level dataset from the 1998-1999 National Family Health Survey. The analysis indicates that institutions are important in understanding why social divisions affect access to tap water in rural India.

CHAPTER 2

DISTRIBUTIONAL CONSEQUENCES OF ACCESS TO SAFE WATER AND SANITATION ON CHILD NUTRITION IN INDIA: A QUANTILE REGRESSIONS APPROACH

2.1 INTRODUCTION

Malnutrition is a major health problem for women and children in developing countries.[1] It is one of the major causes of childhood ill-health and leads to five million deaths among children each year (Aturupane et al., 2008, World Hunger Facts, 2008). Malnutrition occurs when a person lacks some or all nutrients essential for good health.[2] Healthy people have better immune systems, are more productive, and are more likely to fight poverty and find opportunities to sustain a living (World Health Organization (WHO), 2008). Lack of access to safe water and sanitation is a major factor that has led to more than 4,500 children under the age of five dying each day due to diarrhea and about 800,000 children worldwide under the age of five dying each year due to malaria (WHO, 2008). High disease incidence can reduce the body's ability to take in nutrition. In addition, unsafe water and sanitation can lead to malnutrition (WHO, 2008). Therefore, interventions to ensure safe water and sanitation can help reduce malnutrition(WHO,2008).

[1] According to the Food and Agriculture Organization's 2006 estimate, there are about 854 million malnourished people worldwide, of which 820 million live in developing countries (World Hunger Facts, 2008).

[2] Stunting (low height-for-age z-score), wasting (low weight-for-height z-score), and underweight (low weight-for-age z-score) are measures of malnutrition (Nandy et al. (2005)). I will use these terms interchangeably in the analysis. A z-score is the standard deviation from the median of the reference population.

A few studies find evidence that safe water improves child nutrition outcomes.[3] However, they use the traditional ordinary least squares (OLS) and instrumental variables approach to address the relationship between water and nutritional status. OLS captures how an improvement in water and sanitation access affects the nutritional status of an average child. However, the effect of safe water and sanitation is likely to differ across the conditional nutritional status distribution.[4] Quantile regressions (QR) estimation is a novel approach, which examines the distributional impact of safe water and sanitation across various points on the conditional nutritional status distribution. For this reason I employ the QR estimation in this analysis.

I use the National Family Health Survey for 2005-2006 to analyze the distributional impacts of safe water and sanitation on child nutrition in India. This analysis identifies which group in the distribution is most affected by improvements in safe water and sanitation, thereby informing policymakers about the most cost-effective groups to target for intervention. To this extent, this paper is related to Borooah (2005), Aturupane et al. (2008), and Bassole (2007) who examine the distributional impacts of access to water on child nutrition.[5] However, this study varies from the existing literature in several important dimensions.

---

[3] Thomas and Strauss (1992) find that child height is significantly affected by the availability of water, modern sewerage and electricity in Brazil. Esrey (1996), using data from Demographic Health Surveys (DHS) for eight countries, finds that improved water quality increases child height and weight when sanitation is also improved. Finally, David et al. (2004) find that in Western Honduras access to tap water had a positive impact on only the height-for-age outcomes. However, the community variables like the proportion of households having access to tap and toilets did not affect the height-for-age, weight-for-age and weight-for-height among children in Nicaragua.

[4] For example, the QR estimators β^q for each value of q (quintiles) examine how different parts of y depend on the covariate vector X^j (Koenker and Bassett (1978)).

[5] Borooah (2005) examines the determinants of height among children under the age of five in India. Using height-for-age as a health outcome measure, he estimates the model using both OLS and QR and finds that the OLS underestimates the effects of improvements in safe water and hospitals on child nutrition. Aturupane et al. (2008) use the DHS data for Sri Lanka to find that improvements in piped water positively affect child nutrition for all quantiles. Finally, Bassole (2007) uses a household survey to find the effect of access to public infrastructure on child nutrition in Senegal using an instrumental variables quantile approach to account for the potential endogeneity of household expenditure. He finds that access to safe water improves child nutrition (measured by height for age z-scores) of the lowest (10^{th}) quantile.

Even though several studies analyze the direct effect of water and sanitation on child nutrition, they fail to test for indirect effects such as diarrhea. The indirect effects capture morbidity, and can influence the health status of a child. This is the first study to test for indirect effects using the quantile regressions approach. Second, I use richer and more precise measures of safe water and sanitation than those used previously. While I measure access to water and sanitation at the household level, some papers treat water access at a more aggregated level.[6] I also analyze the impact of water and sanitation on child nutrition based on the different sources of water and sanitation facilities such as, access to piped water, well water, or a public tap and access to flush toilets or pit latrines within the household.

This analysis uses several measures of child nutrition namely, weight-for-height, height-for-age, and weight-for-age z-scores that may capture a more comprehensive picture of the direct and indirect effects of safe water and sanitation.[7] Finally, I separately analyze the distributional impacts on child nutrition in urban and rural areas. The mortality rate for an average child under the age of five who lives in rural areas with inadequate basic health services and sanitation is 105 per 1000 live births compared to 69 deaths per 1,000 live births in urban areas (Doherty, 2008). Distributional impacts may be very different by geographical location, and this breakdown will allow resources to be further targeted.

The main results of the paper are: First, access to sanitation significantly improves the weight-for-height, height-for-age, and weight-for-age z-scores. For example, a 1 percent

[6] For example, Borooah (2005), measures access to water based on whether the child lives in a village that has safe water.

[7] Recent studies mention that using measures like height-for-age, weight-for-age or weight-for-height z-scores individually to assess child nutrition fail to capture the true number of undernourished people. These individual measures ignore other dimensions of undernourishment, since weight-for-age is a product of stunting and wasting and not the sum. An alternative measure of malnourishment called the composite index of anthropometric failure (CIAF) captures all dimensions of malnourishment when the objective is to capture the proportion of the population suffering from anthropometric failure (Svedberg (2000)). Since the objective of this paper is to use anthropometrics for predicting health outcomes, I cannot use the CIAF index, which is a form of classification rather than an index.

increase in access to flush toilets is associated with an increase of .126 standard deviations in the weight-for-height z-scores at the 20th quintile, .086 standard deviations in the 40th, and .096 standard deviations in the 50th quintiles. Second, an incidence of diarrhea reduces the weight-for-height, height-for-age, and weight-for-age z-scores. In addition, I find evidence from the QR estimations that this effect varies significantly along the conditional nutrition distribution. For example, diarrhea reduces the height-for age z-scores in the 80th quintiles only, while diarrhea reduces the weight-for-age z-scores in all quintiles, and it reduces the weight-for-height z-scores in the 40th, 60th, and quintiles. Third, there is evidence of location effects (as explained earlier). For example, an incidence of diarrhea significantly reduces the weight-for-height z-scores for rural children in upper quintiles, but there is no such evidence for urban children. On the other hand, diarrhea reduces the height-for-age z-scores in the 80th quintile only for rural children, and in the median, 60th, and 80th quintiles for urban children. The secondary results of the study are that mother's literacy significantly determines child health outcomes. In this respect, I find that educated mothers improve the nutritional status of children, and the strength of this effect depends on the geographical location.

The rest of the paper is organized as follows. Section 2.2 describes a stylized model of child health. Section 2.3 discusses the methods, use of quantile regressions, the reduced-form equation, and definition of the variables used in the analysis, while section 2.4 characterizes the data and descriptive statistics. Section 2.5 discusses the results and some underlying intuitions, while Section 2.6 concludes.

2.2 A STYLIZED MODEL OF CHILD HEALTH

Following Becker (1981), I use a model of family allocation decisions commonly used in the child nutrition literature (Esrey et al. (1992), Behrman and Deolalikar (1998), Thomas and

Strauss (1992), Strauss and Thomas (1998), Bassole (2007)). In this model, parents choose the quantity and quality of child's health H, leisure L, and consumption of goods and services C. The household maximizes the following quasi-concave utility function:

$$\max_{H,L,C} U = U(H, L, C; X_h, \psi) \tag{1}$$

X_h consists of the parents' characteristics and ψ represents the unobserved heterogeneity of preferences. The parents maximize the utility based on two constraints, namely, the child health production function and the budget constraint. The child's health production function is

$$H_i = f(M_i, c_i, c_h, c_c, \xi_i) \tag{2}$$

where M_i is the vector of health inputs including duration of breastfeeding, age at which introduced to foods, nutrient intake, and disease incidence. c_i is a vector of child characteristics like sex and age. c_h is a vector of household characteristics such as the wealth holdings, access to piped water into dwelling, public tap, well water and access to pit latrine and flush toilets, and mother's characteristics.

Theoretically, for example, we expect that educated mothers can specifically provide better care for children's health. c_c is a vector of local community characteristics and ξ_i represents the unobservable child, household and community characteristics that affect a child's anthropometric measures. [8] The parents' budget constraint is

$$p_m M + p_c C = \underbrace{w(T - L) + y}_{z} \tag{3}$$

Let z denote the right-hand side of equation (3), where T is the total time endowed , $w(T-L)$ is the labor income, y is the non-labor income and p_m, p_c and w are the

[8] Anthropometry is defined as "the study of human measurement for use in anthropological classification and comparison" (thefreedictionary.com).

price vectors for health inputs, consumption goods and leisure, respectively. Solving the system (1)-(3), we derive the reduced-form function describing parents' demand for child health which depends only on exogenous child, household and community characteristics:

$$H_i = g(c_i, c_h, c_c, Z, p_c, p_m, \upsilon_i) \tag{4}$$

where υ_i is the unobserved heterogeneity of health outcomes. Equation (4) has been the focus of most empirical literature that examines child health outcomes.

2.3 METHODS

2.3.1 Reduced-Form Equation

The unit of analysis is individual children between the ages of six months and two years, because this is the period during which the human body builds immunity and has rapid growth. It also corresponds to a transition period in the life of children who start eating foods other than their mother's milk (Sheperd, 2007). Therefore, children in this age group are most vulnerable to water and airborne diseases. The reduced form equation (4) can be empirically estimated as follows

$$H_i = \alpha + \lambda W_i + \eta S_i + \beta X_i + \chi X_h + \delta RC_i + \phi X_c + \varepsilon_i \tag{5}$$

where H_i is the vector of anthropometric measures of children (height-for-age, weight-for-height and weight-for-age z-scores used individually). W_i captures whether the household has access to piped water, public tap or well water, and S_i captures whether the household has access to sanitation facilities. X_i is a vector of individual child characteristics that include sex, age, birth order, and incidence of diarrhea, X_h is a vector of other household characteristics that includes several mother's characteristics (see section 1.3.4), and a wealth index.[9] RC_i is a vector of religion and caste variables. X_c is a vector of community characteristics that includes whether the child lives in a city or a slum, and whether the child has access to a nearby healthcare facility.[10] ε_i is the error term.

2.3.2 Hypotheses

The primary variables of interest include all the water and sanitation (WS) variables, and diarrhea. I expect the signs of the WS coefficients λ and η to be positive, because higher access to water and sanitation assists better health outcome. On the other hand, I expect an incidence of diarrhea to decrease the weight-for-height, height-for-age, and weight-for-age z-scores. In addition to these interest variables, for the urban and rural sub-sample analysis, I focus on mother's education variables. Theoretically, educated mothers tend to be more aware and knowledgeable about child healthcare than illiterate mothers. Therefore, I expect literacy to improve the health outcomes.

2.3.3 Quantile Regression

The QR approach examines the impact of access to water and sanitation at several points on the conditional distribution of the child's nutrition status. This approach is particularly useful for this analysis because, for example, improved access to water in the household might improve the status of children in the lowest quantile of the weight for height distribution, while having a small effect on children in the higher quantiles. β_j^q, the QR coefficient captures the change in

---

[9] To account for gender inequality I include a gender variable (Sen, 2001 and Borooah, 2005). I include birth order of a child, because as the household size increases, there is competing use of parents' time for childcare. I control for age and age square, to allow for non-linear effects of age on the health outcomes.

[10] A slum, according to the United Nations, UN-Habitat Agency, is a run-down area of a city characterized by substandard housing and squalor and lacking in tenure security (Wikipedia).

the qth conditional quantile value of the dependent variable, y, with respect to the jth regressor, or $\beta_j^q = \partial quantile_q(y \mid X^j)/\partial x_j$.[11]

2.3.3 Definition of Variables used in the analysis

A) Dependent variables

Weight-for-height, height-for-age and weight-for-age z-scores denote the number of standard deviations that a child is above or below the growth standards provided by the World Health Organization/National Center for Health Statistics (WHO/NCHS). WHO/NCHS growth standards are international gold standards describing how children should grow when measured by weight and height. A low height-for-age z-score indicates long-term change in growth and a low weight-for-height z-score indicates short-term change in growth. Typically, children whose anthropometric measurements fall below -2 z-scores from the reference population median are considered malnourished.

B) Independent variables

I use binary variables to measure access to water and sanitation within the household. The dummy variables take on the value one if the household has access to these facilities, else they are zero. I include a dummy variable for whether the child had an incidence of diarrhea in the past 2 weeks. The other independent variables include child-level characteristics such as age, age-squared, birth order and sex. Age and age-squared are measured in months. The model includes mothers characteristics such as age in years and dummy variables for primary, secondary and higher educational attainment. DHS survey defines educational attainment as follows: a) primary: when a mother has five years of education, b) secondary: when a mother has between six and twelve years of education and c) higher: when a mother more than twelve years

[11] For example, improved access to water in the household might improve the status of children in the lowest quintile of the weight-for-height distribution, while having a small effect on children in the other quintiles.

of education. Finally, I include laborers and cultivators as controls for occupation, which are binary variables.[12]

The DHS survey calculates the wealth index using the principle component analysis, which assigns different weights on a household's various asset holdings (for example, television, radio, etc.) and living conditions (for example, type of roof materials, etc.). I use the wealth index, to minimize the endogeneity problem, because this index includes the assets possessed by a household over time and is less likely to be jointly determined with child health. All religion and caste variables are dummy variables that represent whether the child belongs to a particular religion and caste.[13] I also include dummy variables to represent whether the child lives in a city or a slum, and whether the household has access to a nearby healthcare facility.

2.4 DATA AND DESCRIPTIVE STATISTICS

I use data from the 2005-2006 National Family Health Survey (NFHS-3) for India. The survey covers 123,485 women between the ages of 15 and 49 across 29 states from a total of 109,041 households in both rural and urban areas.[14] The survey interviewed more than one eligible woman per household and asks several questions including child health, anthropometric measures of children, family planning practices, fertility and the socioeconomic status of women. The survey covers a total of 51,555 children under the age of five.

---

[12] A cultivator prepares a proper seedbed for the crop to be planted into, buries crop residue in the soil, controls weeds, and mixes and incorporates the soil to ensure the growing crop has enough water and nutrients to grow well during the growing season (Wikipedia).

[13] Caste is a form of social stratification of the population in India. Scheduled Caste and Scheduled Tribes are terms the government of India uses 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*. There are several other backward castes in India.

[14] Data on union territories are not collected. A union territory is a sub-national administrative division of India. Unlike the states, which have their own elected governments, union territories are ruled directly by the federal national government (Wikipedia).

I include children from households who are usual residents.[15] I include one child per household and exclude multiple births in the analysis, because children from the same household share the household and mothers characteristics. After omitting observations due to missing values, I include 11,466 children between six months and two years of age in this analysis.[16] There are 4,267 children in urban areas and 7,199 children in rural areas. The household response rate is 98 percent for India as a whole, 97 percent in urban areas, and 99 percent in rural areas. Table 2.1 presents the summary statistics for the variables, including rural and urban sub-samples. The WHO standard considers a child to be stunted, wasted and underweight if the respective z-score falls below -2.[17] The mean height-for-age, weight-for-age and weight-for-height z-scores are -1.56, -1.76 and -1.00, respectively for the total sample, which shows that children on average are close to being underweight. On the other hand, the mean height-for-age, weight-for-age and weight-for-height z-scores for the urban sub-sample are -1.38, -1.53, and -0.86 respectively and -1.69, -1.90, and -1.08 respectively for the rural sample. These figures show that children in rural areas have lower z-scores than urban children and are also close to being underweight. The summary statistics shows that on average, the age of the child and mother, and incidence of diarrhea is similar across urban and rural areas. In contrast, there is large variation in other variables such as mother's educational attainment, access to water and sanitation across urban and rural areas. For example, in rural areas, the mean access to piped water is only 13 percent, whereas in urban areas it is about 50 percent.

[15] The survey interviews both women who are usual residents of the household, and those who are visitors.
[16] I conduct a two-sample t-test on each of the variables used in the analysis to check for sample selection bias. None of the variables is statistically significant at the 5 percent level.
[17] $$z-score = \frac{\text{Measured Value - Average Value in the Reference Population}}{\text{Standard Deviation of the Reference Population}}$$

Table 2.2 reports that about 29 percent of the children are stunted and underweight, 16 percent of the children are underweight and wasted, and wasted and 7 percent of the children are stunted, wasted and underweight.

2.5 RESULTS

2.5.1 OLS and QR results for the entire sample

Table 2.3 shows the regression results when the weight-for-height z-score is the dependent variable. I report estimation results for the variables of interest in addition to some controls.[18] Column (1) presents the OLS results that include the entire set of controls.[19] Access to well water decreases the weight-for-height z-scores, although I include "protected" well water in the analysis and it is statistically significant at the 1 percent level.[20] For a 1 percent increase in access to well water, the weight-for-height z-scores decrease by 0.13 percent.[21] Columns (2)-(6) present the QR estimation. There is strong evidence that the weight-for-height z-scores of children decrease in all quintiles when access to well water increases, and this result is statistically significant at the 1 percent level in all but the 80th quintile. This indirectly captures water quality, even though I use protected well water in these regressions. Nevertheless, a common feature in India is that even protected water may seldom meet the quality standards. However, there is no evidence that access to piped water or a public tap improves the weight-for-height z-scores both in the OLS and QR results.

---

[18] The estimation results for the entire set of controls are available upon request. In each table, the controls are categorized as follows: Household characteristics, Child characteristics, Mother's characteristics, Religion/ Caste variables, and Community characteristics. Specifically, social disparities can influence child health (see Braveman and Tarimo, 2002). I do find evidence that social disparities influence weight-for-height, height-for-age, and weight-for-age z-scores (results available upon request).

[19] I cluster the standard errors by household identification number and report the heteroskedastic-consistent t-statistics in all OLS regressions.

[20] The DHS survey includes protected well or spring as an improved source of drinking water.

[21] A lower weight-for-height z-scores means the child is getting closer to the threshold (as defined by WHO, see section 1.3.3) of being malnourished.

Access to pit latrines and flush toilets, on the other hand, improves an average child's weight-for-height z-scores. The QR shows a differential impact of access to sanitation on the weight-for-height z-scores. Access to pit latrines improves the weight-for-height z-scores especially in the lower end of the weight-for-height distribution.

Access to flush toilets improves the weight-for-height z-scores but is statistically significant only in the 20th, 40th and 50th quintiles. A 1 percent increase in access to flush toilets is associated with an increase of .126 standard deviations in the weight-for-height z-scores at the 20th quintile, .086 standard deviations in the 40th, and .096 standard deviations in the 50th quintiles. Children in the top quintiles are healthy, and access to sanitation may contribute minimally to maintaining their health status. On the other hand, children in the other quintiles, especially in the 20th quintile, are malnourished for a wide variety of reasons, and therefore access to flush toilets in addition to a broader set of conditions is required to improve nutritional status first, and then maintain adequate nutrition. Under these circumstances, access to flush toilets has negligible impacts on weight-for-height z-scores in higher quintiles. Finally, the wealth index shows a strong positive association with the weight-for-height z-scores in both the OLS and QR estimations and is statistically significant at the 1 percent level.

Under child characteristics, the OLS estimates provide evidence that an average child is wasted during a recent incidence of diarrhea and the coefficient estimate on diarrhea is negative. Since wasting (low weight-for height z-score) reflects short-term growth change, morbidity will have an immediate impact on wasting. A 1 percent increase in an incidence of diarrhea decreases the weight-for-height z-scores by 0.08 percent. The QR results show that incidence of diarrhea reduces weight-for-height z-scores in children, especially those in the top 60th and 80th quintiles and not in the lowest quintile. For example, the results in the 80th quintile imply that an incidence of diarrhea reduces the weight-for-height z-scores by .144 standard deviations. Children in the top quintiles are healthy, and morbidity due to diarrhea contributes to the short-term change in their health status. On the other hand, children in the other quintiles, especially in the 20th quintile, are malnourished for a wide variety of reasons, and therefore it mutes the effect of morbidity, because a broader set of conditions such as living conditions may contribute to their weight-for-height z-scores.

The mother's literacy level is positively correlated to the child's weight-for-height z-scores. The QR estimates show that having a primary education does not improve the weight-for-height z-scores in almost all quintiles. Mothers age, which proxies knowledge about health care (Borooah, 2005), does not affect the weight-for-height z-scores perhaps because weight-for-height z-scores are a short-run change in growth.

The OLS coefficient estimate on cultivators is -.079 (t-stat: -2.00) and is negatively associated with weight-for-height z-scores. On the other hand, the QR result shows that mothers who are cultivators reduce the weight-for-height z-scores among children in the 40th, median, and 60th quintiles. This implies that working mothers have a higher time trade-off that leaves them with less time to provide health inputs to their children, and this is significant only for children in the center of the distribution, because for children in the extreme ends of the distribution (top and bottom quintiles), mothers time may alone not influence their current health status. Overall, the results for the weight-for-height z-score outcomes indicate that it is useful to investigate the effects based on QR before we make inferences about the OLS estimations.

Table 2.4 presents the regression results where the height-for-age z-scores are the dependent variable. Even though the OLS coefficient estimates of public tap and well water are positive, the result is statistically significant at the 1 percent level for well water only. However, access to well water improves height-for-age z-scores in all but the 20th quintile. Children in the lowest quintile are malnourished, and therefore, access to water alone may not solve their height-for-age z-scores. Access to pit latrines and flush toilets improves the height-for-age z-scores. In addition, access to pit latrines and flush toilets improves the height-for-age z-scores for children in all quintiles. This implies that access to sanitation is a major factor that influences health outcomes among all children irrespective of their position in conditional height-for-age distribution. This also indicates that policies that target better hygiene can improve child health outcomes.

The OLS coefficient estimate of diarrhea reduces the height-for-age z-scores. In addition, the QR results show a differing impact of incidence of diarrhea, which reduces the height-for-age z-scores by .135 standard deviations in the 80th quintile. The wealth index shows a strong positive association with the height-for-age z-scores in both the OLS and QR results and is statistically significant at the 1 percent level.

The OLS coefficient estimates on mother's age and higher education are positive (.026 and .363, respectively). Being a laborer reduces the child's height-for-age z-scores. In addition, the QR estimates indicate that a 1 percent increase in the probability of being a laborer is associated with a decrease of .274 standard deviations in the height-for-age z-scores at the 20th quintile, and .365 standard deviations at the 80th quintile. For children in the lowest end of the distribution, mothers care is critical in improving their height-for-age z-scores. On the other hand, children in the top quintile may also seek mothers care to maintain their health status.

Table 2.5 shows the regression results when weight-for-age z-scores are the dependent variable. Access to flush toilets and pit latrines improves the weight for age z-scores both in the OLS estimation and all quintiles. Access to sanitation is essential for better health outcomes for all children irrespective of their position on the weight-for-age distribution. There is no evidence that access to water improves the weight-for-age z-scores. The wealth index is positive and is statistically significant at the 1 percent level in the OLS and QR estimations.

Under child characteristics, the OLS estimate of diarrhea reduces the weight-for-age z-scores. The QR estimates show that an incidence of diarrhea reduces a child's weight-for-age z-scores, which is statistically significant at the 1 percent level in all quintiles. This is plausible, because an incidence of diarrhea in the recent past indicates morbidity and may influence the weight of children. Mother's age and literacy level improves the child's weight-for-age z-scores. Even in the QR estimations, her age and educational attainment raises weight-for-age z-scores in all quintiles.

In sum, the results for each of the health outcomes indicate that it is useful to investigate the effects based on QR before we make inferences about the OLS estimations. In addition, we cannot assume that access to water and sanitation automatically would improve child health outcomes. The results suggest that access to well water alone (among the different sources of water) significantly influences the weight-for height, height-for-age and weight-for age z-scores. Next, I sub-divide the sample into urban and rural areas. As explained earlier, it is worthwhile to investigate if the influence of the variables of interest varies across the nutritional distribution of the health outcomes based on location. For this reason I compare the OLS and QR estimation results for urban and rural samples in section 2.5.2

2.5.2 Urban Vs. Rural Areas

According to the 2001 Census of India about 94 percent of the rural population and about 91 percent of the urban population have access to safe drinking water (Kharuna and Sen, 2008). Overall, the results in this paper underscores this point that access to water and sanitation has a stronger influence on the weight-for-height, height-for-age, and weight-for age z-scores in rural than urban areas. Mother's education influences all measures of nutrition across rural and urban areas. However, this effect is lower on the weight-for-height z-scores of rural children, which suggests that quality of education may differ across geographical location and in turn, affect weight-for-height z-scores differently. More importantly, policies that target higher education may assist urban mothers to improve weight-for-height z-scores but if the same policies are implemented in rural areas, they may not achieve their objectives. This is a typical example of why "one-size-fits-all" policies may not be effective under these circumstances. I also find that an incidence of diarrhea has varying influence not only based on the position in the weight-for-height, height-for age or weight-for-age distributions but also across geographical location.

Table 2.6 shows the regression results for the urban and rural sub-samples when weight-for-height z-scores are the dependent variable. In rural areas, access to well water significantly reduces the weight-for-height z-scores. Both in the OLS and QR estimations access to well water significantly reduces weight-for-height z-scores. This suggests that the quality of well water may be poor and therefore, leads to lower height-for-weight z-scores. On the other hand, for the urban sub-sample there is no evidence that access to water significantly improves the weight for height z-scores both in the OLS and QR estimations.

In rural areas a 1 percent increase in access to flush toilets is associated with an increase of .203 standard deviations in the weight-for-height z-scores at the 20^{th} quintile, .144 standard deviations at the 40^{th}, and .147 standard deviations at the 50^{th} quintiles. Children in the upper quintiles are healthy and for them access to sanitation has negligible impact on their weight-for-height z-scores. For urban children, access to flush toilets has negligible influence on their weight-for-height z-scores irrespective of their position in the conditional weight-for-height distribution. This strongly suggests that geographical location can influence health outcomes in addition to other factors.

For a rural child in the 20th and 80th quintiles of the weight-for-height distribution, an incidence of diarrhea reduces his or her weight-for-height z-scores. A 1 percent increase in incidence of diarrhea is associated with a decrease of .084 standard deviations in the weight-for-height z-scores at the 20th quintile, .158 standard deviations at the 80th quintile. In rural areas, morbidity has a strong association with wasting outcomes for both healthy and malnourished children. On the other hand, the coefficient estimates of diarrhea do not show evidence that urban children are wasted due to an incidence of diarrhea. This is counter-intuitive, because morbidity is associated with weight loss.

The QR results show that urban mothers who have higher education improve their children's weight-for-height z-scores in all quintiles. On the other hand, rural mothers who have higher educational attainment significantly influence the weight-for-height z-scores in the 40th quintile and median only. A one percent increase in mothers having higher educational attainment is associated with an increase of .182 standard deviations in the weight-for-age z-scores at the 40th quintile, and .172 standard deviations at the median.

Table 2.7 reports the regression results when height-for-age z-scores are the dependent variable in urban and rural areas, respectively. Access to well water improves the height-for-age z-scores for urban children in the median and 60th quintiles only, while it improves the height-for-age z-scores in the 40th, 60th and 80th quintiles for rural children in the OLS and QR estimations. A 1 percent increase in access to well water is associated with an increase of .113 standard deviations in the weight-for-height z-scores at the 40th quintile, .139 standard deviations at the 60th, and .193 standard deviations at the 80th quintiles. Mothers may take precautionary measure to treat well water, and as a result children may have access to purified water. Children in the 20th quintiles are malnourished for a variety of reasons, and therefore, purified water may have a negligible influence both in rural and urban areas. In urban areas, access to flush toilets alone improves the height-for-age z-scores among urban children in the OLS, and 20th, median, 60th and 80th quintiles. In rural areas, both the sanitation variables improve the height-for-age z-scores both in the OLS and QR estimations.

In the QR estimation, diarrhea reduces the height-for-age z-scores among urban children in the 60thand 80th quintiles are stunted due to diarrhea. In contrast, in rural areas a 1 percent increase in incidence of diarrhea is associated with an increase of .121 standard deviations in the height-for-age z-scores at the 80th quintile. This shows that for healthy children (in upper quintiles) both in the rural and urban areas, morbidity affect their height-for-age z-scores (see Assis et al. (2005)).

While urban mothers who have higher education improve the height-for-age z-scores, it is interesting to note that this effect is statistically significant only for children in the upper quintiles. This implies that children in the lower quintiles are malnourished for various reasons and therefore, mother's education alone may not solve the height-for-age z-scores for children in the 20th and 40th quintiles. In rural areas, literate mothers have a positive influence on the height-for-age z-scores among rural children in all the quintiles.

Finally, for rural children who fall in all but the 20th quintile access to a health facility improves their height-for-age z-scores and is statistically significant. For rural children who are malnourished (20th quintile), interventions to prevent diseases may assist better health outcomes rather than curative measures such as access to a health facility. In contrast, access to a health facility does not significantly improve the urban children's height-for-age z-scores.

Table 2.8 presents the results for weight-for-age z-scores (dependent variable) in urban and rural areas, respectively. In rural areas access to well water decreases the weight-for-age z-scores in the OLS estimation and in the median quintile. On the other hand, in urban areas, there is no evidence that access to any type of water affects the weight-for-age z-scores in both the OLS and QR estimations. Access to flush toilets improves the weight-for-age z-scores. A 1 percent increase in access to flush toilets is associated with an increase of .110 standard deviations in the weight-for-age z-scores at the 20^{th} quintile, and .125 standard deviations at the 40^{th} quintile. For children in the lower quintiles, safe sanitation in addition to other factors is crucial for better weight-for-age z-scores. In rural areas, however, access to flush toilets improves weight-for-age z-scores in all quintiles significantly.

Even though in urban areas, the average and the differential effect of morbidity on the weight-for-age z-scores are not statistically significant at the 10 percent level, in rural areas incidence of diarrhea reduces the weight-for-age z-scores in the OLS and QR estimations. Rural children are more vulnerable to diseases (such as diarrhea) and therefore, this may contribute to higher mortality rates in these areas compared to urban ones (see Bhalhotra, 2008). A mother's level of education is an important determinant of weight-for age z-scores in almost all quintiles in both rural and urban areas.

The results indicate that geographical location in addition to other factors significantly influence child health outcomes. It also provides evidence that rural children who experience ill-health (such as diarrhea) face greater risk of wasting and underweight than urban children.

2.6 CONCLUSIONS

I examine the distributional impacts of access to water and sanitation on children's stunting, wasting, and underweight outcomes in India. In contrast to most studies, which analyze only some of these measures of malnutrition, I analyze all of these measures individually. Consequently, this approach better explains the varying distributional impacts of access to water and sanitation on each measure of malnutrition. The traditional ordinary least squares estimation may undermine important information regarding the partial effect of the interest variable on the health outcome variables. Compared to the OLS results, the QR estimations may better inform policymakers to make inferences based on the results from the quintile distribution rather than the average effect.

The empirical analysis suggests that the direct and indirect effects (such as diarrhea) of water and sanitation have significant but varying impacts on children's nutritional status based on their position in the conditional nutrition distribution. I confirm that incidence of diarrhea has an inverse and differential effect on children's nutrition distribution. I also find that between the water and sanitation variables, sanitation is a key factor to improving the nutritional status of children. Access to water especially, well water significantly influences the weight-for-height and height-for-age z-scores. In addition, the sub-analysis of rural and urban samples provides evidence that the living conditions differ across geographical locations, and as a result, the health outcomes vary significantly. For example, even though the incidence of diarrhea reduces the weight-for-height z-scores in the OLS estimation, the QR results for rural and urban areas show a varying yet interesting pattern. In urban areas, only for children in the top quintiles an incidence of diarrhea reduces the weight-for-height z-scores, while in rural areas children in almost all quintiles are wasted. This provides evidence that rural children irrespective of being healthy are more vulnerable to diseases and therefore, have relatively higher morbidity rates. Other factors like mother's education, age and occupation have significant influence on child health.

This analysis also suggests that the distributional impact of various health interventions may better inform policymakers to target policies based on the conditional distribution of the outcome variable rather than a one-size-fits-all approach. For example, even though "educating all mothers" is a good policy, it may not be the most effective, because even though the results indicate that educated urban mothers improve weight for height z-scores, higher education is not a significant factor to influence the weight-for-height z-scores in rural areas. Therefore, if policymakers take into account the geographical location when setting policies, the policies may be more cost-effective, and in the process improve overall health outcomes.

2.7 REFERENCES

Assis, Ana Marlucia. O, Mauricio. L Barreto, Leonar Maria Santos, Rosemeire Fiaccone, and Gecynalda Soares da Silva Gomes. 2005. "Growth Faltering In Childhood Related To Diarrhea: A Longitudinal Community Based Study," European Journal of Clinical Nutrition, 59(11), 1317-1323.

Aturupane, Harsha, Anil B. Deolalikar, and Dileni Gunewardena. 2008. "The Determinants of Child Weight and Height in Sri Lanka: A Quantile Regression Approach," UNU-WIDER working paper no. 2008/53.

Bassole, Leandre. 2007. "Child Nutrition in Senegal: Does Access to Public infrastructure really Matter? A Quantile Regression Analysis," working paper.

www.uneca.org/aec/documents/Leandre%20Bassole.pdf

Becker, Gary S. 1981. "A Treatise on the Family." Harvard University Press, Cambridge, MA, USA.

Behrman, Jere R, and Anil B. Deolalikar. 1988. "Health and Nutrition." In Hollis Chenery and T.N. Srinivasan (ed.), Handbook of Development Economics, volume 1, chapter 14, 631-711 Elsevier Science.

Bhalhotra, Sonia. 2008. "Fatal Fluctuations? Irreversible Health Consequence of recession in India," Research in Public Policy, Bulletin of the Centre for Market and Public Organisation, University of Bristol.

Borooah, Vani K. 2005. "The Height-for-Age of Indian Children," Economics and Human Biology, 3(1), 45–65.

Braveman, Paula and Eleuther Tarimo. 2002. "Social Inequalities in Health within Countries: Not Only an Issue for Affluent Nations," Social Science and Medicine, 54(11), 1621-1635.

David, Vincent, Marco Moncada, and Fidel Ordonez. 2004. "Private and Public Determinants of Child Nutrition in Nicaragua and Western Honduras," Economics & Human Biology, 2(3), 457–488.

Doherty, Hannah. 2008. "Child Mortality," World Watch Institute. http://www.worldwatch.org/node/5875#notes

Esrey, Steven. 1996. "Water, Waste, and Well-being: A Multicountry Study," American Journal of Epidemiology, 143(6), 608-623.

Esrey, Steven, Jean-Pierre Habicht, and George Casella. 1992. "The Complementary Effect of Latrines and Increased Water Usage on the Growth of Infants in Rural Lesotho," American Journal of Epidemiology, 135(6), 659-666.

Khurana, Indira and Romit Sen. 2008. "Drinking Water Quality in Rural India: Issues and Approaches,"

http://www.indiawaterportal.org/sites/indiawaterportal.org/files/Drinking_Water_Quality_in_Rural_India_Issues_and_Approaches_Jan_2008.pdf

Koenker, Roger, and Gilbert Bassett. 1978. "Regression Quantiles," Econometrica, 46(1), 33–50.

Nandy, Shailen, Michelle Irving, David Gordan, S.V. Subramaniam, and George Smith S. 2005. "Poverty, Child Undernutrition and Morbidity: New Evidence from India," WHO Bulletin, 83, 210-216.

Sen, Amartya. 2001. "Many Faces of Gender Inequality," Frontline (18), 4-14.

Shepherd, Susan. 2007. "MSF Programs in Maradi, Niger: We Want to See Fewer Children Dying of Malnutrition,"

http://doctorswithoutborders.org/news/article.cfm?id=2082&cat=voice-from-the-field&ref=tag-index

Strauss, John and Duncan Thomas. 1998. "Health, Nutrition and Economic Development," Journal of Economic Literature, 36(2), 766–817.

Svedberg, Peter. 2000. "Poverty and Undernutrition: Theory, Measurement, and Policy." Oxford: Oxford University Press, New York, USA.

Thomas, Duncan and John Strauss. 1992. "Prices, Infrastructure, Household Characteristics and Child Height," Journal of Development Economics, 39(2), 301–331.

World Health Organization (WHO). 2008. "Nutrition for Health and Development," http://www.who.int/nutrition/nhd/en/index.html

World Hunger Facts. 2008. "World Hunger Education Service," http://www.worldhunger.org/articles/Learn/world%20hunger%20facts%202002.htm

Table 2.1: Summary Statistics

Variables	Total Sample N=11,466 Mean (Std. Dev)	Urban Sample N=4,267 Mean (Std. Dev)	Rural Sample N=7,199 Mean (Std. Dev)
Dependent Variables			
Height-for-age z-scores	-1.57	-1.38	-1.69
	(1.55)	(1.19)	(1.57)
Weight-for-age z-scores	-1.76	-1.53	-1.90
	(1.21)	(1.19)	(1.20)
Weight-for-height z-scores	-1.00	-.86	-1.08
	(1.23)	(1.25)	(1.21)
Independent Variables			
Household Characteristics			
Water/Sanitation amenities			
Piped water into dwelling/yard or plot	.26	.50	.13
	(.44)	(.50)	(.33)
Public Tap	.15	.18	.13
	(.36)	(.38)	(.34)
Well	.13	.05	.17
	(.33)	(.28)	(.37)
Flush Toilet	.43	.78	.23
	(.49)	(.41)	(.42)
Pit Latrine	.08	.07	.10
	(.29)	(.25)	(.29)
Wealth index	.02	.64	-.34
	(1.01)	(.92)	(.87)
Child Characteristics			
Girl	.48	.48	.48
	(.50)	(.50)	(.50)
Birth order	2.60	2.23	2.81
	(1.79)	(1.51)	(1.91)
Age (in months)	14.83	14.95	14.76
	(5.42)	(5.45)	(5.40)
Age Squared (in months)	249.29	253.13	247.02
	(162.96)	(164.43)	(164.05)
Child had Diarrhea	.15	.15	.15
	(.36)	(.36)	(.36)
Mother's Characteristics			
Mother's age (in years)	25.73	25.82	25.67
	(5.22)	(4.75)	(5.48)
Mother is literate (Primary Education)	.14	.11	.15
	(.34)	(.31)	(.36)
Mother is literate (Secondary Education)	.40	.49	.33
	(.49)	(.50)	(.47)
Mother is literate (Higher Education)	.08	.17	.03
	(.28)	(.38)	(.18)
Mother is a laborer	.02	.01	.03
	(.16)	(.11)	(.18)

Table 2.1 Continued

Variables	Total Sample N=11,466 Mean (Std. Dev)	Urban Sample N=4,267 Mean (Std. Dev)	Rural Sample N=7,199 Mean (Std. Dev)
Mother is a cultivator	.05 (.21)	.00 (.06)	.07 (.26)
Religion/Caste Variables			
Hindus	.71 (.45)	.67 (.47)	.73 (.44)
Muslims	.14 (.36)	.19 (.40)	.11 (.31)
Christians	.10 (.31)	.10 (.29)	.11 (.31)
Sikhs	.02 (.14)	.01 (.11)	.02 (.15)
Buddhists	.01 (.11)	.01 (.11)	.01 (.10)
Others	.02 (.12)	.01 (.11)	.02 (.13)
Scheduled Caste	.19 (.39)	.17 (.38)	.19 (.40)
Scheduled Tribe	.17 (.38)	.10 (.31)	.21 (.41)
Other Backward Castes	.33 (.47)	.33 (.47)	.34 (.47)
Community Characteristics			
Child lives in a city	.23 (.42)	.61 (.49)	-
Child lives in a slum	.05 (.22)	.14 (.37)	-
Access to a health facility	.71 (.45)	.69 (.46)	.72 (.45)

Table 2.2 Classification of Malnutrition (N=11,466)

Multiple Outcomes	Percentage of Children in the Sample
Wasted and Underweight	16.27%
Wasted and Stunted	0%
Stunted and Underweight	29.08%
Stunted, Underweight, and Wasted	7.32%

Table 2.3: OLS and Quantile Regressions (Total Sample=11,466)

Dependent Variable: **Weight-for-height z-scores**	OLS (1)	Q20 (2)	Q40 (3)	Q50 (4)	Q60 (5)	Q80 (6)
Household Characteristics						
Water/Sanitation amenities						
Has access to piped water into dwelling & yard/plot	.015	.011	-.012	.011	.015	.037
	(0.52)	(0.27)	(-0.35)	(0.33)	(0.38)	(0.79)
Public Tap	-.015	-.057	-.037	-.016	-.014	.006
	(-0.43)	(-1.28)	(-1.01)	(-0.51)	(-0.35)	(0.14)
Well	-.132***	-.156***	-.142***	-.098***	-.136***	-.130**
	(-4.01)	(-3.28)	(-3.39)	(-2.67)	(-3.48)	(-2.28)
Has Pit Latrine	.121***	.170***	.130***	.149***	.089*	.146**
	(2.88)	(2.91)	(2.60)	(3.24)	(1.70)	(2.15)
Flush Toilet	.075**	.126***	.086**	.096***	.051	.014
	(2.28)	(2.84)	(2.33)	(2.70)	(1.44)	(0.33)
Wealth index	.113***	.094***	.123***	.125***	.138***	.142***
	(5.23)	(4.16)	(5.87)	(6.90)	(6.71)	(5.73)
Child Characteristics						
Child had Diarrhea	-.087***	-.050	-.069*	-.049	-.072*	-.144***
	(-3.15)	(-1.19)	(-1.79)	(-1.33)	(-1.73)	(-3.01)
Mother's Characteristics						
Mother's age (in years)	.001	-.000	.002	-.000	.003	.001
	(0.56)	(-0.07)	(0.79)	(-0.20)	(0.98)	(0.34)
Mother is literate (Primary)	.061*	.000	.022	.044	.070*	.070
	(1.72)	(0.00)	(0.55)	(1.04)	(1.71)	(1.14)
Mother is a cultivator	-.079**	-.074	-.134**	-.126**	-.139**	-.049
	(-2.00)	(-1.22)	(-2.20)	(-2.29)	(-2.16)	(-0.56)
Constant	-.363	-1.33***	-.697***	-.408***	-.139	.580***
	(-3.77)	(-8.27)	(-6.03)	(-3.77)	(-1.07)	(2.86)

Heteroskedasticity- consistent t- statistics (clustered by household identification number) included in parentheses for the OLS and the standard errors are bootstrapped with 100 replications for QR; *, ** and *** represent 10, 5 and 1% significance levels respectively. All columns include other child variables such as age, age squared, birth order and gender; Mother characteristics such as secondary and higher educational attainment, laborer; Religion variables such as Christians, Muslims, Sikhs, Buddhists, others; Caste variables include scheduled castes, scheduled tribes, and other backward castes; Community variables include access to a health facility and whether child lives in a slum or a city.

Table 2.4: OLS and Quantile Regressions (Total Sample=11,466)

Dependent Variable: **Height-for -age z-scores**	OLS (1)	Q20 (2)	Q40 (3)	Q50 (4)	Q60 (5)	Q80 (6)
Household Characteristics						
Water/Sanitation amenities						
Has access to piped water into dwelling & yard/plot	-.022	-.013	.012	.007	-.020	-.048
	(-0.60)	(-0.31)	(0.29)	(0.21)	(-0.59)	(-0.93)
Public Tap	.038	.007	.019	.046	.058	-.005
	(0.96)	(0.16)	(0.42)	(1.05)	(1.22)	(-0.09)
Well	.107***	.067	.128***	.142***	.154***	.209***
	(3.37)	(1.15)	(2.80)	(2.91)	(3.00)	(3.44)
Has Pit Latrine	.244***	.211***	.232***	.226***	.210***	.171**
	(4.36)	(3.31)	(3.81)	(3.84)	(3.66)	(2.31)
Flush Toilet	.162***	.168***	.179***	.169***	.187***	.122*
	(3.55)	(3.85)	(4.00)	(3.66)	(3.99)	(1.91)
Wealth index	.208***	.260***	.224***	.215***	.194***	.170***
	(10.33)	(11.80)	(9.39)	(8.84)	(8.35)	(5.51)
Child Characteristics						
Child had Diarrhea	-.067*	.017	-.009	-.050	-.058	-.135**
	(-1.85)	(0.42)	(-0.22)	(-1.17)	(-1.40)	(-2.50)
Mother's Characteristics						
Mother's age (in years)	.026***	.030***	.025***	.020***	.020***	.026***
	(6.40)	(6.49)	(6.05)	(5.41)	(4.95)	(4.56)
Mother is literate (Higher)	.363***	.304***	.357***	.385***	.431***	.378***
	(5.55)	(4.08)	(5.67)	(5.89)	(7.31)	(4.58)
Mother is a laborer	-.258***	-.274***	-.102	-.089	-.164*	-.365**
	(-3.09)	(-2.76)	(-0.77)	(-1.04)	(-1.89)	(-2.54)
Constant	.103	-.945***	.024	.326**	.631***	1.25***
	(0.77)	(-5.85)	(0.16)	(2.43)	(4.34)	(6.28)

Heteroskedasticity- consistent t- statistics (clustered by household identification number) included in parentheses for the OLS and the standard errors are bootstrapped with 100 replications for QR; *, ** and *** represent 10, 5 and 1% significance levels respectively. All columns include other child variables such as age, age squared, birth order and gender; Mother characteristics such as primary and secondary educational attainment, cultivator; Religion variables such as Christians, Muslims, Sikhs, Buddhists, others; Caste variables include scheduled castes, scheduled tribes and other backward castes; Community characteristics include whether child lives in a city or a slum, access to a healthcare facility.

Table 2.5: OLS and Quantile Regressions (Total Sample=11,466)

Dependent Variable: **Weight-for-age z-scores**	OLS (1)	Q20 (2)	Q40 (3)	Q50 (4)	Q60 (5)	Q80 (6)
Household Characteristics						
Water/Sanitation amenities						
Has access to piped water into dwelling & yard/plot	-.005	-.020	.004	-.003	-.014	-.029
	(-0.16)	(-0.54)	(0.11)	(-0.11)	(-0.48)	(-0.61)
Public Tap	.007	.013	.036	.044	-.018	-.028
	(0.20)	(0.37)	(1.14)	(1.19)	(-0.64)	(-0.62)
Well	-.053*	-.025	-.048	-.047	-.061	-.073
	(-1.84)	(-0.61)	(-1.34)	(-1.22)	(-1.48)	(-1.58)
Has Pit Latrine	.242***	.258***	.179***	.214***	.232***	.194***
	(5.24)	(5.02)	(3.50)	(4.12)	(5.08)	(2.92)
Flush Toilet	.152***	.191***	.146***	.155***	.119***	.103**
	(5.33)	(5.66)	(4.19)	(4.59)	(3.49)	(1.96)
Wealth index	.221***	.206***	.241***	.235***	.243***	.237***
	(11.14)	(12.26)	(13.53)	(12.19)	(13.58)	(9.43)
Child Characteristics						
Child had Diarrhea	-.099***	-.102***	-.088***	-.092***	-.083***	-.123***
	(-3.88)	(-3.20)	(-2.97)	(-2.93)	(-2.62)	(-2.88)
Mother's Characteristics						
Mother's age (in years)	.017***	.013***	.012***	.013***	.013***	.022***
	(6.14)	(4.23)	(3.87)	(3.75)	(4.00)	(6.31)
Mother is literate (Higher)	.372***	.288***	.344***	.365***	.419***	.412***
	(7.94)	(5.25)	(6.22)	(6.05)	(7.51)	(5.60)
Constant	-.576***	-1.36	-.531***	-.302**	-.215*	.248
	(-5.14)	(-10.03)	(-4.38)	(-2.51)	(-1.83)	(1.63)

Heteroskedasticity- consistent t- statistics (clustered by household identification number) included in parentheses for the OLS and the standard errors are bootstrapped with 100 replications for QR; *, ** and *** represent 10, 5 and 1% significance levels respectively. All columns include other child variables such as age, age squared, birth order and gender; Mother characteristics such as primary and secondary educational attainment, laborer; Religion and caste variables such as Sikhs, Buddhists, others and other backward castes, respectively; Caste variables include scheduled castes, scheduled tribes and other backward castes; Community variables include access to a health facility and whether child lives in a slum or a city.

Table 2.6: OLS and Quantile Regressions; Dependent Variable: Weight-for-height z-scores

	Urban (N=4,267)					Rural (N=7,199)				
Variable	OLS	Q20	Q40	Q60	Q80	OLS	Q20	Q40	Q60	Q80
Piped Water	.03 (0.60)	.00 (0.05)	.03 (0.67)	.01 (0.19)	-.00 (-0.03)	-.03 (-.09)	.00 (0.02)	-.08 (-1.34)	-.03 (-0.45)	.00 (0.03)
Public Tap	-.07 (-1.18)	-.12 (1.39)	-.05 (-0.71)	-.07 (-1.10)	-.14* (-1.68)	.02 (0.45)	-.023 (-0.43)	-.00 (-0.05)	.04 (0.78)	.08 (1.14)
Well	-.03 (-0.41)	-.15 (-0.92)	-.02 (-0.16)	-.02 (-0.24)	-.02 (-0.15)	-.2*** (-3.97)	-.16*** (-3.23)	-.15*** (-3.44)	-.15*** (-3.49)	-.11* (-1.66)
Flush Toilet	-.01 (-0.20)	.03 (0.43)	-.02 (-0.30)	.03 (0.46)	-.05 (-0.56)	.13*** (3.47)	.20*** (4.03)	.14*** (3.06)	.10** (2.18)	.10 (1.31)
Diarrhea	-.04 (-0.67)	.05 (0.68)	-.05 (-0.94)	-.05 (-0.86)	-.10 (-1.25)	-.1*** (-3.56)	-.08* (-1.71)	-.06 (-1.49)	-.07 (-1.45)	-.2*** (-2.77)
Mother is Literate (Higher)	.23** (2.18)	.22* (1.94)	.26*** (2.77)	.30*** (3.14)	.39*** (3.30)	.06 (0.73)	.04 (0.34)	.18* (1.64)	.09 (1.09)	.04 (0.34)

Heteroskedasticity- consistent t- statistics (clustered by household identification number) included in parentheses for the OLS and the standard errors are bootstrapped with 100 replications for QR; *, ** and *** represent 10, 5 and 1% significance levels respectively. All columns include other child variables such as age, age squared, birth order and gender; access to pit latrines; Mother's characteristics such as her age, primary and secondary educational attainment, cultivator, and laborer; wealth index; all the religion and caste variables; all the community characteristics.

Table 2.7: OLS and Quantile Regressions; Dependent Variable: Height-for-age z-scores

Variable	Urban (N=4,267) OLS	Q20	Q40	Q60	Q80	Rural (N=7,199) OLS	Q20	Q40	Q60	Q80
Piped Water	.00 (0.01)	-.07 (-1.04)	-.03 (-0.52)	-.02 (-0.43)	.06 (0.64)	.02 (0.05)	.09 (1.34)	.09* (1.73)	.00 (0.07)	-.12 (-1.44)
Public Tap	.09 (1.28)	.07 (0.64)	.11 (1.26)	.08 (1.05)	.08 (0.87)	.02 (0.48)	-.04 (-0.72)	-.00 (-0.02)	.07 (1.04)	-.00 (-0.01)
Well	.24*** (3.05)	.03 (0.27)	.18 (1.34)	.26*** (2.53)	.16 (1.00)	.08** (2.33)	.06 (0.90)	.11** (2.06)	.14** (2.21)	.19** (2.52)
Flush Toilet	.15** (2.09)	.15** (2.04)	.14 (1.50)	.20** (2.54)	.26*** (2.71)	.16*** (3.17)	.15** (2.47)	.15*** (2.96)	.18*** (3.12)	.11 (1.34)
Diarrhea	-.07 (-1.21)	.09 (1.22)	-.04 (-0.69)	-.16*** (-2.98)	-.18* (-1.88)	-.06* (-1.76)	-.03 (-0.52)	-.03 (-0.60)	-.02 (-0.38)	-.12* (-1.79)
Mother is literate (higher)	.23** (2.24)	.15 (1.43)	.15 (1.31)	.27*** (2.80)	.35** (2.41)	.52*** (5.87)	.57*** (4.75)	.55*** (4.51)	.51*** (4.25)	.35** (2.08

Heteroskedasticity- consistent t- statistics (clustered by household identification number) included in parentheses for the OLS and the standard errors are bootstrapped with 100 replications for QR; *, ** and *** represent 10, 5 and 1% significance levels respectively. All columns include other child variables such as age, age squared, birth order and gender; access to pit latrines; Mother's characteristics such as her age, primary and secondary educational attainment, cultivator, and laborer; wealth index; all the religion and caste variables; all the community characteristics.

Table 2.8: OLS and Quantile Regressions; Dependent Variable: Weight-for-height z-scores

	Urban (N=4,267)					Rural (N=7,199)				
Variable	OLS	Q20	Q40	Q60	Q80	OLS	Q20	Q40	Q60	Q80
Piped Water	.03 (0.53)	-.02 (-0.36)	-.04 (-0.84)	-.02 (-0.34)	.02 (0.28)	-.02 (-0.54)	.00 (0.07)	.02 (0.34)	.00 (0.00)	-.08 (-1.21)
Public Tap	.02 (0.33)	.05 (0.66)	-.02 (-0.29)	-.01 (-0.15)	-.07 (-0.87)	.01 (0.27)	.01 (0.23)	.06 (0.97)	-.00 (-0.08)	.02 (0.28)
Well	.10 (1.44)	.13 (1.23)	.02 (0.26)	-.00 (-0.04)	.04 (0.42)	-.09*** (-2.63)	-.07 (-1.58)	-.07* (-1.64)	-.07 (-1.56)	-.08 (-1.53)
Flush Toilet	.07 (1.32)	.11* (1.66)	.13** (2.05)	.06 (1.01)	-.04 (-0.53)	.20*** (5.71)	.25*** (5.13)	.18*** (3.84)	.15*** (3.04)	.21*** (3.16)
Diarrhea	-.07 (-1.56)	-.08 (-1.39)	-.06 (-1.02)	-.06 (-1.00)	.02 (0.28)	-.11*** (-3.92)	-.11** (-2.56)	-.12*** (-2.78)	-.13*** (-2.95)	-.16*** (-3.24)
Mother is literate (Higher)	.32*** (4.15)	.25** (2.56)	.24*** (2.89)	.29** (3.87)	.43*** (3.68)	.38*** (4.85)	.32*** (3.00)	.44*** (3.62)	.48*** (5.99)	.30** (2.14)

Heteroskedasticity- consistent t- statistics (clustered by household identification number) included in parentheses for the OLS and the standard errors are bootstrapped with 100 replications for QR; *, ** and *** represent 10, 5 and 1% significance levels respectively. All columns include other child variables such as age, age squared, birth order and gender; access to pit latrines; Mother's characteristics such as her age, primary and secondary educational attainment, cultivator, and laborer; wealth index; all the religion and caste variables; all the community characteristics.

CHAPTER 3

GOT WATER? SOCIAL DIVISIONS AND ACCESS TO PUBLIC GOODS IN RURAL INDIA[22]

3.1 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 many developing countries, the government acts as the sole provider for water services. India represents a prime example of a fast-growing developing country where water is a “public” good, i.e., whose provision by the government is designed to provide universal and free access to its pre-dominantly poor and rural population. The data, however, reveal wide variation in access to drinking (tap) water across rural India, with some districts having no access at all. The central objective of this paper therefore is to understand the extent to which social and economic factors in rural India determine access to water services.

Our study focuses on the roles played by caste and religion in determining access to publicly provided water, which are important in India’s social setting. To this extent, our paper is related to recent work by Banerjee and Somanathan (2007), who examine how access to public goods across rural India changed between 1971 and 1991. However, our study differs from existing work in several very important dimensions.

Most studies use an aggregate index of social fragmentation, which is made up of several socio-economic characteristics such as ethnicity, race, language, religion, caste, etc. We argue

---

[22] Balasubramaniam, Divya, Santanu Chatterjee and David B. Mustard. Submitted to Economica, 08/28/2009.

that the aggregate fragmentation index is not useful for designing public policy, as it does not convey 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 that sense, our study represents a new approach whereby we use disaggregated measures of social fragmentation: do measures of diversity based on caste, religion, and political preferences move in the same or different directions when determining access to water services in rural India? What are their individual effects on access to drinking water? In essence, we examine how fragmentation within a religion (e.g. caste system among Hindus) compares with fragmentation across religions (Hindus, Muslims, Christians, etc) in determining access to public goods.

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.[23] We use three measures of tap 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 either aggregated measures of fragmentation or 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 of the issue at hand. The main result of this paper is that caste heterogeneity reduces and religious heterogeneity increases tap water access in rural India.

Districts that are more heterogeneous in terms of caste have significantly lower access to both total tap water and within-residence tap water.[24] A 10 percentage-point increase in caste-based heterogeneity in a district reduces a household's probability of total tap water access by

[23] 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.
[24] 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.

3.9 percentage points. On the other hand, districts that are more heterogeneous in terms of religion have significantly higher access to total tap water and within-residence tap water: a 10 percentage-point increase in religion-based heterogeneity in a district increases a household's probability of total tap water access by 3.1 percentage points.[25]

Another contribution of this paper is that disaggregating the standard measure of social fragmentation helps resolve a puzzle reported by Banerjee and Somanathan (2007), who find that 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 were able to mobilize themselves politically, which diminished the importance of social fragmentation over time. However, our results provide an alternative interpretation: when employing an aggregate measure of social fragmentation, the effects of the individual components may offset each other. Indeed, even 10 years beyond Banerjee and Somanathan's period of study (1971-91), we find that social fragmentation does matter, with its different 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 indicate that public policy must be designed to target different aspects of social divisions, rather than the one-size-fits-all approach often adopted.

Finally, our results also highlight the extent to which social factors influence the source of drinking water: for access to tap water within the residence, the role of caste heterogeneity is crucial; by contrast, for access to tap water outside the residence, what really matters is the concentration of different caste groups.

---

[25]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.

The rest of the paper is organized as follows. Section 3.2 explains the institutional set up and water crisis in India. Section 3.3 discusses the nature and depth of social divisions in India and the possible mechanisms through which they affect access to drinking water. Section 3.4 describes the empirical specification, while section 3.5 characterizes the data and discusses some econometric issues such as endogeneity and selective migration. Section 3.6 analyzes the results and the underlying intuition, while Section 3.7 concludes with a brief discussion of how private provision of water services might overcome the barriers imposed by social divisions.

3.2 INDIA'S INSTITUTIONAL SETUP AND WATER CRISIS

3.2.1. WATER INSTITUTIONS

Since independence in 1947, the provision of water services in India has predominantly been under the control of 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). Each state is responsible for the delivery of water within its state boundary. The state may, however, vest this responsibility to the Panchayat Raj Institutions (PRI) in the rural areas or to the municipalities (urban local governments) in urban areas.[26] The delivery of water in each state is overseen by the Department of Public Health Engineering, Rural Development Engineering, or a Water Board. Even though the states generally plan, design, and execute water supply schemes, final decision-making and approvals are dictated by the central government, which coordinates investment in the water sector though it's Five Year Plans.[27] Several institutions within the central government, such as the Central Water Commission, closely coordinate and regulate surface water use in various sectors like

[26] As per the Indian Constitution, the PRI have certain powers and authority at the village level to devise plans that ensure economic development and social justice.

[27] The Five Year Plan, designed and implemented by the Planning Commission of India, is a statement of the economic development targets and objectives for five years from the date of implementation.

industry, irrigation, drinking water, etc. They act as mediators in inter-state water disputes as well. The Central Ground Water Board (CGWB), created in the year 1972, is responsible for ground water management.

At the national level, the Ministry of Rural Development delivers water to rural areas through the Department of Drinking Water Supply. The department controls issues related to water policies, provides funds for investment in water projects, and manages and assists states in water supply to rural areas. In the urban sector, the Ministry of Urban Development (MoUD) is responsible for water supply. Responsibilities include planning, setting standards, monitoring and support state programs by providing financial and technical expertise. The Ministry of Housing and Urban Poverty alleviation is responsible for urban water supply in the union territories. However, there is no autonomous body to regulate the water sector in India (Government of India, 2002).

3.2.2 THE WATER CRISIS IN INDIA

Several recent studies indicate a looming water crisis in India, with demand far outstripping supply. India has experienced enormous growth in its agriculture, industry, and service sectors since independence. Particularly, economic liberalization since the early 1990's and the sustained economic growth India has experienced in the last two decades has resulted in a dramatic increase in the demand for water from all sectors of the economy: consumers, industry, agriculture, and services. However, investments in the water supply infrastructure have failed to keep pace with the rapid growth elsewhere in the economy, thereby increasing the scarcity of a critical resource (Bajpai, 2007). Moreover, the lack of private provision in the water sector, especially in the rural areas, has further exacerbated the problem (Swiss Agency for Development and Cooperation, 2007).

Due to rapid urbanization, industrialization, and a lack of enforceable environmental laws, most Indian rivers are polluted and non-potable (Ramachandran, 2006). The World Bank estimates that unsafe water is responsible for 21 percent of communicable diseases in developing countries (World Bank, 1993). Although the Indian government made huge investments in the water infrastructure during the 1960's, since then it has allocated very little for operations and maintenance, which has resulted in an aging infrastructure. Haarmeyer and Mody (1997) review how governments could draw on private capital to address inefficiencies in the water sector. They argue that water sectors controlled by governments in developing countries are financially and operationally weak, because they collect revenues that cover only 35 percent of the total cost of water production. They highlight the need for a more efficient institutional arrangement. Because the Indian government lacks the ability to supply adequate water and no law restricts the amount of ground water extracted by a land owner (Saleth, 2005), many middle-income people in urban areas extract private ground water. Over-extraction of ground water has depleted water tables and magnified the crisis. Inter-state legal disputes about water sharing have further worsened this crisis. Another issue is global warming, which several environmentalists believe causes erratic climatic conditions and aggravates the water crisis (Brooks, 2007). All these trends have spurred significant concerns about the current provision of water services.

3.3 SOCIO-RELIGIOUS STRATIFICATION AND ITS EFFECTS ON ACCESS TO DRINKING WATER

The data reveal a stark story about water access for rural India: according to the 2001 Census of India, the average share of households in a rural district that had access to either inside or outside tap water 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 1.1). Despite its geo-tropical position and being endowed with one of the world's largest river networks, the supply of water in India has been woefully inadequate (Sengupta, 2006). 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. The wide variation in the share of rural households with access to tap water is a strong reflection of this point. Many economists believe that the government's inadequacy in providing an adequate and uniform supply of water to India's rapidly growing economy poses one of the biggest threats to its potential for progress and prosperity.

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). 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 play a pivotal role in both politics and the allocation of scarce public resources more than six decades after independence.

The predominant form of social stratification in India is the caste system, which has deep historic roots in the majority Hindu religion. Historically, the objective to create caste divisions was to facilitate the identification of social groups based on their skill levels. However, over time, certain groups gained power over others, forming the basis for social discrimination. Over time, stratification based on caste became the determinant of an individual's position in the social hierarchy. Even though discrimination based on caste is illegal in India, there is still a sharp "social" difference between "high" and "low" castes. 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.[28]

Another dimension of social division in India is religion. Even though the Indian Constitution provides for a secular state with equal tolerance for all religions, more than 80 percent of the population is Hindu, 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 along with high levels of poverty and inequality, underscore the nature and depth of its social fragmentation.

Why might social (caste) and religious divisions determine access to public goods such as water? 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. In our context, therefore, social divisions based on caste and religion can, in theory, either impede or facilitate this coordination mechanism. On the one hand, social fragmentation may lead to explicit or implicit competition for scarce public goods and, over time, may lead the dominant castes and religious groups to appropriate much of the benefits of public goods. Water, being

[28]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. These are terms the Government of India uses 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*.

essential to economic activity, is therefore at the center of this competition. On the other hand, strong secular and legal institutions could also lead to more tolerance and cooperation amongst social or religious groups over time, so that communities may benefit from this cooperation in the form of higher access to public goods. Therefore, it is not clear how different dimensions of social divisions (within a religion and across religions) affect access to public goods. Our paper sheds new light on this important public policy issue and highlights the underlying mechanisms through which different social divisions affect access to drinking water in rural India.[29]

3.4 EMPIRICAL SPECIFICATION

The central focus of this paper is to identify the determinants of access to tap water in rural India. The specification we test is given by the following cross-section regression:

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

$Y_{i,k}$ 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 Hindu caste and/or religion homogeneity index in a given district of a state. The index measures 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. This is similar to a Herfindahl index and is constructed by calculating

$$HI_j = \sum_j^n s_j^2$$

where s_j is the population share of the j-th caste or religious group. Therefore, a positive coefficient estimate indicates that higher social homogeneity (based on caste or religion) increases tap water access, or conversely, higher social heterogeneity decreases tap water access,

[29] In this context, our paper is related to research on the importance of social heterogeneity for public good provision; see Dayton-Johnson (2000), Miguel and Gugerty (2005), and Khwaja (2009) for some recent contributions.

since the corresponding heterogeneity index is simply equal to 1-homogeneity index. $S_{i,k}$ represents a set of social variables, including population shares of the 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, including 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 cast that were received by the winning party and a political heterogeneity index.[30] $X_{i,k}$ 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; see 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).[31]

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

[30] The political heterogeneity index estimates the probability that two individuals randomly drawn from a population will belong to different political parties.

[31] Banerjee and Somanathan (2007) and Rushton (2008) also adopt a similar procedure.

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 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 for each caste group.[32] Second, they appear as squared terms in the caste homogeneity index, with a single coefficient. 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.[33]

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 will also change simultaneously. We follow the same procedure to analyze and interpret the effects of religious heterogeneity. We estimate the model with ordinary least squares, and discuss related econometric issues such as endogeneity, in detail in the next section.

[32]Because we use 180 caste groups, the inclusion of each caste group separately in the model may be 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.

[33] 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.

3.5 DATA

We use district-level data from the 2001 Census of India, the latest year for which data are available. Our dataset includes 436 of the 593 rural districts in India, since data for some of the explanatory variables were not available for districts in several states. Table 3.1 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.

3.5.1 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 in our study: (i) total tap water access, (ii) within-residence tap water access, and (iii) outside-residence tap water access.[34] Table 3.2 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 the state of 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 not only reflect large variation across Indian states, but also document that none of these states have 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 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.

---

[34] Total access to tap water includes both within and outside the residence access. Outside-residence tap water access refers to a household traveling 100 meters or more in rural areas to fetch drinking water.

3.5.2 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 methodology described in Banerjee and Somanathan (2007).[35] Because the caste data are from the 1931 Census, and a significant 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.[36] 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 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

---

[35] 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.

[36] 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).

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.[37]

Data on the various religious groups (Hindus, Muslims, Christians, and others) are from the religion tables of the Census of India 2001. The religion index may have potential problems because a person can potentially hide his or her religion to avoid oppression. Individuals may change from one religion to another while it is historically less likely that people change from one caste to another (Alesina, et al. 2003). We address this issue in detail in section 3.3.2.

Table 3.1 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 0.37and 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 religiously homogeneous.

3.5.3. ECONOMIC CONTROLS

Private wealth, an indicator of economic status, might be an important determinant of access to water. To this end, per-capita bank deposits across rural districts in 2001 are obtained 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 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 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 land

[37] 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.

on average has been owner-cultivated (Banerjee and Somanathan, 2007). Table 2.1 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.

3.5.4 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. Data on general elections for 1999 are from the Election Commission of India (1999) website.[38] We use two political variables, namely (i) the vote share of the winning party, and (ii) a political fragmentation index. The index is calculated using:

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

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 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 2.1).

3.5.5 GEOGRAPHY CONTROLS

Since average rainfall and temperature affects access to water, we also control for these measures. Data on average annual rainfall and average temperatures are from two sources: (i) The Indian meteorological department (IMD) and (ii) rainfall and average temperature maps

[38] 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.

released by the Maps of India website.[39] The share of wastelands in each district determines the level of water access as well. The shares of land that are steep, barren, and sandy were obtained from the Wasteland Atlas of India (Ministry of Rural Development) for 2003.[40] Finally, we also include a coastline dummy.

3.5.6 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, since population composition may affect access to tap water.

3.5.7 CORRELATION BETWEEN DEPENDENT AND CONTROL VARIABLES

Table 3.3 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 has the largest positive correlation with the Hindu caste homogeneity index (0.295). The total tap water share and outside residence tap water share have a very high correlation of 0.91, which implies that most of the tap water access for the given sample is from outside the residence. The Brahman share is positively correlated with tap water share access, because Brahmans are historically 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 does not bear any consistent correlations with the dependent variables.

Table 3.4 shows the correlation coefficients between the dependent variables and economic and political variables. It reveals consistent signs for each of the variables. One

---

[39] 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.

[40] According to the Ministry of Rural Development of India sandy areas have stabilized accumulation of sand, in coastal, riverine, or inland areas, and steepy lands are steep sloping wasteland areas.

expects a negative sign for the land Gini, which implies that the higher is land inequality, the lower is the 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.

3.5.8 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 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. A more serious issue is the endogeneity of the religion homogeneity index. Data on religious conversion rates over time may solve the issue, but are not available for India. However, the religion data between 1961 and 2001 show that the population proportions across religions have been very stable.[41] Therefore, it is also unlikely that the religious homogeneity index is endogenous.

3.5.9 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. This can influence the caste group shares and the

---

[41]The Census of India (2001) 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.

size of the village. 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. 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. We also include a wide range of other controls. Since we restrict the analysis to only rural areas, this minimizes the effects of urbanization on water access. 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.

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).

3.6 RESULTS AND DISCUSSION

3.6.1 THE HINDU CASTE HOMOGENEITY INDEX

Table 3.5 shows the results from the ordinary least squares regressions where the share of total tap water is the dependent variable. Since the test for homosckedacticity in all the regressions was rejected, we cluster errors by state and calculate Huber-White standard errors. Column (1) in Table 3.5 shows the coefficient estimates on the social variables, column (2) adds the religion, economic and political controls, and column (3) contains the entire set of controls in the regression. The coefficient estimate of the Hindu caste homogeneity index is positive and statistically significant at the 10 percent level in all three columns: a household's probability of access to (total) tap water increases by 3.9 percentage points when the household's own caste share of the population increases by 10 percentage points (column 1). The magnitude of this coefficient estimate changes only by 0.3 percentage points between columns (2) and (3). The estimated coefficient of Scheduled Tribes is statistically significant at the 5 percent level in all the columns. The point estimates imply that districts with a high concentration of Scheduled Tribes have relatively lower access to total tap water.

Inclusion of the entire set of controls improves the model's fit, increasing the overall R^2 from 0.009 to 0.051. Adding the geography, population, and wasteland controls only marginally changes the magnitude on the coefficient estimates of the social variables. The geography controls, especially the coefficient estimates of average rainfall and coastline controls, are negative and statistically significant at the 5 percent level. This is counter-intuitive because we expect that higher rainfall and proximity to a coast will increase access to water. One plausible reason for this may be the rainfall and temperature calculations. Since the calculations are based on visual mapping of weather stations to districts, the values may not be accurate. The religion, population density, economic, and political controls are not statistically significant.

Table 3.6 reports the results of the ordinary least squares regression where within-residence tap water access share is the dependent variable. Even after controlling for all covariates, the Hindu caste homogeneity index remains positive and statistically significant at the 5 percent level: a 10 percentage-point increase in a households' own caste share of the population increases the household's probability of within-residence tap water access by 2.5 percentage points. The differences in magnitude are small across the three columns. The vote share of the winning political party and the political fragmentation index are both positive and statistically significant at the 5 percent level. The religion, economic, and population controls are not statistically significant.

Table 3.7 shows results using outside-residence tap water access as the dependent variable. The most interesting result here is for the Scheduled Tribes, which are among the most disadvantaged groups. The coefficient estimate is negative and statistically significant at the 5 percent level in all three columns. 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. The coefficient estimate on the Hindu caste homogeneity index is positive but statistically insignificant. Therefore, in the case of access to outside-residence tap water, the concentration of different groups matter more than the number of groups. The religion, economic, political, and population controls are statistically insignificant. Inclusion of the entire set of controls improves the model's fit, increasing the overall R^2 from 0.02 to 0.15.

3.6.2 THE RELIGION HOMOGENEITY INDEX

Table 3.8 provides the regression results for the three types of tap water access, but with the religion homogeneity index as the main explanatory variable. All three regressions include the entire set of controls discussed earlier. The results show that more homogenous districts in terms of religion have lower access to tap water. In other words, controlling for other factors, households in districts that are more fragmented on religious lines have higher access to tap water than those in more religiously homogeneous districts. Column (1) shows the results for total tap water access: a 10 percentage point increase in a household's own religion group in the population decreases its probability of access to total tap water by 3.1 percentage points. The estimated coefficient is statistically significant at the 5 percent level. The results for within-residence tap water access, in column (2), are also similar and statistically significant at the 1 percent level. In column (3), for outside-residence tap water access, the effect of the religion homogeneity index is negative but is statistically insignificant. In all three columns, the share of Scheduled Tribes is negative and statistically significant at the 10 percent and 5 percent level in columns (1) and (3), respectively.

The results for tap water access for different religious groups are interesting. In all three regressions the relative tap water access for Muslims is lower compared to Hindus and is statistically significant at the 5 percent level for both total and within-residence access to tap water and at the 10 percent level for outside-residence tap water access. These results are plausible because Muslims are a prominent minority religion group and there are more frequent social conflicts between Hindus and Muslims than between any other religious groups. The more intriguing results are for the Christians. Their relative access to all three types of tap water is lower and statistically significant at the 5 and 10 percent levels in columns (1) and (2), respectively. In general, social tensions between Hindus and Christians are less frequent and therefore we do not expect any form of access problems for Christians. But the results indicate that they too have lower relative access to tap water. On the other hand, Sikhs, who represent another minority religion (breakaway from Hindus), have higher access to both total and within-residence tap water (statistically significant at the 10 and 5 percent levels, respectively).

Finally, we also include both the Hindu caste homogeneity and religion homogeneity indices with the shares of castes and religion groups in the same regression (table not reported).[42] The estimated coefficients on all of the social and religion variables are similar to the ones previously reported and consistent.

---

[42] Results are available upon request.

3.6.3 ROBUSTNESS CHECKS

We perform a number of sensitivity checks to ensure that the results from using the two measures of social divisions are robust. Table 3.9 reports the coefficient estimate on the Hindu caste homogeneity index from a range of sensitivity tests. Specification (1) shows the baseline estimates. Specification (2) shows the coefficient estimates when the population density variables are excluded from the regressions. The results are robust to dropping the population density variables. Particularly, the coefficient estimate of Hindu caste homogeneity index is positive and is statistically significant at the 10 percent level. We also conduct many other specification checks. Since the correlation between the two political control variables is high (-0.9), in some specifications we include either the political index or the vote share of the winning party. These specifications seldom change the results of the variable of interest. Further, both the political variables were also excluded from the specification due to the possibility of endogeneity. But the results remain intact.

Because the population of a village influences caste composition in a given district, we include an interaction term between the Hindu caste homogeneity index and the average village population, to check if the magnitude of the average village population affects the partial effect of the Hindu caste homogeneity index. However, there is no evidence from this sample that the magnitude of the average village population influences the effect of caste heterogeneity on tap water access.

Table 3.10 reports the results of an analogous sensitivity analysis for the religion homogeneity index. All the results, excluding the interaction between the religion index and the average village population, are robust to specification changes. The interaction term shows evidence that the magnitude of the average village population influences the partial effect of religion heterogeneity on tap water access.

3.6.4 CASTE VERSUS RELIGION: THE ROLE OF HISTORY AND INSTITUTIONS

One of the striking findings of our empirical exercise is that while social divisions based on the Hindu caste system reduces access to tap water in rural India, those based on religion seem to improve access. A natural question at this point is: why do the results go in opposite directions? A possible explanation may lie in the role played by history and institutions in determining the impact of social divisions on the coordination mechanism needed by a community for gaining access to public goods.

As discussed earlier, the caste system in India has been historically pervasive and has been 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 with those from the "higher" castes, provided the foundation for this segregation and lack of tolerance. For example, a person belonging to a lower caste was prohibited from entering the house of a person who was higher up in the caste hierarchy. These historically pervasive social divisions create social norms over time, which not only tend to be very persistent, but also are critical in determining economic outcomes; see Ray (1998, chapter 5). Our results on the effects of caste heterogeneity therefore indicate that, even with the caste system being constitutionally illegal in India, its historical barriers have prevented the cooperation necessary within communities to get access to publicly provided drinking water.

Religious diversity, on the other hand, plays a very different role in affecting the coordination mechanism needed for access to public goods. India has a long history of trade with Persia and Europe, as well as external conquests from the very same regions. These also exposed the country very early on to the world's predominant religions, such as Islam and Christianity. The Mughal Empire ruled India for almost 400 years, and was followed by British colonization for 200 more (which also coincided with some areas being under French, Dutch, and Portuguese occupation). The existing social order and the need for external commerce perhaps created a degree of tolerance among religions in India that eventually became historically persistent. The secular institutions guaranteed by the Indian Constitution (Freedom of Religion being a Fundamental Right), further strengthened the tolerance and cooperation among religions since Independence. Our result on the religion heterogeneity index underscores this point.

3.7 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 aggregate measures of social fragmentation that are comprised of several socio-economic characteristics such as ethnicity, race, language, religion, and caste, we employ disaggregate measures. Consequently, our approach better allows for individual measures of diversity to have heterogeneous effects on outcomes.

The empirical analysis suggests 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. By contrast, communities that are fragmented across religions are likely to have more access to tap water than correspondingly homogeneous communities. In essence, even though Hindus are a large majority among religious groups in India, representing more than 80 percent of the population, our results indicate that religious diversity fosters better access to public goods, perhaps because secular institutions generate greater tolerance between people across religions. These are interesting results, since they indicate that though 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 results also indicate that Scheduled Tribes and minority religious groups such as Christians and Muslims have relatively lower access to tap water in rural India compared to Hindus. Therefore, while caste-based and religious heterogeneity is important for understanding public goods access, the existence of minority groups in the economy (in terms of economic or religious classifications) also matter. Finally, the interaction between social factors and the source of tap water is also important: while caste-based fragmentation is crucial for tap water access within the residence, the concentration of caste groups matters for tap water outside the residence.

These results point to the need for public 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 Hyderbabad. 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 the water (Mulford, 2006).

We end 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.

3.8 REFERENCES

Agricultural Census of India. 2001. Department of Agriculture & Cooperation, *Agricultural Census Division*, Ministry of Agriculture, India.

Alesina, Alberto, Baqir, Reza and Easterly, William. 1999. “Public goods and ethnic divisions.” *Quarterly Journal of Economics,* 114 (4): 1243–1284.

Alesina, Alberto, Devleeschauwer, Arnaud, Easterly, William, Kurlat, Sergio and Wacziarg, Romain. 2003. “Fractionalization.” *Journal of Economic Growth*, 8, 155–194.

Bajpai, Nirupam. 2007. “India’s Growing water crisis.” *Hindu Business Line*, 12, January.

Banerjee, Abhijit, Iyer, Lakshmi and Somanathan, Rohini. 2005. “History, Social Divisions, and Public Goods in Rural India.” *Journal of the European Economic Association*, 3(2-3), 639–647.

Banerjee, Abhijit and Somanathan, Rohini. 2007. "The Political Economy of Public Goods: Some Evidence from India." *Journal of Development Economics*, 82, 287-314.

Brooks, Nina. 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, Jenna, Gary White, Said Damodaran, and Rich Thorsten. 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.

Easterly, William, and Levine, Ross. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions," *The Quarterly Journal of Economics*, 112(4), 1203–1250.

Election Commission of India. 1999. Statistical Reports of Lok Sabha Elections. http://www.eci.gov.in/StatisticalReports/ElectionStatistics.asp

Government of India. 2002. India- Assessment-Water and Sanitation. Planning Commission, Government of India.

Haarmeyer, David and Mody, Ashoka. 1997. "Private Capital in Water and Sanitation." *Finance and Development.*

Haub, Carl and Sharma, O.P. 2006. "India's Population Reality: Reconciling Change and Tradition." *Population Bulletin,* 61(3), 1-20.

Jackson, Ken. 2007. "Why Does Diversity Affect Public Good Provision? An Empirical Analysis of Water Provision in Africa." Dissertation Chapter, University of British Columbia.

Khwaja, A. 2009. “Can good projects succeed in bad communities?” *Journal of Public Economics*, 93, 899-916.

Miguel, Edward, and Gugerty, Mary Kay. 2005. “Ethnic Diversity, Social Sanctions and Public Goods in Kenya.” *Journal of Public Economics*, 89, 2325–2368.

Mulford, David. 2006. “Public-Private Partnership can help solve India’s water and sanitation challenges.” *The Hindu*, March 22 issue.

Ramachandran, Asha. 2006. “Any Hope for India’s Water Woes.” India/Kerala News. http://www.greentreks.org/woodchuckcafe/features/feature_indiaswater.asp

Ray, Debraj. 1998. Development Economics. Princeton University Press, Princeton, NJ.

Rushton, Michael. 2008. “A Note on the Use and Misuse of the Racial Diversity Index.” *Policy Studies Journal*, Forthcoming.

Saleth, Rathinasamy Maria. 2005. “Water Institutions in India: Structure, performance, and change.” In Gopalakrishnan, C.; Biswas, A. K.; Tortajada, C. eds., Water Institutions: Policies, Performance and Prospects. New York, NY. pp: 47-80.

Sengupta, Somini. 2006. “In Teeming India, Water Crisis Means Dry Pipes and Foul Sludge.” *New York Times*, 29, September.

Swiss Agency for Development and Cooperation (SDC). 2007. “Retention and sustainable use of water.” http://www.sdcindia.in/en/Home/Activities_in_2007/Water_NRM

Vigdor, Jacob. L. 2002. “Interpreting Ethnic Fragmentation Effects.” *Economic Letters*, 75, 271-276.

Vigdor, Jacob. L. 2004. “Community Composition and Collective Action: Analyzing Initial Mail Response to the 2000 Census.” *The Review of Economics and Statistics*, 86, 303–312.

World Bank. 1993. World Development report: "Investing in health". New York: Oxford University Press.

World Bank Report. 2005. "India's Water Economy: Bracing for a Turbulent Future."

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

Table 3.1: Summary Statistics

Variables	No. of Obs.	Mean	Std. Dev.	Min.	Max.
Dependent Variables					
Total tap share	436	.27	.24	.01	.91
Total Tap Share (Within)	436	.11	.11	0	.57
Total Tap Share (Outside)	436	.17	.18	0	.80
Independent Variables					
Caste Variables					
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
Religion Variables					
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
Economic Variables					
Land Gini	436	.71	.09	.41	.86
Per Capita Bank Deposits	436	.02	.02	.002	.23
Political Variables					
Political Index	436	.59	.09	.14	.80
Vote Share of Winning Party	436	.48	.08	.27	.71
Population Variables					
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
Geography Variables					
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 3.2: Mean Share of Households with Access to Tap Water

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

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

Table 3.3: Correlations between Tap Water Access and Hindu Caste Variables

	Tap share (total)	Tap share (within)	Tap share (outside)	Hindu Caste Index	Brahman	SC	ST	Religion Index
Tap share (total)	1							
Tap share (within)	0.71	1						
Tap share (outside)	0.92	0.36	1					
Hindu Caste Index	0.28	0.29	0.21	1				
Brahman	0.13	0.13	0.1	0.42	1			
SC	0.01	0.08	-0.03	0.25	0.3	1		
ST	0	0.24	-0.14	0.29	0.06	0.24	1	
Religion Index	0.1	-0.03	0.16	0.2	0.28	0.19	0.14	1

Note: SC- Scheduled Caste; ST- Scheduled Tribe

Table 3.4: 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						
Tap share(within)	0.71	1					
Tap share (outside)	0.92	0.36	1				
Land Gini coefficient	-0.11	-0.1	-0.08	1			
PCBD	0.32	0.17	0.32	-0.17	1		
Vote share	0.15	0.04	0.17	0.08	0.14	1	
Political Index	-0.22	-0.06	-0.25	-0.04	-0.19	-0.91	1

Note: PCBD- Per Capita Bank Deposits.

Table 3.5: Caste Composition and Total Tap Water Access
Dependent Variable: Total Tap water access share

Effect [Independent Variable]	(1)	(2)	(3)
Social Variables			
Mean within group affinity [Caste Homogeneity Index]	0.385*	0.350*	0.297*
	(1.84)	(1.82)	(1.91)
Share of Brahmans	-0.574	-0.520	-0.339
	(-1.58)	(-1.45)	(-1.19)
Share of Scheduled castes	0.131	0.149	0.073
	(0.79)	(0.93)	(0.57)
Share of Scheduled Tribes	-0.277**	-0.272**	-0.296**
	(-2.32)	(-2.23)	(-2.47)
Religion Variables			
Share of Muslims		0.039	0.087
		(0.25)	(0.63)
Share of Christians		-0.022	-0.041
		(-0.11)	(-0.31)
Share of Sikhs		0.195	0.229
		(0.58)	(0.83)
Economic Variables			
Per Capita Bank Deposits		0.0253	0.207
		(0.04)	(0.31)
Land Gini		-0.0411	-0.043
		(-0.32)	(-0.43)
Political Variables			
Vote share		0.203	0.197
		(1.38)	(1.36)
Political Index		0.046	-0.011
		(0.26)	(-0.08)
Geography Variables			
Rainfall (in meters)			-0.042***
			(-2.54)
Temperature (Celsius)			-0.004
			(-1.03)
Coastline			-0.0920**
			(-2.93)
Constant	0.278***	0.170	0.485**
	(21.47)	(0.80)	(2.43)
N	436	436	436
Population Control Variables	NO	NO	YES
Wasteland Variables	NO	NO	YES
State Fixed Effects	YES	YES	YES
Overall R^2	0.009	0.004	0.051

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 religions.

Table 3.6: Caste Composition and Within-residence Tap Water Access
Dependent Variable: Tap water access share (Within)

Effect [Independent Variable]	(1)	(2)	(3)
Social Variables			
Mean within group affinity [Caste Homogeneity Index]	0.217	0.266**	0.249**
	(1.22)	(2.41)	(2.07)
Share of Brahmans	-0.332	-0.334**	-0.249
	(-1.41)	(-1.96)	(-1.34)
Share of Scheduled castes	0.055	0.0698	0.047
	(0.67)	(1.37)	(0.84)
Share of Scheduled Tribes	-0.123	-0.126	-0.126
	(-1.26)	(-1.32)	(-1.37)
Religion Variables			
Share of Muslims		0.123	0.134
		(1.16)	(1.26)
Share of Christians		0.091	0.076
		(1.05)	(0.99)
Share of Sikhs		0.265	0.271
		(1.28)	(1.57)
Political Variables			
Vote share		0.205**	0.203**
		(2.31)	(2.27)
Political Index		0.195***	0.173**
		(2.85)	(2.44)
Constant	0.109	-0.172*	-0.024
	(11.43)	(-1.95)	(-0.20)
N	436	436	436
Economic Control Variables	NO	YES	YES
Population Control Variables	NO	NO	YES
Geography Control Variables	NO	NO	YES
Wasteland Control Variables	NO	NO	YES
State Fixed Effects	YES	YES	YES
Overall R^2	0.000	0.000	0.008

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 all other religion groups.

Table 3.7: Caste Composition and Outside-residence Tap Water Access
Dependent Variable: Tap water access share (Outside)

Effect [Independent Variable]	(1)	(2)	(3)
Social Variables			
Mean within group affinity [Caste Homogeneity Index]	0.167	0.083	0.048
	(0.71)	(0.40)	(0.28)
Share of Brahmans	-0.242	-0.186	-0.091
	(-0.97)	(-0.69)	(-0.46)
Share of Scheduled castes	0.077	0.079	0.027
	(0.46)	(0.50)	(0.20)
Share of Scheduled Tribes	-0.155**	-0.146**	-0.170**
	(-1.98)	(-2.00)	(-1.99)
Religion Variables			
Share of Muslims		-0.083	-0.047
		(-1.23)	(-1.02)
Share of Christians		-0.114	-0.117
		(-0.90)	(-1.53)
Share of Sikhs		-0.069	-0.041
		(-0.46)	(-0.32)
Constant	0.168***	0.342	0.509***
	(24.85)	(1.59)	(2.63)
N	436	436	436
Economic Control Variables	NO	YES	YES
Political Control Variables	NO	YES	YES
Geography Control Variables	NO	NO	YES
Population Control Variables	NO	NO	YES
Wasteland Variables	NO	NO	YES
State Fixed Effects	YES	YES	YES
Overall R^2	0.021	0.159	0.152

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 all other religion groups.

Table 3.8: Religious Composition and Tap Water Access
Using the Religious Homogeneity Index

Effect [Independent Variable]	Tap water Share (Total)	Tap water Share (Within)	Tap water Share (Outside)
Social Variables			
Mean within Group affinity [Religion Homogeneity Index]	-0.309**	-0.260***	-0.049
	(-2.29)	(-3.18)	(-0.60)
Share of Brahmans	-0.157	-0.096	-0.062
	(-0.48)	(-0.54)	(-0.29)
Share of Scheduled castes	0.116	0.082*	0.034
	(0.87)	(1.80)	(0.28)
Share of Scheduled Tribes	-0.223*	-0.065	-0.158**
	(-1.75)	(-0.65)	(-1.99)
Religion Variables			
Share of Muslims	-0.212**	-0.118**	-0.094*
	(-2.24)	(-2.16)	(-1.64)
Share of Christians	-0.427**	-0.250***	-0.178
	(-1.99)	(-2.00)	(-1.57)
Share of Sikhs	0.249*	0.287***	-0.038
	(1.79)	(3.66)	(-0.35)
Constant	0.829***	0.266*	0.563***
	(3.35)	(1.92)	(2.77)
N	436	436	436
Economic Control Variables	YES	YES	YES
Political Control Variables	YES	YES	YES
Geography Control Variables	YES	YES	YES
Population Control Variables	YES	YES	YES
Wasteland Control Variables	YES	YES	YES
State Fixed Effects	YES	YES	YES
Overall R^2	0.028	0.012	0.142

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

Table 3.9: Robustness Check for the Hindu Caste Homogeneity Index

Specification	Tap water share (Total) (1)	Tap water Share (Within) (2)	Tap water share (Outside) (3)
(1) Baseline (including All controls)	0.297*	0.249**	0.048
	(1.91)	(2.07)	(0.28)
(2) Excluding the population density variables	0.314*	0.249**	0.066
	(1.91)	(2.13)	(0.36)
(3) Excluding Political Index	0.297*	0.234**	0.064
	(1.85)	(1.96)	(0.35)
(4) Excluding Vote share of the winning party	0.287*	0.242**	0.048
	(1.84)	(2.05)	(0.28)
(5) Excluding both the political controls	0.313*	0.238*	0.075
	(1.78)	(1.93)	(0.40)
(6) Using Interaction Terms:			
Average Village Population * Caste Homogeneity Index	0.290	0.094	0.197
	(1.52)	(0.72)	(0.85)
N	436	436	436
State Fixed Effects	YES	YES	YES

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

Table 3.10: Robustness Check for the Religion Homogeneity Index

Specification	Tap water share (Total) (1)	Tap water Share (Within) (2)	Tap water share (Outside) (3)
(1) Baseline (including all controls)	-.309**	-.260***	-0.048
	(-2.29)	(-3.18)	(-0.60)
(2) Excluding the population density variables	-0.319**	-0.250***	-0.059
	(-2.31)	(-3.04)	(-0.74)
(3) Excluding Political Index	-0.307**	-0.263***	-0.044
	(-2.30)	(-3.15)	(-0.54)
(4) Excluding Vote share of the winning party	-0.311**	-0.261***	-0.049
	(-2.28)	(-3.11)	(-0.60)
(5) Excluding both the political controls	-0.301**	-0.261***	-0.040
	(-2.19)	(-3.13)	(-0.47)
(6) Using Interaction Terms: Average Village Population * religion homogeneity Index	-0.399**	-0.314***	-0.084
	(-2.34)	(-3.31)	(-0.79)
N	436	436	436
State Fixed Effects	YES	YES	YES

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

CHAPTER 4

INSTITUTIONS VERSUS PREFERENCES APPROACH: WHICH UNDERLYING MECHANISM DETERMINES TAP WATER ACCESS IN RURAL AND URBAN INDIA? EVIDENCE FROM THE NATIONAL FAMILY HEALTH SURVEY

4.1 INTRODUCTION

One of the major factors that contribute to economic development and long run economic growth is the adequate provision of public goods. In developing countries, the inadequate provision of these goods hinders development. Several factors could affect the provision of public goods in a society. The related literature emphasizes the role of ethnic divisions, and political and economic heterogeneity in determining the allocations of public goods (Alesina, et al. (1999), Miguel (2000) and Banerjee and Somanathan (2007)). Banerjee and Somanathan (2007) examine the importance of relative power of different ethnic groups in determining the level of access to different public goods. They find that areas with a high concentration of scheduled castes (SC) and scheduled tribes (ST) have relatively poor access to public goods in rural India.[43]

Several theories explain why there is a negative effect of ethnicity on public good provision. Alesina et al. (1999) focus on how differences in preferences across various ethnic groups affect the level of public goods provision. This is the preferences theory where the

[43] Scheduled caste and Scheduled tribe are terms used by the Government of India to classify the poorest and most disadvantaged communities in India; scheduled caste refers to the dalit community, and scheduled tribal to the tribal communities or adivasis.

optimal provision of public goods is lower in an ethnically heterogeneous community.[44] On the other hand, the theory of institutions emphasizes the role of transaction costs associated with inter-ethnic group coordination.[45] (Miguel and Gugerty (2005) and Vidgor (2004) Miguel and Gugerty underscore that "social sanctions" are an effective tool for collective action in homogeneous communities in rural Kenya, while social sanctions in heterogeneous communities are ineffective due to free-riding problems. Identifying factors that affect access to public goods is crucial for policy reforms: if a highly ethnically fragmented area has less access to water because of high coordination costs, then this substantiates the case for policy reforms in the water sector. And it is therefore, the aim of this paper to analyze which of these theories explains the negative effects of caste heterogeneity underlying on access to tap water across India.

This study deviates from the existing literature in an important dimension. Since most of the empirical literature focuses on aggregate measures of public goods provision, the underlying mechanism that drives the effects of ethnicity on public goods provision is not revealed.[46] By contrast, this study uses household-level data from the II National Family Health Survey of 1998-99 to analyze the channel through which these effects happen. No other studies analyze why caste heterogeneity affects tap water access a household- level dataset for India. The results from this household analysis confirm that the theory of institutions explain the negative effect of social divisions on tap water access in rural India. That is, local institutions may allow

---

[44] Based on the similar idea of differing preferences, Alesina and Ferrara (2000) examine the degree of participation in heterogeneous groups of population. They specifically analyze the tolerance level of individuals in the presence of other groups. They find that the effects of heterogeneity are important when there is very low excludability of individuals in public goods access.

[45] They posit that the within-group coordination is more feasible than between groups due to the free-riding problems associated with them

[46] A number of empirical papers on ethnic heterogeneity use aggregate data such as city level, district level dataset to test for the negative relationship between ethnicity and public goods provision (Alesina et al. (1999), Banerjee and Somanathan (2007)). The aggregate measures fail to predict which mechanism drives the negative effects.

coordination among inter-ethnic groups and thereby improve access to tap water in rural India. However, I do not find similar results for urban India.

The paper is organized as follows. Section 4.2 explains the theoretical framework and its implications. Section 4.3 will present the empirical model to test the theory. Section 4.4 will explain the data used in the analysis, and some important econometric issues. Section 4.5 presents the results at the household level and finally Section 4.6 concludes.

4.2 THEORETICAL FRAMEWORK

The paper uses Jackson's (2007) model to analyze the underlying mechanism that contributes to the negative effects of ethnicity on tap water access in India. The model assumes that a community provides a public good funded by the households in that community and this good can be influenced by both institutions and preferences. The total population in the community N is divided among E ethnic groups with p_e, representing the share of the e^{th} ethnic group. Each household derives utility from private consumption and the public good.

$$u_i = W_i - x_i + \mu \ln X_i \tag{1}$$

where, W_i is household wealth, x_i is household's contribution to the public good and X_i is the benefit from the public good. μ accounts for local preference for the public good and the costs of installation and maintenance. It assumes that ethnic groups have no internal free-riding, identical decision making within the group and they maximize the group utility.[47]

[47] Due to symmetry within groups, we denote x_i as x_e

$$X_e = p_e N x_e + \alpha \sum_{e' \neq e} p_{e'} N x_{e'} \qquad (2)$$

Equation (2) implies that the benefit the individual receives from the public good (X_e) depends on both the contributions from one's own ethnic group and from all other individuals. However, the magnitude of the benefits would be greater if the contributions come from one's own ethnic group. The model accounts for differing preferences through $\alpha \in [0, 1]$, where $\alpha = 1$ represents no variation in preferences. It also incorporates the role of inter-ethnic institutions to enable the varying incentives faced by different ethnic groups to contribute that would maximize the welfare of the society. Each group chooses their contributions taking the actions of all other groups as given. Therefore the model is:

$$U_e = \max_{x_e}[(1-\tau)p_e N(W - x_e + \mu \ln X_e) + \tau \sum_{j=1}^{E} p_j N(W - x_j + \mu \ln X_j)] \qquad (3)$$

where the parameter $\tau \in [0, 1]$ captures the effectiveness of local institutions to manage inter-ethnic coordination.

The two extreme cases highlighted in this model are a) divergent preferences, which implies $\alpha = 0$ and no effect of institutions, τ. Each group contributes to the extent it maximizes the group's welfare and the contribution of one group has no impact on other groups' welfare. The goods thus supplied are called "ethnic goods"; b) Complete spillovers are when $\alpha = 1$ and it implies that the role of institutions are important. If $\tau = 0$, then only the largest group will contribute and no other group contributes. However, if $\tau = 1$, then all groups contribute. These goods are referred to as "community goods." "If spillovers are less than complete an individual in the largest group will receive greater value from the public good than an individual in any other group. For $\alpha < 1$, if $p_1 > p_j$, $X_1 > X_j$. For $\alpha = 1$, $X_1 = X_j$."(pp: 12 Proposition: 2, Jackson (2007)).The proposition implies that in case of the preferences approach (α <1) individuals from the largest ethnic group should receive larger benefit as compared to individuals in other ethnic groups. On the other hand, if the institutions approach is true, then individuals receive equal benefit irrespective of the who contributes, then any negative effects of ethnicity on public good provision is due to poor institutions. I use this proposition to test for the channel that causes the negative effects of diversity and access to public goods.

4.3 METHODS

To identify which of these two mechanisms cause the negative effect of caste on tap water access, I use the following household specification:

$$X_i = \delta_0 + \delta_1 SLCG_c + \delta_2 HHLG_i + \beta_1 Z_c + \beta_2 H_i + \varepsilon_i \quad (4)$$

where X_i takes on the value 1 if the household has access to tap water and 0 otherwise. $SLCG_c$ represent the share of the largest caste group within each state (explained in section 3.4). $HHLG_i$ measures whether the household belongs to the largest caste group (as defined above).[48] Z_c is a set of geography and economic controls and H_i is a vector of household characteristics. If preferences drive the negative effect of access to tap water then individuals in the largest caste group will have better access to tap water and therefore, in addition to δ_1>0, the results should indicate δ_2>0. Only if there is no changing preferences ($\alpha = 1$) the model predicts δ_2=0 and it implies that it is the role of institutions that affect the access to tap water. In this analysis I find that it is the role of institutions that help explain the variation is access to tap water. These results are similar to Jackson (2007).

4.4 DATA, DESCRIPTION, AND ECONOMETRIC ISSUES

[48] A crucial assumption is that the household is more likely to access tap water when their own caste group members invest and influence the resulting public good in the community.

4.4.1 DATA AND DESCRIPTIVE STATISTICS

I use data from the 1998-1999 National Family Health Survey (NFHS-2) for rural and urban India. The main reason to use the 1998-99 dataset is because elsewhere I study the effects of social divisions on access to tap water in rural India at the district level using the 2001 Census data from India (Balasubramaniam et al. (2009)). To address the objective of this paper I use the household-level data and 1998-99 is the closest year available from the National Family Health Survey. The survey covers 89,199 women between the ages of 15 and 49 across 29 states from a total of 91,196 households in both rural and urban areas.[49] The survey asks several questions including child health, anthropometric measures of children, family planning practices, fertility and the socioeconomic status of women.

I use the state average rainfall and temperature maps from Maps of India website to measure the geography controls. Wealth index measures the economic status of the household. The wealth data are from the NFHS dataset. The NFHS survey calculates the wealth index using the principle component analysis, which assigns different weights on a household's various asset holdings (for example, television, radio, etc.) and living conditions (for example, type of roof materials, etc.).

The NFHS provides the name of the caste to which each household belongs. To calculate the share of the largest caste group I sort the data by caste groups within each state and the group that has the maximum number of households represented in the sample is the largest group within each state. Based on this information, I measure whether each household belongs to the largest caste group in the state. I restrict the analysis to Hindu castes to be consistent with Balasubramaniam et al. (2009). Other controls include information of the household head, and

---

[49] Data on union territories are not collected. A union territory is a sub-national administrative division of India. Unlike the states, which have their own elected governments, union territories are ruled directly by the federal national government (Wikipedia).

age of the household head. I include 29,180 rural households and 12, 967 urban households in this analysis.[50] Tables 4.1 and 4.2 show that on average, 10 percent of the rural households and 52 percent of urban households have access to tap water in India and the share of the largest caste group is about 13 percent in rural areas and 14 percent in urban areas. About 14 percent of the rural households and 12 percent of urban households belong to the largest caste groups.

4.4.2 ECONOMETRIC ISSUES

There are some important econometric issues in this analysis. It is important to control for wealth since wealthier states may fund more projects that in turn increase access to tap water in the community. Since I use the wealth index, there are some endogeneity issues. First, access to tap water may be indirectly associated with increases in the wealth measure. Second, better access to water may leave households with more time for more productive activities (Jackson, 2007). To solve this problem, I instrument for wealth using a stock variable, namely, whether the household has a color television. It is highly unlikely that this instrument is directly associated with the source of access to water. In addition, in India, households do not frequently change their TV, which further reduces the endogeneity issue.[51]

Another concern is migration. More diverse regions may have recent migrants and therefore, change the composition of the region. I therefore include households who have at least lived in the residence for one year or more in the sample. Jackson (2007) includes households who have lived in the same residence for over 20 years. I could not use this threshold in this analysis as because it substantially reduced the sample.

---

[50] I include only the following states: Andhra Pradesh, Arunachal Pradesh, Kerala, Karnataka, Tamilnadu, Uttar Pradesh, West Bengal, Gujarat, Punjab, Haryana, Madhya Pradesh, Maharastra, Orissa, Rajasthan,Himachal Pradesh. Balasubramaniam et al. (2009) includes most of these states in their district-level analysis.

[51] There is a possibility that color TV (that requires electricity) and in turn may be associated with the source of water access. However, I could not find a better instrument in the given dataset.

4.5 RESULTS

Tables 4.3 and 4.4 present the regression results for a household's access to tap water in rural and urban areas respectively. I cluster the analysis by district number to account for the unobserved heterogeneity within districts where the households reside.[52] Column (1) is the base regression using OLS and column (2) shows the IV results when I instrument for wealth measure. The OLS coefficient estimate of share of the largest caste group is positive and statistically significant at the 1 percent level for the rural and urban sample. This implies that when there is a concentration of a single caste group it assists access to public goods within the community. However, the OLS coefficient estimates of the household that belongs to the largest group ($\hat{\delta}_2$) is not statistically significant both in the rural and urban areas. The model predicts that if α =1, then δ_2=0. This result implies that the institutions explain why heterogeneous communities have lower access to public goods.

The result is crucial for the analysis because, this implies that the household in the largest group is no more likely to access tap water relative to other groups. This clearly shows that it is the institutions that cause the negative effects of social divisions on access to public goods. The IV regressions also present similar results for the rural sub-sample confirming that institutions are the channel through which social divisions can play an important role in access to public goods. I use the F-stat to determine the strength of the instrument. As a rule of thumb if the F-stat is greater than 10 then we can conclude that the instrument is strong. In this analysis, I find that the F-stat both for the rural and urban sample is greater than 10 and therefore, conclude that the instrument is strong.

---

[52] It is very possible that the access to tap water within each district in which the household live may not be independent, and this could lead to residuals that are not independent within districts. Therefore, I use the cluster option to indicate that the observations are clustered into districts and that the observations may be correlated within districts, but would be independent between districts.

These results for the rural sample are consistent with Jackson's (2007) findings for Sub-Saharan Africa. However, the IV result for the urban sub-sample does not show this pattern implying that there is no strong evidence to suggest which mechanism drives the ethnic effects across the urban sample.

4.6 CONCLUSIONS

Social divisions may affect access to public goods. The literature explains the relationship between ethnic diversity and public goods provision by two competing mechanisms. One is that people have different preferences in having the public good. The other posits that weak institutions may make free-riding feasible across different groups and in turn, lead to lower public goods in diverse communities. The analysis suggests institutions have a greater influence on how social divisions affect access to tap water in both in India, specifically in rural areas.

4.7 REFERENCES

Alesina, Alberto, and Eliana La Ferrara. 2000. "Participation in Heterogeneous Communities." *Quarterly Journal of Economics*, 115(3), 847-904.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public goods and ethnic divisions." *Quarterly Journal of Economics* 114 (4): 1243–1284.

Balasubramaniam, Divya, Santanu Chatterjee, and David B. Mustard. 2009. "Got Water? Social Divisions and Access to Public Goods in Rural India." (March 19, 2009). Available at SSRN: http://ssrn.com/abstract=1365393

Banerjee, Abhijit and Rohini Somanathan. 2007. "The Political Economy of Public Goods: Some Evidence from India." *Journal of Development Economics*, 82, 287-314.

Jackson, Ken. 2007. "Why Does Diversity Affect Public Good Provision? An Empirical Analysis of Water Provision in Africa." Dissertation Chapter, University of British Columbia.

Miguel, Edward, and Mary Kay Gugerty. 2005. "Ethnic Diversity, Social Sanctions and Public Goods in Kenya." *Journal of Public Economics*, 89, 2325–2368.

Miguel, E. 2000. "Ethnic Diversity and Education Funding in Kenya," Mimeo, Harvard University.

Vigdor, Jacob. L. 2004. "Community Composition and Collective Action: Analyzing Initial Mail Response to the 2000 Census." *The Review of Economics and Statistics*, 86, 303–312.

Table 4.1: Summary Statistics, **N= 29,180 (Rural)**

Variables	Mean	Std. Dev.	Min.	Max.
Dependent Variables				
Household has access to Tap water	.10	.31	0	1
Independent Variables				
Caste Variables				
Share of the largest caste group	.13	.07	.06	.34
Household belongs to the largest caste group	.14	.34	0	1
Household Characteristics				
Female is the household head	.08	.28	0	1
Age of the household head	43.69	13.41	2	95
Years of residency	23.45	25.89	1	95
Economic Variables				
Wealth index	-.46	.71	-1.52	2.60
Geography Variables				
Avg. Rainfall (centimeters)	94.97	43.84	38.46	244.92
Avg. Temperature(celsius)	25.45	2.73	15	28.24

Table 4.2: Summary Statistics, **N= 12,967 (Urban)**

Variables	Mean	Std. Dev.	Min.	Max.
Dependent Variables				
Household has access to Tap water	.52	.50	0	1
Independent Variables				
Caste Variables				
Share of the largest caste group	.14	.07	.06	.34
Household belongs to the largest caste group	.12	.33	0	1
Household Characteristics				
Female is the household head	.09	.29	0	1
Age of the household head	44.41	12.43	13	95
Years of residency	25.25	30.19	1	95
Economic Variables				
Wealth index	.82	.93	-1.42	2.79
Geography Variables				
Avg. Rainfall (centimeters)	97.34	42.75	38.46	244.92
Avg. Temperature(Celsius)	25.67	2.55	15	28.24

Table 4.3: OLS and IV regressions; Dependent Variable: Household has access to tap water (Rural Sample; N=29,180)

Variables	Base Regression (1)	IV Wealth (2)
Share of the largest caste grou	.424***	.260**
	(3.34)	(2.30)
Household in the largest caste group	.006	.007
	(0.51)	(0.56)
Household wealth	.168***	.148***
	(18.82)	(21.01)
Female headed household	-.014**	-.012*
	(-2.21)	(-2.05)
Age of the household head	-.000**	-.000
	(-2.23)	(-1.12)
Years of residency	-.000	-.000
	(-0.51)	(-0.85)
Average rainfall	-.006***	-.000***
	(-6.47)	(-4.21)
Temperature	.006**	.002**
	(2.11)	(0.82)
First Stage Results		
F-Stat (Household wealth)	-	117.14
Observations	29,180	29,180
Clusters	63	63
R^2	0.199	.202

t-statistics reported in parentheses

Table 4.4: OLS and IV regressions; Dependent Variable: Household has access to tap water (Urban Sample; N=12,967)

Variables	Base Regression (1)	IV Wealth (2)
Share of the largest caste grou	.762***	.285
	(2.82)	(0.71)
Household in the largest caste group	-.001	.007
	(-0.09)	(0.30)
Household wealth	.269***	.251***
	(40.44)	(29.89)
Female headed household	-.000	-.006
	(-0.00)	(-0.33)
Age of the household head	-.000	.000
	(-0.57)	(0.56)
Years of residency	.000	-.000
	(0.75)	(-0.33)
Average rainfall	-.002***	-.002***
	(-11.25)	(-7.26)
Temperature	.008*	.000**
	(1.65)	(0.12s)
First Stage Results		
F-Stat (Household wealth)	-	291.82
Observations	12,967	12,967
Clusters	50	50
R^2	0.371	.363

t-statistics reported in parentheses

CHAPTER 5

CONCLUSIONS

The first chapter examines the distributional impacts of access to water and sanitation on children's stunting, wasting, and underweight outcomes in India. In contrast to most studies, which analyze only some of these measures of malnutrition, I analyze all of these measures individually. Consequently, this approach better explains the varying distributional impacts of access to water and sanitation on each measure of malnutrition. The traditional ordinary least squares estimation may undermine important information regarding the partial effect of the interest variable on the health outcome variables. Compared to the OLS results, the QR estimations may better inform policymakers to make inferences based on the results from the quintile distribution rather than the average effect.

In the second chapter 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 aggregate measures of social fragmentation that are comprised of several socio-economic characteristics such as ethnicity, race, language, religion, and caste, we employ disaggregate measures. The empirical analysis suggests 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. By contrast, communities that are fragmented across religions are likely to have more access to tap water than correspondingly homogeneous communities.

Social divisions may affect access to public goods. The literature explains the relationship between ethnic diversity and public goods provision by two competing mechanisms. One is that people have different preferences in having the public good. The other posits that weak institutions may make free-riding feasible across different groups and in turn, lead to lower public goods in diverse communities. The analysis suggests institutions have a greater influence on how social divisions affect access to tap water in rural India. Overall, this research provides interesting results that are useful for public policies targeted to provide better water access in India.