DRAFT – FOR SUBMISSION TO WORLDBANK/IZA 2008 CONFERENCE

Providing Education for Adolescent Girls in China – The Power of Water

Maimaiti Yasheng[1] W Stanley Siebert
University of Birmingham

Abstract:

This study investigates why dropout rates from secondary education are high for girls in China and brings forward a variable which is novel for economists: access to water. We hypothesise that girls' education suffers when their greater water need for female hygiene purposes with the onset of periods is not met. For testing we use six waves of the China Health and Nutrition Survey, 1989 - 2004. We find that the onset of periods is indeed associated with a marked increase in the secondary school dropout rate, but only for girls with poor access to water.

1. Introduction

In the developing world, the education of girls falls far behind that of boys, with widespread evil ramifications, reducing child health and womens' earnings (see, e.g., Schultz, 2002), and even women's role within the political system, since there is a strong link between education and political participation (Hannum and Buchmann, 2005). It is therefore important to find out why women are failing to be educated properly in developing counties, and our paper seeks to make a contribution with a special dataset from rural China. In particular, we hypothesise that girls' education suffers when there is a lack of adequate water and sanitation both at home and at school. Lack of water for girls to clean themselves during their periods can lead to them staying away from school (Burrows et al, 2004). It is also possible that lack of water affects girls more than boys because girls "traditionally" do the water-carrying, which takes time out of school. Thus, we test whether, and to what extent, lack of water affects the education of girls more than boys. We find large effects.

The argument that lack of water has particular adverse effects on girls education has been put forward before (Bista, 2004, Nahar, 2006; Kirk and Sommer, 2006; Singh, 1999; Snel, 2005), but data limitations have prevented a test of the magnitude of the effect. We use six waves of the China Health and Nutrition Survey (CHNS), 1989 – 2004, to make the test. The CHNS dataset is well-adapted for this inquiry, because it has information on the household's access to water, as well as detailed information about time spent on household tasks such as water-carrying, and information about the onset of periods. It also has variables measuring family resources and income, which are important determinants of education, and clearly need to be controlled for the test. Admittedly, the dataset does not contain information about availability of clean water at the school which respondents attend, and school water facilities can counteract bad home facilities (Nahar, 2006). Nevertheless, any finding of impacts from household access to water is of interest in itself.

We restrict the analysis to children in rural areas, since access to water is primarily a rural problem in China: over 90% of families in urban China have direct

[1] Corresponding author – mxy548@bham.ac.uk

access to clean (safe) water compared to only 47% in rural areas. However, as of 2005, nearly 60% of the Chinese population live in rural areas. Clearly, our investigation will be relevant to women's education in other rural communities in Africa, Asia, and South America to which most of the literature on watrer and sanitation currently relates.

As we will see, a survival analysis of the duration of children's schooling supports the hypothesis. We find that poor household access to water has a significant adverse impact on both boys and girls education, but the impact is much worse for girls, beginning after the onset of their periods. There is no adverse effect of periods for girls having good household access to water.

The paper is organised as follows. A literature review is presented in section 2. Section 3 describes relevant costs associated with poor access to water. Section 4 presents the model and its empirical specification. We describe the data used in this study in section 5. The empirical results are given in section 6. Concluding remarks appear in section 7.

2. Gender education gaps in China

In China, primary and secondary education takes 12 years to complete, divided into primary, junior secondary and senior secondary stages. In general, primary education lasts 6 years. At junior secondary stage, most have 3 years schooling. The 9-year schooling period in primary and junior secondary schools pertains to 'compulsory' education. General senior secondary education lasts a further 3 years.

In secondary schools in rural China, substantial gender gaps exist, which has generated much research (Conelly & Zheng 2003, Song et al. 2005, Brown & Park 2002, Hannum 1998, 2005). The literature on the reasons behind lower female educational attainment can be categorised as follows:

Opportunity cost: Some researchers argue that girls drop out of school more during their middle school, because their opportunity cost of staying in school is higher compared to boys. According to Li and Tsang (2002), in the past two decades the transition to a market oriented economy has allowed many privately owned enterprises to hire a large number of young female workers with limited education in the manufacturing and service sectors, especially in the booming coastal cities. Furthermore, rural villages and towns have developed various small-scale factories and enterprises that hire young women with limited education. These developments raise the opportunity cost of sending girls to school considering the earnings that they must forego. Connelly & Zheng (2003) and Song et al. (2005) also argue that girls have a higher opportunity costs due to the rise of light manufacturing jobs. In addition, Knight & Li (1996) argue that girls' education has a higher opportunity cost, since "traditionally" girls are family helpers. However, this argument should be true for girls from households with good access to water as well as with bad access, so long as we control for other determinants of education.

Household income and spending on education: this variable is important for education, as theorised by Gary Becker in his work on the economics of the family

(Becker, 1981). However, the existing literature didn't come to a solid conclusion about the effects of income and spending on the creation of gender gaps in education in rural China. Brown & Park (2002) show the importance of household income in determining educational outcomes, but there is no indication in their empirical findings that poverty affects girls more severely than boys. Connelly and Zheng (2003) find county per capita income has larger effects for girls than boys for initial enrolment of primary schools (the significance of the difference is not provided), but the difference of the effect has disappeared in junior secondary school where the real gender gaps of school enrolment start to appear. In addition, Song et al. (2005) didn't find any statistically different coefficients in terms of household educational spending on girls vs. boys. Yueh (2006) using data from China's urban household survey, finds that household income seems not to matter for primary school enrolment, but there are conflicting effects for boys and girls at the various secondary stages. So the final outcome is unclear. Nevertheless, household income is important for our investigation to ensure that our access to water variable is not simply picking up the poverty of families which have poor water access, and therefore we lay emphasis on controlling carefully for it.

Future earnings and dependence: Song et. al. (2005) argue that patrilocal marriage traditions in China mean that the long term returns on investments in daughters are more likely to be realised by marital, rather than natal, families, while the reverse is true for sons. Thus, the education of a son is likely to be perceived as a necessary investment for support in old age. Admittedly they find that educating males appears only to bring small long term benefits in terms of household income - an extra year of education raises future expected income by only 0.9%. Still, these benefits are positive, and should predispose parents to prefer to educate sons. Li & Tsang (2002) also argue that, in rural Chinese society, married daughters are expected to contribute to the husband's family – which again means a son's education should be more profitable. It is true that the contrary seems to be the case in urban China (Yueh 2006) where favourable assortative mating may generate high returns from investment in daughters. However, we are interested in the rural areas. Again, this argument should be true for girls from households with good access to water as well as with bad access.

Family characteristics: higher status of parental education and occupation is in general expected to have a positive impact on children's education, because educated parents in good jobs have direct experience of the benefits of schooling, which may make them develop a 'taste' for their children's schooling (Song et al., 2005, Emerson & Souza, 2007). Of course, higher parental education and better occupation will also mean higher household income, so our control for these variables can also serve to control for income.

Siblings: Different number of siblings and different sibling structure may also have different impacts on parents' education decisions (Conelly & Zheng 2003, Yang 2006, Tsui et al. 2002). While the findings remain mixed, we agree that detailed controls for sibling structure are necessary.

Geographical location: Children living in remote areas lack nearby schools, adequate transportation and information. All these may have negative impacts on their school enrolment. And the negative impact may be bigger for girls, especially for their secondary school enrolment. However, Connelly & Zheng (2003) find no evidence that living in a hilly county has a significant negative impact on school attendance of girls in all school levels. Li and Tsang (2002) assert that safety concerns clarify why parents from remote rural settings have lower educational expectation for their daughters – schools are simply too far away, given the poor transport, giving boys the advantage. But they provide no direct measure to test this hypothesis. We discuss below whether our access to water variable is picking up this location effect – since poor access is likely to overlap with remote geography. However, our controls for family income, and also for travel time to school will hopefully sweep out this effect.

'Culture': Many researchers address traditional and cultural aspects in China that have potential impacts on the school enrolment of boys and girls. Li & Tsang (2002) describes how "families without sons are recorded as having died out". This rigid lineage system, along with the economic, social and political advantages of educating a son discussed earlier, generates the concept of pro-son bias in schooling decisions. Song et al. (2006) argue that a son's education is more of an 'investment' good in rural China, whereas a daughter's is often taken as 'consumption', even 'luxury' good. Moreover, Knight and Li (1996) find that while traditional values favouring boys' education appear to have been eroded in urban areas, they have not in rural areas – which are our focus.

The difficulty with arguments relying on culture is that we need to know where the "culture" itself comes from. As Yueh (2006) argues, cultural traditions are often the historical product of practical necessity, so that rational acts under one set of circumstances, such as in rural China, will change when the context is altered, as in urban China with a different set of household needs and constraints. In any case, the pro-son bias, if it exists in the rural areas, will exist irrespective of access to water. Any finding that girls with good access to water are not disadvantaged compared to boys will cause arguments based on culture to fail.

In sum, previous research aiming at addressing gender gaps in education in China has come to mixed conclusions. The CHNS provides a large dataset to re-investigate this inquiry, and with detailed information on individual, household and community characteristics relevant to our research. In particular, with the CHNS, we have special data on access to water and personal hygiene including the onset of periods. Both factors are likely to be important for girls' schooling, as we now describe.

3. New variables: Access to water and girls' periods

Much economic and educational research has been conducted in terms of access to clean water and its direct and indirect impacts on girls' schooling in other developing countries where having access to clean water remain a big social problem (Behrman et al. 1999; Holmes 2003; Colclough et al. 2000; Hill & King 1995; Davison et al. 1992). In rural China, access to clean water has also been regarded as a major social issue. For example, at least 300 million rural residents in China have no access to safe, clean drinking water in the year 2006 (source: Xinhua, Wednesday, August 30, 2006). Will schooling outcomes of girls in rural China be similarly affected by poor access to clean water as found for other rural communities (Burrows et al, 2006)? Here we categorise the literature which links access to water with girls' education.

First, time costs generated by fetching water may have adverse impacts on girls' schooling as girls are considered to be the main water-fetchers in many African and South Asian Countries (Hill & King 1995; Behrman 1997). The production of schooling involves time costs, so does fetching water. But are girls in rural China also considered to be the main water-fetchers? If that is the case, is the distance to fetch water long enough to affect their time for schooling?

Secondly, health costs generated by poor access to safe and clean water are expected to be more for girls. There are many studies that report significant associations between child health and child schooling performance (see for example, Behrman 1996). We assume households allocate their resources rationally (Foster 2002). Assuming also that marginal benefits of educating a son and a daughter is the same, if marginal cost (here is the health cost) of educating the daughter exceeds that of boys, parents will invest more in their sons' schooling. In addition, Girls' health will suffer more from poor access to safe and clean water after they reach puberty. This is an interesting link, and may explain partly why girls drop out of school more after they reach puberty (usually after the age of 11). Furthermore, period-related symptoms partly resulted from lack of water and timely cleaning may also lead to discomfort for girls, which will disturb their schooling. For example, in Poureslami & Osati-Ashtiani (2002), 15% of the subjects stated that dysmenorrhea had interfered with their daily life activities and caused them absent from school from one to seven days a month. Besides, even if girls suffering from pain go to school, they will lose concentration (Huerta 1994).

Finally, psychic costs may arise for girls during their period if they are unable to be clean (Bista2004, Nahar, 2006; Kirk and Sommer, 2006; Singh, 1999; Snel, 2005). In schools, toilets may be totally absent or few in number, with broken doors or defective water supply and sewerage (El-Gilany 2005, Behrman at al. 1997). Rose Lidonde (2005) has explained the problems that girls face in detail in an African context. Snel & Shordt (2005) also claim that school drop-out rates and low literacy levels, especially among adolescent girls, can be attributed in part to inadequate sanitation conditions in schools. In addition, Cairncross et al. (1998) find that a school sanitation programme in Bangladesh increased girls' enrolment by 11 per cent.

The arguments above boil down to the belief that that girls' education suffers when greater time, health and psychic costs associated with poor access to clean and safe water are generated. Table 1 provides motivation for our hypothesis. We take the enrolment rate for the rural population in age groups 6 to 18, and distinguish four groups of females by whether or not access to water is good (defined in more detail below), and whether periods have begun. For comparison, we also show results for boys, distinguishing those with and without good access to clean and safe water.

As can be seen, the initial enrolment rate is high for all groups, but girls with poor access to water (no access to tap water) tend to drop out, particularly after periods have begun. Thus, the school enrolment rate of girls aged 11-15 whose periods have begun, and who have poor access to water, is only 74%, whereas girls without periods, and boys, have an enrolment rate of 90% or more, irrespective of access to water. The Kaplan-Meier survival curves in Graph 1 also shows that girls with poor access to water have a much lower survival curve than the other three groups, having only about a 30% chance of surviving until period 9, which is the end of junior secondary school. These basic data seem to back up our hypothesis, and we will now test it using appropriate statistical methods.

4. The statistical model

The nature of our dependant variable requires the use of survival models to take account of the fact that many observations are right-censored, that is, children will leave school after the survey is conducted (Hosmer & Lemeshow, 1999, Kalbfleisch, & Prentice 1980). We therefore estimate the probability of dropping out of school conditional on school enrolment until the previous grade. The simplest approach is the Cox proportional hazard model, in which the hazard rate, $h_i(t)$, that is the exit probability from schooling, conditional on being at school up to time t, is specified as follows:

$$h_i(t) = h_0(t)\mathbf{exp}\,(\beta_1 x_{1i} + ... + \beta_k x_{ki}) \qquad (1)$$

where $h_o(t)$ is the baseline hazard, and the x's are covariates. Here the baseline hazard is not given a parametric form (Cox & Oaks, 1984). However, we modified this approach from the outset, since tests using the scaled Schoenfeld residuals rejected Cox's proportional hazards assumption (for girls: df=33; Chi-square=88.73; p=.000. for boys: df=31; Chi-square=72.85; p=.00).

Consequently we use models which give a parametric specification to the baseline hazard. The Weibull accelerated failure time (AFT) model fits best, as can be seen from Table 3's test results. Compared to the Log-logistic and Log-normal models, the Weibull has the largest log likelihood and the smallest Akaike Information Criterion (AIC) values. The AFT coefficient gives the relative probability of duration of schooling, conditional on being at school up to time t, and is specified as follows (Cameron & Trivedi, 2005, 592):

$$\lambda(t \mid \mathbf{x}) = \lambda_0(t \exp(-\mathbf{x'\beta}))\exp(\mathbf{x'\beta}) \qquad (2)$$

where $\lambda(t \mid \mathbf{x})$ is the hazard rate, t is survival time, the $\mathbf{x}$'s are covariates and the $\boldsymbol{\beta}$'s are time ratios which need to be estimated. This is an acceleration of the baseline hazard $\lambda_0(t)$ if $\exp(-\mathbf{x'\beta}) > 1$ and a deceleration if $\exp(-\mathbf{x'\beta}) < 1$. The time ratios can be interpreted as survival time multipliers. For example, a time ratio of 1.5 means that, conditional on being at school up until the instant past moment, the relative time of surviving the school is 1.5 times (50%) longer if the independent variable increases by one unit. Time ratios greater than one correspond to positive coefficients and time ratios less than one correspond to negative coefficients.

A variation on this model is to include a further parameter, α, to allow for unobserved heterogeneity (Hosmer and Lemeshow, 1999, 318). The following modification of the model (2) is meant to take into account of such heterogeneity :

$$\lambda(t \mid \mathbf{x}, \alpha) = \lambda_0(t \exp(-\mathbf{x'\beta}))\exp(\mathbf{x'\beta})\alpha \qquad (3)$$

where α is a random positive quantity and assumed to be distributed as gamma with mean 1 and variance θ. The further the value of θ deviates from 0 the greater the effect of heterogeneity. It reflects the impacts of unobserved factors on the probabilities of individual's schooling duration. We tried this specification, and generally θ is greater than zero.

Our specification of variables to be included in the model is as follows. First, the dependent survival time variable is accumulated years of schooling'. Not being at school when the survey is conducted is specified as the "failure" event. As there are multiple entries and failures per subject, a subject ID is specified and robust standard errors are obtained.

Our independent variables of interest are access to water, and the onset of the period. As regards the access to water variable, this effect is obviously allowed to be different for boys and girls. For both sexes, we expect poor access to water to decrease the likelihood of children's school enrolment, not only because of the general water related health problems, but also because of the increased time required to fetch water. In this respect, poor access to water equally reduces both boys' and girls' schooling. However, we expect poor access to water to have a worse impact on girls' schooling after periods have begun, due to the hygiene related economic and psychological problems that girls face as described above. These considerations point to an interaction between the access to water and period variables.

We also control for household income, which is an important variable, as explained above. In our case, controlling for household income is even more important, given the fact that high income is likely to be associated with good access to water. Hence, we need to measure income well in order to separate its effect from the effects of access to water. While measuring household income is difficult in a rural society where much income does not arise through the market, the CHNS offers good data. In particular, (see below) the CHNS asks respondents about a comprehensive set of income generating activities. To supplement this measure of household income, we also have variables for father's and mother's education and occupation. Generally, higher status of parents' education and occupation always go together with the wealth of the family.

A further consideration is children's work, both market and household, which needs to be included to control for the alternative uses of children's time. A possible problem is that children's market and household work will be chosen by the household jointly with schooling, so these variables are endogenous. We need to find instruments for these variables which of course is difficult. Fortunately, however, even in rural China few children report that they help much with market work or household work (see below for details). Hence, when we experiment by dropping the two work variables from the regression, similar coefficients remain for most of the other explanatory variables, with the impact of the work variables mainly changing the impact of the age dummies.

5. Data

The data used in this study come from the Chinese Health and Nutrition Survey (CHNS), jointly conducted by the University of North Carolina and the Chinese Academy of Preventive Medicine, Beijing. The CHNS is designed to examine "how the social and economic transformation of Chinese society and family planning programs implemented by national and local government affect the economic, health and nutritional status of its population" (CHNS , 2007). The survey drew a sample from nine provinces:, Heilongjiang, Liaoning, Henan, Shangdong, Jiangsu, Hubei, Hunan, Guizhou, and Guangxi. These provinces stretch from the North-East to the South-West, and vary substantially in geography and economic development (Yang, 2006). Households in each province are selected using a stratified multistage cluster design that includes approximately 20 households in each of some 190 urban and rural communities.

The CHNS has interviewed households six times between 1989 and 2004. While this period is quite long, the main change in education law, requiring 9 years of compulsory education, was in 1986 (Yang, 2006), well before our first survey. Hence, we will simply rely on intercept dummies for each survey to pick up trend effects.

Our analysis uses information only from rural settings, where "urban" and "rural" are defined according to household registration. For each wave of the survey, we identify children graded 1-12 and usually aged 6-18. We do not consider tertiary education, after age 19, since different factors such as marriage enter the schooling decision. In addition, our hypothesis relates to the interaction of poor access to water and the onset of periods for girls. In our sample, the mean age of onset of periods is 12.7 (compare 13.5 in Singh (1999), and 12.9 in El Gilany (2005)), and thus, with the 6-18 age span, we have adequate observations from both before and after.

In table 2, the means and standard deviations are shown for variables used in the analysis. About 81% of the individuals in the sample are at school at the time of survey, 47% of the sample are girls. And 47% of the girls for whom we have the period information have begun periods .

The access to water variable is derived from the question 'how does your household get drinking water?'. In the original sample, 30% of the children had tap water at home, 17% tap water in the courtyard, 34% other water sources in the courtyard and 19% other water sources outside their courtyard. Thus, 53% of the children have no access to tap water, which we use as our definition of "poor" access to water.

The CHNS provides a detailed per capita income estimate for rural households, which is not usually available from other sources. Gross household income in cash or kind is created for different categories and then expenses are deducted to create a net income value, deflated using the appropriate price deflators. The main problem is that much income is not received via the market, and a market value has to be imputed. To measure income in-kind, the CHNS relies on the respondent's (usually the household head's) estimation about the market value of the goods consumed and received as gifts. For home gardening income, the total value of household food consumed at home or sold is measured. Income from farming, raising livestock/poultry; collective and household fishing; and the value of income from other household business is obtained by same calculations. The CHNS also takes into account welfare subsidies including housing subsidy, child care subsidy and gifts. In our sample, the mean per capita rural household income is 1,245 *yuan* (1988 community CPI) for the period of 1989-2004. This figure is generally in line with the estimates from other sources. For example, mean per capita household income for rural residents in China is 1,067 *yuan* (1990 CPI) for 1987-2001 in Benjamin et al (2005).

The dataset also allows us to use most of the conventional control variables for determinants of children's schooling. Educational qualifications of parents are grouped into four categories: 0 being no qualification (16% for fathers and 38% for mothers), 1 being graduation from primary school (27% for fathers and 22% for mothers), 2 from junior secondary school (38% for fathers and 26% for mothers) and 3 from senior secondary school or college (21% for fathers and 14% for mothers). We also include parents' work status, grouped into 4 categories: 1 being the jobs which relates to high salary or power (9% for fathers and 3% for mothers); 2 to stable salary but no power (20% for fathers and 14% for mothers); 3 to no stable salary and no power (mostly farmers, and 56% for fathers and 64% for mothers) and 4 to others (including unemployed, and 13% for fathers and 18% for mothers).

As for sibling structures, we see that 30% households have a single child, 42% have two children and 28% have more than two children. We also account for sibling structure in the households who have two children, to measure the overall impacts of 'gender bias' or 'son preference' if any. We see that within this group 12% of children have one older brother, 9% one older sister, 11% one younger brother; and 10% one younger sister. Wide-spaced and close-spaced birth intervals may also have different impacts, and we have constructed wide-space and close-space birth variables.

Finally, we also have household and market work variables measured as hours worked per day to capture the impact of work opportunities on children's schooling. But, in fact, only 16% of girls aged 13 or over report more than 1 hour/day household work (including washing clothes, cooking, laundry and caring for younger bothers and sisters), and even fewer boys, about 4%. In addition, only 10% of girls report more than 1 hour/day market work (farming, livestock raising, working in local business and gardening), and similarly 10% of boys. Admittedly, these figures show girls are still doing a relatively greater amount of household work than boys, but their time doing market work is similar. Below, we control for these variables, but expect their overall impacts to be small for education choices.

6. Empirical results and discussion

Our empirical strategy is as follows. First, we run separate Weibull AFT regressions for girls with good and poor access to water. We also run separate regressions for the boys. These results are presented in table 4. However, the Chi square test results after seemingly unrelated estimations allow us to pool the two groups of girls. (chi2(31) = 33.40, Prob > chi2 = 0.3514). A similar results hold for the boys (chi2(30) =31.93, Prob> chi2 = 0.3709). We present this simpler specification in Table 5.

Taking first Table 4, we see that the onset of periods has a large adverse impact on girls with poor access to water, but little impact on girls with good access. Thus, the period variable for girls with good access to water indicates an insignificant 0.06 (=1 - 0.94) proportionate fall (i.e., 6%) in schooling duration once periods begin. The corresponding figure for girls with poor access to water to water is much larger, 0.16 (= 1 - 0.84), and significant[2]. These coefficients are also significantly different at the 1% level. The survival distributions shown at the bottom of the table bear out this impression. For example, half of boys and girls with good access to water reach grade 12. But with poor access to water, only 35% of boys and 25% of girls reach grade 12.

The interaction of periods and access to water can be more clearly shown in the pooled regressions of Table 5. Here we see an interesting pattern. Poor access to water, of itself, has an adverse impact on both boys and girls. This effect can be seen for girls whose periods have not yet begun, whose conditional schooling duration is reduced by 11% (=1-0.89), and for boys who suffer a similar reduction, 10% (=1-0.90). Equally, the onset of periods, of itself, has an adverse impact. This effect can be seen by the 10% (=1-0.90) decline in school duration for girls with the onset of periods in good access to water areas. However, it is the combination that really matters. Thus, we see that girls beginning periods in poor access to water areas suffer a 0.22 (=1-0.78), or 22% fall in schooling duration, ceteris paribus.

The other variables in the model behave much as expected, and generally are not significantly different as between boys and girls, and so cannot explain the gender education gap. Looking first at household resources, we see that per capita household income plays important role in girls' schooling, with a significant coefficient of 1.06. Thus, increases in family income lengthen schooling duration, as we would expect. Admittedly the coefficient is smaller for boys, 1.01, and insignificant. However, likelihood ratio tests cannot reject the hypothesis that the income coefficients are statistically different between boys and girls. Moreover, several of the other proxies for household income such as parental educational qualification and job status are significant in the appropriate direction for boys.. In particular, parents' school qualifications have strong and statistically similar impacts on boys and girls schooling. In general, then, increases in household resources significantly increase schooling duration for both boys and girls. Below, we will assess to what extent better-off households can make up for the adverse impact of poor access to water.

As for other variables, we see that sibling structure variables have neutral impacts on children's schooling. This result validates our suspicion that after controlling for socio-economic factors, so called 'pro-son' bias does not exist, or at least its existence is insignificant. Furthermore, household work and market work opportunities have significant, though small, negative impacts on girls' and boys' schooling. At the same time, we find that the impacts for the different genders are the same statistically. Finally, the survey wave dummies indicate that girls' education has improved a lot compared to the boys throughout the years of reform period. This finding contradicts the argument that female-centred job places which emerged rapidly in 1990's have had adverse impacts on girls schooling.

---

[2] Unobserved heterogeneity is found to be significant in the model of girls with poor access to water (chibar2(01) = 11.13, Prob>=chibar2 = 0.000) but not in the model of girls with good access to water (chibar2(01) = 0.62, Prob>=chibar2 = 0.215). Unobserved heterogeneity is found to be significant in both of the models for boys. (chibar2(01) = 9.82, Prob>=chibar2 = 0.001 for boys with good access to water, and chibar2(01) = 10.44, Prob>=chibar2 = 0.001 for boys with bad access to water).

Moreover, although the impacts of household income, parental job status and parents' educational qualifications have strong impacts on children's schooling, their impacts are not statistically different between different genders, which means that they are not the main factors that generate the gender gaps in education. And household and market work opportunities do not seem to have gender specific impacts on children's schooling either. If the gender difference is not caused by the difference in household income or the opportunity costs of schooling, personal costs of schooling must be worth further investigation, which brings us back to the interaction between periods and poor access to water shown in the regressions.

A possible alternative reason for the adverse effects of poor access to water for girls is that poor access proxies for extreme poverty. However, for this argument to hold, poor access to water would have the same negative impact on girls prior to the onset of periods, as afterwards, or on boys. Another argument is that poor access to water may relates to remote villages, and the water variable simply picks up the extra cost of going to school. But again, if this were the case, girls prior to their periods, and boys, would be similarly affected. Nevertheless, we also have a control for the time to school, and find no significant impact there.

Then there is the culture argument, which pictures parents as too ready to push their daughters into the marriage market. However, this cultural push should have the same effect on girls' education irrespective of poor access to water – unless the access to water variable is picking up geographic isolation, or some such factor. To allow for possible omitted geographic confounding influences, we stratified by region or province (allowing the baseline hazard to vary by region or province), and the results remained similar, with a strong adverse impact of poor access to water with the onset of periods.

Table 6 presents percentiles of some survivorship distributions for different scenarios for access to water and family income. In all cases, we assume that girls have started their periods, so the scenarios do not illustrate the timing of school drop-out, but rather the final impact of access to water, and family income. Scenarios (1) and (2) show the impact of access to water alone, and we see that with poor access, girls at the median lag a 1.2 years (10.1 – 8.9) schooling behind boys.

Scenarios (3) and (4) vary a composite measure of household income alone. Here, the composite income measure is based on a k-mean clustering of households into two categories, "rich" and "poor", based not only on household income, but all 16 father's and mother's occupation and education variables listed in the regressions. Here we see that being in poor family category pulls both girls and boys down about the same amount, as might be expected. Taking the median, for example, girls with families in the rich category achieve 10.3 years of schooling (similar to boys), but fall to only 8.1 years if family income is in the poor category (boys fall to 8.3 years).

In rows (5) to (8), we vary both access to water and family income category. We see from scenarios (6) and (7) that high income tends to make up for poor access to water. Girls with good access but in a poor family income achieve only 7.9 years at the median, but their counterparts with poor access but rich achieve about one year more (8.8 years).Of course, water still has its effect, as scenarios (5) and (6) indicate. Here all families have high income, but at the median those girls with poor water access achieve only 8.8 years of education, compared to the 10.8 years of their counterparts with good access. Still, the comparison between (6) and (7) shows that high income can make up to some extent for the disadvantages which poor water access poses for girls' education. We imagine that the better-off families without tap water can perhaps store water for washing, and also buy pads and protection which make school-going tolerable for the girls.

It is also interesting that it is only in scenario (8), with poor water and poor families that girls really lag behind boys. In the other situations, either water access is good, so girls are not disadvantaged, or the family is rich, and can overcome the access to water problem for its daughters. Of course, unfortunately, it is scenario (8) which is most prevalent in the developing world.

For the reasons mentioned above, it is necessary to run a regression for girls by allowing the impacts of period and poor access to water to change by different income category. We use the clustered income variable to allow for such a change. The new estimation results are provided in table 7. As have been discussed, poor accesses to water and girls' period have their own impacts. The clustered income variable turns out to have a very strong positive impact which is reasonable, as this is a 'joint' impact of higher household income, and higher status of parental education and occupation. However, we are interested in the cross-impacts. Time ratios can be multiplied to compute the cross-impacts of the variables. For example, the impact of the poor access to water with the onset of period is equal to 0.78= 0.925*0.909*0.929, which is an exact impact estimated in previous regressions. The negative impact of period (0.93) can be alleviated with the higher income (1.18=0.93*2*0.64). And the adverse impact of poor access to water with the onset of period is also lessened if the households are richer (0.92 = 0.925*0.906*2*0.929*0.489*0.639*1.89).

6. Conclusion

Narrowing the gender gap in education is profoundly important for economic and social development. Women cannot play their proper role in public affairs unless they are educated as well as men. Democracy itself suffers. In rural China, there are gender gaps in school enrolment in secondary education. However, previous research addressing the reasons for the gap give mixed results. Our results are clear: poor access to water is what drives the education gap between girls and boys. Moreover, because we have data on the onset of periods, we can point to the reason why poor access to water has this adverse effect. It is primarily women who have periods that suffer the consequences of poor access to water. Our findings simply underline the fact that periods require water for cleanliness, without which school-going becomes very difficult.

Our research gives estimates of the weight of the disadvantage that girls with poor access to water face. Not having good access to water is never pleasant, as we can see from the effects on boys, or girls prior to their period – both of whom suffer a 10% loss in expected school duration. For girls after their periods, it is particularly unpleasant, and we can add a further 12% loss, a doubling, in other words. Our results have a striking policy implication. A major benefit of policies to improve water supplies may not be the obvious household or industrial benefit, but rather an unseen benefit, the improvement in the position of women.

Reference:

Becker, G, 1981, A Treatise on the family, Cambridge: Harvard University Press.

Behrman, J., R., 1996, "Impact of health and nutrition on education", *World Bank Research Observer,* 11(1): 23-37

Behrman, J., R., Khan, S., Ross, D. & Sabot, R., 1997, "School Quality and Cognitive Achievement Production: A Case Study for Rural Pakistan", *Economics of Education Review*, 16 (2): 127-142

Behrman, J., R., Foster, A., D., Rosenzweig, M., R. & Vashishta, P., 1999, "Women's Schooling, Home Teaching and Economic Growth", *Journal of Political Economy,* 107(4): 682-714

Benjamin, D, Brandt, L, Giles, J, Wang, S, 2005, "Income Inequality During China's Economic Transition", working paper, http://www.chass.utoronto.ca/~benjamin/BBGW.pdf

Bista M B, 2004, "A Review of Research Literature on Girls' Education in Nepal", Report prepared for UNESCO Bangkok, http://unesdoc.unesco.org/images/0013/001386/138640e.pdf

Brown, P, & Park, A, 2002, "Education and poverty in rural China", *Economics of Education Review*, 21: 523-541

Burrows, G, J Acton and T Maunder, 2004, Water and Sanititation: the Education Drain, London: Wateraid, http://www.wateraid.org/other/startdownload.asp?DocumentID=113

Cairncross, S. et al., 1998, "The public and domestic domains in the transmission of disease", *Tropical Medicine and International Health*, 1: 27–34.

Cameron, A. C., Trivedi, P. K., 2005, Microeconometrics – Methods and Applications. Cambridge University Press

CHNS, 2007. China Health and Nutrition Survey, Chapel Hill, North Carolina: Carolina Population Center, http://www.cpc.unc.edu/projects/china

Colclough, C., Rose, P., Tembon, M., 2000, "Gender inequalities in primary schooling, the role of poverty and adverse cultural practice", *International Journal of Educational Development*, 20: 5-27

Conelly, R, & Zheng, Z, 2003, "Determinants of school enrolment and completion of 10 to 18 year olds in China", Economics of education review, 22: 379-388.

Cox, D.R., & Oakes, D., 1984, Analysis of Survival Data. London: Chapman Hall

Davison, J., Kanyuka, M., 1992, "Girls participation in basic education in southern Malawi", *Comparative Education Review,* 36(4): 446-466

El-Gilany, A. H., Badawi K., El-Fedawyb, S, 2005, "Menstrual Hygiene among Adolescent Schoolgirls in Mansoura, Egypt", *Reproductive Health Matters*, 13(26): 147-152

Emerson, P.M., Souza, A.P., 2007, Child Labor, School Attendance, and Intrahousehold Gender Bias in Brazil, The World Bank Economic Review Advance Access published March 8, 2007, http://wber.oxfordjournals.org/cgi/reprint/lhm001v1

Foster., E.,M., 2002, "How Economists Think About Family Resources and Child Development", *Child development,* 73(6), 1904-14

Hannum, E, 1998, "Educational inequality: hidden consequences of the reform era in rural China: 1949-1985. *Research in Social Stratification and Mobility*, 13, 73-98

Hannum, E, 2005, "Market transition, educational disparities, and family strategies in rural China: new evidence on gender stratification and development", *Demography*, 42(2): 275-298

Hannum E and C Buchmann, 2005, "Global education expansion and socio-economic development: An assessment of findings from the social sciences", World development, 33 (3); 333-354.

Hill, M., A., King, E., M., 1995, "Women's education and economic well-being", *Feminist Economics*, 1(2): 21-46

Holmes, J., 2003, "Measuring the determinants of school completion in Pakistan, Analysis of censoring and selection bias", *Education of Economics Review*, 22, 249-264

Hosmer, D, & Lemeshow, S, 1999, Applied Survival Analysis, New York: Wiley.

Kalbfleisch, J.D., Prentice R.L., 1980, The Statistical analysis of failure time data, Wiley and Sons, New York.

Kirk, J. and Sommer, M. 2006, "Menstruation and body awareness: critical issues for girls' education", *EQUALS*, 15: pp 4-5.

Knight, J, & Li, S, 1996, "Educational attainment and the rural-urban divide in China", *Oxford Bulletin of Economics and Statistics*, 58: 83-117

Li, D, & Tsang, M.C, 2003, "Household education decisions and implications for gender inequality in education in rural China", *China: An International Journal*, 2003, 1(2): 224-248.

Lidonde, R., 2005, "Why gender is important in SSHE", *WaterLines,* 23(3), 6-9

Nahar, Q, 2006, "Addressing special needs of girls – challenges in school", Second South Asian Conference on Sanitation, 20-21 September, Islamabad, Pakistan

Poureslami M, Osati-Ashtiani F, 2002, "Attitudes of female adolescents about dysmenorrhea and menstrual hygiene in Tehran suburbs", *Archives of Iranian Medicine*, 5(4): 219-224

Schultz, T, 2002, "Why governments should invest more to educate girls", World Development, 30 (2): 207-225.

Singh MM, et al., 1999, Awareness and health seeking behaviour of rural adolescent school girls on menstrual and reproductive health problems, Indian Journal of Medical Science, 53(10):p 439-43

Snel, M, & Shordt, K, 2005, "The evidence to support hygiene, sanitation and water in schools", *WaterLines,* 23(3), 2-5

Song, L., & Appleton, S., & Knight, J. , 2006, "Why do girls in Rural China have lower school enrolment?" *World Development, 34*(9), 1639_1653

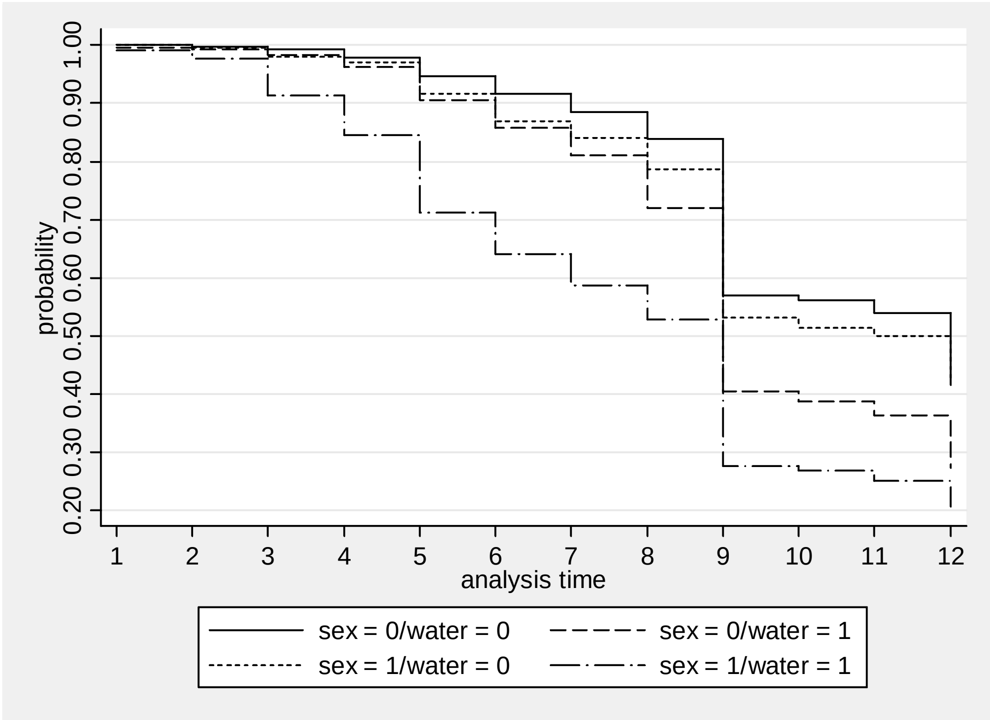
Tsui, M, & Rich, L, 2002, " The only child and educational opportunity for girls in urban China", *Gender and Society,* 16(1): 74-92

UNICEF, 2008, "Water, environment and sanitation", http://www.unicef.org/wes/index_womenandgirls.html

Yang J H, 2006, "Has the One-Child Policy Improved Adolescents' Educational Wellbeing in China?" __ CHNS discussion paper 2006.

Yueh, L, 2006, "Parental investment in children's human capital in urban China", *Applied Economics,* 38: 2089_2111

Graph 1. Survival distribution of schooling for children with good vs. poor access to water



Notes: analysis time corresponds to accumulated years of schooling: 0-6 being primary, 7-9 junior secondary (the end of compulsory education), and 10-12 senior secondary school; boys = 0; girls = 1; good water=0; poor water=1 (for definitions, see Table 2); for girls, the analysis is restricted to those who have started periods.

Table 1. Average School Enrolment Rates % (Rural China, CHNS: 1989-2004)

	Girls				Boys	
	Good Water		Poor Water		Good	Poor
Age	Period	No period	Period	No period	Water	Water
6-10		99		99	99	99
		(679)		(800)	(844)	(1037)
11-15	88	96	***74***	91	93	90
	(423)	(467)	(432)	(382)	(1136)	(1233)
16-18	57		***34***		56	40
	(534)		(494)		(534)	(634)

Notes: see Table 2 for definitions of good/poor access to water. Sample numbers are in brackets.

Table 2. Descriptive Statistics

Variables	N	Mean	S D	Min	Max	Description
edu_year	11125	5.87	3.11	0	16	accumulated years of schooling
at_school	10966	0.81	0.39	0	1	whether at school (=1) or not (=0)
sex	11125	0.47	0.5	0	1	gender (female=1)
period	6398	0.47	0.5	0	1	if having period (=1) or not (=0)
water	10939	0.53	0.5	0	1	access to good (=0) or poor (=1) water. Good water is defined as having tap water at home/courtyard. Poor water refers to other water sources.
ln_income	11055	6.79	0.87	1.63	9.82	log of per capita household income
mom_job	11049	3.04	0.57	1	4	4 job status representing: 1= high income/power; 2= no power/ average constant income; 3= non-constant income; 4= other
dad_job	10790	2.81	0.72	1	4	same as above
mom_edu	10423	1.03	1.03	0	3	4 educational qualifications representing: 0= no qualification; 1= primary school; 2= junior middle school; 3= high school/college
dad_edu	10214	1.58	0.96	0	3	same as above
single_kid	11125	0.3	0.46	0	1	single child at home
one_sibling	11125	0.42	0.49	0	1	child having two siblings
old_brother	11125	0.12	0.32	0	1	having one older brother
old_sister	11125	0.09	0.28	0	1	having one older sister
yo_brother	11125	0.11	0.31	0	1	having one younger brother
yo_sister	11125	0.11	0.31	0	1	having one younger sister
two_more_siblings	11125	0.28	0.45	0	1	child having three or more siblings
wide spaced birth	11125	0.76	0.42	0	1	having wide birth space (>3 years)
close spaced birth	11125	0.24	0.42	0	1	having close birth space (<= 3 years)
household_work	11125	0.22	0.89	0	12	number of hours spent for household work per day last week
market_work	11125	0.25	1.18	0	12	number of hours spent for market work per day last year
time to school						
age	11125	12.55	3.67	6	19	age of children
agegroup_6-11	11125	0.41	0.49	0	1	age 6-11
agegroup_12-16	11125	0.42	0.49	0	1	age 12-16
agegroup_17-19	11125	0.17	0.38	0	1	age 17-19
eastern region	11125	0.27	0.44	0	1	provinces in eastern region
central region	11125	0.44	0.5	0	1	provinces in central region
western region	11125	0.29	0.45	0	1	provinces in western region
survey waves	11125	1994	5.06	1989	2004	survey year

Table 3. AIC and log likelihood values

		parametric models (AFT with heterogeneity)		
		Weibull	Log-logistic	Log-normal
Girls	Log-likelihood	-695.32	-704.77	-724.43
	AIC	1460.63	1479.54	1518.86
Boys	Log-likelihood	-791.46	-802.26	-833.23
	AIC	1652.91	1674.52	1736.46

Note: Akaike Information Criterion (AIC) =-2(log-likelihood)+2(c+p+1) where *c* is number of covariates excluding constant, *p* is number of ancillary parameters.

Table 4. Schooling duration for children with good and bad access to water

Weibull AFT Model with Heterogeneity (Time Ratio coefficients shown)

Dependent variable: schooling duration

variable	Girls				Boys			
	good access		poor access		good access		poor access	
	TR	z	TR	z	TR	z	TR	z
period	0.94	-0.74	0.84	-2.28	--	--	--	--
ln_income	1.03	1.04	1.08	3.32	1.00	-0.17	1.02	0.91
dadjob_1 (ref)								
dadjob_2	0.90	-1.39	0.91	-0.95	0.88	-2.07	1.04	0.45
dadjob_3	0.80	-3.04	0.87	-1.55	0.83	-2.92	0.93	-1.00
dadjob_4	0.77	-2.27	0.72	-2.23	0.88	-1.29	0.86	-1.26
momjob_1	1.11	0.37	1.14	0.44	1.26	1.09	0.94	-0.31
momjob_2 (ref)								
momjob_3	0.88	-1.74	0.93	-0.66	0.92	-1.48	0.99	-0.05
momjob_4	0.84	-2.08	1.02	0.17	0.88	-1.88	1.02	0.18
dadedu_1	0.77	-3.04	0.84	-2.19	0.95	-0.69	0.76	-3.92
dadedu_2	0.80	-2.96	0.97	-0.39	0.86	-2.54	0.81	-3.16
dadedu_3	0.82	-2.75	0.99	-0.13	0.97	-0.49	0.89	-1.80
dadedu_4 (ref)								
momedu_1	0.84	-1.99	0.65	-3.17	0.81	-2.46	0.83	-1.86
momedu_2	0.94	-0.73	0.68	-2.77	0.84	-2.02	0.79	-2.25
momedu_3	0.96	-0.51	0.67	-2.89	0.87	-1.69	0.92	-0.75
momedu_4 (ref)								
household work	0.95	-2.64	0.97	-3.03	0.92	-3.08	0.96	-1.08
market work	0.88	-5.12	0.93	-5.97	0.92	-7.29	0.91	-4.62
agegroup_1 (ref)								
agegroup_2	1.08	0.60	0.92	-0.70	1.07	0.68	0.83	-1.78
agegroup_3	0.98	-0.15	0.81	-1.56	0.98	-0.18	0.71	-3.21
p	3.94		2.89		3.90		3.72	
θ	0.66		0.10		0.33		0.49	
N	1840		1817		2163		2540	
Log-likelihood	-288.26		-387.01		-322.71		-452.18	
Survival Distribution quartiles	Education duration		Education duration		Education duration		Education duration	
0.25			12					
0.35			11				12	
0.50	12		9		12		11	
0.75	9		7		9		8	
0.95	6		4		6		6	

Notes: The regressions are controlled for family and sibling structure, location and survey time. Underlined figures in the survival distributions are estimates for girls who have started their periods.

Table 5. Schooling duration with pooled samples

Weibull AFT Model with Heterogeneity (Time Ratio coefficients shown)
Dependent variable: schooling duration

	Girls		Boys	
	Tm. Ratio	z	Tm. Ratio	z
good water – no period (ref)				
poor water – no period	0.89	-1.35	0.90	-3.76
good water – period	0.90	-1.34		
poor water – period	0.78	-3.10		
ln_income	1.06	3.05	1.01	0.45
dadjob_1 (ref)				
dadjob_2	0.87	-2.22	0.92	-1.52
dadjob_3	0.81	-3.53	0.86	-3.16
dadjob_4	0.73	-3.31	0.86	-1.98
momjob_1	1.09	0.39	1.07	0.51
momjob_2 (ref)				
momjob_3	0.85	-2.51	0.95	-1.07
momjob_4	0.87	-2.00	0.93	-1.30
dadedu_1	0.79	-3.88	0.83	-3.77
dadedu_2	0.89	-2.13	0.84	-3.86
dadedu_3	0.90	-2.01	0.93	-1.65
dadedu_4 (ref)				
momedu_1	0.75	-3.76	0.82	-2.99
momedu_2	0.81	-2.68	0.81	-3.21
momedu_3	0.82	-2.50	0.89	-1.73
momedu_4 (ref)				
household work	0.97	-3.69	0.94	-2.83
market work	0.91	-7.57	0.92	-8.65
agegroup_1 (ref)				
agegroup_2	0.98	-0.20	0.91	-1.34
agegroup_3	0.88	-1.33	0.79	-2.99
wave_1	0.67	-5.66	0.82	-3.64
wave_2	0.77	-3.55	0.91	-1.58
wave_3	0.76	-3.80	0.88	-2.33
wave_4	0.92	-1.21	0.95	-1.07
wave_5	0.83	-2.76	0.92	-1.64
wave_6 (ref)				
p	3.29		3.70	
θ	0.37		0.40	
N	3657		4703	
Log-likelihood	-695.32		-791.46	

Notes: The regressions are controlled for family and sibling structure, location.

Table 6. Percentile of survival distributions using Weibull – various scenarios

	Survival Time	0.95	0.75	0.50	0.35	0.25	0.05
Good water (1)	Girls	5.1	9.1	11.6			
	Boys	4.9	8.9	11.4			
Poor water (2)	**Girls**	**3.4**	**6.4**	**8.9**	**10.4**	**11.8**	
	Boys	**4.7**	**7.8**	**10.1**	**11.8**		
High income (upper cluster) (3)	Girls	4.2	7.5	10.3	11.4	12.0	
	Boys	4.6	7.7	10.2	11.2	12.0	
Low income (lower cluster) (4)	Girls	3.3	5.8	8.1	9.4	10.7	11.7
	Boys	3.8	6.3	8.3	9.6	10.8	11.4
Good water High income (5)	Girls	4.6	7.8	10.8	12.0		
	Boys	4.8	8.2	11.0	12.0		
Poor water High income(6)	**Girls**	**3.3**	**6.3**	**8.8**	**11.7**	**12.0**	
	Boys	**4.6**	**7.6**	**9.9**	**11.5**	**12.0**	
Good water Low income (7)	Girls	3.5	5.9	7.9	9.3	10.4	12.0
	Boys	3.6	6.1	8.3	9.8	10.9	12.0
Poor water Low income (8)	**Girls**	**2.8**	**5.1**	**7.2**	**8.7**	**9.2**	**10.4**
	Boys	**3.6**	**6.0**	**8.3**	**9.6**	**10.4**	**11.6**

Notes: Weibull AFT with heterogeneity specification used to create the distributions; variables specified for a typical value in model (1) – (8) are listed in the first column. Other variables in the model are set to their mean values. Girls are those who have started their periods.

The high and low income categories are based on k-means clustering using per capita household income, parental job status and parental educational qualifications. The cut-off point of per capita household income is set at its median value, while job status and educational qualifications variables are kept with their four original categories. We classified the data into two clusters "poor" (7751 subjects) and "rich" (2198 subjects).