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

Céline Nauges (Toulouse School of Economics)
and Jon Strand (DECEE, World Bank)

May 26, 2011

Outline of the talk

- Introduction
- Related (academic) literature
- Presentation of the data
- Empirical strategy
- Estimation methodology and main results
- Conclusion

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)

- Because of lack of tap water access...
  - time and effort are needed to bring water to the home
  - volumes of water may be inadequate to support basic personal hygiene and food preparation
  - getting water can be financially costly (e.g. tanker trucks)
- Water hauling is often the responsibility of women and girls
- Possible social and economic consequences: women's work, girls' school attendance

This paper

- Purpose: to investigate further the relationship between water hauling and girls' school attendance
- Tools: econometric analysis using micro data from four Demographic and Health Surveys (DHS) rounds in Ghana
- Main result: a 15-minute reduction in hauling time would increase the proportion of girls aged 5-15 attending school by 8 to 12%

Related literature

(i) Koolwal and van de Walle (WP, 2010)

- cross-sectional micro data from 9 countries (4 from SSA)
- relationship between time to the water source and a) working decisions, b) school attendance and c) children's health
- main finding: a one hour reduction in time to water increases girls' enrolment rate by about 10% in Yemen and 12% in Pakistan
- no significant relationship between time to water and school attendance in African countries

Related literature (cont'd)

(ii) Ilahi and Grimard (EDCC, 2000)

- cross-sectional data (1991) for Pakistani women above 15 years of age
- relationship between distance to water source and women's time allocation
- main finding: a shorter distance increases the time that women allocate to income-generating activities

Related literature (cont'd)

(iii) Nankhuni and Findeis (Ag. Eco., 2004)

- cross-sectional data (1997-98) from Malawi ($\simeq$ 10,000 HH)
- relationship between resource collection (fuel wood + water) and school attendance
- main finding: "children in districts with severe fuel wood deficits are about 9% less likely to attend school than those from fuel wood surplus districts"
- also: hours of collection work impact more significantly girls than boys
- HH composition, education of the HH head and wealth are significant determinants

Related literature (cont'd)

Positive relationship between electrification rates and...

- educational outcome (Kularni, Barnes and Parodi, 2007, in Nicaragua and Peru)
- literacy rate and school enrolment (Asaduzzaman, Barnes and Khandker, 2010, in Bangladesh; Kanagawa and Nakata, 2008, in Pakistan)
- women's time at work (Grogan and Sadanand, 2009, in Guatemala; Dinkleman, 2009, in South Africa)

+ studies on the impact of road construction, railroad, etc.

Presentation of the data

- Ghana: West African country, on the Gulf of Guinea
- Population: 24 million, GNI per capita: USD 1,190
- 10 administrative regions (poorer regions in the North)
- Access to an improved water source: 82 percent in 2008
- Sanitation coverage: 13 percent


GHANA
Burkina Faso
Upper East
Upper West
Northern
Togo
Cote d'Ivoire
Brong-Ahafo
Volta
Ashanti
Eastern
Western
Central
Greater Accra
Gulf of Guinea
0 30 60 120 Kilometers

- Four rounds of the Demographic and Health Surveys: 1993-94, 1998-99, 2003, 2008
- 5,822 HH surveyed in 1993-84 ... 11,778 HH surveyed in 2008
- Each HH belongs to a unique cluster, identified by GPS coordinates data
- 400 clusters in first two rounds and 412 clusters in last two rounds
- Surveyed HH and clusters differ from one round to the other
- Data on main drinking water source, time to haul water, school attendance and socio-demographic characteristics of each member

Summary statistics on water infrastructure and school attendance

Survey year	HH with source on premises (%)	Av. collection time (minutes)	Girls 5-15 in school (%)
1993-94	17%	18	65
1998-99	18%	23	63
2003	18%	19	59
2008	n.a.	19	84

- About 18% of HH have access to water in residence or in the yard (tap or well)
- For the rest: public taps (20-27%), boreholes or public wells (29-39%), surface water (11-26%)
- For HH without access on premises, the average round collection time is 18-23 minutes
- Significant variation in terms of water infrastructure and school attendance across the 10 regions

Heterogeneity across regions (2003 DHS round)

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

Purpose: to identify the impact of water infrastructure on school attendance

Two potential endogeneity biases:

- Infrastructure placement (community level): infrastructure spending is not random in general and may be targeted either to growth centres or towards areas that are lagging behind
- Spurious correlation between water infrastructure and school attendance (household level): wealthier and more educated households are likely to have a better access to water and to have stronger preferences for education

- Identification strategy (bias related to infrastructure placement):
  - instrumental variables (e.g., land gradient for electrification in Dinkleman, 2009)
  - geographical characteristics supposed to be correlated with infrastructure placement (Koolwal and van de Walle, 2010)
  - panel data of communities/clusters (this paper) - panel data specific methodologies allow us to control for cluster-specific unobserved characteristics
- Identification strategy (bias related to households' preferences): the model is estimated at the community level by calculating community-averages from household data (Dinkleman 2009, Koolwal and van de Walle 2010, this paper)

GPS coordinates are used to build a panel of communities

- We have a total of 405 GPS-identified clusters in 2008
- We identify the closest cluster in each of the other three rounds by calculating great distance circles
- We obtain a panel of 405 clusters followed over 4 periods of time
- The average distance between two matched clusters varies from 1.1 miles to 14.0 miles, depending on the region
- In general, higher density of clusters in Southern (wealthier) regions

- Panel of 405 communities over 4 years (1,617 observations overall)
- We calculate community-averages of all variables
- Dependent variable: average proportion of girls aged 5-15 attending school in the community
- Independent variables:
  - Round time to the water source (minutes)
  - Wealth index (1 to 5)
  - Number of children 5 and under
  - Number of women 16 to 65
  - Number of men 16 to 65
  - HH head is a male (0/1)
  - Regional dummies + dummies for the month of interview

Specific issues to deal with:

- (i) The dependent variable is a proportion bounded by 0 and 1
- (ii) Unobserved cluster-specific effects may be correlated with some explanatory variables
- (iii) Endogeneity of the time to haul water and the number of children aged 5 and under

Solution: Papke and Wooldridge (JoE, 2008)

- (i) Probit functional form
- (ii) Chamberlain (1980)'s approach: cluster-specific effects written as a linear function of cluster means
- (iii) Control function approach: first-stage regressions with excluded instruments

Proportion of girls attending school - conditional mean equation:

$$E(s_{it}|x_{it}, c_i, v_{it}) = \Phi(x_{it}\beta + c_i + v_{it})$$

- s_{it}: the proportion of girls 5-15 attending school in community i and time t
- x: the vector of explanatory variables
- c_i: unobserved cluster-specific effects
- $\Phi(.)$: the standard normal cumulative distribution function.

Following Chamberlain (1980), we make the following assumption:

$$c_i = \psi + \bar{x}_i\xi + a_i$$

where $\bar{x}_i \equiv \frac{1}{4}\sum_{t=1}^{4} x_{it}$ is the vector of cluster means and $a_i|x_i \sim Normal(0, \sigma_a^2)$.

After some transformations...

- Main equation - proportion of girls attending school:

$$(1) \quad E(s_{it}|x_i, u_{it}, v_{it}) = \Phi[\psi_a + x_{it}\beta_a + \bar{x}_i\xi_a + \rho_a u_{it} + v_{it}]$$

where u is the residual of the first-stage regression, and the subscript a denotes division of the original coefficient by $(1 + \sigma_a^2)^{1/2}$.

- First-stage regression: to control for endogeneity of variable y_{it}:

$$(2) \quad y_{it} = \psi_2 + x_{it}\delta_1 + z_{it}\delta_2 + \bar{x}_i\eta_1 + \bar{z}_i\eta_2 + u_{it}$$

where z is the vector of (excluded) instruments.

- Assumptions: $v_{it} = \rho u_{it} + e_{it}$ and $e_{it}|z_i, u_{it} \sim Normal(0, \sigma_\varepsilon^2)$.

- First stage: estimation of equation (2) using OLS
- Second stage: estimation of equation (1) with residuals from first-stage model using QMLE (i.e. maximization of the pooled probit log-likelihood)
- Bootstrapped standard errors for the second-stage equation
- Specification tests (endogeneity tests): $\xi = 0$ and $\rho = 0$
- Calculation of Average Partial Effects for each explanatory variable

- Correlation of unobserved cluster-specific effects with explanatory variables
- Endogeneity of the time to haul water and the number of children aged 5 and under
- Factors driving school attendance:
  - Time to the water source: (-)***
  - Wealth index: (+)**
  - Number of children less than 5: (-)***
  - Number of children 5 to 15: (-)**
  - Number of men 16 to 65: (+)**
  - HH head is a male: (-)*
  - + some dummies for year, region, month of interview

Average partial effects:

- Time to the water source: -0.0080***
- Wealth index: 0.0248**
- Number of children less than 5: -0.0670***

Impact on the proportion of girls 5-15 attending school:

- a 15-minute reduction in collection time: $+12\%$ $(-0.0080 \times 15 = 0.12)$
- a marginal increase in the wealth index: $+2.5\%$
- a marginal increase in the number of children aged 5 and under: -6.7%

Sensitivity analysis

- Is the partial effect of fetching time on school attendance the same regardless of the distance to the source?
- We test different threshold effects: 10 min, 20 min, 30 min
- Main finding: a 15-minute reduction in collection time increases school attendance...
  - by 8% when time to the source is lower than 20 minutes
  - by 12% when time to the source is greater than 30 minutes
  - the difference between the two is statistically significant

Sensitivity analysis (cont'd)

- New rule: two clusters can be matched only if the distance between them is $\leq$ 5 miles
- Final sample: 1,136 observations
- The weight of the poorest regions is lower $\Rightarrow$ less variation in collection time
- Main findings:
  - A 15-minute reduction in collection time increases school attendance by 9%
  - Wealth index still positive but no longer significant
  - Comparable effect for the number of children 5 and under

Conclusion

- Statistical evidence of a significant relationship between hauling time and girls' school attendance
- Main finding: a 15-minute reduction in collection time increases proportion of girls attending school by 8-12%
- HH composition and wealth are also important
- Threshold effects: the impact of hauling time on school attendance is stronger above 30 minutes
- Possible extensions:
  - Duplicate the analysis for boys
  - Relationship between hauling time and women's activities

Thank you for your attention