The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach

Shanti Gamper-Rabindran
Shakeeb Khan
Christopher Timmins

February 26, 2009

Abstract

We examine the impact of piped water on the under-1 infant mortality rate (IMR) in Brazil using a recently developed econometric procedure for the estimation of quantile treatment effects with panel data. The provision of piped water in Brazil is highly correlated with other observable and unobservable determinants of IMR – the latter leading to an important source of bias. Instruments for piped water provision are not readily available, and fixed effects to control for time invariant correlated unobservables are invalid in the simple quantile regression framework. Using the quantile panel data procedure in Chen and Khan (2007), our estimates indicate that the provision of piped water reduces infant mortality by significantly more at the higher conditional quantiles of the IMR distribution than at the lower conditional quantiles (except for cases of extreme underdevelopment). These results imply that targeting piped water intervention toward areas in the upper quantiles of the conditional IMR distribution, when accompanied by other basic public health inputs, can achieve significantly greater reductions in infant mortality.

JEL Codes: I18, H41, Q53, Q56, Q58

Keywords: Infant mortality, piped water supply, quantile regression with panel data, heterogenous program impact, distribution of public goods

1 Assistant Professor, University of Pittsburgh, and Associate Professors, Duke University, respectively. Corresponding author: Shanti Gamper-Rabindran. Email: shanti@gspia.pitt.edu. Phone: 412-648-8266 Fax 412-648-2605. Acknowledgments: We thank colleagues at IPEA, John Briscoe, Maureen Cropper, Daniel Kammen, Margaret Kosek, Grant Miller, Anita Millman, Kara Nelson, Narayan Sastry, Werner Troesken, Tara Watson, seminar participants at University of Pittsburgh, Yale School of Forestry and Environmental Studies, AERE Health and Environment summer workshop, the Population Association of America and the Allied Social Science annual meetings. Errors are ours.
1 Introduction

The Millennium Development Goals aim to reduce by two-thirds the under-five child mortality rate by 2015 from the base year 1990 (United Nations, 2005). In 2000, diarrhea caused approximately 22% of these deaths worldwide (Black et al, 2003). About 1.5 million child deaths (or 88% of those from diarrhea) are caused by ingestion of unsafe water, inadequate availability of water for hygiene, and lack of access to sanitation (Black et al, 2003). A proposed strategy to achieve the Millennium Development Goals of reducing child mortality is to improve access to safe drinking water. Indeed, the Brazilian government has announced its goal to achieve universal coverage for piped water (World Bank, 2003). These proposals raise an important policy question – can provision from piped water from the network, hereafter “piped water”, reduce the infant mortality rate (IMR). For those populations at greatest risk (i.e. in areas that suffer severe infant mortality rates) can this provision reduce infant mortality rates, or is the provision of piped water effective only when accompanied by complementary income-related inputs at the household or community levels?

In situations involving extreme inequality, it is possible for simple conditional mean estimates to mask the answers to these questions. Quantile estimation, which recovers the marginal impact of piped water on various quantiles of the conditional distribution of the IMR, can address this problem. Quantile regression is, however, not easily adaptable to dealing with problems of endogenous regressors. This presents a difficulty for most policy analyses, since policies are seldom applied randomly. When valid instruments are available, endogeneity can be addressed with instrumental variable quantile techniques (Abadie et al., 2002; Arias et al., 2001; Chernozhukov and Hansen, 2005, Khan

2 These figures are for the 42 countries with 90% of the worldwide under-5 deaths in 2000 (Black et al., 2003).
3 Our study is limited to the following measure: percentage of households that receive piped water from a network. We do not have data on the quality of that piped water. We also do not have data on the type of connection from the network to the home (i.e., whether through plumbing internal to the house or through standpipes external to the house). Furthermore, we do not evaluate the effectiveness of piped water interventions relative to other water-related interventions. Mintz et al. (2001) argue that “decentralized approaches to making drinking water safe, including point-of-use chemical and solar disinfection, safe water storage, and behavioral change merit far greater priority for rapid implementation.”
and Tamer, 2008). The practical problem is that there are often no good instruments for many policies. The usual statistical approach in mean regression is to exploit panel variation and estimate fixed effects to control for time invariant sources of correlated errors. However, this approach is not applicable using standard quantile techniques.⁴

Using a new approach to quantile regression with panel data developed by Chen and Khan (2007), we examine the impact of provision of piped water on the under-1 infant mortality rate at various quantiles of the conditional IMR distribution using panel data for 3568 census units in all Brazil. We describe the effect of the treatment on various quantiles of the outcome distribution, making no assumption about the joint distribution of the treated and untreated distributions. Our interpretation follows that of Abrevaya (2001) and Bitler et al. (2005).

We find that an increase of one percentage point in the number of households receiving piped water in the group of counties with poor development indicators in the period 1980-1991 causes a decline of 1.25 deaths per 1,000 live births at the 90th percentile of the conditional IMR, but a decline of only 0.54 deaths at the 10th percentile.⁵ The marginal effect at the mean (i.e., 0.72 deaths per 1,000 live births) turns out to provide a poor indication of the effect of water on much of the IMR distribution. The most important implication of this result is that the impact of a piped water provision policy is determined in large part by how those piped water connections are distributed. There is tremendous payoff to targeting water provision to the areas with the highest IMR (both conditional and unconditional). In practice, however, piped water interventions have tended to be in places with good indicators of development and which are low in the conditional IMR distribution.

Our paper makes two methodological contributions to the program evaluation literature in developing countries. First, by using novel quantile

---

⁴ Differenced regression cannot be applied in the quantile regression context, and simple fixed effects estimation suffers from incidental parameters bias unless the panel is very long in the time dimension.

⁵ These indicators are described in Section 4 and Section 5.2.
techniques, we examine whether the provision of piped water can reduce infant mortality in the upper tail of the conditional IMR distribution. A priori, it is unclear whether the provision of piped water, without sufficient complementary health inputs, will yield a reduction in IMR at these quantiles. Previous studies’ focus on the impact at the mean of the conditional distribution may obscure this policy-relevant heterogeneity.

Second, by applying panel data techniques to quantile estimation, we can estimate the impact of piped water on IMR while controlling for potential time invariant confounders. Areas with fewer piped water connections are also high IMR areas. These areas may suffer from systematic underreporting of infant deaths (Victora and Barros, 2001). At the same time, areas with more piped water connections are likely to benefit from other superior health inputs (these inputs are unobservables in our study, such as access to medical care, nutritional supplements, and public health infrastructure) (Jalan and Ravallion, 2003; Weinreb, 2001). Our estimates will not suffer from the downward bias arising from the systematic underreporting of deaths or the upward bias arising from these time invariant inputs. At present, only a few, albeit important papers, have applied quantile regression to program evaluation in developing countries (Djebbari and Smith, 2005) and fewer still have applied strategies to address time invariant confounders within the context of quantile regression.

While instrumental variables may not always be available, the proliferation of quality panel data means that our methodological approach can be widely applied to the evaluation of other programs that provide health inputs or other public goods in developing countries. Our task of evaluating the impact of piped water on IMR shares two key characteristics with the evaluation of programs that provide health inputs in developing countries, such as the provision of nutritional supplements or medical assistance to populations at risk. First, from a policy perspective, it is important to understand the impact of these programs on the subpopulations that are most at risk; if unobservables are important determinants

---

6 The quantile panel data technique we employ, like other fixed effect models, cannot correct the bias arising from time varying unobservables (Ahn, Lee, and Schmidt, 2001).
of the outcome variable, these subpopulations will tend to occupy the tails of the conditional outcome distribution. Mean impacts will fail to capture heterogeneous impacts across the conditional distribution. Second, the evaluation of these programs is complicated by their systematic placement in areas that receive superior health inputs. If these inputs are unobserved by the econometrician, they will cause an upward bias in the measurement of positive program impacts. At the same time, the systematic underreporting of outcome variables (e.g., mortality in higher mortality areas) (Victora and Barros, 2001), may attenuate the relationship between health inputs and mortality.

2 Piped water and infant mortality in Brazil

2.1 Infant mortality

Piped water supply reduces infant mortality directly by reducing the incidence of diarrhea that arises from the ingestion of contaminated water and food, and indirectly when caregivers are able to devote more time to childcare instead of water collection activities. Brazil serves as a case study for the impact of piped water on infant mortality for three reasons. First, diarrheal diseases are an important cause of infant mortality, accounting for 8% of infant death in Brazil in 1995-7 (Victora, 2001). In Northeast Brazil, the poorest area in the country, diarrhea accounted for 15% of infant mortality (Victora, 2001). Second, under-1 infants in Brazil are susceptible to water-borne diseases due to the relatively short duration of breastfeeding (Sastry and Burgard, 2005). Diarrhea is likely to increase when the infant is first exposed to supplemental liquids or solids, usually at ages below 1 year old (Sastry and Burgard, 2005). The 1989 Brazilian National Health and Nutrition Survey indicates that only 29.5% of infants aged 0-5 months were exclusively breastfed and 36.3% of those aged 0-23 months were breastfed (Senauer and Kassouf, 2000). In 1996, the Brazil-wide estimate of the duration for breastfeeding (both exclusive and supplemental) was 8.2 months.
Third, our results from Brazil, particularly the Northeast, are potentially transferable to other developing countries.

### 2.2 Patterns of piped water provision (1970-2000)

The provision of piped water by regions is tabulated in Table 1A. Specifically for this table, we classify counties as urban in a given year if 50% or more of their population live in urban areas. As evident from Table 1A, Brazil’s policies (described in detail in Gamper-Rabindran, Khan, and Timmins (2008)) have resulted in superior piped water coverage in urban counties situated in the more prosperous regions i.e., the Southeast and the South, but poor coverage in other less prosperous regions, i.e., the Northeast and the North. Coverage lags in rural counties across all regions, but is especially scant in the North and Northeast. In the 1970s, piped water coverage in the urban counties was low in the Southeast (51%), extremely low in the South (27%) and sparse the Center-West, North and Northeast (15-20%). Between 1970 and 1991, piped water coverage in urban areas grew to moderate levels in the Southeast and South (62-66%), to low levels in the Center-West and Northeast (39-44%), but remained extremely low in the North (27%). By 2000, coverage had grown to high levels in the South and Northeast (74-77%) and to moderate levels in the Center-West and Northeast (62-69%), while coverage lagged in the North (50%).

Piped water coverage in rural counties has lagged behind that in urban counties. In rural counties in the Southeast, South, and Center-West, piped water coverage grew from scant levels in 1971 (4-20%) to moderate levels (44-48%) by 2000. In contrast, the coverage in the rural counties in the North and Northeast, which has been sparse even as late as 1991 (13% and 19%), remains low in 2000 (31% and 42%).

---

7 The average duration of breastfeeding did not differ dramatically between the Northeast and the rest of Brazil (Sastry and Burgard, 2005).
3 Econometric issues in estimating the impact of piped water on IMR

3.1 Marginal effects of piped water – mean versus quantiles

We use quantile techniques to recover the marginal impact of piped water on various parts of the conditional IMR distribution. In contrast, previous studies of piped water have focused on the conditional mean of that distribution (Sastry, 1996; Merrick, 1985; Jalan and Ravallion, 2003). Only under the assumption that the marginal effect of piped water is a simple "common effect" or "location shift" will the impact at the mean be the same as the impact for the entire distribution (Heckman et al., 1997; Abadie et al., 2002). In other words, under the “common effect” assumption, the piped water intervention has the same impact on everyone with the same observed characteristics (Heckman et al., 1997).

In this section, we illustrate why estimating mean effects can be significantly different from estimating quantile effects. To keep the discussion simple, we focus on the cross-sectional case. In particular, consider the linearly heteroskedastic model: \( y_i = \beta_0 + x_i \beta_1 + x_i \psi \varepsilon_i \), where \( y_i \) measures the infant mortality rate in county \( i \) and \( x_i \) measures the percentage of households there with access to piped water. We assume for this discussion that \( \varepsilon_i \) is independent of \( x_i \) (although relaxing this assumption with panel data is a major focus of the rest of the paper). Let \( \mu_\varepsilon \) denote the mean of \( \varepsilon_i \) (i.e., zero) and let \( \rho_\theta \) denote the \( \theta \)th quantile of the \( \varepsilon_i \) distribution. The variance of \( \varepsilon_i \) is one.

The marginal effect associated with the conditional mean function (which would be estimated were we to use simple OLS) is of the form \( (\beta_1 + \psi \mu_\varepsilon) = \beta_1 \), whereas the marginal effect associated with the \( \theta \)th quantile is \( (\beta_1 + \psi \rho_\theta) \). The differences between these two measures will generally depend upon the skewness of the distribution of \( \varepsilon_i \). For example, if \( \psi \) is positive and the distribution of \( \varepsilon_i \) is skewed toward the right, then the marginal effect of \( x_i \) associated with the mean will exceed that associated with the median and the lower quantiles. On the other hand, if the distribution is skewed toward the left, the reverse will be true – marginal effects associated with the median and higher quantiles will exceed the marginal effect attained from OLS.
Papers on health inputs have shown that estimates at the mean may obscure heterogeneous impacts at the various quantiles of the conditional distribution. Moreover, the heterogeneity in the conditional distribution of the outcome variable is relevant for public policy. For example, Abrevaya (2001) finds that prenatal care in the US has a significantly higher impact at lower quantiles of the conditional distribution of infant birthweight than at the higher quantiles. Furthermore, he finds that the black-white differential in birthweight is larger at the lower conditional quantiles of birthweight.

Heterogeneity in the impact of piped water is relevant for policy decisions regarding piped water placement. On the one hand, targeting piped water to vulnerable households may improve their welfare significantly. Households or communities with low income typically have the fewest public resources for children’s health. In such cases, we would expect piped water to have greater protective effect among households or communities with lower incomes. On the other hand, targeting piped water to vulnerable households may be necessary but not sufficient to improve their welfare. In particular, their limited income or education may constrain their ability to exploit the benefits from piped water supply. In that case, water supply placement would need to be accompanied by other interventions (Jalan and Ravallion, 2003).

In exploring the impact of piped water on IMR, our study explores two types of heterogeneity that call for distinct policy responses: (1) heterogeneity along observable dimensions such as income and (2) heterogeneity due to unobserved factors. The policy response for the first type of heterogeneity is to target along observables such as income and education. The policy response to the second type of heterogeneity is more challenging. It would not be sufficient to simply consider income, education, and sewage in defining “vulnerable populations”. Instead, in their task of allocating water, policy-makers need to look for other factors (i.e., unobserved factors in our analysis) that make IMR high. In this paper, we also seek to determine the return to targeting these

---

8 Thomas and Strauss (1992) make this argument for maternal education.
unobservables in the placement of piped water, and we describe how such targeting might be accomplished.

The first type of heterogeneity can be explored by standard techniques in the literature, i.e., by allowing the marginal impact of water to vary by the income variable. However, a mean regression with interaction terms would not address the second type of heterogeneity. In contrast, quantile techniques allow us to explore the second type of heterogeneity. A priori, it is unclear whether the marginal effect of water is greater in higher or lower income communities. Similarly, a priori, it is unclear whether, controlling for observables such as income, education, and sewage network, the marginal effect of water is greater at higher or lower percentiles of the conditional distribution of the IMR.

Previous studies suggest a complex relationship between health status, water supply and socio-economic status. Shuval et al. (1981) propose a four-stage threshold-saturation model to explain the relationship between health status, water supply, and socioeconomic levels reported in several empirical studies with seemingly contradictory results (Esrey et al., 1985; Cairncross, 1991, 2008; Curtis et al., 1995). They propose that at the first stage, i.e., below a threshold of socioeconomic development, the provision of water does little to improve the health status of the community. Individuals have low disease resistance due to their extremely poor nutrition and personal hygiene and their exposure to multiple and simultaneous routes of disease transmission. The provision of water alone, which addresses only one route of disease transmission, does not have a strong impact on health. Shuval et al.’s (1981) argument echoes that of Briscoe (1984) – i.e., the improvements in drinking water supply in Matlab, Bangladesh did not cause major reductions in cholera incidence because complementary interventions were not undertaken to eliminate other important, albeit secondary, routes of cholera transmission (e.g., the ingestion of polluted water during bathing). Similarly, Esrey et al. (1992) find that water supply had a significant health impact only when accompanied by the presence of latrines in their study of infants in Lesotho.
At the second stage, above that threshold but below the saturation point, socioeconomic development improves the standard of living and reduces the exposure to infection (Shuval et al., 1981). At this level of socioeconomic development, communities have a strong health response to investments in water supply. At the third stage, as communities develop further, they move towards a saturation point, whereby improvements in water supply have only a small impact on health. At the fourth stage, beyond the saturation point, communities have reached high levels of socioeconomic development. Improvements to water supply would not cause further improvements in health status (Shuval et al., 1981). The practical problem in testing this theory is that it is not clear what variables should be used to define “socioeconomic development”. Our quantile approach allows us to measure the sensitivity of IMR to determinants of development not explicitly included in the analysis.

3.2 Non-random program placement and measurement error

Piped water is systematically placed in areas with superior health inputs (Jalan and Ravallion, 2003), giving rise to the econometric issue of non-random program placement. Additionally, the systematic underreporting of IMR in areas with greater infant mortality rates (Vicora and Barros, 2001, citing Simões, 1999) results in measurement error in the IMR variable. Details of both issues are discussed in Gamper-Rabindran, Khan, and Timmins (2008). To overcome the issue of non-random program placement, quantile studies from developed countries have been able to rely on experimental design such as in the evaluation of welfare reform or job training programs (Bitler et al., 2005 and 2006) or instrumental variables such as in evaluating the impact of childbearing on income9 (Abadie et al., 2002), the returns to education (Arias et al., 2001), and returns to job training programs (Chernozhukov and Hansen, 2005). In contrast, only a few quantile studies from developing countries have been able to rely on experimental design or instrumental variables. Djebbari and Smith (2005)

9 “Childbearing reduces the lower tail of the income distribution considerably more than other parts of the income distribution.” (Abadie et al., 2002).
use random assignment experimental data to examine the distributional impact of Mexico’s program of education, health, and nutrition (PROGRESA). They find that the program had a smaller impact on wealth and nutrition for households in the lower tail of the wealth and nutrition distribution.

A few studies, looking only at developed countries, have begun to explore the use of panel data in the context of quantile regressions. For example, Abrevaya and Dahl (2006) examine the impact of prenatal care and smoking on infant birthweight using panel data on maternally-linked births. They assume a correlated random effects model as is done here and in Chen and Khan (2007), but also impose the additional restriction of a linear structure on the individual specific effect. In this and other important policy contexts, randomized placements and instrumental variables are not readily available. In these situations, the use of panel techniques has the potential to correct the estimation bias from selective placement and systematic measurement error.

4 Data

We use newly available census data published by the Brazilian Institute for Economic Analysis (IPEA). These data are reported at the level of minimally comparable areas (MCA’s) for the years 1970, 1980, 1991, and 2000. Previously, census data were available at the municipio or county level, which is the smallest political division in Brazil (Alves and Beluzzo, 2004). Changes in county boundaries between the decades had limited the comparability of the census data. To overcome this limitation, IPEA created the MCA dataset, in which geographical units share common boundaries across the decades. The MCA boundaries correspond to county boundaries for those counties whose borders did not change between 1970 and 2000. For those counties that changed their borders between 1970 and 2000, neighboring counties were dissolved into one larger MCA. Data from households were then aggregated up to the MCA level for 1970, 1980, 1991, and 2000.
The MCA dataset divides Brazil into 3568 MCAs, a number which compares favorably with the 4500 counties in Brazil in 1998 (Mobarak et al., 2004) and 5560 in 2000 (Alves and Beluzzo, 2004). While the MCA dataset is imperfect in that it sometimes aggregates several counties which may differ in their policy and institutional context, we believe that this dataset represents the best demographic panel dataset currently available for Brazil. The finer resolution of the MCA data relative to other available Brazilian panel census data lessens the degree of within unit heterogeneity.

Table 1C presents summary statistics. The mean infant mortality rate declined from 125 deaths per 1000 live births in 1970, to 87 deaths in 1980, to 49 deaths in 1991, and to 34 deaths in 2000. At the same time, we see improvements in other development indicators. The percentage of households with piped water has increased fourfold from a mere 15% in 1970 to 62% by 2000. The percentage of households connected to the sewage network, starting from a lower baseline of 5% in 1970, has increased six-fold to 29% by 2000. Total fertility rate has more than halved from 5.9 births in 1970 to 2.8 births by 2000. Both the income-related Human Development Index and the education-related Human Development Index show improvement between 1970 and 2000.

5 Method
5.1 Estimation

Our dependent variable is the number of deaths of infants under one year of age per thousand live births. Reviews of active surveillance of developing areas and of studies published between 1990 and 2000 indicate that the under-1 age-group experiences the highest diarrhea specific mortality rates (Kosek et al, 2003). Our analysis focuses on all-cause infant mortality. Brazilian vital statistics

---

10 In the 1980s, Brazil had 4088 municipalities, with an average population of 29,800 and an average area of 2118 km² (Sastry, 1996). We drop one observation in our analysis because of missing values.

11 Potter et al. (2002) use the previous version of decennial data (terminating in 1991) that divides Brazil into 518 microregions. Another data source, the PNAD, suffers from municipio boundaries that are not consistent from one survey to another.
data (except when the information is specifically collected by researchers) are notoriously unreliable on cause-specific deaths, and the unreliability is worse in high mortality areas (Sastry and Burgard, 2002). By focusing on infant mortality, we avoid the potential bias inherent in studies that examine child health. Studies that use child health (e.g., height-for-weight scores) need to correct for the selection on surviving children in order to avoid underestimating the overall impact of piped water on child health (Lee et al., 1997). 12 We interpret the coefficient on piped water to capture the impact of piped water on infant mortality, typically through reduced risk of death from diarrheal diseases.

Our study is limited to the analysis of one aspect of the quantity of piped water. The definition for the water variable is the percentage of households with piped water from the general network. 13 As in Sastry (1996) we focus on households' source of water, i.e., from the network, and not on the type of connection. Our focus on this variable has two limitations. First, we are not able to provide separate estimates for piped water delivered to the household through external plumbing (e.g., communal standpipes) and for piped water delivered to the household through internal plumbing. Second, we do not study the quality of piped water. See Gamper-Rabindran, Khan, and Timmins (2008) for additional discussion of the mode of water provision and water quality.

In addition to piped water, we include several covariates to account for other time-varying factors that influence the IMR. Income-based Human Development Index (which we refer to simply as “income”) is added as a covariate as higher income levels are associated with improved chances for child survival 14 (Sastry, 1996, citing Merrick. 1985, Thomas et al., 1990, and Victora et al., 1986).

---

12 Child health data from the PNAD, 1996 Demographic and Health Survey and 1989 the Brazilian Health and Nutrition fail to provide municipal-level information. Cause-specific vital statistics data are not publicly available for all of Brazil.

13 The IPEA definition is “numero domicilios com água canalizada de rede geral”. We divide this figure by the total number of households in that MCA.

14 The definition for HDI_income = ln (observed value of RFPC) – ln (lower limit of RFPC) / [ln (upper limit of RFPC)-ln(lower limit of RFPC) where RFPC is the family per capita income.
Our regression model should control for improved sanitation,\(^{15}\) as the latter influences infant mortality rates (Habicht et al., 1988). We explicitly include in our model one type of improved sanitation – i.e., the percentage of households with network sewage.\(^{16}\) This type of sanitation is considered the only adequate kind in urban areas (UNICEF, 1997). Network sewage is the predominant method of improved sanitation in urban areas (WHO/UNICEF, 2000). We rely on the panel method of our analysis to control for cross-sectional variation in other types of improved sanitation in Brazil. Nevertheless, the panel method can adequately control for the cross-section variation in improved sanitation other than network sewage only to the extent that that variation is fairly constant over the decade. In rural areas, an estimated one-fifth to one-half of households use basic latrines (WHO/UNICEF, 2000). Use of pit latrines, while not as effective as the sewage network, is correlated with some declines in morbidity (Esrey et al, 1991). To address the shortcoming in sanitation data for rural areas, we repeat our regression analysis with the urban only sample.

Maternal education, by improving mother’s access to health-related information and her ability to make better use of health inputs, influences the reduction in the infant mortality rate (Sastry, 1996 citing Barrera, 1990, Rosensweig and Schultz, 1982, and Thomas et al., 1991). In the absence of women-specific education or literacy data, we use the education-based Human Development Index (which we refer to simply as “education”) variable. The education variable has been constructed by IPEA from a 2:1 weighting of the index for literacy rate and the index for school attendance rate.\(^ {17}\) As seen in Table 1B, while men’s and women’s literacy rates are positively correlated, one limitation in using the non-gender specific education variable proxy variable is the

---

\(^{15}\) Improved sanitation is defined by the World Health Organization as connection to a public sewer, connection to a septic system, a pour-flush latrine, a simple pit latrine or a ventilated improved pit latrine (United Nations, 2005).

\(^{16}\) The IPEA definition is “numero domicilios instalações sanitárias de rede geral”. We divide this figure by the total number of households in that MCA. Victora et al. (1986) and Victora et al. (1988), both cited in Sastry (1996), find that household toilet facilities are related very weakly to child mortality risks.

\(^{17}\) The HDI_education variable includes current schooling, which captures MCA-level investment in education of children. The index of literacy rate or the index of school attendance rate = (observed rate – minimum rate) / (maximum rate – minimum rate).
presence of some regional variation in the gap between men’s and women’s literacy rates.

Finally, our panel data procedure controls for county-specific time-invariant unobservables. One such time-invariant characteristic that influences infant mortality is the climate – greater seasonality in temperature and precipitation is associated with greater infant mortality from infectious diseases (Sastry, 1996). Some variables that vary in the cross-section and that we aim to control for using the panel procedure (e.g., access to healthcare and breastfeeding behavior), are not strictly speaking time-invariant. There are, however, no county-level data maintained on breastfeeding behavior. While there are data available that describe the number of doctors, nurses, and hospitals at the county level, we found that these variables had no explanatory power after controlling for the county effect, \( \alpha_i \).

The basic panel data model to be estimated is of the form:

\[
y_{i,t} = \alpha_i + x'_{i,t}\beta + \varepsilon_{i,t} \quad t = 1, 2
\]

where \( y_{i,t} \) denotes the under-1 infant mortality rate in county \( i \) and year \( t \), defined as the number of deaths for every 1000 live births before the end of the first year. \( x_{i,t} \) includes the percentage of households with piped water from the network, the percentage of households with sewage connection, the income variable, the education variable, and the interaction between income and the water supply variable.\(^{19}\)

\( \alpha_i \) denotes the (unobserved) county effect, which controls for time-invariant sources of unobserved heterogeneity. Without this control, we would expect piped water to be correlated with the error in (1), leading to biased estimates.

---

\(^{18}\)Victora et al. (1996) report that ORT played a larger role than income, education, and access to water in the sharp decline in infant deaths due to diarrhea in the 1980s. Nevertheless, Sastry and Burgard (2005) raise questions about this conclusion. “There is considerable uncertainty regarding trends in mortality by cause, because death registration is not complete and information on death certificates that are filed is often missing or inaccurate” (Sastry and Burgard, 2005).

\(^{19}\) Colinearity in these covariates makes it difficult to estimate their distinct effects when they are included within the same regression model.
estimates. Indeed, we show this to be the case with a series of cross-sectional regressions below. If unobserved determinants of IMR do not vary in a county over the course of a decade, the county effect will control for them non-parametrically. Similarly, measurement error in infant mortality may vary by county. As long as the rate of measurement error is stable over the course of a decade, \( \alpha_i \) will control for its impact on the reported IMR.

With a constant coefficient vector \( \beta \) and a mean zero restriction on the error term, the typical approach to identifying \( \beta \) with panel data is to estimate the first-differenced model:

\[
y_{i,2} - y_{i,1} = (x_{i,2} - x_{i,1})' \beta + (\varepsilon_{i,2} - \varepsilon_{i,1})
\]

by simply regressing the differenced dependent variable on the differenced covariates. Unfortunately, such an approach will not be valid in the quantile regression setting. To see why, we return to the basic model introduced by Koenker and Bassett (1978) and Koenker and Hallock (2001), which allowed marginal effects to vary by quantile. They considered a (cross-sectional) linearly heteroskedastic model of the form:

\[
y_i = \alpha_i + x_i' \beta + (x_i' \psi \varepsilon_i)
\]

which implies that the \( \theta \)th conditional quantile of the dependent variable has the following form:

\[
q_{\theta} = \alpha_i + x_i' \beta + x_i' \psi \rho_{\theta}
\]

\[
= \alpha_i + x_i' (\beta + \psi \rho_{\theta})
\]

\[
= \alpha_i + x_i' \beta_{\theta}
\]

where \( \rho_{\theta} \) denotes the \( \theta \)th quantile of the distribution of \( \varepsilon_i \). We now demonstrate that this model cannot carry through to the panel data model by first-differencing. In the linear heteroskedastic framework, differencing equation (3) yields:
Taking conditional quantiles of both sides of equation (5) yields:

\[ q_\theta(y_{i,2} - y_{i,1} \mid x_{i,1}, x_{i,2}) = (x_{i,2} - x_{i,1})' \beta + q_\theta(x_{i,2}' \epsilon_{i,2} - x_{i,1}' \epsilon_{i,1}) \]

Since the quantile and difference operators cannot typically be interchanged (unlike the mean and difference operators), the last term in the above expression is not equal to \((x_{i,2} - x_{i,1})' \psi \rho_\theta \). \(^{20}\)

We therefore apply the approach described in Chen and Khan (2007). In particular, we impose non-parametric structure on the county effect:

\[ \alpha_i = \phi(x_{i,1}, x_{i,2}) \]

Where \( \phi(\cdot) \) is an unknown function that allows for arbitrary dependence on the covariates.\(^{21}\) In particular, \( \phi(\cdot) \) expresses \( \alpha_i \) as a function of \( i \)'s covariates in both years \( t = 1, 2 \). This structure generalizes the typical random effects approach, which does not permit \( \alpha_i \) to depend upon covariates. It also generalizes approaches which impose parametric specification on \( \alpha_i \), such as

\(^{20}\) We also note that if we did not allow for the heteroskedastic component, \( x_{i,2} \psi \), then the quantile difference function would be a linear function of \( \beta \) plus an additive constant that varied with the quantile. In this restricted setting, marginal effects would not be allowed to vary across quantiles.

\(^{21}\) In practice, data limitations (in particular, a high degree of correlation between many of our regressors) will restrict us to using a second-order polynomial in this stage of the estimation. Depending upon the specifics of the application, this could be expanded to a higher-order polynomial or even a non-parametric bin estimator.
Chamberlain (1982) and Abrevaya and Dahl (2006). Consequently, we have the following functional form for the conditional quantile functions: 

\begin{equation}
q_\theta(y_{i,t} | x_{i,1}, x_{i,2}) = \phi(x_{i,1}, x_{i,2}) + x'_{i,t} \beta + x'_{i,t} \psi \rho_\theta
\end{equation}

This implies that the first differences in the conditional quantile functions are of the form:

\begin{equation}
q_\theta(y_{i,t} | x_{i,1}, x_{i,2}) - q_\theta(y_{i,t-1} | x_{i,1}, x_{i,2}) = \\
\phi(x_{i,1}, x_{i,2}) - \phi(x_{i,1}, x_{i,2}) + (x_{i,2} - x_{i,1})' \beta + (x_{i,2} - x_{i,1})' \psi \rho_\theta
\end{equation}

which, with some simplification, yields:

\begin{equation}
q_\theta(y_{i,t} | x_{i,1}, x_{i,2}) - q_\theta(y_{i,t-1} | x_{i,1}, x_{i,2}) = \\
(x_{i,2} - x_{i,1})' (\beta + \psi \rho_\theta)
\end{equation}

This implies an ability to estimate quantile-varying marginal effects. Of course, the above equations do not translate directly into a feasible estimation procedure since the conditional quantile functions, \( q_\theta(y_{i,t} | x_{i,1}, x_{i,2}) \) and \( q_\theta(y_{i,t-1} | x_{i,1}, x_{i,2}) \), are unknown. The approach can be implemented, however, by following a simple two-step procedure. First, non-parametrically estimate the conditional quantile

22 Chen and Khan (2007) show that, despite this generalization, there is no curse-of-dimensionality associated with estimating \( \beta \).

23 We can also allow for an additive unobserved term in our structure if we impose certain conditions such as independence of the regressors for both this term and \( \varepsilon_{i,t} \) as well as stationarity on \( \varepsilon_{i,t} \) on this term. Specifically, we can define \( \alpha_i \)

\[ \alpha_i = \phi(x_{i,1}, x_{i,2}) + \eta_i + \varepsilon_{i,t} \]

if we assume the "composite" error term, \( \eta_i + \varepsilon_{i,t} \), is distributed independently of the regressors and identically across individuals, as well as strongly stationary. In this way, the quantiles of the composite error are constants that vary across quantiles but do not depend on \( i \) nor \( t \). Equation (8) would then be of the form

\[ q_\theta(y_{i,t} | x_{i,1}, x_{i,2}) = \phi(x_{i,2}, x_{i,1})' + c_\theta + x'_{i,t} \beta + x'_{i,t} \psi \rho_\theta \]

where \( c_\theta \) is a constant that gets differenced out, maintaining the form of equation (9).
functions in (8), \( q_\theta(y_{i,t} \mid x_{i,t}, x_{i,t+1}) \) for \( t = 1, 2 \). It is important that the function \( \phi() \), which controls non-parametrically for the county fixed effect \( \alpha_i \), include data from both time periods. When estimating \( q_\theta(y_{i,t} \mid x_{i,t}, x_{i,t+1}) \), the equation also includes observables from period \( t \) in linear form. Denote these fitted values from each of these quantile regressions as \( \hat{q}_\theta(y_{i,t} \mid x_{i,t}, x_{i,t+1}) \). In the second step, we regress the differenced fitted values, \( \hat{q}_\theta(y_{i,t+1} \mid x_{i,t+1}, x_{i,t+2}) - \hat{q}_\theta(y_{i,t} \mid x_{i,t+1}, x_{i,t+2}) \), on the differenced regressors, \( (x_{i,t+2} - x_{i,t}) \). As seen in equations (9) and (10), the proxies for the county effects difference out, yielding an estimate of \( \beta_\theta \) — i.e., the marginal effect for the \( \theta \)th quantile. As discussed in Chen and Khan (2007), this procedure is very simple to implement, requiring little more than STATA or comparable statistical software.\(^{24}\)

We implement this panel data procedure separately for three time periods: 1970-1980, 1980-1991, and 1991-2000. For the weighted regressions, we weight the observations by the average county-level population over the two years. We also estimate unweighted regressions in order to check the robustness of our results. Finally, we use 2000 bootstrap simulations to recover standard errors for our estimates.

See Gamper-Rabindran, Khan, and Timmins (2008) for additional information on how to interpret the coefficients from quantile regressions.

5.2 Simulation – Marginal effects of piped water and averted infant deaths

After conducting the estimations described above, we simulate the policymaker’s expectation of averted infant deaths resulting from the additional provision of piped water. We make this calculation using the estimates from both

\(^{24}\) Since this approach attains identification off of variation in the regressors without varying the individual specific effect, it cannot be applied to estimate coefficients of time invariant regressors. This is also the case with standard fixed effect estimation. Nonetheless, the change in coefficients on time invariant variables (e.g., climate) can be estimated using our procedure by interacting them with time dummies.
the mean and quantile regression specifications. We apply these estimates to a simulated change of one percentage point in the number of households receiving piped water in each county.

Counties are grouped as being high or low in each of these four covariates, or development indicators: piped water, piped sewage, income, and education. We therefore have 16 groups of counties corresponding to 16 possible combinations of high or low values for the four indicators. We calculate for each group of counties their intra-group mean income. Next, for each group of counties, we calculate the intra-group distribution of the infant mortality rate. A county will therefore occupy the $\theta$th percentile of the conditional infant mortality rate distribution (i.e., within a group of counties that are similar in their four development indicators).

For each county in a given group, we calculate the marginal effect of piped water on its infant mortality rate (measured as deaths per 1000 live births) using estimates from the appropriate quantile regression and accounting for local conditions as captured by income. We then simulate the effect of an increase of one percentage point in the number of households with piped water.25

6 Regression results

Table 2 reports the results from the regressions weighted by county-level population.26 Results from the mean regression are in column 1, while those from the quantile regressions are in columns 2 to 10. Panels A, B, and C present results from the regressions for 1970-1980, 1980-1991, and 1991-2000, respectively. To calculate the marginal impact of piped water, we use the coefficients from the weighted quantile panel data regressions for 1970-1980,

---

25 For counties whose intra-group IMR is below the 10th percentile, we use the estimates from the 10th quantile regression. For counties whose intra-group IMR is between the 10th and 20th percentile, we use the estimates from the 20th percentile regression, and so on. We use the estimates from the 90th percentile for counties whose intra-group IMR is between the 80th and 90th percentiles, as well as for counties whose intra-group IMR is above the 90th percentile.

6.1 Marginal impact of piped water

We report the impact of a one percentage point increase in the number of households with piped water supply. As seen in Table 2, the coefficients for water and the interaction term between water and income for the years 1970-1980 and 1980-1991 are generally statistically significant at or below the 10% level. For the estimates in 1991-2000, these coefficients are statistically significant only in the regressions at the upper quantiles.

We find four main results for piped water. First, our results are consistent with Shuval et al.’s (1981) theory. Piped water has a sizable impact in reducing IMR after counties exceed a minimal threshold of socioeconomic development, but it has little impact after counties cross a saturation threshold. Consider Figure 1, which plots these impacts in counties that measure low in all their development indicators. In 1970-1980, the increased piped water supply reduced IMR by 0.61 to 0.82 deaths per 1000 live births. By 1980-1991, the counties at the upper tail of the conditional IMR distribution have exceeded the threshold of development, and we see piped water having a very strong impact. At the 80th to the 90th conditional quantiles of the IMR distribution, piped water reduced IMR by 1.25 to 1.28 deaths per 1000 live births. These reductions are sizable when compared to the mean of 86.8 and 49.2 deaths per 1000 live births in 1980 and 1991, respectively. In that same time period, we see some evidence of counties in the lower tails of the conditional IMR distribution moving towards

---

26 The results from the un-weighted regressions are similar to those from the weighted regressions and are available on request from the authors.
27 Results using the intra-group mean from 1980, 1991, and 2000 are qualitatively similar.
28 We describe reductions in infant mortality resulting from a one percentage point increase in households with piped water. In practice, the mean increase in households with piped water is 8.9 percentage points between 1970 and 1980, 18.2 percentage points between 1980 and 1991, and 20.2 percentage points between 1991 and 2000.
29 As discussed above, each year we separate counties into sixteen groups that rank similarly in their development indicators. For brevity, we present two graphs only, the first, from the group of counties that rank low in their development indicators and the second, from the group of counties that rank high in at least two of their development indicators.
the saturation point for the impact of water. At the 10th and 20th quantiles of the conditional IMR distribution, increased piped water reduced IMR by 0.54 to 0.55 deaths per 1000 live births. Finally, corresponding to the saturation threshold, by 1991-2000 we see that piped water has very little impact at any point across the conditional IMR distribution. By this point, increased piped water reduced IMR by only 0.03 to 0.21 deaths per 1000 live births.

Second, we find that piped water has a larger impact in reducing IMR in counties that rank low in their development indicators than in counties that rank high, particularly in 1980-1991. Consider Figures 1 and 2, which plot these impacts in counties that measure low in all of their development indicators and high in at least two of their indicators, respectively. In the first set of counties, a one percentage point increase in piped water reduces IMR by 0.54 to 1.28 deaths per 1000 live births. In contrast, in the second set of counties, the increase in piped water reduces IMR by only 0.36 to 0.72 deaths per 1000 live births. This pattern of piped water having a stronger protective effect in counties that measure low in observable development indicators than in those that measure high hold for most of the other group of counties as well.30

Third, we find evidence that the impact of piped water varies across the conditional IMR distribution. In particular, in the group of counties that measure similarly low or similarly high in their development indicators, piped water exerts a stronger protective effect at the upper tail of the conditional IMR distribution (particularly in 1980-1991). Consider counties that measure low in all of their development indicators. As seen in Figure 1, additional piped water reduces IMR by 1.25 deaths per 1000 at the 90th percentile but by only by 0.55 deaths per 1000 live births at the 10th percentile of the conditional IMR. This pattern of a stronger protective impact of water in the upper quantiles of the conditional IMR distribution in 1980-1991 is also evident in the group of counties that measure high in their development indicators. As seen in Figure 2, the additional piped water reduces IMR by 0.68 deaths per 1000 at the 90th percentile, but by 0.36 deaths per 1000 live births at the 10th percentile of the conditional IMR.

30 Figures for these other groups of counties are available on request from the authors.
Fourth, we find that (particularly in 1981-1990) the estimates from the mean panel regression model severely understate the protective impact of piped water for the populations occupying the upper quantiles of the conditional IMR distribution. The mean estimates suggest that a one percentage point increase in the number of households with piped water reduces IMR by 0.72 deaths per 1000 live births. In contrast, for the group of counties that measure low in their development indicators, the quantile panel model finds that an increase in piped water reduces IMR by 1.25 deaths per 1000 live births at the 90th conditional quantile of the IMR.

Results describing similar marginal effects of income can be found in Gamper-Rabindran, Khan, and Timmins (2008). The positive interaction effect between piped water and income suggest that these two inputs into infant health are substitutes. Moreover, the larger size of the coefficients at the upper quantiles suggests that income can be more effectively used as a substitute for water at those upper quantiles.

We also examine a naïve cross-sectional specification. Estimation results can be found in Gamper-Rabindran, Khan, and Timmins (2008). The association between a one percentage point increase in the number of households with piped water and the infant mortality rate at the θth quantile is plotted in Figure 3. The cross sectional estimates, particularly in 1970 and 1980, indicate that greater provision of piped water is correlated with larger infant mortality rates. We see a particularly strong correlation between water supply and increased mortality at the higher conditional quantiles of the IMR, although the size of the bias diminishes in the latter years of the analysis. This counterintuitive result is likely to be an artifact from the systematic underreporting of infant mortality rates in areas on the upper tails of the IMR distribution that tend to receive less water.

6.2 Other variables: Sewage

The marginal impact of sewage at the θth quantile is given by \( \beta_{θ,\text{sewage}} + \beta_{θ,\text{water} \times \text{sewage} \times \text{WATER}} + \beta_{θ,\text{water} \times \text{income} \times \text{INCOME}}, \) where WATER and INCOME represent the intra-group mean percentage of households with piped...
water supply and the intragroup mean income. Table 2 indicates that this impact is not statistically significant. These results stand in contrast with our earlier results on piped water.

Our results are consistent with Victora et al (1988), in a population-based study in Porto Alegro and Pelotas, Brazil, which finds that piped water supply is associated with reduced infant mortality, but no association is detected for measures of sanitation facilities (flush toilets or pit latrine). In contrast, Barreto et al. (2007) find that a city-wide sanitation program in Salvador, Brazil reduced diarrheal prevalence by 22% on average and by 43% in the poorest neighborhoods.\(^ {31, 32}\)

Nevertheless, our results should be treated with caution in light of two limitations in our study. First, as our variable measures networked sewage (i.e., only one type of sanitation facilities), our study provides limited information on the potential impact of sanitation facilities on IMR. Second, we have little cross-sectional variation in our explanatory variable (sewage \(_{t2}\)-sewage \(_{t1}\)), so we have limited ability to discern related variation in our dependent variable, (IMR \(_{t2}\) – IMR \(_{t1}\)). Thus, the corresponding estimated coefficient has a large standard error.

### 6.3 Other variables: Education\(^ {33}\)

Our results indicate that education is more effective at reducing IMR where the overall circumstances are worse. The marginal impact of education is indeed larger in 1970-1980 and 1980-1991 and its asymmetry is more pronounced in these years than in the latter years. In 1970-1980, a 0.01 increase in education reduces infant mortality by 1.8 deaths in the 10\(^{th}\) percentile and by almost one-and-a-half times that amount (i.e. 2.8 deaths) at the 90\(^{th}\) percentile. In 1980-1991, the gap is larger with 0.74 avoided deaths at the 10\(^{th}\)

\(^{31}\) The sanitation program aimed to increase the number of households with an adequate sewer system from 26% to 80%. The program focused on sewage connection, but included some improvements in water supply (Barreto et al. 2007).

\(^{32}\) The contrasting results between diarrheal prevalence and mortality may be caused by care-seeking behavior, case management, and nutritional status (Caincross, pers. comm. May 21 2008).
percentile and 2.0 deaths at the 90th percentile. In 1991-2000, the gap declines to 0.77 avoided deaths at the 10th percentile and 1.2 deaths at the 90th percentile. In all three decades, the mean estimate falls between the estimates from the 10th and 90th quantile regressions.

### 6.4 Other models: Urban counties only

As described earlier, our basic model includes only one type of sanitation (i.e., network sewage), which is considered the only adequate sanitation method for urban areas. Our model therefore omits other types of sanitation, such as pit latrines, that are considered to be adequate in rural areas. To address this limitation, we re-run our analysis restricting our sample to only urban counties. We define urban counties to be those with 50% or more of their households living in urban areas.\(^{34}\)

Our results (discussed in Gamper-Rabindran, Khan, and Timmins (2008)), particularly from 1980-1991, indicate that additional piped water supply has a sizable impact in reducing the infant mortality rate. Figures 4 and 5 show a comparison of the marginal impact of water estimated using the full sample and that estimated using the urban only sample. As seen from these figures, the marginal impact of water remains sizable for the urban sample, though it is smaller in magnitude than in the full sample. It may well be that, in our earlier estimates, the failure to control adequately for latrines in rural areas led us to attribute too much of the impact in reduction of IMR to piped water supply. Nevertheless, our results from the restricted urban only sample continue to show the four patterns observed earlier. First, the marginal impact of water follows the pattern suggested by Shuval et al. (1981). Second, comparing Figure 4 and 5, we see that piped water has a stronger effect in the set of counties that measure low in their development indicators. Third, piped water has a stronger protective

---

\(^{33}\) Gender-specific education information are not available at the MCA-level. Therefore, we are not able to examine the question of the interaction between maternal education and piped water (Barrera, 1990a).

\(^{34}\) Urbanized counties in the 1970-80, 1980-1991, and 1991-2000 panels are defined based on 1970, 1980, and 1991 values. Results (available from the authors upon request) are similar using a 70% urbanization cutoff.
effect in the upper tails of the conditional IMR distribution. Fourth, we find that the mean panel regression model underestimates the protective impact of income for the populations occupying the upper quantiles of the conditional IMR distribution. While the mean estimate suggests that a one percentage point increase in the number of households with piped water reduces IMR by 0.63 deaths, the quantile estimates at the 90th percentile suggest a reduction of 0.87 deaths per 1000 live births.

6.5 Other models: Under-5 child mortality rate

We extend our analysis to the under-5 child mortality rate. Our analysis is limited to one panel model (i.e., 1991-2000) because the under-5 child mortality rates at the MCA level are available only for those two census years. Our results (discussed in Gamper-Rabindran, Khan, and Timmins (2008)) indicate that piped water did not have a statistically significant impact on under-5 child mortality. Recall that piped water has a strong protective impact on under-1 mortality in 1980-1991 but its effects are attenuated by 1991-2000.

Our results showing that piped water has a stronger protective impact on infants than on children under the age of five are consistent with those in the previous literature. Several studies document a negative association between the presence of piped water and infant mortality, but no statistically significant association between that presence and child mortality (Rutstein, 2000; Abou-Ali, 2003; Woldemicael, 2000). Butz et al. (1984) suggests the following

35 A large number of studies, reviewed in Fewtrell et al. (2005), examine the relationship between water and sanitation provision and diarrhea incidences among infant and children aged 0-60 months.
36 Rutstein (2000), in her meta-analysis of the 89 DHS surveys between 1986 and 1998, finds that a statistically significant negative association between piped water and the mortality of children aged 1, but no statistically significant association between piped water and the mortality of children aged 1-4. Abou-Ali (2003), examining the 1995-6 Egyptian Demographic and Health Survey (DHS), finds a statistically significant negative association between presence of piped water (whether in residence or in the neighborhood) and post-neonatal mortality (2-12 months) but no statistical significant association between piped water and child mortality (12-60 months). Woldemicael (2000) finds, in the 1995 Eritrean DHS, a statistically significant negative association between good household environment (neither piped water nor flush toilet or only one of the two) and neonatal mortality (1-13 months) but no statistically significant association between piped water and child mortality (14-60 months).
explanation for this age pattern: “As infants mature (i.e., their immunity systems mature), they become less susceptible to enteric pathogens. Infant maturation explains why, in spite of increasing exposure with age to the environment and its pathogens, improved water and sanitation prevent fewer deaths in the later months.”

6.6 Averted deaths

Section 1 of Table 3 shows the number of averted deaths as a result of an increase of one percentage point in the number of households with piped water in each county. Section 2 shows the number of averted deaths as a result of an increase of 0.01 in income. Calculations using the quantile panel data procedure are tabulated in column (1) while those using the mean fixed effect regressions are tabulated in column (2). The difference between the two columns reflects the asymmetry of the distribution of marginal effects in each decade. There is a strong right-skew pulling up the mean marginal effect of both water and income estimated in the 1980-1991 regression. In contrast, the opposite skew pulls down the mean effect for water estimated in the 1970-1980 regression. By the period 1991-2000, aggregate effects (particularly for water) are very similar under both measures. These results emphasize the fact that it is difficult to predict which direction asymmetric marginal effects will pull mean estimates, highlighting the role played by quantile estimation in recovering the full distribution of effects.

6.7 Cost-effectiveness of piped water provision

The more cost-effective intervention achieves the same policy goal, i.e., averts the same number of infant deaths, at lower costs. We provide a back-of-the-envelope calculation on the relative cost-effectiveness of piped water provision in the areas with high and low conditional IMR. We estimate the costs of providing piped water in order to avert one infant death, respectively, in these two areas. We use the average costs of piped water provision from World Bank
Brasilia, i.e., BRL 2000 per household.\textsuperscript{37} The drawback of this exercise, i.e., the use of \textit{average} costs of piped water provision, can be rectified in future analysis should location-specific costs information become available. We use the estimates from our regression analysis describing the impact of a one percentage point increase in households with piped water on averted infant deaths per 1000 live births (i.e. 1.25 and 0.55 per 1000 live births in the 90\textsuperscript{th} and 10\textsuperscript{th} conditional IMR areas, respectively, within counties with poor development indicators). Using these estimates and the estimated total births in those counties, we calculate the total number of averted infant deaths resulting from this intervention. We also calculate the costs of providing piped water to the additional households. Finally, we calculate the cost effectiveness of the piped water provision. We conclude that providing piped water to these high conditional IMR areas is more cost-effective than similarly providing it to low conditional IMR areas: BRL 1010 (US$630) versus BRL 3240 (US$2010) per averted infant death, respectively.

\section*{7 Discussion and Policy Implication}

For those populations at greatest risk, can the provision of piped water reduce the infant mortality rate or are complementary inputs such as income and other public health infrastructure required? Our results are consistent with Shuval et al.’s (1981) threshold-saturation hypothesis, in which the relationship between water supply and IMR varies with changing socioeconomic levels. Assuming that our differencing procedure can adequately control for correlated unobservables and measurement error, we find that water has a small effect in the most undeveloped places (i.e., when we look at the high conditional quantiles in 1970-80). As counties start to develop (i.e., the higher quantiles in 1980-91), the protective effect of water on IMR starts to rise rapidly. As counties become more developed (i.e., low quantiles in 1980-91) the protective effect of water

\footnotesize{\textsuperscript{37} Dr. Marcos Thadeu Abicalil, Senior Water and Sanitation Specialist, World Bank Brasilia Office (pers. comm., 2008).}
declines. Finally, when very developed (i.e., low quantiles in 1991-00), the effect of water on IMR is very small.

In 1980-1991, the marginal impact of piped water is greatest in those counties with poorest performance in their observable development indicators. For counties with poor development indicators averted deaths at the 90th percentile is 1.25 per 1000 live births, while for counties with good development indicators only 0.68 are averted. In addition, among those counties that share common development indicators, particularly in 1980-1991, we find that piped water exerts a stronger protective effect in those counties that occupy higher positions in the conditional IMR distribution (i.e., counties that are worse in unobservable development indicators).

Our results therefore show that (1) piped water provision can cause a significant reduction in the IMR (when accompanied by a basic level of other public health inputs); and (2) the impact of a piped water provision policy is determined in large part by how those piped water connections are distributed. Ignoring costs of provision, our results suggest that, from the perspective of health outcomes, new piped water resources should be targeted to the most disadvantaged communities.38, 39

What can policy-makers learn from our study? In addition to recognizing the role of particular observed characteristics that influence the effectiveness of piped water in reducing IMR, policymakers also need to take into account the role of unobserved characteristics – i.e., characteristics that cannot be easily summarized with available data. In practice, policymakers can control for these unobservables by implementing the following strategy, which allows one to recover their distribution up to a scale and location normalization. Reconsider

38 Policymakers may consider other factors in piped water placement such as population density. We acknowledge that the provision of piped water may be cheaper in areas with good development indicators and/or low conditional IMR. These locations may already have a minimal level of existing infrastructure. New outlays of pipelines may have to be undertaken in disadvantaged areas.

39 “It has been suggested that piped water disproportionately benefits the better-off people of a village” (Mohan, 2005). Further interventions would have to be undertaken to overcome social constraints and connection costs that prevent the vulnerable households from accessing the network.
equation (10), from which we know that the estimated value of $\beta$ is equal to $\beta + \psi \rho_\theta$. Making the location normalization that the median of $\varepsilon_{i,t}$ is zero (i.e., $\rho_{50} = 0$), we immediately identify $\beta$ (i.e., $\beta = \beta_{50}$). Next, making a scale normalization (e.g., $\rho_{75} = 1$) we can further identify $\psi$ (i.e., $\psi = \beta_{75} - \beta$). With estimates of $\beta$ and $\psi$, we can then recover the distribution of $\varepsilon_{i,t}$ (i.e., different values of $\rho_\theta$) from equation (9) simply by observing the conditional quantiles of $y_{i,t}$. The resulting ranking of residuals can then be used to help determine where to target piped water interventions. Our results suggest that there will be statistically and economically significant, policy-relevant differences in the effectiveness of piped water over these indicators of unobservable determinants, with the biggest effects coming high in their distribution.

Methodologically, these results highlight the importance of applying the quantile regression framework to recover the marginal effects of water at various parts of the conditional distribution of the IMR. The marginal effects at various parts of the conditional IMR distribution differ substantially from those at the mean of the distribution. Indeed, focusing on the mean of the distribution can lead to an underestimate of the potential impact of piped water intervention in higher percentiles of the conditional IMR distribution. Our results for piped water intervention correspond with the growing literature on the heterogeneity of program impacts across the quantiles of the conditional distribution of the outcome variable and the insufficiency of mean estimates to represent this policy-relevant heterogeneity.

Quantile estimation for the evaluation of policy is, however, quite difficult. Policies are not often allocated randomly, and good instruments may not be available. Traditional quantile regression is not generally feasible in the panel data context. In contrast, our quantile panel data approach can be widely applied to the evaluation of other programs that provide health inputs or public goods in developing countries. This method allows policymakers to understand the impact of these programs on the subpopulations that are most at risk, and these subpopulations tend to occupy the tails in the conditional distribution. Amidst the
scarcity of random assignment and viable instruments, but with the growing availability of panel data in developing countries, the panel data approach provides a promising strategy to address the issue of bias arising from unobservables (albeit only time invariant ones) within the context of quantile regressions.
Bibliography


Alves D, Belluzzo W. Infant mortality and child health in Brazil. Economics and Human Biology 2004; Special Issue 2(3); 391-410.


Black RE, Morris SS, Bryce J. Where and why are 10 million children dying every year? Lancet 2003; 361 (9376); 2226-2234.
Briscoe J. Intervention studies and the definition of dominant transmission routes. American Journal of Epidemiology 1984; 120(3); 449-455.


Cairncross S. Handwashing with soap – a new way to prevent ARIs? Tropical Medicine and International Health 2003; 8(8); 677–679.


Chernozhukov V, Hansen C. An IV model of quantile treatment effects. Econometrica 2005; 73(1); 245-261.


Djebbari H, Smith J. Heterogeneous Program Impacts in PROGRESA, mimeo, Laval University and University of Michigan working paper; 2005.


Fewtrell L, Kaufmann R, Kay D, Enanoria W, Haller L, Colford Jr J. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed
countries: a systematic review and meta-analysis. The Lancet Infectious Diseases 2005; 5(1); 42–52.


Heckman J, Smith J, Clements N. Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. In: The Review of Economic Studies; Special Issue: Evaluation of Training and Other Social Programmes 1997; 64(4); 487-535.


Mohan P. Inequities in coverage of preventive child health interventions: The Rural Drinking Water Supply Program and the Universal Immunization Program in Rajasthan, India. American Journal of Public Health 2005; 95(2); 241-244.

Potter JE, Schmertmann CP, Cavenaghi SM. Fertility and Development: Evidence from Brazil. Demography 2002; 39(4); 739-761.

Rosenzweig MR, Schultz TP. Child mortality in Colombia: individual and community effects. Health Policy and Education 1982; 2(3-4); 305-348.


Sastry N. Community characteristics, individual and household attributes, and child survival in Brazil. Demography 1996; 33(2); 211-229.


Victora CG. Potential interventions to improve the health of mothers and children in Brazil. Revista Brasileira de Epidemiologia 2001; 4(1); 3-69 (In Portuguese).

Victora CG, Barros FC. Infant mortality due to perinatal causes in Brazil: trends, regional patterns and possible interventions. Sao Paolo Medical Journal 2001; 191(1); 33-42.

Victora CG, Olinto MTA, Barros, FC, Nobre LC. Falling diarrhea mortality in Northeastern Brazil: did ORT play a role? Health Policy and Planning 1996; 11(2); 132-141.


### Table 1A: Provision of piped water by region (1970-2000)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Urban</td>
<td></td>
<td>Rural</td>
<td></td>
<td>Urban</td>
<td></td>
<td>Rural</td>
<td></td>
<td>Urban</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>Northeast</td>
<td>15</td>
<td>2</td>
<td>29</td>
<td>8</td>
<td>39</td>
<td>19</td>
<td>62</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMR</td>
<td>Northeast</td>
<td>192</td>
<td>180</td>
<td>138</td>
<td>136</td>
<td>73</td>
<td>78</td>
<td>50</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>North</td>
<td>17</td>
<td>4</td>
<td>22</td>
<td>8</td>
<td>27</td>
<td>13</td>
<td>50</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMR</td>
<td>North</td>
<td>118</td>
<td>113</td>
<td>72</td>
<td>70</td>
<td>54</td>
<td>59</td>
<td>40</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>Center-West</td>
<td>20</td>
<td>4</td>
<td>21</td>
<td>7</td>
<td>44</td>
<td>25</td>
<td>69</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMR</td>
<td>Center-West</td>
<td>93</td>
<td>92</td>
<td>60</td>
<td>60</td>
<td>33</td>
<td>36</td>
<td>25</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>South</td>
<td>27</td>
<td>8</td>
<td>43</td>
<td>16</td>
<td>62</td>
<td>32</td>
<td>74</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMR</td>
<td>South</td>
<td>84</td>
<td>85</td>
<td>55</td>
<td>53</td>
<td>32</td>
<td>31</td>
<td>19</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>Southeast</td>
<td>51</td>
<td>20</td>
<td>52</td>
<td>23</td>
<td>66</td>
<td>33</td>
<td>77</td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMR</td>
<td>Southeast</td>
<td>96</td>
<td>98</td>
<td>60</td>
<td>62</td>
<td>31</td>
<td>40</td>
<td>22</td>
<td>32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Water denotes the percentage of households in a county with piped water.
IMR denotes the under-1 infant mortality rate in deaths per 1000 live births.
Urban counties in a given year are those with 50% or more of their population living in urban areas.

### Table 1B: Literacy rates by region

<table>
<thead>
<tr>
<th>Year</th>
<th>1981</th>
<th></th>
<th>1991</th>
<th></th>
<th>2000</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Northeast</td>
<td>42.3</td>
<td>40.9</td>
<td>38.9</td>
<td>34.1</td>
<td>26.3</td>
<td>22.4</td>
</tr>
<tr>
<td>North</td>
<td>14.2</td>
<td>16.7</td>
<td>11.8</td>
<td>13.1</td>
<td>11.5</td>
<td>11.0</td>
</tr>
<tr>
<td>Central-west</td>
<td>19.9</td>
<td>22.7</td>
<td>16.6</td>
<td>17.1</td>
<td>10.2</td>
<td>10.3</td>
</tr>
<tr>
<td>South</td>
<td>13.5</td>
<td>17.7</td>
<td>10.1</td>
<td>13.1</td>
<td>6.4</td>
<td>7.7</td>
</tr>
<tr>
<td>Southeast</td>
<td>12.5</td>
<td>17.6</td>
<td>9.6</td>
<td>13.0</td>
<td>6.7</td>
<td>8.3</td>
</tr>
</tbody>
</table>

**Source:** IPEA region-level data
<table>
<thead>
<tr>
<th>Table 1C: Summary statistics</th>
<th>Year</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infant mortality rate (in deaths per 1000 live births)</td>
<td>1970</td>
<td>125.3</td>
<td>52.7</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>86.8</td>
<td>45.2</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>49.2</td>
<td>24.4</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>33.7</td>
<td>18.1</td>
</tr>
<tr>
<td>Percentage households with piped water (%) (Water)</td>
<td>1970</td>
<td>15.1</td>
<td>19.5</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>24.0</td>
<td>21.8</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>42.2</td>
<td>23.9</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>62.4</td>
<td>20.5</td>
</tr>
<tr>
<td>Percentage households with sewage connections (%) (Sewage)</td>
<td>1970</td>
<td>5.3</td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>10.6</td>
<td>19.1</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>18.0</td>
<td>26.6</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>29.4</td>
<td>30.4</td>
</tr>
<tr>
<td>Human Development Index Income (income)</td>
<td>1970</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>0.54</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>0.56</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>0.61</td>
<td>0.10</td>
</tr>
<tr>
<td>Human Development Index Education (education)</td>
<td>1970</td>
<td>0.40</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>0.47</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>0.65</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>0.78</td>
<td>0.09</td>
</tr>
<tr>
<td>Total fertility rate</td>
<td>1970</td>
<td>5.9</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>4.5</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>3.6</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>2.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Female population</td>
<td>1970</td>
<td>12,796</td>
<td>68,110</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>16,366</td>
<td>93,898</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>20,322</td>
<td>110,204</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>23,542</td>
<td>123,214</td>
</tr>
<tr>
<td>Population</td>
<td>1970</td>
<td>25,460</td>
<td>132,485</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>32,534</td>
<td>182,528</td>
</tr>
<tr>
<td></td>
<td>1991</td>
<td>40,138</td>
<td>211,990</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>46,418</td>
<td>235,553</td>
</tr>
</tbody>
</table>
Table 2: The influence of piped water on infant mortality rates: panel regression weighted by county-level population

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>water</td>
<td>-0.48 **</td>
<td>-0.81 **</td>
<td>-0.75 **</td>
<td>-0.69 **</td>
<td>-0.75 **</td>
<td>-0.83 **</td>
<td>-0.79 **</td>
<td>-0.94 **</td>
<td>-0.91 **</td>
<td>-0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(0.17)</td>
<td>(0.17)</td>
<td>(0.15)</td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(0.16)</td>
<td>(0.16)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>sewage</td>
<td>0.14</td>
<td>0.33</td>
<td>0.44</td>
<td>0.23</td>
<td>0.38</td>
<td>0.38</td>
<td>0.46 *</td>
<td>0.41</td>
<td>0.20</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.11)</td>
<td>(0.30)</td>
<td>(0.28)</td>
<td>(0.27)</td>
<td>(0.27)</td>
<td>(0.27)</td>
<td>(0.28)</td>
<td>(0.29)</td>
<td>(0.48)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9.90)</td>
<td>(21)</td>
<td>(21.6)</td>
<td>(21)</td>
<td>(22)</td>
<td>(23)</td>
<td>(25.5)</td>
<td>(28.9)</td>
<td>(30.5)</td>
</tr>
<tr>
<td>water x income</td>
<td>0.46 **</td>
<td>0.90 **</td>
<td>0.68 **</td>
<td>0.68 **</td>
<td>0.76 **</td>
<td>0.80 **</td>
<td>0.77 **</td>
<td>1.15 **</td>
<td>1.04 **</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(0.22)</td>
<td>(0.21)</td>
<td>(0.18)</td>
<td>(0.18)</td>
<td>(0.19)</td>
<td>(0.20)</td>
<td>(0.22)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>sewage x income</td>
<td>-0.08</td>
<td>-0.16</td>
<td>-0.48 *</td>
<td>-0.24</td>
<td>-0.37</td>
<td>-0.47 *</td>
<td>-0.44 *</td>
<td>-0.43</td>
<td>-0.21</td>
<td>-0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.10)</td>
<td>(0.30)</td>
<td>(0.27)</td>
<td>(0.26)</td>
<td>(0.25)</td>
<td>(0.24)</td>
<td>(0.26)</td>
<td>(0.28)</td>
<td>(0.35)</td>
</tr>
<tr>
<td>water x sewage</td>
<td>-0.003 **</td>
<td>-0.012 *</td>
<td>-0.002 *</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.008</td>
<td>-0.007</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.001)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.008)</td>
</tr>
</tbody>
</table>

Notes: No. obs. 3568. ** statistically significant at the 5% level. * statistically significant at the 10% level.
Standard errors for quantile regressions are from 2000 bootstrap repetitions.
Table 2 (continued) : The influence of piped water on infant mortality rates: panel regression weighted by county-level population

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>water</td>
<td>-1.36 **</td>
<td>-0.76 **</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>sewage</td>
<td>0.32 **</td>
<td>-0.31</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.31)</td>
</tr>
<tr>
<td></td>
<td>(4.48)</td>
<td>(5.52)</td>
</tr>
<tr>
<td></td>
<td>(5.51)</td>
<td>(7.57)</td>
</tr>
<tr>
<td>water</td>
<td>1.26 **</td>
<td>0.76 **</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>sewage</td>
<td>-0.19</td>
<td>0.50 *</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>water</td>
<td>-0.004 **</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>

Notes: No. obs. 3568. ** statistically significant at the 5% level. * statistically significant at the 10% level.
Standard errors for quantile regressions are from 2000 bootstrap repetitions.
Table 2 (continued) : The influence of piped water on infant mortality rates: panel regression weighted by county-level population

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Quantile Percentile</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel C: 1991-2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>water</td>
<td>-0.20 **</td>
<td>-0.10</td>
<td>-0.22 **</td>
<td>-0.19</td>
<td>-0.10</td>
<td>-0.08</td>
<td>-0.21 *</td>
<td>-0.31 **</td>
<td>-0.40 **</td>
<td>-0.47 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.13)</td>
<td>(0.11)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.13)</td>
<td>(0.16)</td>
<td></td>
</tr>
<tr>
<td>sewage</td>
<td>0.05</td>
<td>0.15</td>
<td>0.20</td>
<td>0.22 *</td>
<td>0.11</td>
<td>0.16</td>
<td>0.17</td>
<td>0.13</td>
<td>0.27 *</td>
<td>-0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.15)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.16)</td>
<td>(0.22)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.87)</td>
<td>(16.6)</td>
<td>(14.3)</td>
<td>(15.3)</td>
<td>(15.2)</td>
<td>(16.2)</td>
<td>(16.7)</td>
<td>(18.4)</td>
<td>(19.7)</td>
<td>(22.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.54)</td>
<td>(7.17)</td>
<td>(6.18)</td>
<td>(6.72)</td>
<td>(6.76)</td>
<td>(7.24)</td>
<td>(7.19)</td>
<td>(6.97)</td>
<td>(7.59)</td>
<td>(7.94)</td>
<td></td>
</tr>
<tr>
<td>water</td>
<td>0.17 **</td>
<td>0.15</td>
<td>0.33 *</td>
<td>0.30</td>
<td>0.15</td>
<td>0.10</td>
<td>0.31</td>
<td>0.43 **</td>
<td>0.54 **</td>
<td>0.59 **</td>
<td></td>
</tr>
<tr>
<td>x income</td>
<td>(0.08)</td>
<td>(0.22)</td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.21)</td>
<td>(0.21)</td>
<td>(0.21)</td>
<td>(0.23)</td>
<td>(0.28)</td>
<td></td>
</tr>
<tr>
<td>sewage</td>
<td>-0.07</td>
<td>-0.01</td>
<td>-0.10</td>
<td>-0.21</td>
<td>-0.11</td>
<td>-0.33</td>
<td>-0.31</td>
<td>-0.13</td>
<td>-0.13</td>
<td>-0.15</td>
<td></td>
</tr>
<tr>
<td>x income</td>
<td>(0.08)</td>
<td>(0.30)</td>
<td>(0.27)</td>
<td>(0.26)</td>
<td>(0.26)</td>
<td>(0.25)</td>
<td>(0.24)</td>
<td>(0.25)</td>
<td>(0.30)</td>
<td>(0.41)</td>
<td></td>
</tr>
<tr>
<td>water</td>
<td>-0.001 **</td>
<td>-0.002 **</td>
<td>-0.002 **</td>
<td>-0.001</td>
<td>-0.0004</td>
<td>-0.0001</td>
<td>-0.001</td>
<td>-0.003 *</td>
<td>-0.003 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x sewage</td>
<td>(0.0004)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: No. obs. 3568. ** statistically significant at the 5% level. * statistically significant at the 10% level.
Standard errors for quantile regressions are from 2000 bootstrap repetitions.
Figure 1: Marginal impact of one percentage point increase in households with piped water supply on the under-1 infant mortality rates (IMR)

Group of counties that measure low in their development indicators

Figure 2: Marginal impact of one percentage point increase in households with piped water supply on the under-1 infant mortality rates (IMR)

Group of counties that measure high in their development indicators
Figure 3: The association between one percentage point increase in households with piped water and the under-1 infant mortality rates (IMR), as estimated in the cross sectional regressions

Group 1: Counties that measure low in their development indicators
Figure 4: Marginal impact of one percentage point increase in households with piped water on the under-1 infant mortality rates (IMR): A comparison of the full sample with the urban only sample

Group of counties that measure low in their development indicators

Figure 5: Marginal impact of one percentage point increase in households with piped water on the under-1 infant mortality rates (IMR): A comparison of the full sample with the urban only sample

Group of counties that measure high in their development indicators
Table 3: Estimated number of averted deaths from simulated changes in health inputs

<table>
<thead>
<tr>
<th>Regression:</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quantile</td>
<td>Mean</td>
</tr>
</tbody>
</table>

**Section 1: Simulation: One percentage point increase in households with piped water**

Source of coefficients:

<table>
<thead>
<tr>
<th>Panel A: 1970-1980</th>
<th>95,000</th>
<th>67,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel B: 1980-1991</td>
<td>104,000</td>
<td>166,000</td>
</tr>
<tr>
<td>Panel C: 1991-2000</td>
<td>26,000</td>
<td>26,000</td>
</tr>
</tbody>
</table>

**Section 2: Simulation: 0.01 increase in income-related Human Development Index**

Source of coefficients:

<table>
<thead>
<tr>
<th>Panel A: 1970-1980</th>
<th>97,000</th>
<th>119,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel B: 1980-1991</td>
<td>14,000</td>
<td>45,000</td>
</tr>
<tr>
<td>Panel C: 1991-2000</td>
<td>16,000</td>
<td>11,000</td>
</tr>
</tbody>
</table>

Notes: Coefficients are from Table 2