

Water Hauling and Girls' School Attendance

Some New Evidence From Ghana

Céline Nauges

Jon Strand

The World Bank
Development Research Group
Environment and Energy Team
May 2013



Abstract

In large parts of the world, a lack of home tap water burdens households as the water must be brought to the house from outside, at great expense in terms of effort and time. This paper studies how such costs affect girls' schooling in Ghana, with an analysis based on four rounds of the Demographic and Health Surveys. Using Global Positioning System coordinates, it builds an artificial panel of clusters, identifying the closest neighbors within each round. The results indicate

a significant negative relation between girls' school attendance and water hauling activity, as a halving of water fetching time increases girls' school attendance by 2.4 percentage points on average, with stronger impacts in rural communities. The results seem to be the first definitive documentation of such a relationship in Africa. They document some of the multiple and wide population benefits of increased tap water access, in Africa and elsewhere.

This paper is a product of the Environment and Energy Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at jstrand1@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Water Hauling and Girls' School Attendance: Some New Evidence From Ghana*

Céline Nauges and Jon Strand

Key words: Household water access; girls' schooling; panel data; Africa.

JEL classification: O13; O15; O55; Q25.

Sector board: Water

*Nauges: The University of Queensland, School of Economics, St Lucia, 4072, QLD, Australia, e-mail: c.nauges@uq.edu.au. Strand: Development Research Group, Environment and Energy Team, the World Bank, e-mail: jstrand1@worldbank.org. We thank Michael Toman and Dominique Van De Walle for helpful comments. This research has been supported by a grant from the World Bank's Research Support Budget. Views expressed in this paper are those of the authors and do not necessarily represent the World Bank, its management or member countries.

Water hauling and girls' school attendance: some new evidence from Ghana

Céline Nauges and Jon Strand

“Women and girls are the “water haulers” of the world. On average, women and girls in developing countries walk 6 kilometers a day, carrying 20 litres of water, greatly reducing the time they have for other productive work or for girls to attend school.” UNICEF Children and water, global statistics.¹

1. Introduction

In large parts of the world, most households have no access to tap water at home. The lack of tap water deprives the household of a number of goods and amenities. First, as focused on in this study, the household is burdened as the water must be brought to the house from outside, at great expense of effort and time. Second, a number of potential uses of water, which are taken for granted by “modern” households, are then excluded, including sanitary and food preparation services, and operation of many appliances (washers, dishwashers, showers, etc.) by households with relatively higher incomes. Third, not having tap water access can be financially costly: convenient alternative services, such as water delivery from trucks, can be far more expensive than taking water from the tap inside the house. Fourth, while tap water is treated and made suitable for human consumption, the same is often not the case for water brought from the outside to the home; its quality is typically much more uncertain.

This study focuses on the first of these four factors. Traditionally in developing countries, water hauling is carried out largely by women and girls. This activity could easily take up a substantial fraction of those household members' time budget, given that the nearest or more relevant water source is far from the home. This could in turn have a number of undesirable social and economic consequences for the women in these households.

One such consequence could be to reduce the ability of children to attend school, depriving them of the education necessary to later obtain gainful employment and raise their economic and social status. The purpose of this paper is to gain understanding of this issue, by investigating further the question of water hauling and its relationship with school attendance for girls in Ghana (Africa). Sub-Saharan Africa is a key region for such a study since, on

¹ http://www.unicef.org/wash/index_31600.html

average for this region, only 5% of the rural population in the region gets water piped to the premises. It is also estimated that more than a quarter of the population in several countries in the region uses more than 30 minutes to make one water collection round trip (WHO-UNICEF, 2010).²

We base our data on four rounds of Demographic Health Surveys (DHS) for Ghana, which provide information on access to water infrastructure along with socio-demographic data for a large number of households. We use spatial identification through GPS coordinates to build a panel of 405 clusters/communities followed over four periods of time: 1993-94, 1998-99, 2003, and 2008. We use panel data techniques for fractional dependent variables to estimate the impact of time to haul water on girls' school attendance. Our findings indicate that reducing by half the time to haul water would increase the proportion of girls aged 5-15 who attend school by 2.4 percentage points on average, with stronger impacts in rural than in urban communities.

In Section 2 we discuss related work where the impact of infrastructure on children's school enrollment or work activities is measured. We describe the model and empirical strategy in Section 3. In Section 4 we present the data along with descriptive statistics for main variables of interest. Estimation methodology and results are discussed in Sections 5 and 6, respectively. Section 7 concludes.

2. Related literature

The role of infrastructure in development, and in particular for school attendance and educational achievement, has been the focus of a number of studies. A recent paper, closely related to ours, is by Koolwal and van de Walle (2013; forthcoming). These authors investigate the relationship between the time to walk one-way to the source of drinking water, participation in income earning market-based activities, and children's school attendance and health (as measured by anthropometric indices of growth status). Their empirical study uses cross-sectional surveys from Sub-Saharan Africa (Madagascar, Malawi, Rwanda, and Uganda), South Asia (India, Nepal, Pakistan), North Africa (Morocco), and the Middle East (Yemen). They find evidence that both boys' and girls' school enrollment is higher when access to water is better. More specifically, a one hour reduction in the time spent to walk to the water source increases girls' school enrollment rates by about 10 percentage points in

² In such situations, the quantity of water collected is often below five litres per capita per day (l/c/d) which implies that consumption requirements are likely not to be fulfilled and hygiene is not possible (unless it is practised at the source), Howard and Bartram (2003).

Yemen, and by about 12 percentage points in Pakistan. They do not find any significant effect in African countries, though. Interestingly, in their study the impact of better water access on school attendance is similar for boys and girls.

Nankhuni and Findeis (2004), using a cross-sectional survey of 10,698 households in Malawi, investigated whether time spent by children (aged 6-14) to collect water and fuelwood impacted on the likelihood of attending school. Each child in the survey was assigned the median value of hours spent on fuelwood and water collection for the area of residence (the survey covers 136 areas), which is likely to induce some measurement error. They found evidence that the probability that a given child is involved in fuelwood and water collection is reduced in households with more women, and more members beyond school age. They also found a significant relationship between time spent collecting fuelwood and school attendance so that “children in districts with severe fuel wood deficits (all of the south and most of the central region) are about 9% less likely to attend school than those from fuel wood surplus districts.” They however provided no direct measure of the impact of the time to collect water on school participation.

Ilahi and Grimard (2000) measured how the quality and quantity of water-supply infrastructure (calculated from the distance to the source of water averaged over all households in the community that report collection but excluding the household in question) affect the allocation of time among Pakistani women to various activities, using a cross-sectional survey from 1991 of Pakistani women above 15 years of age. Improvements in the water-supply infrastructure were here found to increase the time that women allocate to income-generating activities (although no magnitudes are given). Ilahi and Grimard did not study impacts on children’s school enrollment.

Lokshin and Yemtsov (2005) used community-level data from Georgia to assess the welfare impact of infrastructure rehabilitation projects (school rehabilitation, improvements in road infrastructure, and water system rehabilitation projects) in rural areas between 1998 and 2000. Using propensity score-matching and difference-in-difference methods, they measured the impact of water projects (including installing new or repairing existing communal water tanks, installing water treatment equipment, fitting new pumps, repairing or installing pipes, and rehabilitating wastewater management networks) on various outcomes including female wage employment and incidence of water-borne diseases. Their results indicate that water rehabilitation projects significantly reduce the incidence of water-borne diseases. The effect on female employment was however found not to be significant (which, they argue, could be explained by the low number of such projects in the sample).

Other work has measured similar impacts of electrification. Kularni et al. (2007) have found that higher electrification rates lead to better educational outcomes in Nicaragua and Peru. Barkat et al. (2002) similarly found that electrification leads to higher literacy rates and increased school enrollment in Bangladesh. These results are confirmed more recently by Asaduzzaman et al. (2010), who also found that school children's study times increase substantially for families who gain access to electricity. Asaduzzaman and Latif (2005) found that electricity demand by families with electricity access rises substantially with the number of school children. This suggests a strong effect on household electricity demand from school children's need for evening light access. The two latter studies however indicate that schooling and electricity access may be endogenous in a more comprehensive model, thus calling for caution in interpreting results. Kanagawa and Nakata (2008) found that literacy rates in the state of Assam, Pakistan, among children above six years of age would increase substantially, from 63 to 74%, given complete electrification of the state (which was claimed to be reachable by 2012). Grogan and Sadanand (2009) modeled how changes in home production technology might affect fertility and women's time use in Guatemala. Using difference-in-difference type estimators, household electrification substantially reduces cooking times and increases working outside the home among Guatemalan women (the probability of outside employment increases by nine percentage points). Dinkelman (2009) assesses the impact of the national electrification program in South Africa using two waves of community-data from rural areas in KwaZulu-Natal. Results indicate that female employment rises by a significant 9.5 percentage points in communities that receive an electricity project. This author also shows that electrification has larger effects on female employment in middle-poor communities and for women in their thirties and forties, who are less likely to live with young children requiring full-time care.

A thematically related paper, focusing as we do on children's school attendance in Ghana, is Lavy (1996), which is based on data from the Ghana Living Standards Survey in 1987-1988. The focus was here more on schooling costs, assumed to be a positive function of distance to the nearest public school. Greater distance to the school is found to reduce school attendance; more in secondary than in primary school.

Other related studies have dealt with social impacts of improved access to other types of public goods. Banerjee et al. (2009) considered the impact of railroads on wages in China. Akee (2006) estimated the effects of road construction on wage employment and agricultural employment in the Republic of Palau.

Overall, there is still relatively little reliable evidence on the socio-economic consequences of improved water access for households. The few existing articles on the topic, just reviewed, however all indicate a significant influence of increased water hauling times on either children's schooling or women's work. Further evidence is needed to quantify the magnitude of these effects, which motivates our study.

3. Model specification and empirical strategy

School attendance is known to be driven by a range of individual, household, and community characteristics. Sex and age of the individual, composition and size of the household, the education level and occupation of other household members, and household income and assets are factors typically taken into account. These factors tend to reflect or indicate household preferences for education, and their budgetary constraints. Children's school attendance may also depend on community characteristics such as distance to the school, and access to other infrastructure, more specifically to drinking water which is our focus. Identification of causal relationships between infrastructure and school attendance is however difficult. Confounding factors may induce a spurious correlation between infrastructure access and school enrollment: wealthier and more educated households are likely to have better water supply access, and at the same time have stronger preferences for their children to attend school. Certain household and/or individual characteristics may then simultaneously explain both water access and children's schooling. If some of these characteristics are not observed, it may lead to endogeneity bias. But one must also be aware that infrastructure spending is likely not random: it may be targeted either at growth centers, or at areas that lag behind, depending on policy objectives. Hence infrastructure can be endogenous, and community-level infrastructure quality correlated with other community characteristics (average household income, distance to school, road access, etc.) that may also determine school attendance. To identify a causal relationship between access to water and school attendance, one may then need to control for two possible sources of endogeneity bias.

Various approaches to this issue have been pursued in the literature. Dinkelman (2009) uses a community land gradient to instrument for project placement since this gradient was a primary factor in prioritizing areas for electrification. Koolwal and van de Walle (2013) incorporate geographical characteristics presumed to be correlated with infrastructure

placement.³ They argue that infrastructure placement can be presumed exogenous in the equations measuring women's work or school attendance, once these geographical characteristics have been included as additional explanatory variables.

We here adopt a novel approach, by building a (pseudo) panel of clusters/communities based on their spatial GPS coordinates. The use of panel data specific techniques will then allow us to account for cluster/community-specific unobserved effects and hence to control for the endogeneity of infrastructure placement.

In order to control for potential endogeneity bias due to unobserved individual/household characteristics, we propose to estimate school attendance at the cluster/community level, by calculating community averages from household data, along the lines of Dinkelman (2009) and Koolwal and van de Walle (2013). As argued by these authors, under the assumption that endogeneity arises from individual choices within communities, access to water can be treated as exogenous if community averages are used instead of household data.

4. Data and descriptive statistics

The Republic of Ghana (hereafter only Ghana) is a country in West Africa, on the Gulf of Guinea. Situated at only a few degrees northern latitude, it enjoys a warm climate throughout the year. Ghana is divided into ten administrative regions, and has a population of about 24 million (a map of the ten regions is provided in Appendix A). Gross National Income (GNI) was assessed at USD 1,190 per capita in 2009 (World Bank, 2011).⁴ Agriculture contributed about one-third of the country's gross domestic product. Although Ghana has been classified by the World Bank as a middle-income country, it struggles with a water deficit and widespread lack of sanitation. Access to an improved water source rose from only about half of the population in 1990 to 82% in 2008.⁵ By contrast, access to sanitation has increased from 7% to only 13% in nearly 20 years, one of the lowest rates in Africa (World Health Organization/UNICEF (2010)).

³ Among other factors, these authors consider access to roads, schools, banks, health centres and markets; price levels for food and other important commodities; male and female daily agricultural and non-agricultural wage rates.

⁴ The average GNI per capita in Sub-Saharan Africa was USD 1,126 in 2009. These figures were obtained from the document *Ghana at a glance* published by the World Bank in 2011 and available at http://devdata.worldbank.org/AAG/gha_aag.pdf.

⁵ "Access to an improved water source" refers to the percentage of the population with reasonable access to water from an improved source such as a household connection, public standpipe, borehole, protected well or spring, and rainwater collection. Unimproved sources include vendors, tanker trucks, and unprotected wells and springs. "Reasonable access is defined as the availability of at least 20 litres per person per day from a source within one kilometre of the dwelling"; see <http://data.worldbank.org/indicator/SH.H2O.SAFE.ZS>.

4.1. The dataset

We use data collected during four rounds of the DHS in Ghana: 1993-94 (5,822 households surveyed), 1998-99 (6,003 households), 2003 (6,251 households), 2008 (11,778 households).⁶ The four rounds include data for each household on the main source of drinking water and the time it takes to walk to this source, fetch water, and return. It is thus a per-round-trip measure. The number of trips, by each household per day, to collect water is not available in the data. Characteristics of all household members are also available, in particular age, sex and whether each member is or is not in school. The information available for each household however differs somewhat from one round to the other. Each household belongs to a unique cluster which is spatially identified by GPS coordinates data. There were 400 such clusters in 1993-94 and 1998-99, and 412 clusters in 2003 and in 2008.

These four rounds do not represent a pure panel, as the households surveyed in these four rounds are not identical. However, and as will be explained in the next paragraph, the GPS coordinates data still allow us to build a panel of clusters.

4.2. Using the GPS coordinates to build a panel of communities

The DHS data provide GPS coordinates of the surveyed clusters in each of the four rounds (1993-94, 1998-99, 2003 and 2008). For the 2008 sample, GPS coordinates are missing for seven clusters, leaving us with a total of 405 GPS-identified clusters for that year. For each of these clusters we identify the closest neighbor (belonging to the same region) in the other three rounds by calculating great distance circles between all pairs of clusters.⁷ Table 1 reports the average distance (in miles) between two matched clusters in each region.⁸ For example, in the Ashanti region, the average distance between a cluster surveyed in 2008 and the nearest cluster surveyed in 2003 is 3.4 miles. The average distance between matched clusters is small in regions such as Ashanti and Greater Accra with large numbers of clusters, varying here between 1.1 miles and 3.4 miles. The average distance between matched clusters is greater in the North where there are fewer clusters (about 14 miles in the Northern region).

4.3. Descriptive statistics

In Table 2 we report the number of surveyed households, the percentage of households with the main source of drinking water on the premises (in dwelling or in yard), the average

⁶ For greater details on the sampling procedure in each of the four rounds, see the Ghana DHS final reports available at <http://www.measuredhs.com/>.

⁷ The great circle distance is the shortest distance between any two points on the surface of a sphere.

⁸ All Tables 1-7, and Figures 1-2, are found in Appendix F.

time to haul water (for households without source on plot), and the percentage of girls aged 5-15 attending school, for each of the four study periods. Greater details by type of source and region are provided in Appendices B and C.

The analysis of the four DHS rounds indicates that about 18% of surveyed households have access to water in the residence or in the yard, either through a piped access or through a well (see Table 2).⁹ For these households the time spent to haul water is less than one minute (per round of fetching water). For households without any water access on premises, the average round-trip time to the source is between 18 and 23 minutes depending on the survey year. The most important sources located outside the residence are public taps, used by 20% to 27% of the surveyed households (varying by year), boreholes or public wells (29-39% of households), and surface water such as rivers and streams (11-26% of households), see Appendix B. The share of households who rely on surface water has decreased over time. In general these households also spend more time hauling water (20 to 30 minutes) than households relying on other types of sources.

Water infrastructure and average time spent hauling water vary significantly across the ten regions (see Appendix C). The share of households with access to water on the premises varies between 48% and 56% in Greater Accra (the capital region), while it can be less than 10% in Volta, Brong-Ahafo, Northern, Upper West and Upper East regions. The density (histograms) of the time variable in each region (for the 2008 DHS round) is shown in Appendix D. This regional pattern is similar in all four study periods.

School attendance for girls aged 5 to 15 years also varies significantly depending on the type of water source the household relies on and across regions. The higher school enrollment observed for girls living in households with water access in the residence cannot be considered as evidence of a causal relationship, since it could in principle be spurious. As noted, households with good water supply access are likely to also be wealthier and more educated, and to have stronger preferences for their children to attend school.

The higher proportion of girls attending school in 2008 (Table 2) might have been the consequence of the Five Year Action Plan for Girls' Education in Ghana, implemented by the Ghana Education Service from 2003 to 2008. Among other actions, scholarships were offered to girls at Junior and Senior Secondary School levels; and incentives were provided for

⁹ The percentage of households having access to water in residence or in the yard can only be calculated for the first three rounds of the survey. In 2008 the list of sources does not distinguish between wells in residence and well outside the residence. The figure shown in the table (21%) corresponds to the proportion of households reporting a time spent to haul water equal to 0 minute.

female teachers to teach in rural areas and sensitize students, parents, and community members on girls' education (U.S. State Department).¹⁰

5. Estimation methodology

Our data set consists of a four-year panel of 405 clusters. Our dependent variable is the average share of girls aged 5-15 who attend school in the community, measured as a fraction between zero and one. Standard linear regression models are not well suited for such estimation since they may produce predicted values greater than one. We instead follow a method proposed by Papke and Wooldridge (2008), to deal with fractional response dependent variables in a panel data context. This approach will allow us to control for cluster/community unobserved heterogeneity possibly correlated with the explanatory variables, in particular water infrastructure (in our case, time to the water source). We impose bounds on the proportion of girls attending school by using a Probit functional form for the mean proportion. This gives us the following model:

$$E(s_{it} | \mathbf{x}_{it}, c_i, v_{it}) = \Phi(\mathbf{x}_{it} \boldsymbol{\beta} + c_i + v_{it}) \quad (1)$$

where $i = 1, \dots, C$ represent clusters, $t = 1$ to 4 the survey year, and $\Phi(\cdot)$ the standard normal cumulative distribution function. The dependent variable, s_{it} , represents the proportion of girls aged 5-15 attending school, where $0 \leq s_{it} \leq 1$. The \mathbf{x}_{it} vector is the set of explanatory variables, c_i stands for cluster-specific unobserved heterogeneity, possibly correlated with some explanatory variables, and v_{it} is the time-varying error term. Following Papke and Wooldridge (2008) we make the following assumption along the lines of Chamberlain (1980):

$$c_i = \psi + \bar{\mathbf{x}}_i \boldsymbol{\xi} + a_i \quad (2)$$

where $\bar{\mathbf{x}}_i \equiv \frac{1}{4} \sum_{t=1}^4 \mathbf{x}_{it}$ is the vector of cluster means (calculated over the four time periods) and

$$a_i | \mathbf{x}_i \sim \text{Normal}(0, \sigma_a^2).$$

Inserting (2) into model (1) yields:

¹⁰ See <http://www.state.gov/g/drl/rls/hrrpt/2008/af/119004.htm>

$$E(s_{it} | \mathbf{x}_i, a_i, v_{it}) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\xi + a_i + v_{it}). \quad (3)$$

The cluster-specific term, a_i , is here assumed to be independent of \mathbf{x}_i . The parameters in model (3) are identified up to the positive scale factor $(1 + \sigma_a^2)^{1/2}$ (Papke and Wooldridge, 2008):

$$E(s_{it} | \mathbf{x}_i, v_{it}) = E[\Phi(\mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\xi + a_i + v_{it}) | \mathbf{x}_i] = \Phi\left[\frac{(\mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i\xi + v_{it})}{(1 + \sigma_a^2)^{1/2}}\right] \quad (4)$$

Chamberlain (1980)'s approach allows any possible correlation between cluster-specific unobserved heterogeneity and the explanatory variables to be eliminated. However the latter may also be correlated with the time-varying unobservable v_{it} .

The Rivers and Vuong (1988) control function approach allows us to test and correct for such endogeneity. Assume in particular that one of our explanatory variables, called y_i , is endogenous (extension to more than one endogenous variable is straightforward). Assume also that we have some instruments \mathbf{z}_i , which are not among the set of explanatory variables \mathbf{x}_i in model (3). The control function approach implies that we, in a first stage, estimate a linear reduced form for the endogenous variable y_i :

$$y_{it} = \psi_2 + \mathbf{x}_{it}\delta_1 + \mathbf{z}_{it}\delta_2 + \bar{\mathbf{x}}_i\eta_1 + \bar{\mathbf{z}}_i\eta_2 + u_{it} \quad (5)$$

for $i = 1, \dots, C$ and $t = 1$ to 4. Here too we adopt Chamberlain's approach to control for unobserved cluster-specific effects by including in the model the variables $\bar{\mathbf{x}}$ and $\bar{\mathbf{z}}$ where

$\bar{\mathbf{z}}_i \equiv \frac{1}{4} \sum_{t=1}^4 \mathbf{z}_{it}$. Under the assumption that v_{it} given u_{it} is conditionally normal:

$$v_{it} = \rho u_{it} + e_{it}, \quad (6)$$

$$e_{it} | \mathbf{z}_i, u_{it} \square \text{Normal}(0, \sigma_e^2), \quad (7)$$

the (final) version of the model to be estimated in the second stage is as follows:

$$E(s_{it} | \mathbf{x}_i, u_{it}, v_{it}) = \Phi \left[\frac{(\mathbf{x}_{it} \boldsymbol{\beta} + \psi + \bar{\mathbf{x}}_i \boldsymbol{\xi} + \rho u_{it} + v_{it})}{(1 + \sigma_e^2)^{1/2}} \right] \quad (8).$$

In practice model (5) will be estimated first by Ordinary Least Squares (OLS) in order to obtain the residuals \hat{u}_{it} , which are then used in place of u_{it} in model (8). In the second stage, model (8) is estimated using the pooled Bernoulli quasi-Maximum Likelihood Estimator (QMLE), which corresponds to maximizing the pooled probit log-likelihood.¹¹ Because of the panel form of the data one may want to allow for any form of serial dependence across t .¹² Finally, the two-stage estimation procedure should correct for biases in the standard errors. Bootstrap methods are used in the empirical application.

The parameter estimates can then be used to perform several specification tests. In particular, rejection of the null hypothesis $\rho = 0$ will indicate that y_{it} is endogenous; while the joint significance of the $\boldsymbol{\xi}$ parameters will confirm that the observed time-varying explanatory variables and the cluster-specific unobserved effect are correlated.

All parameter estimates are scaled by the same factor which can be calculated as follows:

$$(CT)^{-1} \sum_{i=1}^C \sum_{t=1}^4 \phi(\mathbf{x}_{it} \hat{\boldsymbol{\beta}} + \hat{\psi} + \bar{\mathbf{x}}_i \hat{\boldsymbol{\xi}} + \hat{\rho} u_{it}) \quad (9)$$

where $\phi(\cdot)$ is the density of the normal distribution (Papke and Wooldridge, 2008). This scale factor is unique and corresponds to the scale effect averaged across all time periods and all cross-sections. Hence the average partial effects of the explanatory variables in \mathbf{x}_{it} (that is, the partial effects averaged across the population) are obtained by multiplying the scale factor from (9) by the corresponding estimated parameters $\hat{\boldsymbol{\beta}}$.

6. Empirical analysis

In what follows we describe the estimation of model (8) using data for 405 clusters over four years (1993-94, 1998-99, 2003 and 2008). The dependent variable is the average proportion of girls aged 5-15 attending school in the community/cluster.

¹¹ The Bernoulli log-likelihood function is given by $l_i(b) \equiv y_i \log[G(\mathbf{x}_i \mathbf{b})] + (1 - y_i) \log[1 - G(\mathbf{x}_i \mathbf{b})]$ for $0 < G(\cdot) < 1$; see Papke and Wooldridge (1996).

¹² This model can be estimated using the *glm* command in Stata and the *cluster* option can be used to obtain standard errors robust to any form of serial correlation.

6.1. Description of the variables

The final set of explanatory variables (all community-averages) includes the average time to haul water (in minutes), the average household size, the average number of children below five years of age, the average number of children aged 5 to 15, the average number of women aged 16 to 65, the average number of men aged 16 to 65, the average proportion of male household heads, year and regional dummies, and dummies to control for the month of interview. The education of the household head and the household's wealth index have not been included in the model because of their strong correlation with the variable measuring hauling time (the coefficient of correlation with hauling time is -0.38 and -0.53, respectively).¹³ In 7% of the surveyed clusters, the average time to the water source is zero. In order to take this particular feature of the data into account we include a dummy variable which takes the value 1 if and only if the time to the source is 0, following Battese (1997).

For each cluster, community-averages are calculated using data on households in which at least one girl aged 5 to 15 is present (with no girl aged 5 to 15, the proportion of such girls attending school must be zero). In order to control for possible selection bias, the total number of children aged 5 to 15 is treated as endogenous and instrumented in a first-stage regression. The identifying instruments are the age of the household head (linear and squared versions of this variable) and a dummy variable indicating whether the household head is a widow(er) or is divorced. We also consider the time to haul water as potentially endogenous, and use as identifying instruments the dummy variable indicating whether the household head is a widow(er) or is divorced, whether the household lives in a rural or urban community, and the one-period lagged hauling time. The place of residence (rural or urban community) is assumed exogenous in the sense that water infrastructure was likely not the main factor driving households' choice in terms of residence location. We also argue that average hauling time in the community in the previous period of observation is a good instrument for current hauling time, since the two should be highly correlated, while past-period hauling time should not explain average school attendance in the current period. These two regressions are estimated using OLS but controlling for unobserved cluster-specific effects (Chamberlain,

¹³ The wealth index is provided in the DHS surveys. It is calculated using principal components analysis based on data concerning the household's ownership of a number of consumer items such as a television and car; dwelling characteristics such as flooring material; type of drinking water source; toilet facilities; and other characteristics that are related to wealth status. The resulting asset scores are standardized in relation to a standard normal distribution with a mean of zero and a standard deviation of one. These standardized scores are then used to create the break points that define wealth quintiles as: Lowest, Second, Middle, Fourth, and Highest, see <http://www.measuredhs.com/>.

1980). The estimated residuals are then used as additional explanatory variables to fit girls' school attendance. The use of lagged hauling time as instrument implies that we lose all first-round observations (1998-99). The final model is estimated on a sample of 1,212 observations.

The definition of all variables used in the first- and second-stage models, along with some simple statistics, are shown in Table 3.

6.2. Estimation results

The model is estimated using QMLE following the procedure described in Section 5. We use a bootstrap procedure with 500 replications in order to calculate robust and efficient two-stage standard errors. Estimation results are shown in Table 4.¹⁴ The magnitude of the estimated coefficients is not directly interpretable and partial effects for the main variables of interest will be discussed later.

We find that an increase in the time to haul water lowers the proportion of girls 5 to 15 attending school. Specification tests have revealed that the impact of hauling time on girls' school attendance was not constant across all communities, in particular it was found to depend on the time needed to collect water. Testing different thresholds (10 minutes, 20 minutes, and 30 minutes), we found evidence that the estimated coefficient of the time variable was significantly different for communities where less than 20 minutes was spent on average to haul water (coefficient estimated at -0.010) and for communities where more than 20 minutes was spent per collection round trip (-0.014). The estimated impact is stronger in magnitude and significant only in communities where average hauling time is greater than 20 minutes. These results indicate some sort of threshold effect (whereby the effect of additional hauling time is "strong" and/or "significant" only when hauling time is already above a particular threshold). If so, reduced time costs of water hauling would have little effect on girls' school attendance if these costs are relatively low at the outset (less than 20 minutes). This may have a reasonable explanation: most households could have a "discretionary" budget for water hauling time costs which does not upset other main activities of the household. If so, such disruptions (including not sending girls to school) only occur for ranges of hauling costs that are particularly high.

¹⁴ The estimation results of the first-stage regressions are not shown here but are available upon request.

The composition of the household also plays a role, as expected. A higher number of children below the age of five reduces the proportion of girls who attend school (although not significantly), perhaps because school-age girls are in charge of the younger children. We also find evidence that more members in other demographic groups (children aged 5 to 15, women and men aged 16 to 65) increases school attendance, once household size is controlled for. These results may indicate that when more family members contribute to the household's income, the proportion of girls attending school increases. School attendance for girls is found to be significantly lower for households headed by a male, which may be explained by female heads having stronger preferences for their daughters to attend school. Various regional dummies are found significant as well as dummies for the month of interview. Finally, the null hypothesis that the coefficients of all cluster means are jointly equal to zero cannot be rejected. The first-stage residuals are jointly significant (chi-squared test statistic: 31.75, p-value: 0.000), thus confirming that the number of children aged 5 to 15, and the time to haul water, are endogenous.

6.3. Calculation of partial effects

We calculate partial effects of the main variables of interest following the procedure established by Papke and Wooldridge (2008), see Table 5.

In order to assess the magnitude of the effect of time to haul water on girls' school attendance, we calculate by how much school attendance would increase if times to haul water were reduced by 50% in each community, compared to actual hauling times in 2008. We then find that school attendance would increase on average by 2.4 percentage points across the population, but that the expected increase is lower than 1 percentage point in half of the communities (median partial effect). The density of the calculated partial effect is shown in Figure 1.

Because the estimated impact on school attendance depends positively on the average time to haul water in the community, the magnitude of the effect is larger in rural than in urban areas. More precisely, a reduction by half of the time to haul water would on average increase girls' school attendance by 1 percentage point in urban communities (the median effect is 0.4 percentage points), and by 3.5 percentage points in rural communities (the median is 1.9 percentage points). In Figure 2, we show the estimated partial effect of a 50% reduction in the

time to haul water as a function of collection time as measured in 2008, separately for rural and urban communities.

The estimated impact also varies significantly by region (Table 6). The median effect of a halving of collection time on school attendance is small (one percentage point or less) in the following regions: Western, Central, Greater Accra, Volta, Eastern, Ashanti and Brong Ahafo. In these regions hauling times per round trip are low, less than 17 minutes on average. In the north of the country (Northern, Upper East, and Upper West regions), where hauling times per round trip are longer, 25 to 30 minutes on average, the estimated impact on girls' school attendance is in the range 5-6 percentage points.

The magnitude of the effect may seem strong in the Northern regions, in particular when compared to what was found by Koolwal and van de Walle (2013). By their estimation, a one-hour reduction in time to the water source would lead to a 10 percentage point increase in school enrollment in Yemen and Pakistan. In our case, we find that a 50% reduction in hauling time (corresponding to roughly 15 minutes reduction in time to the water source in the Northern, Upper East and Upper West regions of Ghana) would increase median school attendance among girls in these regions by 5-6 percentage points. Both methodological approach and study area differ between the two studies, though. Importantly also, our measure of hauling time is a per-trip measure as our data contain no information on households' numbers of daily trips to their water sources.

6.4. The case of boys

Water fetching is usually described as an activity undertaken primarily by women and girls. However we test if there is any impact of water infrastructure also on boys' school attendance. We consider boys aged 5 to 15, and use hauling time as the main variable of interest, using as instruments the one-period lagged hauling time, the dummy variable indicating whether the household head is a widow(er) or is divorced, and households' place of residence (rural or urban community). Model (8) is estimated on a sample of 1,174 observations. Estimation results are shown in Table 7. Interestingly the effects of hauling time on school attendance are almost exactly the same for boys as for girls, a result which is in line with findings described in Koolwal and van de Walle (2013). However the impact of household size and composition on school attendance is very different for boys and girls. Girls living in larger households are less likely to attend school in general, whereas household size does not affect boys' overall school attendance. We however find that the higher the

number of children 5 and under, the lower the proportion of boys attending school. The same is observed for girls but the corresponding coefficient is not significant. The number of children 5 to 15 and the number of adults impact girls' school attendance only and having a male as the household head reduces the proportion of girls attending school but has no effect on boys' school attendance. We also observe some differences in terms of regional effects and effects of the month of interview on the proportion of boys attending school.

6.5. Robustness checks

In this section we perform a number of robustness checks. First, in order to check the robustness of our results to the choice of instruments for hauling time, we consider the one-period lagged proportion of households having access to water in the residence/yard as instrument instead of the one-period lagged hauling time. We obtain results similar to the ones reported in Table 4. Hauling times lower than 20 minutes have a negative but non-significant effect on girls' school attendance, whereas hauling times greater than 20 minutes significantly impact the proportion of girls attending school. The corresponding coefficient is -0.009, slightly lower in magnitude than before (-0.014, see Table 4), but not statistically different from the coefficient reported in Table 4.

Second, we replace hauling time as the main variable of interest by the proportion of households having access to water in their residence or their yard (this proportion is 19% on our sample). We would expect to find a significant and positive coefficient for this variable in the model, in terms of explaining the proportion of girls attending school. We instrument this variable using the one-period lagged proportion of households with such access. As expected the corresponding coefficient is positive and significant at the 10% level of significance (0.314, with standard error 0.173).

Third, in order to check the robustness of our cluster matching procedure, we consider a case where two clusters can be matched only if the distance between them is at most five miles. By imposing this constraint, the final sample of matched clusters is reduced to a total of 1,136 observations (instead of 1,617 with no limit on distances between matched clusters). The regional composition of the full and restricted samples is shown in Appendix E (Table E.1).

The weight of some regions has increased in the restricted sample (e.g., Greater Accra from 14.4% to 19.2%, Ashanti from 16.6% to 19.5%). These regions had a better than average water infrastructure (in the sense that the average time to the source was on the whole shorter than for other regions). On the other hand, the weight of the regions in which households have

to spend more time hauling water on average (e.g., Northern, Upper East, Volta) is reduced in the restricted sample. This is not surprising knowing that the average distance between matched clusters was higher in these regions (Table 1). Some statistics on the distribution of the time to the water source in both samples are given in Table E.2 (Appendix E).

In general the estimated coefficients are of the same sign and of similar magnitude in the two models, but fewer coefficients are significant in the restricted sample. In particular the coefficient for hauling time greater than 20 minutes is found not significant at usual levels of significance. This result is probably driven by the lack of variability of hauling times, which itself is a consequence of the removal of the poorest clusters from the restricted sample.

We perform a final robustness check by excluding the last round of observations (2008). As discussed earlier, there was a significant increase in the proportion of girls attending school in this year (Table 2). It is thus important to check that our main findings are not driven (mainly) by what happened at the end of the period. Estimated coefficients of all variables are similar as the ones shown in Table 4. The coefficients of the time variables are found negative and significant. For hauling time greater than 20 minutes, the estimated parameter is now somewhat greater in absolute value, -0.20 (whereas it was -0.14 when the 2008 data were included), not statistically different from the parameter estimated on the full sample.

7. Conclusion

Using four rounds of the DHS from Ghana, we provide some new insights into the relationship between households' water access and girls' school attendance in this African country. Our main methodological innovation is to build a panel of clusters based on GPS coordinates. The latter is combined with the use of panel data techniques suited to the analysis of fractional response data allows us to control for endogeneity of infrastructure placement but also for unobservable cluster-specific effects. As far as we know our paper is the first to find evidence of a statistically significant relationship between time to the water source and girls' school attendance for an African country. Our results indicate that a 50% reduction in the time to haul water would increase the proportion of girls 5 to 15 attending school by 2.4 percentage points on the average, with stronger effects for rural communities. Household composition is also found to be an important determinant of girls' school attendance.

We also find evidence of "threshold" effects, in the sense that the discouraging effect of higher water hauling costs on girls' schooling is particularly salient when this cost is relatively high at the outset (above 20 minutes). We argue that this could be explained

naturally by households having a minimum, “discretionary”, time budget for water hauling, within which other important activities are not disrupted. When this time budget is overrun, however, impacts (here, in girls’ schooling propensity) are more serious.

While it could be argued from principal grounds that effects of water hauling times on school attendance should be most significant for girls, we extend our analysis also to boys aged 5-15, and find quite similar effects. This confirms the findings of Koolwal and van de Valle (2013), who found that the impact of water hauling times on school attendance was indeed similar for boys and girls, for a wider set of countries. Note however that ours are the first such results for an African country. Koolwal and van de Valle, while including several African countries in their study, found no significant effects of water hauling variables on children’s schooling, for any of these.¹⁵

The quantitative findings reported in this paper are specific to Ghana. There is however reason to believe that similar relations might hold for other African countries; although this is left to be shown. Our results should serve to help policy makers and donors in identifying, and quantifying, the potential and actual benefits of water infrastructure improvement in Ghana, and we believe, in the rest of Africa and beyond. A natural extension is to test if there is a significant relationship between time to the water source and women’s time at work or decision to work outside the home.

¹⁵ The African countries included in the Koolwal and van de Walle study were Madagascar, Malawi, Rwanda and Uganda.

References

- Akee, R., 2006. The Babeldaob Road: The Impact of Road Construction on Rural Labor Force Outcomes in the Republic of Palau. IZA Discussion Paper No. 2452.
- Asaduzzaman, M., and A. Latif. 2005. Energy for Rural Households: Towards a Rural Energy Strategy in Bangladesh. Bangladesh Institute of Development Studies, Dhaka.
- Asaduzzaman, M., Barnes, D. and Khandker, S. 2010. Restoring Balance: Bangladesh's Rural Energy Realities. Working Paper No. 181, The World Bank.
- Banerjee, A., Duflo, E., and Qian, N., 2009. On the Road: Access to Transportation Infrastructure and Economic Growth in China. Working Paper, Department of Economics, MIT.
- Battese, G.E. 1997. A note on the estimation of Cobb-Douglas production functions when some explanatory variables have zero values. *Journal of Agricultural Economics* 48(2): 250-252.
- Barkat, A., M. Rahman, S. Zaman, A. Podder, S. Halim, N. Ratna, M. Majid, A. Maksud, A. Karim, and S. Islam. 2002. Economic and Social Impact Evaluation Study of the Rural Electrification Program in Bangladesh. Report to National Rural Electric Cooperative Association (NRECA) International, Dhaka.
- Chamberlain, G., 1980. Analysis of variance with qualitative data. *Review of Economic Studies* 47: 225-238.
- Dinkelman, T., 2009. The Effects of Rural Electrification on Employment: New Evidence from South Africa. Working Paper, Princeton University.
- Grogan, L., and Sadanand, A., 2009. Electrification and the Household. Working Paper, University of Guelph.
- Howard, G., and Bartram, J., 2003. Domestic Water Quantity, Service Level and Health. Geneva, World Health Organization.
- Ilahi, N., and Grimard, F., 2000. Public Infrastructure and Private Costs: Water Supply and Time Allocation of Women in Rural Pakistan. *Economic Development and Cultural Change* 49(1): 45-75.
- Kanagawa, M. and Nakta, T., 2008. Assessment of Access to Electricity and the Socio-Economic Impacts in Rural Areas of Developing Countries. *Energy Policy* 36, 2016-2029.
- Koolwal, G., and van de Walle, D., 2013. Access to Water, Women's Work and Child Outcomes. *Economic Development and Cultural Change*, 61(2), January 2013.

- Kularni, V., Barnes, D. and Parodi, S., 2007. Rural Electrification and School Attendance in Nicaragua and Peru. Working Paper, The World Bank.
- Lavy, Victor, 1996. School Supply Constraints and Children's Educational Outcomes in Rural Ghana. *Journal of Development Economics* 51 (2): 291-314.
- Lokshin, M., and Yemtsov, R., 2005. Has Rural Infrastructure Rehabilitation in Georgia Helped the Poor? *The World Bank Economic Review* 19(2): 311-333.
- Nankhuni, F.J., and Findeis, J.L., 2004. Natural resource-collection work and children's schooling in Malawi. *Agricultural Economics* 31: 123-134
- Papke, L.E. and Wooldridge, J.M., 1996. Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates. *Journal of Applied Econometrics* 11: 619-632.
- Papke, L.E. and Wooldridge, J.M., 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* 145: 121-133.
- Rivers, D., Vuong, Q.H., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 39: 347-366.
- WHO-UNICEF, 2010. Progress on Sanitation and Drinking Water: 2010 update.

Appendices

Appendix A: Map of Ghana (source: DHS report on Ghana, 2008)



Appendix B: Time to haul water and school attendance, by source of drinking water

Table B.1. 1993-94 DHS round

Source of drinking water	# of households	% of households	Time to haul water (min)	Girls 5-15 in school (%)
piped into residence	851	15%	0	74%
well in residence	95	2%	1	60%
public tap	1,211	21%	11	72%
public well	624	11%	16	63%
borehole	1,071	18%	20	61%
spring	43	1%	31	47%
river, stream	1,488	26%	20	61%
pond, lake	110	2%	23	61%
dam	151	3%	40	38%
dugout	78	1%	19	67%
rainwater	48	1%	10	64%
tanker truck	52	1%	22	81%

Table B.2. 1998-99 DHS round

Source of drinking water	# of households	% of households	Time to haul water (min)	Girls 5-15 in school (%)
piped into residence	941	16%	0	80%
well in residence	93	2%	0	71%
public tap/neighbours house	1,254	21%	13	78%
public well	691	12%	22	59%
borehole	1,431	24%	23	55%
spring water	55	1%	29	85%
river/stream	1,007	17%	31	56%
pond/lake	98	2%	30	56%
dam	234	4%	37	44%
dugout	120	2%	30	63%
rainwater	38	1%	0	79%
tanker truck	37	1%	18	54%
bottled water	2	0%	0	-
other	1	0%	5	-

Table B.3. 2003 DHS round

Source of drinking water	# of households	% of households	Time to haul water (min)	Girls 5-15 in school (%)
pipied into dwelling	336	5%	0	78%
pipied into compound/plot	591	9%	0	72%
open well in dwelling	39	1%	0	63%
open well in yard/plot	81	1%	0	49%
protected well in dwelling	37	1%	0	73%
protected well in yard/plot	85	1%	0	71%
public tap	1,258	20%	14	68%
open public well	633	10%	18	55%
protected public well	1,795	29%	18	57%
spring	50	1%	21	42%
river, stream	909	15%	25	53%
pond, lake	112	2%	30	40%
dam	149	2%	36	39%
rainwater	16	0%	0	75%
tanker truck	63	1%	23	69%
bottled water	1	0%	0	-
satchel water	81	1%	0	73%
other	8	0%	14	100%

Table B.4. 2008 DHS round

Source of drinking water	# of households	% of households	Time to haul water (min)	Girls 5-15 in school (%)
pipied into dwelling	687	6%	0	91%
pipied to yard/plot	868	7%	0	90%
public tap/standpipe	3,179	27%	12	88%
tube well or borehole	3,781	32%	22	80%
protected well	558	5%	11	87%
unprotected well	271	2%	24	82%
protected spring	8	0%	19	71%
unprotected spring	77	1%	33	45%
river/dam/lake/ponds/stream/canal	1,256	11%	23	74%
rainwater	56	0%	0	92%
tanker truck	62	1%	21	92%
cart with small tank	16	0%	23	77%
bottled water	43	0%	2	88%
sachet water	914	8%	6	86%
other	1	0%	12	-

Appendix C: Time to haul water and school attendance, by region

Table C.1. 1993-94 DHS round

Region	Time to haul water (min)	Source in residence (#)	Source in residence (%)	Girls 5-15 in school (%)
Western	15	75	14%	69%
Central	14	68	10%	73%
Greater Accra	5	365	52%	73%
Volta	17	22	4%	75%
Eastern	13	84	11%	74%
Ashanti	11	200	19%	69%
Brong-Ahafo	16	34	6%	70%
Northern	29	40	9%	34%
Upper west	28	15	9%	40%
Upper east	25	43	14%	42%

Table C.2. 1998-99 DHS round

Region	Time to haul water (min)	Source in residence (#)	Source in residence (%)	Girls 5-15 in school (%)
Western	15	54	9%	75%
Central	11	63	12%	75%
Greater Accra	6	425	56%	77%
Volta	33	31	6%	73%
Eastern	14	129	17%	79%
Ashanti	17	188	20%	71%
Brong-Ahafo	26	22	5%	68%
Northern	37	46	10%	29%
Upper west	21	21	5%	41%
Upper east	22	55	11%	35%

Table C.3. 2003 DHS round

Region	Time to haul water (min)	Source in residence (#)	Source in residence (%)	Girls 5-15 in school (%)
Western	20	77	13%	69%
Central	18	61	13%	67%
Greater Accra	5	400	48%	74%
Volta	19	64	13%	66%
Eastern	13	91	14%	60%
Ashanti	14	257	24%	69%
Brong-Ahafo	16	85	12%	61%
Northern	19	71	12%	38%
Upper west	18	29	6%	46%
Upper east	21	34	9%	45%

Table C.4. 2008 DHS round

Region	Time to haul water (min)	Source in residence (#)	Source in residence (%)	Girls 5-15 in school (%)
Western	16	115	10%	85%
Central	9	85	8%	84%
Greater Accra	5	575	34%	87%
Volta	16	62	6%	89%
Eastern	10	115	9%	87%
Ashanti	9	359	19%	94%
Brong-Ahafo	13	36	3%	89%
Northern	31	49	5%	55%
Upper west	24	95	12%	85%
Upper east	28	64	7%	77%

Appendix D: Density of the time to haul water, by region (in minutes). 2008 round



Appendix E. Robustness of the cluster matching procedure

Table E.1. Weight of each region in the full and restricted samples

Region	Full sample		Restricted sample	
	Freq.	Percent	Freq.	Percent
Western	147	9.1	86	7.6
Central	136	8.4	107	9.4
Greater Accra	232	14.4	218	19.2
Volta	132	8.2	77	6.8
Eastern	172	10.6	117	10.3
Ashanti	268	16.6	221	19.5
Brong Ahafo	151	9.3	93	8.2
Northern	147	9.1	66	5.8
Upper East	118	7.3	67	5.9
Upper West	114	7.1	84	7.4
Total	1,617	100.0	1,136	100.0

Table E.2. Distribution of the variable 'time to the source' in the full and restricted samples

	First quartile	Median	Third quartile	Mean
Full sample	5.5	12.4	21.1	15.6
Restricted sample	3.6	9.9	18.8	13.3

Appendix F: Further tables and figures

Table 1. Average distance (in miles) between matched clusters

Region	# of clusters in 2008	2008-2003 cluster match (mile)	2008-1998/99 cluster match (mile)	2008-1993/94 cluster match (mile)
Ashanti	67	3.4	3.1	3.4
Brong-Ahafo	38	6.3	7.4	6.8
Central	34	5.3	4.0	4.0
Eastern	43	7.9	4.5	3.9
Greater Accra	58	1.1	1.5	2.9
Northern	37	13.5	14.0	13.5
Upper East	30	4.7	4.3	5.0
Upper West	28	6.4	7.4	9.9
Volta	33	6.1	6.1	6.5
Western	37	7.8	6.3	7.5

Table 2. Summary statistics on water infrastructure and school attendance

Survey year	Number of households	HH ^a with source in dwelling or yard (%)	Average time to haul water for HH without source on plot (min)	Girls 5-15 in school (%)
1993-94	5,822	17%	18	65
1998-99	6,003	18%	23	63
2003	6,251	18%	19	59
2008	11,778	21%	19	84

^a 'HH' stands for 'households'.

Table 3. List and definition of variables (405 clusters and 1,617 observations)

Variable definition	Mean	Std. Dev.	Min	Max
Share of girls aged 5 to 15 attending school	0.70	0.27	0.00	1.00
Time to the water source (min)	16	14	0	124
Time to water source is 0 min (0/1)	0.07	0.25	0.00	1.00
Household size	5.75	1.40	2.50	14.00
Number of children 5 and under	0.96	0.55	0.00	4.00
Number of children 5 to 15	2.42	0.67	1.00	7.00
Number of women 16 to 65	1.43	0.43	0.50	3.40
Number of men 16 to 65	1.00	0.47	0.00	3.67
Household head is a male (0/1)	0.63	0.28	0.00	1.00
Age of the household head	47	6	28	69
Household head is widow or divorced (0/1)	0.19	0.18	0.00	1.00
Household lives in a rural zone (0/1)	0.55	0.50	0.00	1.00

Table 4. QMLE estimation results (1,212 observations)

Dependent variable: proportion of girls 5-15 attending school	Coef.	Bootstrap Std. Err.	P-value
Constant	1.808***	0.214	0.000
Time to the water source (min) (when time lower than 20 min)	-0.010	0.006	0.115
Time to the water source (min) (when time greater than 20 min)	-0.014***	0.005	0.007
Time to water source is 0 min	0.137	0.098	0.161
Household size	-0.668***	0.234	0.004
Number of children 5 and under	-0.098	0.160	0.540
Number of children 5 to 15	0.952***	0.279	0.001
Number of women 16 to 65	0.787***	0.242	0.001
Number of men 16 to 65	0.765***	0.232	0.001
Household head is a male (0/1)	-0.280**	0.142	0.049
<i>Cluster means (Chamberlain's approach)</i>			
Household size – mean	-0.067	0.083	0.422
Number of children 5 and under – mean	0.097	0.148	0.514
Number of women 16 to 65 – mean	-0.107	0.166	0.522
Number of men 16 to 65 – mean	0.245	0.160	0.125
Household head is a male (0/1) – mean	-0.583**	0.254	0.022
<i>Residuals of first stage regression</i>			
Residuals (Number of children 5 to 15)	-1.568***	0.308	0.000
Residuals (Time to the water source)	0.010*	0.005	0.055
<i>Year dummies (2008 as ref.)</i>			
1998-99	-0.892***	0.111	0.000
2003	-0.738***	0.065	0.000
<i>Regional dummies (Greater Accra as ref.)</i>			
Western	0.102	0.099	0.303
Central	0.099	0.112	0.379
Volta	0.250**	0.106	0.018
Eastern	0.068	0.092	0.460
Ashanti	0.140*	0.074	0.057
Brong Ahafo	-0.069	0.086	0.421
Northern	-0.259**	0.126	0.041
Upper East	-0.123	0.117	0.292
Upper West	-0.084	0.097	0.385
<i>Dummies for the month of interview (October as ref.)</i>			
January	0.333**	0.136	0.015
February	0.201	0.171	0.239
July	-0.420***	0.100	0.000
August	-0.209**	0.081	0.010
September	-0.249***	0.062	0.000
November	0.079	0.109	0.471
December	0.407***	0.143	0.004
Scale factor	0.290		

Chi 2 test that all cluster means are equal to 0	7.93	P-value: 0.1603
Endogeneity test (Chi 2)	31.75***	P-value: 0.0000

Note: ***, **, * indicate significance at the 1, 5, and 10% level, respectively.

Table 5. Average Partial Effects for the main variables of interest

	APE	Bootstrap Std. Err.	P-value
Time to the water source (min) (when time lower than 20 min)	-0.003	0.002	0.115
Time to the water source (min) (when time greater than 20 min)	-0.004***	0.001	0.007
Time to water source is 0 min	0.040	0.029	0.161
Household size	-0.194***	0.068	0.004
Number of children 5 and under	-0.029	0.047	0.540
Number of children 5 to 15	0.277***	0.081	0.001
Number of women 16 to 65	0.228***	0.070	0.001
Number of men 16 to 65	0.222***	0.067	0.001
Household head is a male (0/1)	-0.081**	0.041	0.049

Note: ***, **, * indicate significance at the 1, 5, and 10% level, respectively.

Table 6. Estimated partial effect on proportion of girls aged 5-15 in Ghana attending school, due to a 50% reduction in water hauling time, by region

Region	Time to haul water (mean)	Estimated partial effect		
		Median	Minimum	Maximum
Western	16	0.01	0.00	0.06
Central	9	0.01	0.00	0.06
Greater Accra	4	0.00	0.00	0.06
Volta	17	0.01	0.00	0.11
Eastern	10	0.01	0.00	0.04
Ashanti	10	0.01	0.00	0.08
Brong Ahafo	13	0.01	0.00	0.16
Northern	30	0.06	0.00	0.34
Upper East	25	0.05	0.00	0.10
Upper West	30	0.06	0.01	0.21

Table 7. QMLE estimation results (1,174 observations)

Dependent variable: proportion of boys 5-15 attending school	Coef.	Bootstrap Std. Err.	P-value
Constant	2.529***	0.288	0.000
Time to the water source (min) (when time lower than 20 min)	-0.011	0.008	0.196
Time to the water source (min) (when time greater than 20 min)	-0.015**	0.007	0.025
Time to water source is 0 min	0.332**	0.128	0.010

Household size	0.030	0.281	0.914
Number of children 5 and under	-0.586***	0.184	0.001
Number of children 5 to 15	0.064	0.344	0.852
Number of women 16 to 65	-0.031	0.288	0.914
Number of men 16 to 65	0.104	0.284	0.713
Household head is a male (0/1)	-0.140	0.197	0.478
<i>Cluster means (Chamberlain's approach)</i>			
Household size – mean	-0.300***	0.103	0.004
Number of children 5 and under – mean	0.299	0.190	0.117
Number of women 16 to 65 – mean	0.256	0.205	0.211
Number of men 16 to 65 – mean	0.389**	0.182	0.032
Household head is a male (0/1) – mean	-0.317	0.357	0.374
<i>Residuals of first stage regression</i>			
Residuals (Number of children 5 to 15)	-0.555	0.385	0.149
Residuals (Time to the water source)	0.012*	0.007	0.061
<i>Year dummies (2008 as ref.)</i>			
1998-99	-0.563***	0.115	0.000
2003	-0.596***	0.072	0.000
<i>Regional dummies (Greater Accra as ref.)</i>			
Western	0.142	0.141	0.315
Central	-0.024	0.151	0.876
Volta	0.029	0.147	0.845
Eastern	0.016	0.136	0.909
Ashanti	0.309**	0.127	0.015
Brong Ahafo	0.105	0.135	0.437
Northern	-0.130	0.152	0.392
Upper East	-0.260*	0.138	0.059
Upper West	-0.286**	0.137	0.036
<i>Dummies for the month of interview (October as ref.)</i>			
January	0.095	0.146	0.513
February	0.111	0.183	0.544
July	-0.516	0.342	0.131
August	-0.163	0.113	0.148
September	-0.214***	0.070	0.002
November	0.084	0.111	0.452
December	0.245	0.154	0.111
Scale factor	0.259		
Chi 2 test that all cluster means are equal to 0	10.94*	P-value:	0.0526
Endogeneity test (Chi 2)	5.98*	P-value:	0.0502

Note: ***, **, * indicate significance at the 1, 5, and 10% level, respectively.

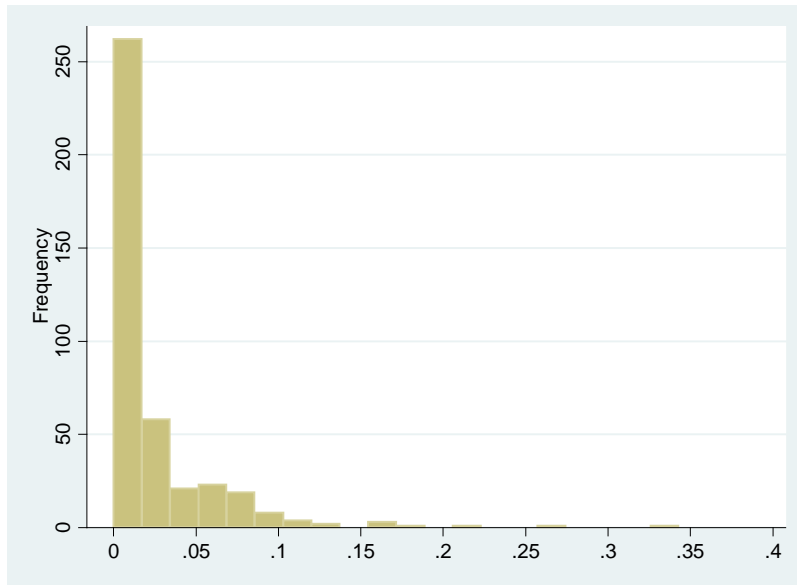


Figure 1. Estimated partial effect of a 50% reduction in the time to haul water on girls' school attendance (histogram)

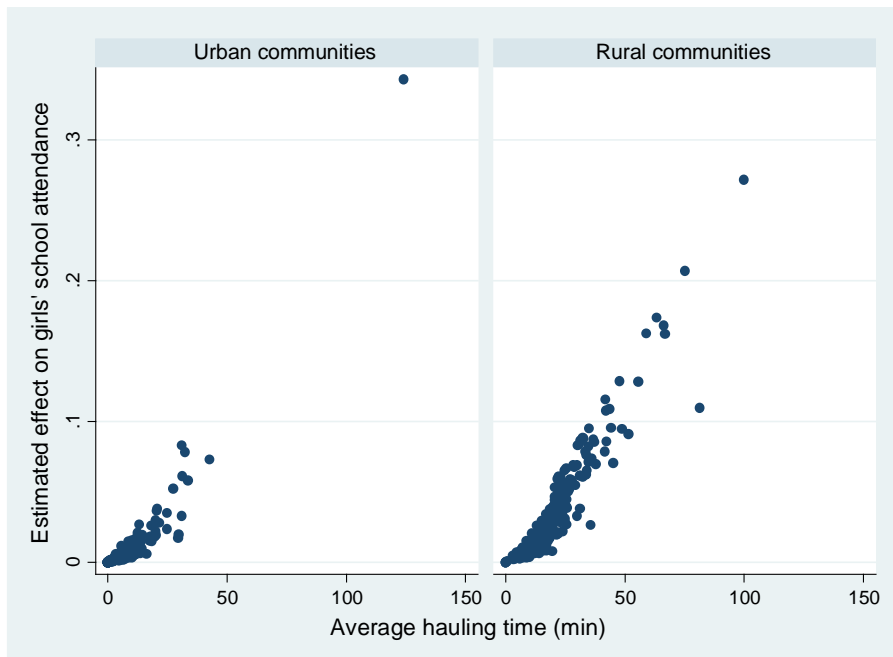


Figure 2. Estimated partial effect as a function of collection time, for rural and urban communities