Child Malnutrition in Senegal: Does access to public infrastructure really matter? A Quantile Regression Analysis

Léandre Bassolé*
CERDI-CNRS, Université d'Auvergne
Jobmarket Paper

August, 2007

Abstract

In this paper we analyse the effect of access to public infrastructure, i.e. safe water and health facilities, on child nutritional status defined by height-for-age z-scores in Senegal. Quantile regression methods are used to achieve a more complete picture of the infrastructure effect. This technique has an advantage over the traditional ordinary least squares method as it does not assume a constant effect of the explanatory variables over the entire distribution of the dependent variable. To deal with the potential endogeneity of household expenditures in a child health production function, we use instrumental variables methods. To the best of our knowledge, this paper provides the first empirical analysis of the impact of public infrastructure on child health using an instrumental variables quantile regression approach. Contrary to OLS estimates, we find that access to safe water improves the height-for-age of the lowest (10^{th}) quantile and the effect of health facilities is significant for the 10^{th}, 25^{th}, 50^{th} percentiles at the national level. However, in rural areas, only health facilities have a positive and significant effect on child health. The heterogeneity of this effect is shown using quantile regression, and we find that the effect of health facilities is more important to the lowest quantile and is decreasing. Safe water also improves child health up to the 10^{th} percentile.

Keywords: Child Health, Anthropometric, Senegal, Quantile regression, Instrumental Variable.

***JEL* Classification numbers**: C13, H41, I12, I38.

*I would to thank Jean-Louis Arcand for his continued support and encouragement. I also thank him for helpful discussions and detailed comments on this paper. I am grateful for comments from Stephen Bond, Marcel Fafchamps, Murray Leibbrandt, B. Essama Nssah , Christophe Cottet, Yacouba Gnegne, Chrystelle Tsafack, Florent Bresson, Bachir Diallo, Maria Dontcheva, Eric Djimeu, seminar participants at CERDI, and participants at the 12^{th} annual conference of the African Econometric Society. The usual disclaimer applies.
Email:l.bassole@u-clermont1.fr

1 Introduction

Malnutrition seriously affects the survival and early development of children, and the health of pregnant and nursing mothers. It also determines overall resistance to diseases and future performance in school and at work. Nutrition is therefore, a major health sector priority in developing countries. Child health is particularly important because of its link to child poverty and also to the accumulation of adult human capital. Better health and nutrition have been found to pay-off in terms of economic growth as well as equity concerns. The improvement of child health and nutrition of poor children has been regarded as an efficient way of improving school attendance and enhance economic growth because learning translates into gains in long run productivity. Several of the Millennium Development Goals of the United Nations deal with child malnutrition, including the goal to halve by 2015 the number of people living with hunger, and the goal to reduce child mortality by two-thirds.

The current paper explores the potential effects of access to public infrastructure such as safe water and health services on child malnutrition status in Senegal. In the literature on child health determinants, this relation is not clearly addressed. Horton (1986) analyzing the effect of family size on child nutrition in the Philippines, found that access to a public pump improves child nutritional status measured by the height for age z-score. However, the distance to family planning centres is not significant in child health production function. Handa (1999) obtained the same result in Jamaica. Using a large household survey from rural Central Java to address the relationship between formal education and nutrition knowledge, Block and Webb (2003) showed that the distance to water and the proportion of households having an access to tap water do not affect child nutrition in Indonesia. David, Moncada, and Ordonez (2004) used two data sets from Nicaragua and Honduras to analyze the private and public determinants of child nutrition. They found that community variables such as public health infrastructure (measured by the proportion of household with tap water within the house and proportion of households with a toilet or a washable latrine) and health care services (measured by the average distance from the nearest health center) do not appear to have an impact on the nutritional status of children in Nicaragua. However, in Honduras, the proportion of households with tap water is positively correlated with long term malnutrition, i.e height-for-age, but not with wasting (weight for age), and the time to reach a health centre is not significant. Christiaensen and Alderman (2004) found that Ethiopian households who drink water from own tap improve their children nutritional status captured by height for age. But the distance to the nearest health center has no effect on child health. Valdivia (2004) offers empirical evidence on the impact of the expansion in health infrastructure of the 1990s upon child nutrition in Peru using three rounds of the Demographic and Health Surveys (DHS). For health infrastructure, he constructed an index using principal components methods. None of the sanitary variables has an effect upon child nutrition status. However, he found a positive effect of this expansion in urban areas on child long term nutritional status, which is not the case in rural areas. Furthermore, the effect on urban children is highly non-linear and has a pro-poor bias, in the sense that the estimated effect is larger for children of less educated mothers. Alderman, Hoogeveen, and Rossi (2006) used a four round panel data set from Tanzania to estimate the determinants of a child's nutritional status. They found that the distance to the closest health center is not significant as community characteristic on child nutritional status. Galiani, Gertler, and Schargrodsky (2005) found that in Argentina the privatization of water services improved child health. Using the variation in ownership of water provision, they found that child mortality fell by $5-9\%$ in areas where water services are privatized. Fay, Leipziger, Wodon, and Yepes (2005) using data from DHS by exploiting the variability in outcomes and explanatory variables observed within countries between asset quintiles, show that better access to basic infrastructure services plays an important role in improving child-health outcomes.[1] Finally, Linnemayr and Alderman (2006), using data from the baseline survey for a nutrition intervention program in Senegal, found that the presence of sanitary facilities in the household improves child nutritional status.

All these studies addressed the relation between child health and its determinants with traditional ordinary least squares (OLS) and instrumental variables (IV) approaches. Both OLS and IV are designed to estimate the *mean* or average causal effect of public infrastructure on child nutritional status. This provides the researcher with an estimate of how efficient an improvement in access to public infrastructure is at boosting the health of the average child. The alternative quantile regression (QR) approach goes further. It allows the researcher to estimate the marginal effect of a given access to public infrastructure for households at different points in the conditional height-for-age distribution. This makes it possible to assess the equity implications resulting from changes in access to water or health facilities. In this case, the question is more precise, where the researcher asks not what the effect of an access to infrastructure in child health is on average but for whom such effects are significant and how large might be. To this end, the following study applies a QR technique to the access to public infrastructure issue in order to better understand for whom improvement in access to infrastructure counts and how large the effects are across various points of the conditional height-for-age distribution.[2] In the child health determinants literature, we have two papers using QR techniques. Borooah (2005), uses QR regression in order to capture the heterogeneity of child malnutrition determinants on height-for-age in India and found that access to safe water and a good hospital improve child $z-$scores at the lowest end of quantiles. Aturupane, Deolalikar, and Gunewardena (2006) analyze the determinants of child weight and height in Sri Lanka with DHS data with quantile regression. They found that access to piped water improves child nutritional status for nearly all quantiles. But the limitation of these two studies is that they did not account for the endogeneity of household expenditure or income. Although this variable is not our variable of interest in this paper, it is used as control variable. But, the endogeneity of one covariate generally results in inconsistent estimators of all the parameters (Wooldridge 2002). Then, without taking this into account, one can question the robustness of their results as we know that access to those infrastructures depends on household income. In this paper we consider this problem in a QR framework.

The rest of the article is organized as follows. Section 2 presents the model of child health. Section 3 describes the data set used in the paper and presents descriptive statistics of some relevant variables. Next, we outline in section 4 our econometric framework designed to analyse the heterogeneity of the effects of public infrastructure on child nutritional status, using an

[1] See Ravallion (2007) and Fay, Leipziger, Wodon, and Yepes (2007) for the comments on this paper

[2] Low height-for-age index identifies past undernutrition or chronic malnutrition. It cannot measure short term changes in malnutrition. For children below 2 years of age, the term is length-for-age; above 2 years of age, the index is referred to as height-for-age. Deficits in length-for-age or height-for-age is referred to as stunting

instrumental variables quantile regression. This is followed by a discussion of the results in section 5 and concluding remarks in section 6.

2 Basic Model of Child Health

Malnutrition has strong negative effects on children's health. Nearly one-third of children in developing countries are either underweight or stunted, and more than 30 % of the developing world's population suffers from micronutrient deficiencies (World-Bank 2006). In this paper, child nutritional status is quantified using one most commonly used anthropometric indicator: height-for-age.[3] This indicator is expressed as a $z-$score, which compares a child's measurements with the measurements of a similar child in the reference population, from healthy population, which has a $z-$score with mean zero and standard deviation one.[4] The height-for-age is an indicator of long-run health and welfare and it is not subject to transitory shocks (Waterlow et al. (1977) , WHO (1986), Barrera (1990)). The determinants of child health have been extensively studied in the economic literature. Based on these studies, several researches have explored interventions needed to address the causes of malnutrition. According to Strauss and Thomas (1995), interventions could focus on helping households to use their resources more effectively in order to improve the nutritional status of their children. The analysis in this paper is conceptually built on a model widely used in the literature on the demand for child wellbeing. The application of this theoretical framework children's health status is well known, and is discussed in detail in Behrman and Dealalikar (1988) and Strauss and Thomas (1998). This analysis is based on a well known model in the tradition of Becker (1981), in which a household maximizes a utility function. In this case, a household may be assumed to choose child health H, leisure L, consumption of goods and services C. The problem is:

$$\max_{H,L,C} \quad U = U\left(H, L, C; X_h, \mu\right) \tag{1}$$

where X_h is a vector of household characteristics including the education level of the household head and his spouse, and μ unobserved heterogeneity of preferences (as described by Pitt and Rosenzweig (1985)). The household maximizes this utility function subject to two constraints: a health production function for nutritional status and a budget constraint. Child health is generated by the following production function[5]

$$H_i = F\left(Y_i, X_i, X_h, X_c, \psi_i\right) \tag{2}$$

where Y_i is a vector of health inputs which are nutrient intake, health care practices, time spent by parents taking care of children, and disease incidence, X_i, is a vector of child characteristics which are age and gender, X_c is a vector of community characteristics that may have a direct

[3] Anthropometrics indicator is an output in a health production function. There is another nutrition indicator, nutrient intakes, which takes account the inputs aspect (see Strauss and Thomas (1998) for more description).

[4] The reference population is the standard WHO-adopted definition which is based on anthropometric measurements of US population surveys.

[5] There are two types of health production function used in the literature: the mortality production function and the morbidity/anthropometric production function (Behrman and Dealalikar (1988)). Here we use the anthropometric production function

impact on child health, which are the accessibility and quality of health services and safe water, and ψ_i are unobservable individual health endowments. In addition, the full income constraint takes the form:

$$I = P_c C + WL + P_Y Y \tag{3}$$

where P_c and W, P_Y are the price vectors of consumption goods, leisure and health inputs respectively, and I is the full income including the value of the time endowment of the household and non-labor income. In this framework, the reduced form function for child health is:

$$H_i = \Phi\left(X_i, X_h, X_c, I, P_c, P_Y, \eta_i\right) \tag{4}$$

where the particular functional form of the function $\Phi\left(\right)$ depends on the underlying functions characterizing household preferences and the health production function, and η_i represents unobserved heterogeneity in health outcomes.

3 Data and Descriptive Statistics

The data for the empirical analysis are drawn from the 2001 national representative household survey in Senegal (*Enquête Sénégalaise Auprès des Ménages, ESAM2*) conducted by the Senegalese department of statistics and forecasting (*Direction de la Prévision et de la Statistique, DPS*). The survey is based on multi-stage stratified sampling design of nearly $6,600$ households, however our analysis used $2,868$ households and $4,484$ children under 5 years of age for whom we have anthropometric data. The average age of children is 25 months (Table 1). In this paper, we use as child nutritional status indicator, the height-for-age z-scores (HAZ) which represents the long term nutrition deprivation (Trapp and Menken 2005).[6] More specifically the variable HAZ is the difference (expressed in standard deviations) of a child's height for age from the median height of children of the same age and sex in the reference population. The standard reference population recommended by the World Health Organization is that of the U.S National Center for Health Statistics. Other malnutrition anthropometric measures include weight-for-height and weight-for-age z-score. The former is also a measure of long-term nutritional status as the height-for-age but is influenced by recent phenomena (Trapp and Menken 2005). In contrast, a commonly used measure for clinical assessment, weight-for-age, is more indicative of short term conditions; as most regressor in cross sectional studies are stock rather than flow variables it is generally not practical to study this variable with such data (Alderman 2000) .

Our main variables are the child nutritional indicator, i.e height-for-age z-score and the infrastructure variables, i.e the health and water conditions for the household in the community. We aim to measure the child-nutritional effects of access to health and safe water facilities.

[6] When working with z-scores, it is important to consider the issue of cut-offs points, i.e. which observations to exclude from the analysis that stem from wrong measurements or erroneous data entry, as outliers can influence the estimation results in a non-trivial way. The World Health Organization (WHO) has defined two different types of limits for acceptable data: on the one hand, it suggests a flexible exclusion range, defined as +/- 5 z-score units from the observed mean z-score, but with a maximum height-for-age z-score of +3.0. The other recommended filter is a fixed restriction range for observations with a mean z-score of higher than -1.5, and bounded by a lower value of -5.0 for both weight-for-age and height-for-age, and an upper bound of +3.0 for height-forage, and 5.0 for weight-for-age. In this paper, we used the WHO proposition.

As the data does not indicate actual demand of these facilities, the analysis is limited to the presence of the facilities rather than their utilisation. However, according to Strauss (1990) the availability is more interesting than the actual take-up of community services as the latter reflects household choice and would then have to be treated as an endogenous variable. These facilities variables are binary indicating (= 1) whether the household reports that time to access to these facilities is less than 15mn, and zero otherwise. We also control for all classical determinants of child health: child and household characteristics, including mother's education, household income, child age, child sex, household size, etc.

The anthropometric results for children reveal better average performance for girls than for boys (Table 1), a fact that has often been noted in Subsaharan Africa over the past 40 years by Svedberg (1990). There is significant heterogeneity when one breaks down the average by quantile ($\tau = 0.10, 0.25, 0.50, 0.75, 0.90$), with a tendency for the mean $z-$score to be worse at the low end of the distribution. The proportion of households with safe water is 85%, and this varies little between rural and urban areas (Table 1). Access to health facilities in Senegal is very low, only 35% of households in ESAM2 have access to this type of infrastructure. The proportion in rural areas is 27% which means that the lack of health facilities is an important problem in this country. Table 2 shows the relationship between some households, child characteristics and child malnutrition prevalence. We observe that among households with access to safe water, the prevalence of stunting is 30.6% and 36.3% at the national and rural levels respectively.[7] For health facilities, the percentage of stunting is low than the former. Only 25.8% of children are stunting in the family with access to health facilities according to ESAM2. However, in rural areas we have 30.3% of stunting.

A preliminary look at the relationship between access to infrastructure and child anthropometric $z-$ score across the five quantile to be estimated (the $10^{th}, 25^{th}, 50^{th}, 75^{th}$ and 90^{th} percentiles) is provided in Table 3. At the national level, households with access to safe water don't perform better in means of $z-$score than those without access in full sample and all quantiles. However, we observe a statistically significant difference in means for access to health facilities only in the full sample for the users. For all quantiles, the differences are not distinguishable from zero. In rural areas, we have the same observation. Clearly, there seems to be no consistent pattern that emerges with respect to child anthropometrics and access to infrastructure. However, one must be wary drawing conclusions from such statistical analysis. A more in-depth analysis that controls for several observable characteristics is necessary to properly ascertain whether there is a significant effect of access to safe water and health facilities on child nutritional status and whether any such effect is heterogeneous across the $z-$score distribution.

4 Estimation Strategy

> *What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of x's. We could go further and compute several different regression curves corresponding to the various percentage points of the*

[7] A child is considered stunted when his or her height-for-age is more than two standard deviations below the NCHS/WHA reference. However, a child is said to be severely malnourished when the relevant nutritional status indicator is more than three standard deviations below the NCHS/WHO reference.

distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions.

Frederick Mosteller and John W. Tukey (1977)

Without loss of generality, an estimation of equation (4) can be written as:

$$H_i = \alpha + X_i\beta + X_h\delta + X_c\theta + \varepsilon_i \tag{5}$$

where H_i is a vector of anthropometric measures of the children under consideration,X_i, X_h, and X_c are vectors of covariates at the individual, household, and community level, respectively, and ε_i is an error term.

In the present paper child nutritional status will be estimated using the quantile regression methodology. The quantile regression estimates will then be compared to the OLS regression estimates and inference drawn.

4.1 Estimation of the Quantile Regression model

4.1.1 Quantile function

OLS approach is based on the *mean* of conditional distribution of the dependent variable, roughly, base some policy recommendations according to this *mean* slope would implicitly assume that possible differences in terms of the impact of the exogenous variables along the conditional distribution, are unimportant. Yet, if exogenous variables influence the parameters of the conditional distribution of the dependent variable differently, then an analysis that ignores this possibility will be severely weakened (Koenker and Bassett 1978).Unlike OLS, quantile regression models allow for a full characterization of the conditional distribution of the dependent variable. The quantile regression (QR) estimator, introduced by Koenker and Bassett (1978), is an increasingly important empirical tool, allowing researchers to fit parsimonious models to an entire conditional distribution. Before presenting a formal definition of quantile regression we want to highlight the notion of quantile function and give the definition of a sample quantile. Thus, the word "quantile" is a synonym for percentile or fractiles and refers to the general case of dividing the population into fours (or more) segments, a quintile divides the reference population into five sub-groups and a decile divide the population into ten sub-groups. The median divides the population into two groups. Indeed, in a sense, quantiles are related to the process of ordering and sorting the data. Assume that the τ^{th} quantile of a population is m_τ where $0 < \tau < 1$ and F_Y is the cumulative distribution function (cdf), in population, of y then m_τ is defined as:

$$\tau = P\left(y \leq m_\tau\right) = F_Y\left(m_\tau\right) \tag{6}$$

Then for a sample the quantile function m_τ is straightforward defined as:

$$m_\tau = F_Y^{-1}\left(\tau\right) = \inf\{y|F_Y\left(y\right) \geq \tau\} \tag{7}$$

The quantile m_τ should be considered through the value of y below which τ of the values fall. Then for any τ in the interval $(0,1)$, m_τ provides the τ^{th} quantile of Y. Similarly, taking a random sample $Y_1,\ldots,Y_n$ with empirical distribution function $\widehat{F}_Y(\tau)=\frac{\#(Y_i\leq\tau)}{n}$, we can also define the empirical quantile function as:

$$\widehat{m}_\tau=\widehat{F}_Y^{-1}(\tau)=\inf\left\{y|\frac{\#(Y_i\leq\tau)}{n}\geq\tau\right\} \tag{8}$$

It can easily be seen from equation (8) that in order to obtain the desired quantile, one first has to sort and rank the observed sample and then check at which observation the threshold is reached. The most frequently examined are the median ($\tau=0.5$), the 25^{th} and 75^{th} percentiles.

4.1.2 Quantile regression

The idea of quantile regression was first introduced by Koenker and Bassett (1978). This method of estimation if the generalization of the concept of ordinary quantiles in a location model. In their seminal paper, Koenker and Bassett (1978) show that the empirical quantile function (equation 8) is the solution of the minimization problem defined by:

$$\begin{aligned}\widehat{m}_\tau &= \underset{b}{\operatorname{argmin}}\left\{\sum_{i:y\geq b}\tau|Y_i-b|+\sum_{i:y<b}(1-\tau)|Y_i-b|\right\}\\ &= \underset{b}{\operatorname{argmin}}\sum_i\rho_\tau(Y_i-b)\end{aligned} \tag{9}$$

with $\rho_\tau(z)$ expressed as:

$$\rho_\tau(z)=\begin{cases}\tau(z) & \text{if}\quad z\geq 0\\ (\tau-1)z & \text{if}\quad z<0\end{cases}=(\tau-I(z<0))z \tag{10}$$

$\rho_\tau(z)$ is the check function and $I(.)$ is the usual indicator function. Let x_i, with $i,\ldots,n$ a sample, a $K\times 1$ vectors of regressors. Following Koenker and Bassett (1978) we can write a linear quantile regression as follows:

$$y_i=x_i'\beta_\tau+\varepsilon_{\tau_i} \tag{11}$$

where the distribution of the error term ε_{τ_i} is left unspecified and the τ^{th} quantile of the error term conditional upon the regressors is zero:

$$m_\tau(\varepsilon_{\tau_i}|x_i)=0$$

From equations (11) and (12) it follows that the τ^{th} conditional quantile of y_i can be written as:

$$m_\tau(y_i|x_i)=x_i'\beta_\tau \tag{12}$$

In analogy of equation (9) we finally obtain the quantile regression by solving with respect to β_τ.

$$\widehat{\beta}_\tau = \underset{\beta_\tau \in \mathbb{R}^K}{\operatorname{argmin}} \left\{ \sum_{i: y_i \geq x_i' \beta_\tau} \tau |y_i - x_i'\beta_\tau| + \sum_{i: y_i < x_i'\beta_\tau} (1-\tau)\, |y_i - x_i'\beta_\tau| \right\} \tag{13}$$

$$= \underset{\beta_\tau \in \mathbb{R}^K}{\operatorname{argmin}} \sum_i \rho_\tau \left(y_i - x_i'\beta_\tau \right) \tag{14}$$

According to equation (13) it's a plain that all observations above the estimated hyperplane given $X\widehat{\beta}_\tau$ are weighted with τ, all observations below the estimated hyperplane are weighted with $(1-\tau)$. Indeed, the quantiles other than the median are defined as the solution of a problem that minimizes the weighted sum of the absolute value of the residuals. We can also notice that for quantiles above the median, say $\tau = 0.75$, a higher weight is placed on residuals above the quantile than on residuals below the quantile. This pushes the minimization up above the median, which is where one wants it in such cases. The Least absolute deviation (LAD) estimator of β is a special case of quantile regression. The LAD is obtained by setting $\tau = 0.50$ (median regression):

$$\widehat{\beta}_{0.5} = \underset{\beta_{0.5} \in \mathbb{R}^K}{\operatorname{argmin}} \sum_i |y_i - x_i'\beta_{0.5}| \tag{15}$$

In contrast to the OLS approach, the quantile regression procedure is less sensitive to outliers and provides a more robust estimator in the face of departures from normality (Koenker (2005), Koenker and Bassett (1978)). Quantile regression models may also have better properties than OLS in the presence of heteroscedasticity (Deaton (1997)). Because the objective function is not differentiable, standard gradient optimization methods cannot be used. However, the problem can be written in the form of a linear programming problem, and solved using linear programming methods as in Koenker and Bassett (1978). Generalized Method of Moments (GMM) estimation of the quantile is also possible, as shown in **?**

Two general approaches exist for the estimation of the covariance matrix of the regression parameter vector. The first derives the asymptotic standard error of the estimator (Koenker and Bassett (1978)) while the second uses bootstrap methods to compute these standard errors and construct confidence intervals. In this paper, we use the design matrix bootstrap method to obtain estimates of the standard errors for the coefficients in quantile regression (Buchinsky, 1995, **?**). Based on a Monte Carlo study, Buchinsky (1995) recommends the use of this method as it performs well for relatively small samples and it is robust to changes of the bootstrap sample size relative to the data sample size. More importantly, the design matrix bootstrap method is valid under many forms of heterogeneity. This method of bootstrap performs well even when the errors are homoscedastic. In addition to the design matrix bootstrap method, we use the percentile method (see (**?**) for more detail on this method) recommended by Koenker and Hallock (2001). This method enables to construct confidence intervals for each parameter in β_τ, where the intervals are computed from the empirical distribution of the sample of the bootstrapped $\widehat{\beta}_\tau^{BS}$'s. Conceptually, in the design matrix bootstrap, we consider the sample of n observations as if it is were the population of interest. Specifically, let $(y_i^{BS}, x_i'^{BS})$, $i = 1, \ldots n$, be the bootstrap sample obtained by sampling with replacement from the original sample (y_i, x_i'). Applying the simplex algorithm to this sample gives $\widehat{\beta}_{\tau}^{BS}$, a bootstrap estimate of β_{τ}^{BS}. Repeating this process R times yields bootstrap estimates $\widehat{\beta}_{\tau 1}^{BS}, \widehat{\beta}_{\tau 2}^{BS}, \ldots, \widehat{\beta}_{\tau R}^{BS}$. The bootstrap estimate of the asymptotic variance-covariance matrix of β_{τ} is then obtained as follows:

$$\widehat{\Omega}_{\tau}^{BS} = \frac{n}{R} \sum_{j=1}^{R} \left(\widehat{\beta}_{\tau j}^{BS} - \overline{\widehat{\beta}}_{\tau} \right) \left(\widehat{\beta}_{\tau j}^{BS} - \overline{\widehat{\beta}}_{\tau} \right)' \tag{16}$$

where

$$\overline{\widehat{\beta}} = \frac{1}{R} \sum_{j=1}^{B} \overline{\widehat{\beta}}_{\tau}^{BS} \tag{17}$$

The number of bootstrap replication, R, should be large enough to guarantee a small sample variability of the covariance matrix. In this paper we use 500 bootstrap replications to obtain the standard errors.[8]

4.2 Instrumental Variable Quantile Regression Estimation

The problem of endogeneity occurs because household per capita income and child health are likely to be jointly determined because time spent in market work for income may subtract from time spent caring for children (Thomas and Strauss 1992). Obviously, household make decisions about their children's health at the same time that they make decisions about income earning activities, therefore these two decisions could be related. Another problem with household income is that it is often measured with random error, simply because it is difficult for households to report accurately their incomes. As in Glewwe, Koch, and Nguyen (2002), household per capita will be used instead of household per capita income to measure I, for two reasons. First, expenditure data are likely to be more accurate than income data in developing countries (Deaton 1997). Second, expenditure data are more likely to reflect a household "permanent income", which is more appropriate because I represents household's income since the child was born, not just current income. However, expenditures per capita is not our variable of interest, but without taking account this endogeneity problem, all coefficients in the model could be bias (Wooldridge 2002). Suppose that, in equation 5 we have two variables, expenditures per capita X and access to infrastructure I. Thus, equation 5 becomes:

$$H_i = \alpha + X_i \lambda + I_i \varphi + \upsilon_i \tag{18}$$

However, expenditure per capita is endogenous ($Cov(X_i, \upsilon_i) \neq 0$) and correlated with infrastructure variable: $E(X_i|I_i) = \pi_0 + I_i \pi_1$. In this case, the OLS estimates of I_i will be biased by:

$$\begin{aligned} E(H_i|I_i) &= \alpha + E(X_i|I_i)\lambda + I_i\varphi \\ &= \alpha + (\pi_0 + I_i\pi_1)\lambda + I_i\varphi \\ &= \alpha + \pi_0\lambda + I_i(\varphi + \pi_1\lambda) \end{aligned} \tag{19}$$

[8] The problem of choosing the number of bootstrap repetitions has been studied in the literature. See Andrews and Buchinsky (2000) and Davidson and MacKinnon (2001). The problem is that one can obtain "different answers" from the same data merely by using different replications if B is too small, but computational costs can be great if B is chosen to be extremely large.

where $\pi_0\lambda$ and $\pi_1\lambda$ are measures of the OLS bias. The bias will increase as λ and the correlation between X and I increase in absolute value.

As in the OLS, when some of the explanatory variables are determined simultaneously with the response variable, a bias arises in quantile regression estimators due to the dependence between the regressors and the error term. Then, instrumental variable methods can, remove this bias caused by either endogeneity or measurement in the household expenditure variable. However, the difficulty with this method is to find plausible instrumental variables, that is variables that are correlated with household income but uncorrelated with unobserved determinants of child health and uncorrelated with the measurement error in the household expenditure variable.

The issue of endogenous regressors is not straightforward in the quantile regression framework and few empirical papers to date have taken this problem into account in quantile regression.[9] Following Amemiya (1982) and Powell (1983) the 2SRQ procedure is essentially the quantile estimator analog of 2SLS. Then, suppose we are given the simple model with two endogenous variables defined as:

$$Y = X_1\beta + X_2\gamma + \varepsilon \tag{20}$$

$$X_2 = X_1\delta + Z\varphi + \rho \tag{21}$$

where Y is the response variable; X_1 is a matrix of exogenous regressors; X_2 is a matrix of endogenous variables determined simultaneously with Y; Z is a matrix of valid instrumental variables; and ε and ρ are random errors terms. The two steps are (Amemiya 1982): the first stage concerns equation 21 where it consists to regress by OLS the endogenous X_2 variable on the exogenous variable (including the instruments), and the second stage is to estimate by quantile regression the structural Equation 20 where the endogenous variable X_2 is replaced by its predicted values from the first stage, $\widehat{X}_2$[10]. However, the standard errors might be biased. Then, the entire system is estimated with bootstrap techniques, that is, not only the second step in order to account for biased standard errors due to the fact that one of the regressors is estimated in the first stage (Amemiya (1982), Arias, Hallock, and Sosa-Escudero (2001), Maitra and Vahid (2006)).

[9] Ribeiro (2001) analyzing the labor supply behavior of urban males in Brazil, investigates the possible endogeneity of wages and income with respect to hours worked in a quantile regression framework by employing a two-stage least absolute deviation estimator (2SLAD) originally developed by Amemiya (1982), and Powell (1983) has investigated its asymptotic normal. Levin (2001) used quantile regression to estimate the impact of class size and peer effects on scholastic achievements of Dutch students. An instrumental variables quantile regression (tow-stage quantile regression) has been used to account for the potential endogeneity of class size as a policy variable. They find that the class size in general does not have an effect on scholastic achievement whereas the peer effect proves significantly positive in the lower part of the distribution. Also, Arias, Hallock, and Sosa-Escudero (2001) estimated the returns to schooling with instrumental variables quantile regression using data on earnings of identical twins. They tested for individual heterogeneity and found that more able individuals obtain higher marginal benefits of schooling. The subject on the instrumental variable quantile regression estimator has also been investigated by Abadie, Angrist, and Imbens (2002) who develop a quantile treatment effects estimator (QTE) to account for endogeneity of the fertility decision in the estimation of the effect of subsidized training on the quantiles of the distribution of trainee earnings. Recently, a quantile regression with endogenous regressors has also been investigated by Chernozhukov and Hansen (2005) who develop a model of quantile treatment effects in presence of discrete endogenous regressors.

[10] Chen (1988), Chen and Portnoy (1996) provide method for correcting covariance matrix for a given quantile by 2SRQ

5 Estimation Results

Our main results are instrumental variables estimations. But, we will present the results with OLS for comparison. As we mention before, household expenditure is instrumented with an asset index and land ownership (Linnemayr and Alderman (2006), Christiaensen and Alderman (2004)). [11]

For all quantile estimations, in order to check whether the results obtained in the present study are sensitive to the survey design, a bootstrap approach proposed by Deaton (1997) is employed. Deaton's approach implies bootstrapping the clusters rather than the children.[12] Also, as Deaton (1997) points out, treating a two-stage sample as if it was a simple random sample can have serious implications since the sampling variability of the estimates can be affected by the design. It is often the case that clustering increases the inter-cluster variability since households within clusters are frequently similar to one another in their relevant characteristics. Therefore, ignoring the cluster design can lead to standard errors that are too small and t-values that are two large, thereby overstating the precision of the estimates (Moulton (1986), Moulton (1990), Bertrand, Duflo, and Mullainathan (2004)). Therefore, we use clustering for correcting this problem in all our estimations as in quantile regressions.

Our IV results depend on instrument relevance as emphasized by the recent literature on weak instruments (Staiger and Stock (1997), Stock, Wright, and Yogo (2002), Stock and Yogo (2002); Moreira (2003), Andrews and Stock (2005); Andrews, Moreira, and Stock (2007)).[13] If the instrumental variables are only weakly correlated with the endogenous explanatory variables, conventional asymptotic theory no longer holds, and statements about statistical significance and inference may lead to the wrong conclusions. Then, the relevance of the instrumental variables has to be checked to allow for statements about statistical significance. The first-stage regressions (Table 7) supply valuable information about the relevance of the instrumental variables. The F-statistics of the first-stage regressions and the usual partial R^2 in Table 7 point to strong instruments at national and rural level. In addition to these test results, a statistic proposed by Cragg and Donald (1993) is computed, which represents the relevance of the weakest instruments. The Cragg-Donald statistic critical value for one endogenous and two instruments is 19.93. Our statistics equal 414.11 and 46.36 for national and rural level respectively, then all our IV results are not affected by weak-instrument problems. We also present in Table 8 and 9 the Hansen test for overidentification restrictions and Hahn and Hausman (2002) an another weak instrument test. The Hansen test does not reject the one overidentifying restriction. Our last weak instrument test of Hahn and Hausman (2002), which can be applied only for overidentified equations, does not reject the null hypothesis of strong instruments.

First, IV estimates in column 1 of Table 8, 9 indicate that access to safe water has a positif effect on child $z-$score in the full sample (0.023), but with a standard error that renders this

[11] Assets index are constructed using Principal Components Analysis

[12] A list of the n sample clusters is made, a bootstrap sample of size n is drawn with replacement, and the individual cluster-level data merged in. As the data used in the present study has been collected using a two-stage stratified cluster sampling procedure. This two-stage design, in which primary sampling units or clusters (often villages) are drawn first and then households from within each cluster, is very common for household surveys in developing countries. See Deaton (1997) for more details in clustering sample

[13] See Nichols (2006) for an interesting theoretical and technical survey.

effect statistically indistinguishable from zero ($s.e = 0.085$) (Table 8). When we switch to rural areas (Table 9), result remains the same, the estimates of access to safe water is positive , 0.030, but the coefficient is still not significant. We have the same results with OLS estimations (see Table 4, 5). At this stage, we can say that, with an IV result, this type of infrastructure has an effect on child anthropometric but it is not different from zero.

However, the quantile regression paints a different picture for this type of infrastructure. At the lower quantile, i.e. the 10^{th} percentiles, access to safe water improves significantly child nutritional status by 0.272 ($s.e = 0.120$) for the full sample (Table 8). Compare to the others quantiles, its effect is important for the 10^{th}. In Table 4, the OLS results also give the same effect on average. The coefficient of access to safe water in rural areas is also positive (0.329) and significant ($s.e = 0.143$) for IVQR (Table 9). Contrary to OLS, quantile estimation shows that, access to safe water is very important for child nutritional status in the lowest quantile of the distribution. The results of our estimations show that at the national level, using OLS , we can not distinguish the contribution of safe water on height-for-age. But, this is not the case when quantile strategy is applied. We note that access to safe water in household communities improve child health for those whose $z-$scores are very bad, i.e the 10^{th} quantile.

For health facilities, in the full sample, the coefficient is positive but not significant on child $z-$score as for access to safe water (column 1 in Table 8). Meanwhile, in rural areas we observe a strong positive significant effect (column 1 in Table 9), showing that access to health infrastructure is an important problem in this part of the country. The IVQR results show important differences in the effect of health infrastructure at different points in the conditional distribution of height-for-age. We observe in the full sample that, contrary to the OLS average results, at the low end of the distribution, i.e $\tau = 25\%$ and 50%, the coefficient of access to health facilities is positive and significant (Table 8). This means that OLS estimates can mask the heterogeneity of the potential contribution of access to health infrastructure on child nutritional status in the full sample. Child anthropometric is affected positively with an access to health facilities for nearly all quantiles except $\tau = 0.90$ (Table 9), and the level of coefficients is decreasing from the lowest quantiles to the higher in rural areas. In other words, the effect of this type of infrastructure is important for undernourished children who were most in need of help. Health facilities remain the infrastructure which has the most an important positive and significant effect on child $z-$score.

An inspection of the quantile estimate in Figure 1 and 2 reveals that the slopes coefficients estimated at different quantiles are not flat but follow nonlinear patterns, which suggests, at least at an informal level, the existence of parameter heterogeneity across quantiles. Formally, accordingly to **?** the parameter heterogeneity can be examined by the test whether the quantiles are statistically different from each other. This differentiation across quantiles is important for the analysis and the formulation of policies that may alter child nutritional status pattern. The null hypothesis that the coefficient vectors are the same for the different quantiles equation is rejected, both in pair-wise comparisons and jointly (Table 6, 10). The tests confirm the visual impression from Figure 1 and 2.

6 Concluding Remarks and Policy Implication

This paper investigates the effect of access to public infrastructure, i.e. safe water and health facilities on child nutritional status in Senegal. In this country, access to water and health is highly limited and very low in rural areas where most of people are poor.[14] Then, the main results of this work is that improving the access to safe water and health facilities have a positif and significant effect on household welfare, more precisely on child nutritional status in rural areas. However, we find that "classical" conditional mean regression, i.e. OLS, has been insufficient to show this effect of the access to public infrastructure on child height-for-age. In addition, there is much heterogeneity in the marginal effect of such infrastructure on child height-for-age which is not highlighted by OLS estimations. Thus our work shows that OLS does not give a complete picture of the effect of safe water and health facilities on child health.

By using a quantile regression, we show that these public infrastructures are very important for the poorest household in rural areas. In general, quantile regression might be used more often as a complement, not a substitute, to traditional regression analysis since we can use what is known about the distributions in a more complete way than giving an average summary. Policy interventions to address child malnutrition need to be sensitive to this reality.

References

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

Alderman, H. (2000): *Anthropometrics*vol. 1 of *Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience*. Oxford University Press.

Alderman, H., H. Hoogeveen, and M. Rossi (2006): "Reducing Child Malnutrition in Tanzania: Combined Effects of Income Growth and Program Interventions," *Economics and Human Biology*, 4(1), 1–23.

Amemiya, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50(3), 689–711.

Andrews, D. W., M. J. Moreira, and J. H. Stock (2007): "Performance of conditional Wald tests in IV regression with weak instruments," *Journal of Econometrics*, 139(1), 116–132.

Andrews, D. W., and J. H. Stock (2005): "Inference with Weak Instruments," NBER Technical Working Paper No. 313 Issued in August 2005.

Andrews, D. W. K., and M. Buchinsky (2000): "A Three-Step Method for Choosing the Number of Bootstrap Repetitions," *Econometrica*, 68(1), 23–52.

[14]According to the survey of the perception of poverty in Senegal 2001, the incidence of poverty varies between 72 percent to 88 percent in rural areas

Arias, O., K. F. Hallock, and W. Sosa-Escudero (2001): “Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data,” *Empirical Economics*, 26, 7–40.

Aturupane, H., A. B. Deolalikar, and D. Gunewardena (2006): “The Determinants of Child Weight and Height in Sri Lanka: A Quantile Regression Approach,” April 2006.

Barrera, A. (1990): “The role of maternal schooling and its interaction with public health programs in child health production,” *Journal of Development Economics*, 32(1), 69–91.

Becker, G. S. (1981): *A Treatise on the Family.* Harvard University Press, Combridge, MA.

Behrman, J. R., and A. B. Dealalikar (1988): *Health and Nutrition*vol. 1 of *Handbook of Development Economics*, chap. 14, pp. 631–711. Elsevier Science.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): “How Much Should We Trust Differences-in-Differences Estimates?,” *Quarterly Journal of Economics*, 119(1), 249–275.

Block, S., and P. Webb (2003): “Nutrition Knowledge and Parental Schooling as Inputs to Child Nutrition in the Long and Short Run,” Nutrition Working Paper No.3.

Borooah, V. K. (2005): “The Height-for-Age of Indian Children,” *Economics and Human Biology*, 3(1), 45–65.

Buchinsky, M. (1995): “Estimating the asymptotic covariance matrix for quantile regression models a Monte Carlo study,” *Journal of Econometrics*, 68(2), 303–338.

Chen, L. A. (1988): “Regression quantiles and trimmed least squares estimators for structural equations and non-linear regression models.,” PhD Dissertation, University of Illinois at Urbana-Champaign.

Chen, L. A., and S. Portnoy (1996): “Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models,” *Communications in statistics Theory and methods*, 25(5), 1005–1032.

Chernozhukov, V., and C. Hansen (2005): “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73, 245–261.

Christiaensen, L., and H. Alderman (2004): “Child Malnutriton in Ethiopia: Can Maternal Knowledge Augment the Role of Income?,” *Economic Development and Cultural Change*, 52, 287–312.

Cragg, J. G., and S. G. Donald (1993): “Testing Identifiability and Specification in Instrumental Variable Models,” *Econometric Theory*, 9(2), 222–240.

David, V., M. Moncada, and F. Ordonez (2004): “Private and public determinants of child nutrition in Nicaragua and Western Honduras,” *Economics & Human Biology*, 2(3), 457–488.

Davidson, R., and J. G. MacKinnon (2001): “Bootstrap Tests: How Many Bootstraps?,” Queen’s University, Department of Economics in its series Working Papers with number 1036.

Deaton, A. (1997): *The Analysis of Household Surveys. A Microeconomic Approach to Development Policy.* Johns Hopkins University Press, Baltimore, MD.

Fay, M., D. Leipziger, Q. Wodon, and T. Yepes (2005): "Achieving Child-Health-Related Millenium Development Goals: The Role of Infrastructure," *World Development*, 33(8), 1267–1284.

——— (2007): "Achieving Child-Health-Related Millennium Development Goals: The Role of Infrastructure-A Reply," *World Development*.

Galiani, S., P. Gertler, and E. Schargrodsky (2005): "Water for Life: The Impact of Privatization of Water Services on Child Mortality," *Journal of Political Economy*, 113(1), 83–120.

Glewwe, P., S. Koch, and B. L. Nguyen (2002): "Child Nutrition, Economic Growth, and the Provision of Health Care Services in Vietnam in the 1990s," World Bank, Policy Research Working Paper 2776.

Hahn, J., and J. A. Hausman (2002): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70(1), 163–189.

Handa, S. (1999): "Maternal Education and Child Height," *Economic Development and Cultural Change*, 47(2), 421–439.

Horton, S. (1986): "Child nutrition and family size in the Philippines," *Journal of Development Economics*, 23(1), 161–176.

Koenker, R. (2005): *Quantile Regression*, Econometric Society Monographs. Cambridge University Press.

Koenker, R., and G. J. Bassett (1978): "Regression Quantiles," *Econometrica*, 46(1), 33–50.

Koenker, R., and K. Hallock (2001): "Quantile Regression," *Journal of Economic Perspectives*, 15(4), 143–156.

Levin, J. (2001): "For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement," *Empirical Economics*, 26, 221–246.

Linnemayr, S., and H. Alderman (2006): "Determinants of Malnutrition in Senegal: Individual, Household, Community Variables, and Their Interaction," .

Maitra, P., and F. Vahid (2006): "The effect of household characteristics on living standards in South Africa 1993-1998: a quantile regression analysis with sample attrition," *Journal of Applied Econometrics*, 21(7), 999–1018.

Moreira, M. J. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71(4), 1027–1048.

Moulton, B. (1986): "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 32, 385–397.

Moulton, B. R. (1990): "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 72(2), 334–338.

Nichols, A. (2006): "Weak Instruments: An Overview and New Techniques," North American Stata Users' Group Meetings 2006 with number 3.

Pitt, M. M., and M. R. Rosenzweig (1985): "Health and Nutrient Consumption Across and Within Farm Households," *Review of Economics and Statistics*, 67(2), 212–223.

Powell, J. L. (1983): "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 51(5), 1569–1575.

Ravallion, M. (2007): "Achieving child-health-related Millennium Development Goals: The role of infrastructure - A comment," *World Development*.

Ribeiro, E. P. (2001): "Asymmetric labor supply," *Empirical Economics*, 26, 183–197.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

Stock, J. H., J. T. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Methods of Moments," *Journal of Business and Economic Statistics*, 20(4), 518–529.

Stock, J. H., and M. Yogo (2002): "Testing for Weak Instruments in Linear IV Regression," NBER Technical Working Papers with number 0284.

Strauss, J. (1990): "Households, Communities, and Preschool Children's Nutrition Outcomes: Evidence from Rural Cote d'Ivoire," *Economic Development and Cultural Change*, 38(2), 231–261.

Strauss, J., and D. Thomas (1995): *Human Resources: Empirical Modeling of Household and Family Decisions*vol. 3A of *Handbook of Development Economics*, chap. 34, pp. 1883–2005. Elsevier.

Strauss, J., and D. Thomas (1998): "Health, Nutrition and Economic Development," *Journal of Economic Literature*, 36(2), 766–817.

Svedberg, P. (1990): "Undernutrition in Sub-Saharan Africa: Is There a Gender Bias?," *Journal of Development Studies*, 26(3), 469–486.

Thomas, D., and J. Strauss (1992): "Prices, infrastructure, household characteristics and child height," *Journal of Development Economics*, 39(2), 301–331.

Trapp, Erin, M., and J. Menken (2005): "Assessing Child Nutrition: Problems with Anthropometric Measures as a Proxy for Child Health in Malnourished Populations," Working Paper, Research Program on Population Processes, Institute of Behavioral Sciences, University of Colorado, Boulder.

Valdivia, M. (2004): "Poverty, health infrastructure and the nutrition of Peruvian Children," *Economics & Human Biology*, 2(3), 489–510.

Waterlow, J. Buzina, R., W. , Keller, J. , Lane, M. , Nichman, and J. Tanner (1977): "The presentation and Use of Height and Weight data for Comparing the Nutritional Status of Groups of Children under the Age of Ten years," *Bulletin of the World Health Organisation*, 55, 489–98.

WHO (1986): "Use and interpretation of anthropometric indicators of nutritional status.," Bulletin of the World Health Organisation. Geneva.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

World-Bank (2006): *Repositioning Nutrition as Central to Development: A Strategy for Large -Scale Action*. World Bank, The World Bank, Washington, DC 20433 USA.

Table 1: Summary Statistics

	National		Rural	
	Mean	SD	Mean	SD
Child characteristics				
Height-for-age $z-$score	-1.24	1.66	-1.41	1.71
male	-1.34	1.66	-1.52	1.70
female	-1.13	1.66	-1.28	1.71
Age (months)	25.57	14.97	25.62	15.05
Sex (1 = female)	0.47	0.49	0.47	0.49
Household characteristics				
Household size	13.04	6.14	13.79	6.10
Mother's age (years)	28.71	6.98	28.18	6.95
Expenditures per capita	178,001	215,552	119,796	98,809
Mother literate	0.23	0.42	0.12	0.32
Head polygamous	0.35	0.47	0.41	0.49
Ethnic group of head:				
Wolof	0.43	0.49	0.42	0.49
Pular	0.27	0.44	0.32	0.46
Serer	0.13	0.34	0.14	0.34
Diola	0.02	0.16	0.01	0.10
Other	0.12	0.32	0.09	0.29
Village characteristics				
Access to water	0.85	0.35	0.82	0.37
Access to health services	0.35	0.47	0.27	0.44
Access to Toilet = septic tank	0.03	0.19	0.001	0.04
Access to Toilet = flush with septic pit	0.21	0.41	0.05	0.22
Access to Toilet = water sealed	0.01	0.13	0.007	0.08
Access to Toilet = latrine	0.51	0.49	0.63	0.48
Quantiles of Height-of-age				
10% quantile	-4.08	0.43	-4.07	0.43
25% quantile	-3.32	0.74	-3.34	0.74
50% quantile	-1.69	0.27	-1.69	0.27
75% quantile	-0.70	0.32	-0.69	0.32
90% quantile	0.59	0.22	0.58	0.20
# of Observations	4,484		2,612	

Table 2: Prevalence of Child Malnutrition, Moderate stunting height-for-age $z-$score $\leq$ -2, Severe stunting height-for-age $z-$score $\leq$ -3

	National		Rural	
	Moderate	Severe	Moderate	Severe
Sex of child				
male	32.72	18.67	38.57	23.28
female	28.65	17.47	33.18	19.93
Children whose mothers were				
literate	22.55	13.37	30.24	19.61
in polygamy	33.49	19.42	37.20	21.05
Children living in village with				
Access to water	30.60	18.35	36.31	22.15
Access to health services	25.87	15.92	30.31	20.86
Region:				
Dakar	23.84	11.43	14.00	09.33
Ziguinchor	38.97	16.31	23.84	22.24
Diourbel	50.99	22.41	52.76	23.70
St Louis	16.51	14.60	17.50	17.00
Tamba	31.51	24.01	34.47	24.69
Kaolack	28.83	23.66	30.45	27.47
Thiès	36.40	18.87	60.13	24.67
Louga	25.42	17.58	23.09	14.72
Fatick	34.69	20.22	40.59	23.71
Kolda	44.35	17.79	45.55	17.37
Total	30.77	18.10	36.00	21.68

Table 3: Test of Equality of Distribution of Height for-age z-score

	National									
	Water					Health				
	Mean (sd)		H_0 : no difference in means	H_0 : equality of distributions		Mean (sd)		H_0 : no difference in means	H_0 : equality of distributions	
				Kolmogorov	Bartlett				Kolmogorov	Bartlett
	Yes	No	[$p-$value]	[$p-$value]	[$p-$value]	Yes	No	[$p-$value]	[$p-$value]	[$p-$value]
Full	−1.179	−1.238	−0.059	0.035	2.586	−1.058	−1.262	−0.203	0.065	4.481
	(1.632)	(1.709)	[0.397]	[0.443]	[0.108]	(1.593)	(1.669)	[0.000]	[0.000]	[0.034]
$Q10$	−4.071	−4.105	−0.033	0.126	2.949	−4.066	−4.082	−0.015	0.083	0.179
	(0.424)	(0.491)	[0.532]	[0.198]	[0.086]	(0.447)	(0.433)	[0.728]	[0.492]	[0.672]
$Q25$	−3.284	−3.360	−0.076	0.071	0.197	−3.246	−3.320	−0.074	0.072	0.670
	(0.743)	(0.762)	[0.199]	[0.363]	[0.657]	(0.727)	(0.755)	[0.122]	[0.133]	[0.413]
$Q50$	−1.683	−1.685	−0.001	0.054	0.146	−1.674	−1.689	−0.014	0.049	0.000
	(0.269)	(0.263)	[0.950]	[0.748]	[0.702]	(0.268)	(0.268)	[0.369]	[0.502]	[0.995]
$Q75$	−0.699	−0.705	−0.005	0.043	0.222	−0.686	−0.709	−0.022	0.051	0.003
	(0.328)	(0.319)	[0.843]	[0.939]	[0.637]	(0.327)	(0.326)	[0.261]	[0.458]	[0.955]
$Q90$	0.600	0.590	−0.010	0.064	0.265	0.591	0.604	0.013	0.050	0.104
	(0.223)	(0.212)	[0.725]	[0.961]	[0.606]	(0.219)	(0.223)	[0.530]	[0.930]	[0.747]
	Rural									
Full	−1.362	−1.323	0.039	0.047	3.143	−1.174	−1.419	−0.245	0.092	1.108
	(1.669)	(1.776)	[0.649]	[0.309]	[0.076]	(1.645)	(1.701)	[0.000]	[0.000]	[0.292]
$Q10$	−4.076	−4.051	0.024	0.102	1.414	−4.075	−4.069	0.005	0.085	1.196
	(.426)	(0.476)	[0.670]	[0.526]	[0.234]	(0.474)	(0.428)	[0.919]	[0.769]	[0.274]
$Q25$	−3.322	−3.378	−0.056	0.061	0.000	−3.259	−3.355	−0.096	0.086	0.004
	(0.750)	(0.749)	[0.398]	[0.706]	[0.985]	(0.751)	(0.748)	[0.129]	[0.219]	[0.949]
$Q50$	−1.681	−1.686	−0.004	0.060	0.022	−1.668	−1.686	−0.018	0.075	0.792
	(0.267)	(0.264)	[0.874]	[0.878]	[0.880]	(0.255)	(0.270)	[0.462]	[0.486]	[0.373]
$Q75$	−0.697	−0.677	0.019	0.081	0.000	−0.681	−0.698	−0.017	0.055	0.024
	(0.324)	(0.325)	[0.564]	[0.540]	[0.976]	(0.322)	(0.326)	[0.556]	[0.808]	[0.875]
$Q90$	0.587	0.610	0.023	0.108	0.002	0.587	0.594	0.007	0.104	1.501
	(0.205)	(0.204)	[0.481]	[0.705]	[0.964]	(0.188)	(0.213)	[0.811]	[0.590]	[0.220]

Note: Testing the null that the distributions of the response variables are identical between children in household with access and those without access to public infrastructure. Tests of the equality of means, Bartlett and Kolmogorov-Smirnov tests of the equality of the distributions.

Table 4: OLS and quantile regression of child height-for-age z-score, national

	OLS	National: Quantile regressions				
Dep.var:		10%	25%	50%	75%	90%
child height-for-age z-score	(1)	(2)	(3)	(4)	(5)	(6)
Access to water	0.022	0.215	0.143	−0.057	−0.032	−0.119
	(0.084)	(0.119)	(0.085)	(0.067)	(0.088)	(0.093)
Access to health services	0.081	0.105	0.133	0.107	0.043	−0.027
	(0.068)	(0.098)	(0.070)	(0.053)	(0.062)	(0.068)
Included control variables:						
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Village characteristics	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	4484	4484	4484	4484	4484	4484
R^2/Pseudo R^2	0.14	0.08	0.07	0.07	0.10	0.14

Note: OLS, standard errors clustered at the district level in parentheses. For quantile regression, bootstrapped standard errors clustered at the district level in parentheses obtained with 500 replications

Table 5: OLS and quantile regression of child height-for-age z-score, rural

	OLS	Rural Quantile regressions				
Dep.var:		10%	25%	50%	75%	90%
child height-for-age z-score	(1)	(2)	(3)	(4)	(5)	(6)
Access to water	0.028	0.351	0.150	−0.024	0.030	−0.096
	(0.105)	(0.149)	(0.133)	(0.088)	(0.112)	(0.127)
Access to health services	0.189	0.211	0.225	0.184	0.224	0.139
	(0.083)	(0.108)	(0.102)	(0.085)	(0.107)	(0.112)
Included control variables:						
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Village characteristics	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	2612	2612	2612	2612	2612	2612
R^2/Pseudo R^2	0.19	0.11	0.12	0.11	0.13	0.16

Note: OLS, standard errors clustered at the district level in parentheses. For quantile regression, bootstrapped standard errors clustered at the district level in parentheses obtained with 500 replications

Table 6: Quantile regression: Test of equality of coefficients between quantile, National/Rural

	Quantile regressions			
	National		Rural	
	$F(27, 4456)$	$p-$value	$F(27, 2582)$	$p-$value
Null hypothesis				
$Q10 = Q25$	0.92	0.57	0.80	0.76
$Q10 = Q50$	2.57	0.00	1.34	0.11
$Q10 = Q75$	5.83	0.00	3.41	0.00
$Q10 = Q90$	12.95	0.00	8.79	0.00
$Q25 = Q50$	4.21	0.00	2.40	0.00
$Q25 = Q75$	7.26	0.00	5.77	0.00
$Q25 = Q90$	19.68	0.00	11.61	0.00
$Q50 = Q75$	5.34	0.00	3.39	0.00
$Q50 = Q90$	15.48	0.00	9.16	0.00
$Q75 = Q90$	6.25	0.00	3.86	0.00
Joint test of equality of all	$F(108, 4454)$		$F(108, 2584)$	
slopes coefficients	7.51	0.00	5.36	0.00

Table 7: First stage estimates

Dep.var: child height-for-age z-score	National	Rural
Exclude Instruments		
Asset Index	0.135	0.085
	(0.009)	(0.016)
Land ownership	−0.167	−0.089
	(0.029)	(0.039)
Joint significance test		
of child characteristics	0.65	0.01
	[0.582]	[0.000]
household characteristics	57.60	27.75
	[0.000]	[0.000]
village characteristics	11.03	4.48
	[0.000]	[0.000]
Weak-instrument test		
Partial R^2	0.15	0.03
Partial F	129.90	16.08
Cragg-Donald	414.11	46.36
Critical value	19.93	19.93

Table 8: IV and IVQR estimations, National

	IV	National: IV Quantile regressions				
Dep.var:		10%	25%	50%	75%	90%
child height-for-age z-score	(1)	(2)	(3)	(4)	(5)	(6)
Access to water	0.023	0.272	0.130	-0.043	-0.050	-0.105
	(0.085)	(0.120)	(0.103)	(0.069)	(0.091)	(0.096)
Access to health services	0.076	0.102	0.146	0.103	0.037	-0.030
	(0.067)	(0.096)	(0.069)	(0.058)	(0.064)	(0.069)
Included control variables:						
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Village characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Test of the OID restrict.	1.444					
[$p-$value]	[0.229]					
Hahn-Hausman IV validity test						
m_3 test statistic	-0.505					
[$p-$value]	[0.612]					
# of Observations	4484	4484	4484	4484	4484	4484

OLS, standard errors clustered at the district level in parentheses. For quantile regression, bootstrapped standard errors clustered at the district level in parentheses obtained with 500 $\times$ 100 replications

Table 9: IV and IVQR estimations, Rural

	IV	Rural: IV Quantile regressions				
Dep.var:		10%	25%	50%	75%	90%
child height-for-age z-score	(1)	(2)	(3)	(4)	(5)	(6)
Access to water	0.030	0.329	0.142	-0.015	0.013	-0.140
	(0.107)	(0.143)	(0.122)	(0.092)	(0.113)	(0.135)
Access to health services	0.187	0.221	0.215	0.176	0.238	0.185
	(0.082)	(0.108)	(0.103)	(0.081)	(0.101)	(0.110)
Included control variables:						
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Household characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Village characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Test of the OID restrict.	0.069					
[$p-$value]	[0.792]					
Hahn-Hausman IV validity test						
m_3 test statistic	-0.015					
[$p-$value]	[0.988]					
# of Observations	2612	2612	2612	2612	2612	2612

Note: OLS, standard errors clustered at the district level in parentheses. For quantile regression, bootstrapped standard errors clustered at the district level in parentheses obtained with 500 $\times$ 100 replications

Table 10: IVQR: Test of equality of coefficients between quantiles, National/Rural

	National		Rural	
	$F(27, 4454)$	$p-$value	$F(27, 2584)$	$p-$value
Null hypothesis				
$Q10 = Q25$	1.14	0.28	0.77	0.79
$Q10 = Q50$	2.57	0.00	1.06	0.37
$Q10 = Q75$	6.00	0.00	3.21	0.00
$Q10 = Q90$	13.15	0.00	8.22	0.00
$Q25 = Q50$	3.90	0.00	2.18	0.00
$Q25 = Q75$	7.53	0.00	4.28	0.00
$Q25 = Q90$	20.95	0.00	10.93	0.00
$Q50 = Q75$	6.38	0.00	2.62	0.00
$Q50 = Q90$	18.20	0.00	8.40	0.00
$Q75 = Q90$	6.98	0.00	4.63	0.00
Joint test of equality of all	$F(108, 4454)$		$F(108, 2584)$	
slopes coefficients	8.95	0.00	4.73	0.00

Figure 1: Distribution of OLS and QR estimates, National

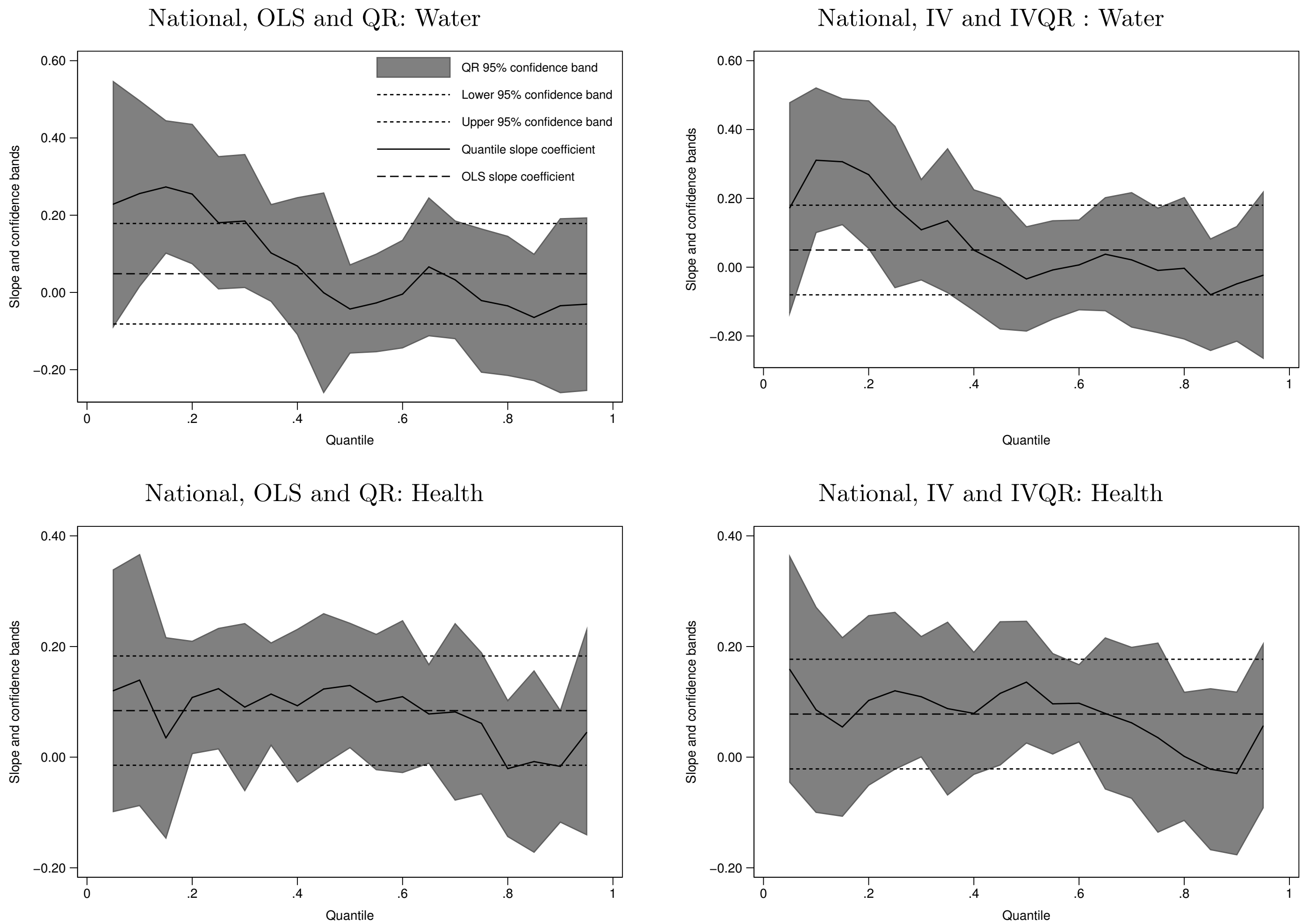

Figure 2: Distribution of OLS and QR estimates, Rural