Students Today, Teachers Tomorrow? The Rise of Affordable Private Schools

Tahir Andrabi (Pomona College) Jishnu Das (World Bank)
Asim Ijaz Khwaja (Harvard University)*

February 2006

Abstract

Private schools comprise an increasingly large and growing share of primary enrollment in low-income countries. For example, in Pakistan, the focus of this paper, 35 percent of children at the primary level are in mainstream private schools. This paper highlights the crucial role of the *public* sector in facilitating private investments in education: Instrumental variable estimates indicate that private schools are three times as likely to be found in villages with a girls' secondary school, an increase of 35 percentage points. There is little or no relationship between the presence of a private school and pre-existing girls' primary, or boys' primary and high schools. Supply-side factors play a role—private school teacher's wages are 20 percent lower in villages with girls' secondary schools. In an environment with low female mobility due to cultural restrictions, and lower wages for women in the labor market, private schools locate in villages with a greater supply of local secondary-school educated women. These findings bring together three related concepts to explain where private schools locate—the inter-generational impact of public schools, the role of cultural labor market restrictions, and the prominent role of women as teachers. They also suggest the continuing importance of the government sector in creating a cohort of women with secondary school education who will become future teachers in private schools.

*Pomona College, Deverlopment Research group, World Bank, and Kennedy School of Government, Harvard University. Email: tandrabi@pomona.edu; jdas1@worldbank.org; akhwaja@ksg.harvard.edu. This paper was funded through grants from the PSIA and KCP trust-funds and the South Asia Human Development Group at the World Bank. We thank Karla Hoff, Hanan Jacoby, Ghazala Mansuri, Sendhil Mullianathan and Tara Vishwanath for long discussions and seminar participants at Lahore University of Management Sciences, Harvard University, The World Bank, Wharton, IUPUI, University of Michigan and University of Maryland for comments. We are specially grateful to Tristan Zajonc for exceptional research assistance. Assistance from the Project Monitoring and Implementation Unit in Lahore is also acknowledged. All errors are our own. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent.

I Introduction

How to deliver education is a central problem in poor countries. A growing literature on the quality of public schooling in low-income countries shows that learning is very poor and teachers are frequently absent (Duflo and Hanna 2005, Chaudhury and others 2006). While this has prompted calls for greater private sector involvement in education (Tooley and Dixon 2005), there is a concern that institutional impediments may limit the spread of private schools, particularly in rural areas. We provide evidence from Pakistan that *government* schooling investments play a large role in facilitating private sector involvement in education. Private schools in rural Pakistan are three times more likely to locate in villages where the government had previously constructed a secondary school for girls.[1]

These location decisions are informative about the overall constraints to better quality education. In contrast to the government sector where teacher hiring is governed by teachers' unions, state-wide hiring regulations and non-transparent processes, private sector investments reflect local market conditions. The construction of girls' secondary schools could lead to greater private sector involvement through increased revenues, most of which are school fees, or decreased costs, most of which are teachers' wages. The effect of government schooling investments was, in part, routed through supply-side channels—wages of private school teachers are 20 percent lower in villages with pre-existing secondary schools for girls. Part of what government girls' secondary schools did was to create a cohort of teachers for future private schools in the village.

We submit that restricted geographical mobility for women is the key institutional impediment that governs this "women as teachers" channel. Constraints on mobility of women across villages create a locally available pool of educated women in villages with girls' secondary schools. Entrepreneurial private school owners can take advantage of this available supply if they locate in the same village; conversely they are severely limited if they located in villages without women with secondary school education.[2] These institutional constraints are likely

[1] The vast majority are co-educational, English medium schools that offer secular education. Contrary to popular views, religious schools play a small role in Pakistan, comprising less than a one percent share and an even lower share in settlements with a private school (Andrabi and others 2006a).

[2] A local entrepreneur desribed the problems eloquently:
"The big problem", he said, *"is teachers. In most villages, I can set up a private school, but who will teach?*

deterrents to private investments in a number of other countries in Middle East, North Africa and South Asia (Bangladesh); more generally, we expect similar results in any context where local teachers have a (quality-adjusted) price advantage (for instance, local teachers are found to be less absent in a number of countries or are better able to teach in the local language).

The data from Pakistan are particularly suited for this exercise. Like in other South-Asian countries, the private sector is large and increasing. It accounts for 35 percent of primary school enrollments in Pakistan compared to 15 and 38 percent for India and Bangladesh. In 1999, 8000 new private schools were set up in 1999 with almost half in rural areas. Despite high enrollments, there is considerable geographical variation—27 of 102 districts report less than 5 percent of enrolled children in private schools; in contrast another 27 districts report shares exceeding 25 percent.

Moreover, while 90 percent of private schools offer only primary classes and 95 percent are coeducational, public schools are segregated both by gender (boys and girls) and by level (primary or secondary). This segregation helps us isolate the particular type of government investment that effects private sector location decisions; we can look at whether the effect of secondary schools and boys' schools is different from that of primary and girls' schools.[3]

Finally, the Federal Bureau of Statistics carried out a high quality census of all private schools in the country in 2001. This census is critical for two reasons. First, secondary schools for girls are present only in 5 percent of villages. The census of private schools, when matched with data on public school locations and village characteristics allows precise estimates of the impact of government schools on private school locations despite the low overall coverage. Second, to address concerns arising from the non-random placement of girls' secondary schools, we propose candidate instruments based on policy rules that restrict the total number of girls' secondary schools in an administrative area. Instruments similar to those proposed here may have wider applicability, but require data on all villages within administrative areas. This

All the men are working and if I pay them what they want, I will never make a profit. I cannot get women from other villages—who will provide the transport for them if it gets dark? How will she be able to work in another village if she is married? The only way we can work is if there are girls who can teach in the village—that is why, I ask if there is an educated girl who can teach. I can pay them Rs.800 ($14) a month and run the school. Otherwise it is not possible."

[3] In some cases, girls are allowed to attend boys' schools in the village if there are no girls' school, but only at the primary level. At the secondary level, there is strict segregation.

requires either a change in standard household sampling techniques, or the use of census-based data such as those available in Pakistan.[4]

A general insight of this paper is that initial government school investments leads to subsequent private sector development. Thus, while government schools may be contemporaneous substitutes for private schools, they are temporal complements. A recent paper from India confirms a similar correlation between private and public school location across communities—private schools will be unlikely to locate in areas without government schooling options (Tilak and Sudarshan 2001).

Furthermore, the market for education is closely linked to local labor market conditions and the labor force participation and availability of educated women plays a particularly important role. Even at a local level, the supply of teachers is not perfectly elastic so that higher levels of education imply higher costs. Banerjee (2005) discusses similar results in a theoretical context, introducing a market for teachers in the standard Barro-Becker framework. One insight is that increasing the returns to education need not increase education. If educated women now choose alternative occupations, the decrease in the supply of teachers pushes up the cost of educating the next generation. Empirically, these concerns have come to the fore in the United States. Recent papers concerned about the decline in teaching quality in US public schools point to greater employment opportunities and increasing pay for women in non-teaching fields. The results from Pakistan suggest that it is precisely the institutional impediments arising from limited geographical and occupational mobility for women that lead to the large causal impact of government schooling investments on private sector development.

The remainder of the paper develops the ideas presented here. Section 2 and 3 provide a background on education in Pakistan, review what we know about private schools in the country, and present the basic bivariate relationships in the data. In brief, there is a strong prima facie case for greater private sector involvement in basic education, perhaps more so than

[4]We thus focus on the existence of private schools rather than their enrollment share. Most variation in the number of children enrolled in private schools (on which we have data from the private school census) is driven by the extensive (that is, whether or not there is a private school in the village) rather than the intensive (variation in private school enrollment conditional on existence) margin. Thus, it is of independent interest to study the existence of such schools. An option would be to look at the share of private schooling in total enrollment. The only data that permits us to construct this variable is based on household surveys. Unfortunately, as discussed, samples in such surveys are too small for us to examine the main hypothesis presented here.

in high-income countries. Private schools improve enrollments, particularly for girls. Compared to public schools, they report higher test-scores and lower per-child costs, the latter primarily due to the (much) lower teachers' wages. For learning outcomes, raw differences and differences conditioning on observed characteristics are higher than in the United States.

Sections 4 and 5 discuss the data and the empirical methodology, with a focus on addressing omitted variable biases associated with the non-random placement of girls' secondary schools (henceforth GSS). We match data from two periods in time (1981 and 2001) and implement linear regression and first-difference specifications. These account for observable differences and differences in initial levels, but assume that time-varying omitted variables are uncorrelated to GSS construction.

Our preferred specification addresses concerns arising from time-varying omitted variables through an instrumental variables approach. The identification problem arises from a positive relationship between village population and GSS construction combined with a similar relationship between village population and private school location. In this context, one option is to base identification on nonlinearity and discontinuities in the function relating the setting up of girl's high schools and village population (Fisher 1976). If these nonlinearities can be justified and understood on the basis of an existing policy, we can identify the causal relationship between girl's high schools and the setting up of private schools by simultaneously controlling for the independent effects of village population.

We use the stipulated policy framework for GSS construction to formulate an eligibility rule. Government rules specify eligibility conditions at a village level, which can be modelled as a "minimum population condition" and a "population rank condition". Eligible villages are those with (a) population above 500 and (b) the largest population within a geographically contiguous administrative unit of 3-4 villages, called a Patwar-Circle. These conditions generate an eligibility rule highly correlated with GSS construction—villages in the eligible group were 60 percent more likely to receive a GSS. Moreover, non-linearities in the eligibility requirement (two villages with same population may be eligible or not depending on their population rank within the administrative unit) allow us to control for linear (and polynomial) effects of village and neighboring villages' populations that might have independent effects on private school placement. Under the assumption that private school placement is not determined in the same non-linear fashion as the eligibility requirements, the instrumental variables estimate is consistent and unbiased.

Section 6 presents results from these different econometric specifications. In tabulations, the numbers are striking: starting with villages without any girls' school (primary or secondary) in 1981, by 2001 there were private schools in 11.5 percent of those did not receive any girls' school in the intervening period, 12.5 percent in those that received a girls' primary school and *30.8* in those that received a GSS. The estimated impact of a GSS decreases to 10 percentage points increase (75 percent) with controls for observed village characteristics, increases to 15 percentage points in the first-difference and to 35 percentage points in the instrumental variables specifications. We discuss the properties and robustness of this instrument to various samples and construct falsification tests to show that the instrument is unrelated to the location of private schools in areas with no GSS construction. The difference in the estimated coefficients suggests that GSS were placed in areas with systematically lower demand for education. We discuss why this may be so later.

The section concludes with results that shed light on the possible channels through which GSS construction could lead to greater private school investments. The "women as teachers" channel is supported by results, which show that (a) private school existence is affected *only* by GSS construction—girls' primary or boys' primary/secondary schools have little or no effect; (b) GSS construction more than doubles the fraction of secondary-school educated women in the village and (c) changes in the percentage of women with secondary school education increase the probability of private school existence while changes in similarly educated men has no effect. Finally, private school teachers' wages are 20 percent lower in villages with a GSS, a difference that is close to the wage differential between male and female teachers controlling for observed characteristics. We discuss the extent to which these results support a role for supply-side factors. Section VII concludes with a tentative policy discussion.

II Background and Institutions

Pakistan has a population of 132 million people living in four main provinces—Punjab, Sindh, North Western Frontier Province and Balochistan. They, along with Islamabad the federal capital, comprise 97 percent of the country's population. Punjab, the focus of our study, is the largest province with 56 percent of the population, a majority of whom are in rural areas. Education provision is managed by provincial governments but recent changes are devolving provision of public services such as education to the local government.

Both the literacy rate and primary school enrollments are low compared to countries in the same income range and other countries in South Asia. Adult literacy in Pakistan is 43 percent compared to 54 percent for the South Asia average. The gender gap in educational enrollment is large, as are differences between rural and urban areas and between the rich and poor. The gross enrollment rate for the top expenditure decile is twice as high as that for the lowest decile. Increasing the stock of human capital is clearly a priority, especially for the poor, and especially for girls.

While returns to education in Pakistan appear to be fairly high at 15.2 percent (Psacharopolous 2002) and higher for women compared to men (World Bank 2005), levels of female education are low, especially in rural areas. In 1981, there were five literate women in the median village in Punjab—the largest and most dynamic province in the country; 60 percent of villages in the province had 3 or less secondary-school educated women and 34 percent had none. A recent household survey (Pakistan Rural Household Survey 2004) shows that 10.5 percent of women between 15 and 49 had primary and 4.5 percent had secondary education. Geographical and occupational mobility is low—more than 70 percent of all women live in the village where they were born; less than 3 percent are engaged in off-farm work and among those with secondary education, 87 percent are teachers or health workers. Wages for women are 30 percent lower than for men after controlling for educational qualifications and experience (World Bank 2005). A recent study documents that safety concerns and a strong patriachal society restrict the ability of women to find work outside the village where they live (World Bank 2005).

The government's position on the role of private schools has oscillated during the last 30 years. Private Schools were nationalized in 1972 amidst a mass government program of nationalization of all industry, but in 1979 the policy was reversed. Private schools were allowed to open, and schools taken over by the government were gradually returned to the original owners. The national policy did not include private schools or provide a subsidy in the form of grants (to parents or directly to schools) unlike in India or Bangladesh (where there are a large number of private *aided* schools). Prior to nationalization, private schools catered primarily to urbanites in large cities. Most private schools were run by missionaries or local schools imitating the missionary model, mainly for the elite.

Within this milieu, the government in concert with international donors tried to increase investment in human capital through a series of "Social Action Programs" or SAPs. The SAP initiated in 1980 called for large investments in the provision of education. School construction was a large part of the education component, and guidelines specified where schools should (and should not) be built.

Figure I shows that this program had an effect on school construction. The figure plots the number of years a school has existed for all villages in Punjab for four types of schools—primary and secondary for boys and girls. The kernel density shows a substantial increase in construction during the SAP III years; in the case of girls, this is the largest construction period since 1930. While most girls' secondary schools were built between 1960 and 1965, there is a definite hump around the SAP III years; in the case of boy's secondary schools, there is no such increase—most such schools appear to have been around since 1930.

A. What are the characteristics of rural private schools in Pakistan?

Recent empirical research makes a strong case for the private sector in education. Expanding school choice leads to better overall educational performance. This is due to both the direct effects of better performance in private schools (Jimenez and others 1999, Tooley and Dixon 2006, Figlio and Stone 1997) as well as improvements in public schools from increased competition and greater per-child resources in the government sector (Hoxby 1994). In low-income countries, greater private schooling also has significant fiscal implications—the unionization and pay-grade of public teachers implies that per-child costs of private schools is half that of their public counterparts (Jiminez and others 1999, Kip and others 1999, Orazem 2000, Hoxby and Leigh 2004). The data suggest similar patterns in Pakistan.

A study on Learning and Educational Achievement in Punjab Schools (LEAPS), involves 112 villages selected randomly in 3 districts of Punjab from a list of villages with at least one private school in 2000. This guarantees significant choice for all households in the sample, and the ability to compare public and private schools within the same geographical area. In these villages, we followed a matched sample of 12000 children, 5000 teachers, 800 schools, and 2000 households, over three years. Several findings from the first year, summarized in Table 1, are relevant for the current paper (all findings are drawn from Andrabi and others 2006b; Andrabi and others 2006c).

First, private schools are affordable. The median fee of a rural private school in Pakistan is Rs.1000 ($18) per year, which represents 4 percent of the GDP per capita. In contrast, private schools (elementary and secondary) in the United States charged $3,524 in 1991. At 14 percent of GDP per capita, the relative cost of private schooling is 3.5 times higher in the US. Perhaps because they are affordable, private schools are increasingly accessed by families in the lowest income decile.

Second, villages with private schools perform significantly better on a number of traditional educational outcomes (Table 1, Panel A). Overall enrollment is significantly higher for villages with private schools (61 percent vs. 46 percent), as is female enrollment (56 percent vs. 35 percent). In villages with private schools, 17 percent of the households in the poorest tercile are enrolled in private schools, which is comparable to the percentage of private school enrollment among the rich in villages *without* such schools (18 percent).

Higher enrollments could reflect a decrease in the distance to school in villages with private schools An important determinant of school enrollment is distance, and the adverse effect of greater distance to school is larger for girls compared to boys (Alderman and others 2001, Jacoby and Mansuri 2006, Holmes 1999, Andrabi and others 2006). Private schools decrease the average distance to school, thus increasing overall enrollment. Kip and others (1999) provide strong evidence for this hypothesis by examining a randomly allocated subsidy for the creation of private schools in rural Pakistan. They find that the subsidy led to increases of 14.6 and 22.1 percentage points in female enrollment for two of three program districts (Kip and others 1999 and Orazem 2000).

Third, private schools, while offering comparable if not better facilities (classrooms per-student, black-boards and seating arrangements), incur half the expenditures per child compared to public schools—Rs. 1,012 per child for the median private school compared to Rs. 2,039 for the median public one. These differences remain large after controlling for parental/village wealth and education. Since salaries are 90 percent of costs in private schools, savings stem primarily from lower wages. Private school teachers' wages are 20 percent of teachers in public schools (Table 1, Panel B). In addition, there is a large and significant discount for female teachers in the private sector (but not in the public sector). These differences remain large and significant after controlling for location (village fixed-effects), training, and education. With school fixed-effects and controls for education, training, experience and residence, private school male teachers earn 21 percent more than females.

Fourth, children's test-scores are higher in private schools. In tests administered through the LEAPS study, students finishing class 3 in private schools students outperform public school students (Table 1, Panel C), scoring 0.9 standard deviations higher in English, and 0.39 standard deviations higher in Mathematics. While selection into private schools could explain part of the difference, there is only a small change in the private-public learning gap after controlling for child, household, and village attributes. The difference for English drops to 0.85 standard deviations and that for Mathematics to 0.34 standard deviations. These results contrast with results from the US, where raw differences between private and public schools tend to be large, but differences are sharply reduced with demographic and location controls (Figlio and Stone).

Establishing causality for each these findings is empirically difficult. Nevertheless, the size of the differences in raw comparisons and with additional covariates presents strong *prima facie* evidence of the benefits of private schools in rural Pakistan. An important question then, is the extent to which they will arise in areas without government schooling options.

III Some Illustrative Patterns

The basic patterns of interest for this paper are laid out in Figures II, III and IV. Figure II shows the date of construction for every private school based on a census in 2001. A quarter of the private schools in 2000 were built in the previous year, and there is an exponential decline as we move further back in time. During the early half of the 1990s, most schools were located in urban areas. This shifted around 1995 with an increase in the share of private schools in rural villages upto 2000 (the last year for which we have data) when approximately one-half of all schools were set up in rural areas.[5]

Figure III shows the relationship between the existence of a private school and various types of government schools. We regress the existence of a private school on the number of years that the village has had a primary or secondary school (both boys' and girls'). The figure plots the predicted marginal probability of a private school against exposure to a public school; these probabilities are the marginal effect of exposure to each type of public school, controlling for other public schools in the village.

Boy's primary schools have an almost flat relationship with the location of a private school over the 20-year horizon. Girl's primary schools and boys' secondary schools do marginally better; the regression implies a 2-3 percentage point increase in the probability of a private school with 10 years of exposure. The role of girls' secondary schools stands out. There is a private school in one-fifth of all villages with a girl's high school, or a 7-8 percentage point increase for every additional 10 years of exposure. The marginal impact of a girl's high school on private school existence is large and significant.

Figure IV is another cut at the data showing the predicted relationship between private school existence and the percentage of men and women with high school education (8 or more years of schooling) in every village. Increasing the percentage of women with high school education in a village increases the probability of a private school from 10 percent (0 percent educated) to almost 100 percent (100 percent educated). Educated males again play an attenuated role—at the mean of the sample for female literacy, increasing high school males from 0 to 100 percent increases the probability of private school existence from 7 to 32 percent. Both these figures suggest that women matter—GSS and educated women in the village are have a qualitatively different role for the setting up of private schools.

[5] Although these data confound school construction and survival, comparisons with the number of schools reported in a 1985 study (Jimenez and Tan 1985) suggest that most of the increase is real. Correcting for the survival rate (around 5 percent of all schools fail every year) still implies a phenomenal growth rate during the 1990s.

IV Data

To examine whether these correlations are causal, we employ four data sources: (a) a complete census of private schools; (b) administrative data on the location and date of construction of public schools and (c) data on village-level demographics and educational profiles from the 2001 and 1981 population censuses. The Federal Bureau of Statistics undertook a census of all private schools in 2001 and data on public school construction is available from provincial Educational Management Information Systems (EMIS). Further, population censuses carried out in 2001 and 1981 (two years after the denationalization of private schools) provide both contemporaneous and baseline data on village-level characteristics.

An important step was matching these data at the village level. Given the high level of data fragmentation (these data are collected by different institutions without a common coding scheme) and variations in the spellings for village names some choices had to be made. By relying on phonetic matching algorithms and a manual post-match, we matched the public school (EMIS) data to both the 1981 and 2001 censuses. Given the availability and nature of the data, we study the province of Punjab and only for rural areas. Our matched sample consists of 18,119 villages out of 25,941 in 2001 in Punjab province with a population of 42.3 million in 1998, representing 84 percent of the total rural population of the province.

From this matched sample, we employ two restrictions to generate our final sample. First, our empirical strategy relies on the availability of village-level baseline data *prior* to the construction of a public school in the village—for villages with pre-existing public schools, we cannot discern whether differences in the baseline data arise from selection into villages or the exposure to a public school. Thus, we first restrict our sample to those villages that had *not* received a girls' secondary or primary school by 1981.

We are also concerned that the presence of pre-existing GSS in *neighboring* villages could bias our results. Pre-existing GSS in nearby villages could affect the demand for education even if the concerned village does not have a school in the baseline data. Moreover, a nearby school could decrease the probability that a village receives a GSS. Our final sample therefore consists of all geographical units, known as patwar-circles, within which *no* villages had a GSS prior to 1981.[6] In the absence of spatial data on the location of villages, patwar-circles, which are typically a group of 2-4 geographically contiguous villages, are a plausible measure of the local availability of schools.[7]

Table II presents summary statistics. Mirroring the construction of public schools (Figure 1), the average village has had a boys' public school for 27 years, a boys' secondary school for 6 years and a girls' primary school for 7 years. In contrast, exposure to a GSS is 0.7 years and this reflects the large number of villages without such a school—only 5 percent of the sample of villages received a GSS between 1981 and 2001. There is a private school in one out of every 7 villages and, continuing with the low availability of educated women, in the average village there are only 13 women (out of a population of 1830) who report secondary or higher education in 1998.

V Methodology and Empirical Framework

A simple framework outlines the private entrepreneur's problem, focusing on the role of the public sector and the econometric and interpretational issues in identifying the impact of a GSS. An entrepreneur opens a private school in village i if the net return, defined as the difference between total revenues and total costs, is positive.[8] For private schools, school fees and teachers' salaries account for 98.4 percent and 89 percent of total revenues and costs respectively (Andrabi and others 2006c). We write net return as:

$$NetReturn_i = Fee_i * N_i - Wage_i * T_i \tag{1}$$

where Fee_i is the average private school fee for a single student, $Wage_i$ is the average private school teacher's salary and N_i and T_i are the number of students enrolled and teachers

[6] We are less worried about girls' primary schools in neighboring village affecting village demand, since there is considerable evidence that younger children do not travel outside their village to go to school (Alderman, Jacoby and Mansuri, Andrabi and others 2005b).

[7] Our results are similar across the restricted sample and the full sample.

[8] This assumes that there is no shortage of entrepreneurs (otherwise not every positive NPV project will be undertaken). While we can incorporate such shortages, doing so will not change the qualitative results. The qualitative results of the model also extend to a dynamic framework provided that the fixed costs of setting up schools is small.

employed. Since the schooling market may be geographically segregated, we allow wages and fees to differ across villages.

The construction of a GSS increases the supply of teachers in the village, thus affecting $Wage_i$; they also increase the potential demand for schooling, reflected in Fee_i. A reduced form expression for net return can then be written as:

$$NetReturn_i = \alpha + (\beta_1 + \gamma_1)GSS_i \ + \beta' X_i^D + \ \gamma' X_i^S \tag{2}$$

where X_i^D and X_i^S are village demographics and characteristics that respectively affect the demand for private schooling and the costs of running such schools. Variables included in X_i^D and X_i^s include village population, measures of village wealth, adult literacy and alternative schooling options. GSS construction has two effects in Equation(2): It affects the demand for private education (negatively by acting as a substitute and positively by creating a more educated populace) through β_1 and the cost of setting up private schools (by affecting the local supply of potential teachers) through γ_1. We are interested in the joint estimation of $(\beta_1 + \gamma_1)$; below we discuss what we can say about the likely sizes of these two coefficients.

Since the net return a private school earns is not observed, we treat net return in equation(2) as a latent variable in a probability model, so that $Prob(PrivateSchoolExists) = Prob(NetReturn_i > 0)$ and estimate a version of Equation(2):

$$\Pr ivate_{it} = \alpha + (\beta_1 + \gamma_1)GSS_{it} + \beta' X_{it} + \ \sum_r \gamma_r' J_{irt} + (v_i + \varepsilon_{it}) \tag{3}$$

where $\Pr ivate_{it}$ is a binary variable that takes the value 1 if a private school exists in village i in time t, GSS_{it} is a binary variable that takes the value 1 if a GSS exists in village i in time t. X_{it}^D observed characteristics village characteristics at time t and J_{irt} are other government schooling options at time t, where each option is indexed by r. The error term, $(v_i + \varepsilon_{it})$ consists of a time-invariant unobserved component, v_i and a random component, ε_{it}. The presence of a program in village i in time period t is likely a function of the latent unobserved components of the region:

$$GSS_{it} = \alpha_1 + \varphi X_{it} + (v_i + \mu_{it}) \tag{4}$$

This simple framework highlights the main empirical issues. The OLS estimate of $(\beta_1+\gamma_1)$ in Equation(3) is biased and inconsistent since $cov(\nu_i+\varepsilon_{it}, GSS_i) \neq 0$. Pitt, Rosenzweig and Gibbons (1995) show that an unbiased estimate of $(\beta_1+\gamma_1)$ is obtained with two periods of data if $cov(\varepsilon_{it}, \mu_{it}) = 0$, by differencing Equation(3) across the two different points in time.[9]

We first estimate Equation(3) with and without lagged village characteristics and geographical dummies, D_j, at the level of the patwar-circle.

$$\Pr iSchool_{it+1} = \alpha + (\beta_1+\gamma_1)GSS_{it+1} + \beta' X^D_{it+1} + \gamma' X^S_{it+1} + \beta' X^D_{it} + \gamma' X^S_{it} + \sum_{j=1}^{M} \delta_j D_j + (\nu_i + \varepsilon_{it}) \tag{5}$$

The estimated $(\widehat{\beta_1+\gamma_1})$ is biased if $cov(\nu_i+\varepsilon_{it}, GSS_i) \neq 0$. We then present the first-differenced specification at the level of the village, with and without aggregate time trends at the level of the Patwar Circle:

$$\Delta_t \Pr iSchool_i = \alpha + (\beta_1+\gamma_1)\Delta_t GSS_i \ + \beta' \Delta_t X^D_i + \sum_r \gamma'_r \Delta_t J_{irt} + (\varepsilon_{it+1} - \varepsilon_t) \tag{6}$$

where Δ_t represents the difference in the variable over the two observed time periods. Finally, we also implement the non-parametric equivalent of Equation(6) through first-differenced propensity score matching techniques. That is, we compare the change in private schools to the change in GSS for matching villages, where the matching is implemented on the baseline data. Differences in the estimated $(\beta_1+\gamma_1)$ between Equation(5) and Equation(6) are also informative about where GSS were constructed. In particular, an increase in the estimated impact of GSS across the two equations suggests that GSS were selectively built in villages where private schools were less likely to arise.

The estimated $(\beta_1+\gamma_1)$ in Equation(6) is still biased if $cov(\varepsilon_{it}, \mu_{it}) \neq 0$. There are several reasons why this may be so: Jalan and Ravallion (1998) suggest that state-dependence leads to a systematic correlation between initial levels and future growth, violating the "parellel" trends assumption of the first-differenced specification. Alternatively, village-level time-varying shocks could both affect the construction of a GSS and a private school: a new road may lead

[9] In the presence of lagged program effects, the estimated coefficient yields an unbiased estimate of the contemporaneous effects of the program, as long as program changes are not correlated to lagged program shocks (Pitt, Rosenzweig and Gibbons 1995).

to better job opportunities, leading to higher demand for a GSS and higher returns to private schools. The particular setting and the program through which the public school construction was undertaken provides a promising instrumentation strategy to address potential correlations in time-varying village attributes; we turn to this next.

A. A Rule-Based Instrumentation Strategy

The instrumental variables strategy follows Campbell [1969] and Angrist and Lavy [1999]. We exploit the fact that the regressor of interest, in our case the construction of a government school, is partly based on a deterministic function of a known covariate, village population. If this deterministic function is non-linear and non-monotonic, it can be used as an instrument while directly controlling for linear and polynomial functions of the underlying covariate itself.

School construction in our sample was a direct consequence of the Pakistan Social Action Program in 1980. The GSS constructed under the SAP were not add-on's to existing primary schools. Under the SAP, a certain number of GSS had to be established, and the government could undertake one of three actions in a village: (a) it could either not construct a girls' primary or secondary school (b) it could construct a girls' primary school or, (c) it could construct a GSS that also included a primary section. In this sense, the construction of a GSS implied that the building and sanctioning of teachers followed both the protocols for a primary and secondary school—there would be more classrooms and more teachers, although in the initial years, the number of children in the secondary classes was likely to be very low. Reflecting this design, out of the 328 villages in our sample that received a GSS between 1981 and 2001, only 31 had a pre-existing girls' primary school; in all the rest, the secondary and primary sections of the school were constructed simultaneously.

Specific guidelines dictated where these schools could (and could not) be built. For GSS the official yardstick for opening a new school was that (i) the population of the village be no less than 500 (ii) there be no GSS within a 10 km radius and (iii) the village would have to provide land for the school, measured as 16 *kanals* in local land-area units.

We construct a binary assignment rule, $Rule_i$, that exploits the first two guidelines. $Rule_i$ is assigned a value 1 if (and only if) the village's population in 1981 was greater than 500 *and* it was the largest village (in terms of 1981 population) amongst nearby villages. The latter captures the radius criteria: if a village is not the largest village amongst its neighbors, the neighbor would receive a GSS first given the stated preference for population in the construction of schools. Provided this school is near enough, the village will be less likely to receive its own public school.[10]

One option is to use the geographical distance from schools to construct the eligibility rule. Such geographical markers are not available for the villages in our sample. Furthermore, it is unlikely that these distances could be accurately computed and used at the time of program implementation. In the absence of geographical data we use the administrative jurisdiction of the "Patwar Circle" (PC) to approximate the radius rule: villages are eligible to receive a GSS if there are no other villages in the same PC that already have a GSS. In terms of actual land area, this is a reasonable approximation—dividing the size of the province by the number of patwar circles, shows that the simple strategy of providing one school in every patwar circle would satisfy the radius requirements of the rule. Our instrument thus defines a village as "eligible for a GSS" if it's population was at least 500 in 1981 and it was ranked first in 1981 population within the PC. Formally:

$$Rule_i = \begin{array}{l} 0 \; if \; Population_i^{81} < 500 \\ 1 \; if \; Population_i^{81} \geq 500 \; AND \; Population_i^{81} = \max\limits_{j \in PC_i} (Population_j^{81}) \\ 0 \; if \; Population_i^{81} \geq 500 \; AND \; Population_i^{81} < \max\limits_{j \in PC_i} (Population_j^{81}) \end{array}$$

The eligibility rule is non-linear (it jumps from 0 to 1 at a fixed population threshold), discontinuous, and non-monotonic—it drops to 0 for larger villages when there is an even larger neighboring village within the patwar circle. In using this rule as an instrument, we can explicitly control for continuous functions of a village and its neighbors' populations since these covariates have a direct impact on the existence of a private school.

[10] Another alternative is to use the radius-rule directly and assign $rule_i = 0$ if there is a village in the patwar-circle that has a GHS. This is problematic since we are worried about the endogenous placement of GHS in the first place.

A-1. Identifying Assumptions

We estimate Equation(3) using $Rule_i$ as an instrument for GSS_{it+1}. $Rule_i$ is a valid instrument if

1. The first stage: $cov(Rule_i, GHS_{it+1}) \neq 0$ and,
2. The exclusion restriction: $cov(Rule_i, \nu_i + \varepsilon_{it}) = 0$, so that the eligibility rule affects the existence of a private school only through the construction of a GSS.

Since $Rule_i$ is necessarily correlated with 1981 population, the exclusion restriction is satisfied only if we explicitly condition on 1981 population in Equation(3). Under these two conditions, the Instrumental Variable (IV) estimate of $(\beta_1 + \gamma_1)$,

$$(\widehat{\beta_1 + \gamma_1})_{IV} = \frac{cov(\Pr ivate_{it}, GSS_{it}|Pop_{1981})}{cov(GSS_{it}, Rule_i)} \tag{7}$$

is unbiased and consistent. To clarify the identifying assumptions, consider how our instrument relates to 1981 population and how population in turn affects our outcome of interest. Figure V illustrates how the existence of private schools and the binary instrument covary with 1981 village population. While all villages below 500 are ineligible ($Rule_i = 0$), there are a substantial number of villages above this threshold that are not in the eligible group—these are villages that are not top-ranked within their patwar circle. We can thus compare two villages with the same population, one of which was eligible to receive the GSS and another that was not, allowing us to exclude the direct effect of population on private school existence.

Figure VI illustrates further. We divided villages into five population quintiles. The top-half of the figure compares the percentage of villages with a GSS in the "eligible" group compared to "not eligible" group; over the entire sample, this difference represents the "first-stage" of the instrumental variables (IV) estimate, $cov(GSS_{it}, Rule_i)$ The bottom-half then compares, over the same population quintiles, the percentage of villages with a private school in the "eligible" compared to the "not eligible" group; this is the reduced form for the IV estimate, $cov(\Pr ivate_{it}, GSS_{it}|Pop_{1981})$. Several features are noteworthy.

First, the instrument varies in every population quintile except for the first, which consists of villages below the 500 threshold—our results are not driven by variation in a single population group. Second, for all population quintiles the first-stage indicates that eligible villages were more likely to receive a GSS. Third, in contrast to the rank rule, the threshold rule (villages less than 500 were not supposed to receive a GSS) was not applied. A number of villages below the 500 threshold received a GSS and, rather than a discontinuity at 500, there is an almost linear increase in the probability of receiving a GSS with population around this threshold. Further comparisons around the 500 threshold confirm the lack of discontinuities that the eligibility rule, if strictly applied, should have implied. Our identification therefore rests on the population rank-rule, rather than a discontinuous jump in GSS construction around the 500 threshold.[11]

The exclusion restriction could fail If private school entrepreneurs search for the highest returns and choose among villages within a PC (but not a broader area), in which case, they will likely choose the village with the highest population. This assumes: (a) that there is a shortage of (local) entrepreneurs, so that even in villages where the net present value of setting up a private school is positive, a school is not set up and (b) that private entrepreneurs need not be resident in the village where they set up the school. Towards ruling out such explanations, we construct a falsification test and show that the instrument has no predictive power in PCs where no GSS were built—the population rank of a village plays a role *only if* the PC also received a GSS.

The exclusion restriction also fails if the government used the same strategy for allocating *other* investments that may affect the return to private schooling in the village. To account for such investments, we control for the presence of all types of government schools in addition to GSS, which could affect the probability of private school existence.[12] Second, we include regressors that are plausible measures of other government investment, such as electrification and water-supply. We find no evidence that the rank of the village affected the level of government investment, either in means-comparisons or in regression specifications. Indeed, other government investments such as water and electricity are slightly *lower* in higher population rank villages, although coefficients are small and insignificant.

[11] In robustness checks, we repeat the IV specifications without the threshold rule and replicate the results reported below.

[12] A potential issue here is that our instrument may also predict the construction of other (than girls high) types of public schools. However, once we condition on a village and its PC's maximum population in the restricted sample the rule predicts post-1981 construction only for GHS.

A final econometric concern is that the existence of a GSS and a private school are both binary variables. Additionally, the percentage of villages that received a GSS is very low. Although at higher treatment and outcome probabilities, estimated coefficients are similar in linear and non-linear models (Angrist 1991), at low treatment probabilities the choice of specification will make a difference. In particular, while the linear IV bias is relatively small, standard-errors will be much larger[13]. Following the literature, we propose two sets of estimates based on a linear and a bivariate probit specification. The bivariate specification requires the assumption of joint normality of the error term in the determination of public and private schools and is formally stated as

$$\begin{aligned} Pub_{it} &= 1(\delta Rule_i + u_{it} > 0) \quad (8) \\ \Pr i_{it} &= 1(X_{it}^{'}\gamma + aPub_{it} + \varepsilon_{it} > 0) \end{aligned}$$

$$\begin{bmatrix} u_{it} \\ \varepsilon_{it} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \quad (9)$$

B. Isolating the Supply Side

The proposed instrumental variables strategy isolates the causal impact of GSS on private school existence. To separate supply from demand-side channels we propose two strategies based on the relative effect of women versus men on the setting up of private schools (the quantity margin) and the costs of operating private schools in villages with and without GSS (the price margin).[14]

On the quantity margin, a supply-side channel suggests several patterns. If the teachers'

[13] Linear specifications assume a uniform error term while non-linear specifications impose a normal or logistic distribution on the error. When treatment probabilities are neither high nor low, the cumulative distributions of the uniform and normal (roughly) match up, yielding very similar results. With low treatment probabilities though, the cumulative distributions differ leading to a dramatic increase in the standard error of linear IV models.

[14] Another possibility is to use variation in the timing of the public school construction. Supplys-side channels suggest that private schools will emerge 5-8 years after the construction of a GHS. Unfortunately, the data are too limited to exploit this variation. We require villages with both private schools and GHS. Since only 328 villages received a GHS, and of these, 30 percent had a private school, we are unable to identify any discontinuities using the 90 or so villages that have both. Another possibility is to check whether there is a difference in the existence of a private school based on years of exposure to a GHS. Here we do find evidence that less than 5 years of exposure has no effect on the likelihood of private school existence. In particular, private schools exist in 18 percent of villages with less than 5 years of exposure to a GHS (compared to 12 percent among the control), and in 33 percent of those with 5 or more years.

supply effect is important compared to parental demand, the impact of *secondary* schools should be larger than that of primary schools. The assumption is that an adult with primary-level education is not a suitable teaching candidate (98 percent of private school teachers have at least secondary education), but the difference in parental demand for primary education does not differ substantially across parents with primary versus secondary education. We examine the differential impact of public schools by level (primary versus high) and gender (boys versus girls). We expect secondary schools to be more important than primary schools, and, since teachers in private schools are primarily women, GSS to have a larger impact than boy's high schools.

We also focus directly on the supply of educated men and women. We examine whether there are more adult women with higher education in villages that received a GSS and, in turn, whether the existence of a private school increases with the fraction of adult women with higher education in the village (both absolutely and relative to adult men with higher education).

While results in the expected direction from these tests lend support to the supply-side channel, alternative explanations based on demand channels are possible: Several studies document the relative importance of women compared to men for investment in the human capital of their children, and secondary education could have a differential impact on parental demand compared to primary education.

An alternative test can be based on the price margin. The price implications of GSS impact through supply and demand differ. If the existence of a private school responds to an increase in the supply of teachers, GSS should reduce the wage-bill of private schools.[15] In the absence of a GSS, the entrepreneur would either have to bring in teachers from outside (if possible) and pay for the compensating differential and travel costs; alternatively she would have to pay higher wages for educated male teachers.[16]

[15] An alternative is to use data on the wages of women with high school education (i.e. the level necessary to become primary school teachers), which would avoid the selection problem of missing data in villages without private schools. However, there are few sources that provide such village level wage estimates. Moreover, very few women work in non-farm activities so that sample sizes are typically small, and selection issues are hard to deal with. While these wages cannot be separated by gender, since the vast majority of private school teachers are women, the wage bill is likely to reflect wages to skilled women.

[16] An alternative would be to look at wages of secondary-school educated women who may or may not be teachers. The only available data source is the Pakistan Integrated Household Survey. Unfortunately, given

We test whether our data support this hypothesis. One complication is that we observe non-missing values of the wage-bill only in villages with private schools; this selection problem will underestimate the true wage-bill differences in the population.[17] We follow two approaches to address the selection problem. We use a Heckman selection model, where the selection stage is the probability of observing a positive wage, which corresponds to having a private school in the village. Another alternative is to use the "control-function" approach , where we condition on the predicted probability of observing a non-missing value of the wage-bill in the wage equation (Angrist 1995). Details of both approaches are in the Appendix.

In both approaches identification is based on the non-linearity of the selection equation (see Duflo 2001 as an example). Augmenting the instrument set with potential candidates that are correlated to the probability of having a private school but uncorrelated to the wage-bill can help in identification and the efficiency of the estimator. Following Dowes and Greenstein (1996), we propose using the *number of public boys primary schools* as an additional instrument in the selection equation. In the presence of competitive schooling effects, private schools should be less likely to setup in villages where there are public boy's primary schools; additionally, such schools are unlikely to affect the wage-bill of the entrepreneur directly. This instrument serves as a "robustness" check to the identification based on non-linearities in the selection equation.[18]

Finally, simultaneous changes in the demand for schooling induced by GSS imply that the estimate of GSS on the wage-bill will be a lower bound—we cannot structurally estimate the size of the supply side effects. For the example in Footnote 11, if GSS increases the demand for private education, the minimum wage-bill at which private schools are profitable in villages with GSS is higher than those without. In this case, $E(WB_i|\text{no GSS})$ - $E(WB_i|\text{GSS})$ recovers the joint-effect of a lower wage-bill and increases in demand.

the small number of villages that received a GHS, the available sample sizes are too small—with the sample restrictions in our paper, we find only 3 villages in the treatment and 31 villages in the control set for these data.

[17] Consider the following wage-bill distributions in villages with and without GHS.

$$[5, 6, 7, 8, 9, 10] \text{ (Without GHS)}$$
$$[3, 4, 5, 6, 7, 8] \text{ (With GHS)}$$

In the absence of any demand effects, suppose that private schools can only afford to set up in villages where the wage-bill is below 7. Thus, where we observe the wage-bill, $E(WB_i|WB_i$ is non missing, no GHS) - $E(WB_i|WB_i$ is non missing, GHS) is lower than the uncensored $E(WB_i|\text{no GHS})$ - $E(WB_i|\text{GHS})$.

[18] This instrument is less than perfect if boys' primary schools are differentially located in villages where the returns to education are low or if men and women compete directly for positions in the labor market for teachers.

VI Results

A. Baseline Differences

For the estimating sample, that is, villages in PCs which did not have any GSS in 1981, we define "treatment" villages as those that received a GSS between 1981 and 2001 and "control" villages as those that did not (these include both villages that did not receive a girls' school at all, or those that received a girls' primary school). For the full set of village characteristics available in the 1981 census including literacy indicators, gender ratios and demographic attributes there were no significant differences between treatment and control villages. Consistent with the stated rule, the population in treatment villages was almost twice as large as in the control. Using measures of infrastructure and public goods in these villages from the 1998 census shows little wealth differences (Appendix Table I). In fact judging from the percentage of households that own land in 1998, treatment villages fare slightly worse than control villages.

B. OLS and First-Difference Specifications

Table IV presents results based on Equation(3). The construction of a GSS increases the probability of a private school in the village by 9.7 percentage points (Column 1, Table 3). Since 12 percent of all control villages have a private school, this represents an 80 percent increase. An equally significant determinant of private school existence is village population; the GSS effect is similar in magnitude to increasing village population by 2000 individuals (coincidentally a one standard-deviation increase in population). The estimated impact remains significant at the 1 percent level with a full set of village-level controls including exposure to other types of public schools, although the point-estimate is somewhat attenuated (Column 2, Table IV). Introducing location dummies for PCs increases the estimate and significance with magnitudes very similar to the first specification (Column 3, Table IV).

Following Equation(6), Columns (4) and (5) control for time-invariant village effects by first-differencing the data at the village level. The effect of a GSS on private school existence increases to 15 percentage points. Adding in aggregate time-trends at the level of the PC (Column 5) further increases the impact of a GSS to 17.4 percentage points. Finally, propensity score estimates yield similar results: A GSS increases private school existence probabilities by 11 to

14 percentage points depending on whether we use local linear regression or kernel matching (results available with authors).

The near-doubling of the estimated impact of GSS as we progressively account for unobserved components of the error suggests that, even though there are no differences in baseline characteristics, GSS were not randomly placed. Indeed, it appears that villages where private schools were *less* likely to setup were disproportionately more likely to receive a GSS. The next section presents the IV results, which show that unobserved time-varying attributes were equally important.

C. Instrumental Variable Specifications

C-1. First-Stage and IV Results

Table V, Columns (1-3) present a series of first stage regressions using the eligibility rule as a predictor for the location of GSS. Without additional controls, a village with population above 500 that was ranked first in population terms within the PC (and is thus "eligible"), was 6.4 percentage points more likely to receive a GSS (Column 1, Table V). Since 4.7 percent of all villages received such a school, this is close to a 130 percent increase.

Part of this is a population effect whereby larger villages are likely to rank high within the patwar circle and more likely to receive a GSS. In Column 2, we condition on linear and quadratic terms of the village's population in 1981 and the maximum village population in the PC in 1981. Although the point-estimate is reduced by half, it remains large and highly significant: Villages that satisfy the eligibility rule were 55 percent more likely to receive a GSS. We obtain similar results with a more exacting first stage that includes a full set of location dummies for administrative jurisdictions, known as "qanoongho halqa" (QH), which include around 10 PCs (Column 3, Table 4).[19] Moreover, the explicit conditioning on polynomial population terms implies that the remaining variation induced by the instrument is non-monotonic and non-linear and therefore likely uncorrelated with omitted variables in Equation(3).[20]

[19]We cannot use PC fixed effects since our instrument and controls rely on PC level population measures (i.e. the maximum village population in a PC).

[20]A useful characterization of the strength of an instrument is the "concentration parameter", inferences for which can be based on the F-statistic from the first-stage. For all three specifications—without additional controls, with population controls and with additional geographical dummies—the F-statistic is greater than 14

Columns (4) to (5) present the corresponding linear IV coefficients. The estimated coefficient of GSS increases dramatically and the significance drops to the 10 percent level. While part of this increase can be attributed to selection on time-varying omitted variables, we think it unlikely that these effects are as large as the estimates suggests. Column (6) assesses whether functional specification plays a role. We implement a bivariate probit specification following equation (8) and report the marginal impact of GSS on the existence of a private school. The standard errors are bootstrapped at sample values of other variables (alternative standard errors calculated at the mean of the sample-value for other variables yield similar results). The point-estimate from the bivariate probit is half that of the linear IV and significant at the 5 percent level. The estimate suggests selection on unobservable variables, and is double what we obtain with the first-differenced specification: Constructing a GSS increases the probability of a private school in the village by 36 percentage points, or over 300 percent.

C-2. Linear IV and Biprobit: Further Robustness Tests

Low treatment probabilities could partly explain the large differences between the biprobit and the linear IV estimator. In Monte-Carlo simulations, we consider various treatment (T) and outcome (Y) probabilities and examine the efficiency of linear IV and biprobit estimates (Das and Lokshin 2006). When treatment (or outcome) probabilities are very low/high, the standard-error of the linear IV estimator increases dramatically and linear IV and biprobit point estimates can differ substantially. In the context of our data, these deviations are indeed large with 87 percent of the sample concentrated in a single cell, (T=0 and Y=0).[21]

Further structure on the manner in which GSS were located can help understand the difference between these two estimators as well as provide evidence on the validity of the instrument. The final sample consists of 6968 villages, of which 2128 are in the "eligible" group (the instrument, $Rule_i = 1$) and 4840 are in the "ineligible" group ($Rule_i = 0$). However, only 328 villages eventually received the treatment of which 60 percent are in the eligible and 40 percent

and exceeds the proposed critical thresholds (approximately 9) for testing the null hypothesis that the instruments are weak (Stock, Wright and Yogo 2002).

[21] The literature on the effects of Catholic schooling on graduation rates is similarly structured (very few children are in Catholic schools) and encounters similar problems. In a recent paper (Altonji and others 2005), the authors suggest that differences between the linear IV and biprobit estimator arise due to the added identification power from the non-linear structure of the latter. They also suggest strategies to understand the validity of the instrument and identification in these different estimators, which motivate the robustness tests discussed here.

are in the ineligible group. Consistent with our interpretation of the policy rules, 96 percent of all PCs with a GSS have a single such school, while the remaining minority have two. However, the vast majority of PCs have none.

There are two ways to interpret these placement patterns. Administrators could either place GSS in villages across the entire sample using the eligibility rule as a guide or, they could first have selected PCs where GSS would be built and then repeated the same exercise. This suggest that the population of PCs can be divided into two sub-groups—"Program PCs", where at least one village in the PC received a GSS and "Non-Program PCs" where no village received a GSS. Even in the absence of knowledge about the *PC* selection rule, comparisons across program and non-program PCs are instructive.

First, the ideal case for a valid instrument, $Rule_i$, is random assignment of $Rule_i$ across villages. If so, we should find no difference in the observable characteristics of villages that are eligible versus those that are not. Table VI compares our set of of observable characteristics in the eligible and non-eligible group; consistent with random assignment, there are no significant differences across the two groups with the exception of the size of the village, which follows directly from the construction of the instrument.

Second, if population rank within the PC has no independent effect on the probability of setting up a private school, we should find a strong relationship between private school existence and eligibility for villages in program PCs, but *not* in non-program PCs. Indeed for program PCs, being in the eligible group increases the probability of a private school by 14 percentage points; conversely in non-program PCs the same eligibility criteria has no impact on private school existence with a small (.008 percentage points) and insignificant coefficient (Columns 5 and 6 in Table 7).

Third, comparing coefficient estimates across the full sample and program PCs sheds some light on the large differences between the linear IV and biprobit coefficients noted above. For a valid instrument, the program impact estimated for program PCs only is identical to the impact estimated on the full sample.[22] However, standard-errors will be smaller when the estimation is

[22] Consider a binary instrument Z, a treatment, T and a binary outcome variable Y in a sample of N villages. Assume further that these N observations can be divided into M administrative blocks, equivalent to patwar

only over program PCs, suggesting that there should be greater agreement between the linear IV and biprobit estimates for this restricted sample.

In Table 7 we repeat the first-stage, linear IV and biprobit estimates for villages in the program PCs only. The eligibility rule increases the probability of a GSS by 30 percentage points, the linear IV estimate drops significantly to 0.47 and there is a slight decline in the biprobit estimate from 0.35 to 0.29. Both estimates have lower standard errors and are significant at 1 percent levels of confidence. Taken together, these results suggest that the large difference obtained in the full sample arose partially from the low treatment probabilities combined with the non-linear structure of the problem; across the two samples, the biprobit estimates imply a causal impact of GSS on private school existence ranging from 29 to 35 percentage points.[23]

The differences between the OLS, first-difference and IV results indicate that both time-invariant and time-varying components of the error term are correlated to GSS placement. Further, GSS were systematically placed in villages where private schools were less likely to arise. One interpretation—advanced for instance, by Pitt, Rosenzweig and Gibbons (1995) in Indonesia—is that governments act altruistically, trying to equalize differences between villages. Villages with lower responsiveness of demand to school construction received GSS and these were also the villages where private schools were less likely to locate.

We are somewhat sceptical of this altruistic argument, given the nature of the state in Pakistan's history. Education department officials who were in the ministry at the time of the program confirm that apart from the SAPP rules, local-level politics also played a role. A more cynical explanation is that these schools were targeted to villages with powerful local landlords

circles in our case. The program operates in a small number of $M_1 << M$ blocks, where we do not know the selection rule determining the choice of the M_1 blocks (the "program areas"). Even in the absence of any knowledge of the selection rule determining the program areas, the Wald estimator $\frac{E(Y|Z=1)-E(Y|Z=0)}{E(T|Z=1)-E(T|Z=0)}$ applied only to program areas is an unbiased estimate of the treatment effect, if the treatment effect is homogeneous across villages. Further, the coefficient is identical to the estimation repeated over the entire sample as long as Z is a valid instrument so that $E(Y|Z=1, Non\Pr ogram) = E(Y|Z=0, Non\Pr ogram)$ Formally, both the numerator and denominator of the Wald estimator are weighted by $\frac{n_1}{n_1+n_2}$ when restricted to program areas where n_1 is the number of observations in the program areas and n_1+n_2 is the size of the full sample. In the presence of covariates the linear IV estimator $\widehat{\beta}_{IV} = \frac{Cov(Y,Z)}{Cov(Y,T)}$ yields similar results.

[23] If treatment effects are homogenous across villages, the effect estimated for program PCs only is the average treatment effect on the treated. If treatment effects are heterogenous across program and non-program PCs and selection of program PCs is non-random, our estimates are valid only for program PCs. The problem is identical to interventions evaluated over a set of villages where an organization has *chosen* to work with the usual caveats regarding the external validity of the estimate to non-program areas.

and officials. The context in Pakistan suggests that these are precisely the villages where the demand for education is lower, and less likely to increase over time. Construction in villages with a lower demand for education could thus reflect political-economy considerations rather than a desire for equity.

D. Potential Channels

In Tables V-VI we examine four questions: (a) the relative importance of GSS for private school existence compared to other types of public schools; (b) the contribution of GSS to the supply of educated women; (c) the contribution of educated females to the existence of a private school, and (d) the relationship between the wage-bill of a private school and the construction of a GSS.

If private schools arise because of the availability of "women as teachers", we expect a larger impact of GSS compared to other types of public schooling. Columns (1)-(2) in Table V present estimates from a probit and linear probability model, where the latter includes PC-level location dummies. Both specifications confirm the importance of GSS relative to other types of public schooling, with coefficients for years of exposure to a GSS almost three times as large an effect as that of the next most important public school type.

Since selection effects are important, Columns (3) and (4) present results from a first-difference specification with and without cluster-specific time-trends. These results magnify the importance of GSS: the change (from 1981 to 1998) in whether a village has a GSS or not is the *only* schooling variable that matters, and the magnitude of the effect is large. Whether a village received a boys' primary/high or girls' primary school between 1981 and 1998 has no affect on the likelihood of a private school setting up in the village.

Columns (5)-(8) present the next logical step. We assess the correlation between educated women and the presence of a GSS for a variety of specifications. In all specifications, a GSS increases the percentage of adult women with higher levels of education (equal to or more than 8 years of schooling) by 1.5-2.2 percentage points and the estimated increases are significant at the 1 percent level of confidence. While this may appear as a small effect, it represents a change in the *stock* of educated women. Since 1.3 percent of all women in an average village in 1981 reported higher levels of education, a GSS more than doubles this percentage.

Columns (9)-(12) in Panel B examines the importance of secondary school educated women for the existence of a private school. While the effect of educated men is only slightly smaller in the basic probit specification in column (9), the difference between educated men and women increases substantially once we control for geographical location, suggesting that part of the estimated coefficient on male education reflects omitted geographical characteristics. For our preferred first-difference specification, the impact of women with 8 or more years of schooling remains as strong while the percentage of similarly educated males has *no* impact on the existence of a private school.

These results suggest that supply-side factors are important, since they constrain the routes through which a demand-side story can work: It must be the case that fathers' education does not stimulate demand for children's' education (since boys' schools have no effect) and that primary schooling for mothers is not enough. Indeed, there must be strong nonlinearity in the demand for children's education and mother's schooling.

Table VI examines the price margin; we compare the wage-bill in private schools in villages with and without GSS using data from the private school census. Column (1) presents the OLS results in the sample of villages for which wage data is not missing.[24] The results are large and significant and in the direction predicted by the supply-side channel. Private schools in villages with a GSS report an 18 percent lower wage-bill.[25]

Columns (2)-(5) correct for selection into the sample. Columns (2)-(3) present results using Heckman's selection model and Columns (4)-(5) use the "control function" approach (see Appendix). Columns (2) and (4) identify from the functional form of the selection equation and Columns (3) and (5) introduce an additional instrument for the selection stage—the number of government boys' primary schools in the village. The results are similar to the OLS estimates, with estimates of 20-21 percent suggesting that selection into the non-zero wage sample is of limited importance.

Together with the results on the quantity margin—that GSS and not other forms of public schooling, and female higher education and not male higher education, affect private school

[24] This is slightly smaller than the number of villages where there is a private school since in a few cases in the PEIP data private schools did not report wages.

[25] Attenuation bias from a noisy measure of women's wages (average wages in private schools), implies that the actual differential may be even higher.

existence—the wage-bill results present direct evidence that the supply-side "women-as-teachers" channel is a factor in determining the location of private schools are setup. Institutional constraints limit the geographical and occupational mobility of women. In this environment, GSS reduce the wage-bills of a private school by increasing the supply of potential teachers; this in turn allows private school to take advantage of an affordable local supply of teachers. Government investment thus alleviates the key logjam implied by restricted mobility.

VII Conclusion

With 2 million adolescents in Pakistan poised to enter the labor-force each year, human capital investments will play a large role in determining where these children will fit into a fast globalizing world. Children in rural private schools will have learnt a lot more by the time they leave the formal schooling system than their counterparts; at the same time, their annual education costs will have been half as much. However, not everyone has access to private schools—children are far more likely to attend private schools if they are located in their own village, and less than one in six villages has a private school.

Government school investments can lead to facilitate private sector involvement and part of the effects are through supply-side channels that increase the availability of local teachers. This is similar in spirit to calls for public investment-led growth in the spirit of "big-push" arguments advanced by Rosenstein-Rodan (1943) and Murphy, Shleifer and Vishny (1989). In contrast to the literature that calls for larger primary school compared to secondary school investment, our findings suggest that both play a role. That the students in today's schools are the potential repositories of human capital for the next generation implies that low-income countries can enter a "virtuous cycle" by investing heavily in the creation of a cohort of educated women.

The results on *how* girls' secondary schools lead to the creation of private schools in the next generation are a testimony to the resilience of the private sector and its ability to convert cultural constraints—the restricted mobility and labor-market opportunities for women—into an advantage. Villages with girls secondary schools are also those with a larger stock of educated women, who can then teach in private schools. With limited mobility, a private school entrepreneur becomes a virtual monopsonist when located in such a village. At one level, this seems like a pernicious outcome: Women receive lower wages in the labor market compared to men for the same job. At the same time, these cost savings are directly passed on to the children who study in these schools; in the absence of this labor market distortion, it is unclear whether these schools could have catered to the children from low and middle-income families that are currently enrolled in them.

The results also provide a fascinating glimpse of education in high-income countries during the early to mid-twentieth century, and particularly the debate on the effect of increasing labor force participation for women on the quality of teachers. The rise of private schools in Pakistan suggest that in low-income countries at least, the "implicit-subsidy" to education from low female labor-force participation is alive and kicking.

Appendix

Selection Issues in the Wage Bill

Since we only observe the wage bill in villages where there is a private school, a concern described in the main text is that simple OLS estimates may be biased if such selection is not accounted for. Here we provide details on two approaches we use in the paper to address such concerns. Following Angrist (1995), the problem can be formally stated as follows. The wage-bill is determined through a linear equation conditional on the existence of a private school

$$WB_i = \alpha + \beta GHS_i + \varepsilon_i \tag{10}$$

and a censoring equation (denoting $WB_i = I$ as the indicator for whether WB_i is non-missing)

$$WB_i = I\{\delta GHS_i - \nu_i > 0\} \tag{11}$$

The instrument Z_i determines a first stage

$$GHS_i = \gamma + \mu Z_i + \tau_i \tag{12}$$

Given the validity of the instrument, Z_i, we assume that $cov(\tau_i, Z_i) = 0$. Then,

$$E(\varepsilon_i|Z_i, WB_i = 1) = E(\varepsilon_i|Z_i, (\delta\gamma + \delta\mu Z_i > \nu_i - \delta\tau_i)$$

so that $cov(\varepsilon_i, Z_i) \neq 0$ in equation(10) above. Thus, although Z_i is a valid instrument for the decision to setup a private school, it is not a valid instrument in equation(10). There are two potential solutions.

Following Heckman (1978) if we assume that $(\varepsilon_i, \nu_i, \tau_i)$ are jointly normally distributed, homoskedastic and independent of Z_i, we obtain the familiar "mills-ratio" as the relevant expectation function conditional on participation. That is,

$$E(\varepsilon_i|Z_i, (\delta\gamma + \delta\mu Z_i > \nu_i - \delta\tau_i) = \lambda(\delta\gamma + \delta\mu Z_i)$$

where $\lambda(\delta\gamma + \delta\mu Z_i) = \frac{-\phi(\lambda(\delta\gamma+\delta\mu Z_i))}{\Phi(\lambda(\delta\gamma+\delta\mu Z_i))}$ and $\phi(.)$ and $\Phi(.)$ are the density and distribution functions of the normal distribution for $\nu_i - \delta\tau_i$. This mills-ratio can is then directly included in equation(10) as the appropriate selection-correction.

An alternative approach, proposed by Heckman and Robb (1986) and developed by Ahn and Powell (1993) uses the "control-function" approach, where we condition on the predicted probability of $WB_i = 1$ in equation(10). In essence, this method proposes to estimate β by using pair-wise differences in WB_i for two villages (in our case) for which the non-parametric probability of participation is very close. The approach is implemented by first estimating equation(11) directly, and then including the predicted probability of participation (and its polynomials) as additional controls in equation(10).

References

[1] Ahn, H. and J.L. Powell. 1993. "Semi-parametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," Journal of Econometrics 58, 3-29.

[2] Alderman, Harold, Peter Orazem, and Elizabeth Paterno. "School Quality, School Cost and the Public/Private School Choices of Low-Income Households in Pakistan," *Journal of Human Resources* 36(2):304-326 Spring 2001

[3] Alderman, Harold, Jooseop Kim, and Peter Orazem. "Design, Evaluation, and Sustainability of Private Schools for the Poor: The Pakistan Urban and Rural Fellowship School Experiments," *Economics of Education Review* 22(3):265-274. 2003

[4] Altonji, Joseph, Todd Elder and Christoper Taber 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools" forthcoming *Journal of Political Economy.*

[5] Altonji, Joseph, Todd Elder and Christoper Taber 2005."An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schools." forthcoming, *Journal of Human Resources*

[6] Andrabi, Tahir, Jishnu Das, Asim Khwaja and Tristan Zajonc, 2006a, "Religious Education in Pakistan: A Look at the Data", forthcoming, Comparative Education Review.

[7] Andrabi and others 2006b. "Report on Learning and Educational Attainments in Pakistani Schools. Working Paper.

[8] Andrabi, Tahir, Jishnu Das and Asim Khwaja 2006c. "Private Schooling in Pakistan: Possibilities and Limits." Working Paper.

[9] Angrist, J. and V. Lavy, 1999. "Using Maimonedes Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, Vol. 114, No. 2, 533-575.

[10] Angrist, J. 1995."Conditioning on the Probability of Selection to Control Selection Bias," NBER Technical Working Paper No. 181. Cambridge, Massachussetts.

[11] *Banerjee, Abhijit V. "Economics of the Family." Journal of Development Economics*

[12] Besley, Timothy, and Stephen Coate, "Public Provision of Private Goods and the Redistribution of Income." The *American Economic Review* 81(4), 979-984. September 1991.

[13] Campbell, D.1969. "Reforms as Experiments," *American Psychologist*, XXIV, 409-429.

[14] *Chaudhury and others "Missing in Action" Journal of Economic Perspectives.*

[15] Downes, Thomas A. and Shane M. Greenstein. 1996. "Understanding the Supply Decisions of Nonprofits: Modeling the Location of Private Schools" *Rand Journal of Economics* Vol. 27(2)

[16] Downes, Thomas A. & Greenstein, Shane M. 2002. "Entry into the Schooling Market: How is the Behaviour of Private Suppliers Influenced by Public Sector Decisions?" *Bulletin of Economic Research* 54 (4), 341-371.

[17] Dulfo, E. (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," American Economic Review, Vol. 91, No. 4.

[18] *Duflo and Hanna "Monitoring Works"*

[19] Figlio, David N. and Joe E. Stone. 1997. "School Choice and Student Performance: Are Private Schools Really Better?" Discussion Paper 1141-97. Institute for Research on Poverty. University of Wisconsin.

[20] Fisher, Gary. 1976. "*The Identification Problem in Econometrics.*" Robert E. Kreiger Publishing Company, Huntington, New York.

[21] Government of Punjab, Department of Education. 2002. "Educational Management Information System for Punjab."

[22] Government of Pakistan, 1983, Sixth Five Year Plan 1983-88, Planning Commission, Islamabad

[23] Government of Pakistan, 1997, Report on the PIHS 1995-96, Islamabad

[24] Holmes, Jessica. 1999. "Measuring the Determinants of School Completion in Pakistan: Analysis of Censoring and Selection Bias." Center Discussion Paper Number 794, Economic Growth Center, Yale University.

[25] Heckman, J. J., and R. Robb. 1986. "Alternate Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in H. Wainer, ed., Drawing Inferences from Self-Selected Samples, New York: Springer-Verlag.

[26] Hoxby, Caroline M. 1994. "Do Private Schools Provide Competition for Public Schools?" National Bureau of Economic Research Working Paper Number 4978.

[27] Hoxby, Caroline M. and Andrew Leigh. 2004. "Pulled Away or Pushed Out? Explaining the Delhine of Teacher Aptitude in the United States." *American Economic Review*, Vol.93(2)

[28] Jacoby, Hanan and Ghazala Mansuri. 2006. "Crossing Boundaries: Social Distance and the Gender Gap in Schooling in Rural Pakistan." Mimeo.

[29] Jimenez, Emmanuel and Jee-Peng Tan. 1985. "Educational development in Pakistan : the role of user charges and private education, Volume 1." Education and training series Discussion Paper Number EDT 16. World Bank. Washington, D.C.

[30] Jimenez, Emmanuel and Jee-Peng Tan. 1987. "Decentralized and private education : the case of Pakistan, Volume 1." Education and training series Discussion Paper Number EDT 67. World Bank. Washington, D.C.

[31] Jimenez, Emmanuel, Marlaine E Lockheed and Vicente Paqueo. 1999. "The relative efficiency of private and public schools in developing countries." *World Bank Research Observer.* Vol. 6, no.2 (July 1991), pp. 205-218.

[32] Kingdon, Geeta. 1996. "The Quality and Efficiency of Private and Public Education: A Case-study of Urban India." *Oxford Bulletin of Economics and Statistics.*

[33] Kim, Jooseop, Harold Alderman, and Peter Orazem. "Can Private School Subsidies Increase Schooling for the Poor? The Quetta Urban Fellowship Program," *World Bank Economic Review* 13(3):443-466. 1999

[34] Murphy, Kevin M & Shleifer, Andrei & Vishny, Robert W, 1989. "Industrialization and the Big Push," *Journal of Political Economy*, University of Chicago Press, vol. 97(5), pages 1003-26.

[35] Orazem, Peter F. 2000. "The Urban and Rural Fellowship School Experiments in Pakistan: Design, Evaluation and Sustainability." The World Bank. Washington, DC.

[36] Pitt, Mark M., Mark R. Rosenzweig and Donna M. Gibbons, 1995. "The Determinants and Consequences of the Placement of Government Programs in Indonesia." In D. van de Walle and K. Nead, eds., *Public Spending and the Poor: Theory and Evidence*. Baltimore: Johns Hopkins University Press for the World Bank.

[37] Psacharopolous, G. 2002. "Returns to Investment in Education: A Further Update." World Bank Policy Research working paper series Number 2881. World Bank, Washington D.C.

[38] Rodenstein-Rodan, Paul. 1943. "Problems of Industrialization of Eastern and South-Eastern Europe." *Economic Journal* 53 (210-211): 202-211.

[39] Stock, James H., Jonathon H. Wright and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." Journal of Business and Economic Statistics, Vol.20(4) 518:529.

[40] Tilak, Jandhyala B.G. and Ratna M. Sudarshan. 2001. "Private Schooling in Rural India." National Council of Applied Economic Research Working Paper #WP010009. NCAER, New Delhi, India.

[41] Tooley, James and Pauline Dixon. 2005. "Private Education is Good for the Poor." Cato Institute. Washington DC.

[42] Tooley, James. "Private Schools for the Poor." *Education Next*. Fall 2005: 22-33.

[43] World Bank. 2005. "Improving Gender Outcomes: The Promise for Pakistan.". World Bank, Washington, D.C.

TABLE I
PRIVATE SCHOOLS IN PUNJAB

	Villages With private schools (Punjab)	Villages Without Private Schools (Punjab)	Difference
PANEL A			
Percentage Enrolled	61	46	15
Percentage Females Enrolled	56	35	21
Percentage Males Enrolled	67	55	12
Private Enrollment Share	23	11	12
Public Enrollment Share	76	88	-12
PrivateEnrollment Share (Poor Only)	17	6	11
Private Enrollment Share (Middle Only	18	11	7
Private Enrollment Share (Rich Only)	34	18	16
PANEL B			
Differences in Wages	**Private Schools**	**Public Schools**	**Difference**
Men	1758.28	6394.18	4635.89
	(-1284.52)	(-2678.37)	(-122.46)
Women	1067.270	5888.480	4821.21
	(761.540)	(2066.280)	(55.58)
All	1231.000	6178.000	4946
	(959.140)	(2447.010)	-55.71
PANEL C			
Differences in Test Scores	**Private Schools**	**Public Schools**	**Difference**
English Scores (Raw Percentage Correct)	41.800	24.400	17.400
	(15.500)	(15.080)	(0.400)
English Scores (Item Response Scaled Scores)	0.640	-0.260	0.900
	(0.630)	(0.910)	(0.020)
Mathematics Scores (Raw Percentage Correct)	43.430	34.560	8.870
	(16.610)	(18.520)	(0.470)
Mathematics Scores (Item Response Scaled Score)	0.360	-0.030	0.390
	(0.660)	(0.820)	(0.020)

Note: All numbers presented in this table are drawn from Andrabi and others (2006b) and Andrabi and others (2006c). The data for Panel A is based on the Pakistan Integrated Household Survey (2001), a representative survey of households in the four main provinces. The data for Panel B are based on a survey of 5000 teachers in public and private schools as part of the LEAPS project, and the data for Panel C are based on tests of over 12,000 children in 800 public and private schools as part of the same project.

TABLE II
SUMMARY STATISTICS

variable	mean	sd	N
Years Exposure - GHS	0.68	3.23	6968
Years Exposure - GPS	7.19	7.24	6960
Years Exposure - BHS	6.34	20.35	6960
Years Exposure - BPS	26.73	20.39	6869
Private School Exists?	0.13	0.34	6968
Number of Private Schools	0.22	0.81	6968
1998 % Enrolled in Private Schools	0.10	0.21	902
1998 Population	1829.09	2023.31	6968
1981 Population	1210.50	1272.31	6968
1998 Number of Women w/ Matric and Above Education	13.07	39.36	6968
1998 % HHs w/ Permanent Housing	0.07	0.05	6723
Village Land Area	1647.79	2340.71	6874
Number of Villages in Patwar Circle	4.38	2.12	6968

Summary statistics are for the sample of villages that (a) did not have girls' high or primary school prior to 1981 and (b) villages whose neighbors did not have a girls' high school before 1981. Land is measured in Kanals.

TABLE III

BASELINE DIFFERENCES IN MEANS

	Treated	Not Treated	Difference
Number of Villages	328	6640	
1981 Female Literacy Rate	0.017	0.015	0.002
	(0.007)	(0.001)	(0.007)
1981 - % adult women with Middle and above Education	0.016	0.012	0.004
	(0.007)	(0.001)	(0.007)
1981 % girls age 0-4	0.154	0.155	-0.001
	(0.020)	(0.004)	(0.020)
1981 % girls age 5-14	0.289	0.285	0.004
	(0.025)	(0.006)	(0.026)
1981 adult Male Literacy Rate	0.184	0.164	0.020
	(0.021)	(0.005)	(0.022)
1981 - % adult men with Middle and above Education	0.135	0.116	0.019
	(0.019)	(0.004)	(0.019)
1981 % boys age 0-4	0.143	0.143	0.001
	(0.019)	(0.004)	(0.020)
1981 % boys age 5-14	0.295	0.293	0.002
	(0.025)	(0.006)	(0.026)
1981 Female/Male Ratio	0.911	0.906	0.005
	(0.016)	(0.004)	(0.016)
1981 Population	2069.69	1168.05	901.63***
	(94.17)	(15.12)	(71.16)

The table shows baseline differences between treatment and control villages. Standard-errors of t-tests or proportion tests (as appropriate) are in parenthesis.

Table IV - Private School Existence and Previous Girls High Schools

	(1)	(2)	(3)	(4)	(5)
	Probit	Probit - All controls	OLS (PC Location Dummies)	First diference	First diference & PC Dummies
Treatment- Received GHS	0.097	0.0646	0.0928	0.1494	0.1739
	(0.0223)	(0.0207)	(0.0247)	(0.0250)	(0.0241)
1998 Population (000s)	0.051	0.0391	0.0905		
	(0.0032)	(0.0075)	(0.0176)		
1998 Population (000s) Sq	-0.0014	-0.0011	-0.0046		
	(0.0002)	(0.0003)	(0.0014)		
1981 Population (000s)		0.0275	0.0134		
		(0.0133)	(0.0281)		
1981 Population (000s) Sq		-0.0013	0.0029		
		(0.0012)	(0.0041)		
% Perm Houses		1.2862	0.9383		
		(0.0821)	(0.1804)		
1998-1981 Population (000s)				0.0795	0.1162
				(0.0070)	(0.0079)
Years Exposure - GPS		0.001	-0.0001		
		(0.0005)	(0.0007)		
Years Exposure - BPS		0.0001	0.0004		
		(0.0002)	(0.0003)		
Years Exposure - BHS		0.0011	0.002		
		(0.0002)	(0.0003)		
With Patwar-Circle Dummies	NO	NO	YES		
With PC cluster-specific time trends				NO	YES
Observations	6968	6761	6761	6968	6968
Pseudo R-sq	0.1	0.18			
Adj R-sq			0.34	0.07	0.3

The table shows the relationship between the existence of a private school and GHS. Columns (1) and (2) estimate non-linear probability models (probit) and column (3) the corresponding linear specification. Columns (4) and (5) present results from the village-level first-differenced specification. Robust standard errors in parentheses.

Table V - Private School Existence - Instrumental Variables

	(1)	(2)	(3)	(4)	(5)	(6)
	First- Stage Probit	First-Stage Probit	First-Stage (QH Location Dummies)	Linear 2nd-Stage	Linear 2nd-Stage- QH Location Dummies	BiProbit - Boostrapped SEs (xx vars are also included but Cioeffs and Ses not reported)
Girls High School Rule	0.0642	0.026	0.0297			
	(0.0067)	(0.0079)	(0.0078)			
Treatment- Received GHS				0.8277	0.7507	0.3567
				(0.4808)	(0.4368)	(0.1331)
1981 Population (000s)		0.0268	0.0335	0.0269	0.0194	xx
		(0.0059)	(0.0066)	(0.0263)	(0.0252)	
1981 Population (000s) Sq		-0.0023	-0.0017	-0.0016	-0.0008	xx
		(0.0006)	(0.0008)	(0.0015)	(0.0015)	
1981 Max Population (000s) in PC		-0.0025	0.0065	-0.0041	-0.0061	xx
		(0.0049)	(0.0063)	(0.0089)	(0.0094)	
1981 Max Population (000s) sq in PC		0.0005	0.0002	0.00001	-0.0004	xx
		(0.0005)	(0.0008)	(0.0014)	(0.0012)	
1998 Population (000s)				0.0403	0.0544	xx
				(0.0105)	(0.0099)	
1998 Population (000s) Sq				-0.0001	-0.0009	xx
				(0.0005)	(0.0005)	
% Perm Houses				1.3055	0.7645	xx
				(0.1053)	(0.1516)	
Observations	6968	6968	6968	6874	6874	6874
Chi-sq/F-Test (GHS Rule = 0)	122.61	12.94	14.67			
Pseudo R-sq	0.05	0.07				
Number of QGH 1998					656	
Prob > chi2	0	0			0	0
Prob > F			0	0		
Adj R-sq			0.07			

The first three columns in the table show the first-stage of the IV strategy. Column (1) shows the bivariate correlation between the eligibility rule an GHS. Columns (2) and (3) are the corresponding first-stages for Columns (4) and (5); Column (6) reports the estimated marginal impact of GHS and bootstrapped standard-errors for a bivariate probit specification (xx represents variables included in the regression, but whose marginal coefficients and standard errors we have not bootstrapped for computational convenience). Standard errors in parentheses.

TABLE VI
BASELINE DIFFERENCES IN MEANS

	Instrument=1	Instrument=0	Difference
Number of Villages	2128	4840	
1981 Female Literacy Rate	0.013	0.015	-0.002
	(0.002)	(0.002)	(0.003)
1981 - % adult women with	0.012	0.013	-0.001
Middle and above Education	(0.002)	(0.002)	-0.003
1981 % girls age 0-4	0.159	0.153	0.006
	(0.008)	(0.005)	(0.009)
1981 % girls age 5-14	0.287	0.285	0.002
	(0.010)	(0.006)	(0.012)
1981 adult Male Literacy Rate	0.161	0.167	-0.005
	(0.008)	(0.005)	(0.010)
1981 - % adult men with Middle	0.110	0.120	-0.010
and above Education	(0.007)	(0.005)	(0.008)
1981 % boys age 0-4	0.145	0.142	0.004
	(0.008)	(0.005)	(0.009)
1981 % boys age 5-14	0.296	0.292	0.004
	(0.010)	(0.007)	(0.012)
1981 Female/Male Ratio	0.905	0.907	-0.002
	(0.006)	(0.004)	(0.008)
1981 Population	2246.05	755.20	1210.50***
	(37.72)	(8.14)	(15.24)

The table shows baseline differences between villages for which the Instrument is one and zero. Standard-errors of t-tests or proportion tests (as appropriate) are in parenthesis.

Table VII - Private School Existence - Instrumental Variables in Restricted Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	First-Stage (QH Location Dummies)	Linear 2nd-Stage	Linear 2nd-Stage- QH Location Dummies	BiProbit - Boostrapped SEs (xx vars are also included but Cioeffs and Ses not reported)	Reduced Form - Program PCs	Reduced Form - Non-Program PCs
Girls High School Eligibility Rule	0.3128				0.1405	0.008
	(0.0528)				(0.0396)	-0.0125
Treatment- Received GHS		0.4718	0.4839	0.2905		
		(0.1377)	(0.1699)	(0.0859)		
1981 Population (000s)	0.2365	-0.0063	-0.1489	xx	0.047	0.0473
	(0.0530)	(0.0811)	(0.1079)		(0.0730)	(0.0174)
1981 Population (000s) Sq	-0.0207	0.0038	0.0197	xx	-0.0031	-0.0021
	(0.0066)	(0.0092)	(0.0128)		(0.0074)	(0.0014)
1981 Max Population (000s) in PC	-0.0985	0.0384	0.1471	xx	-0.0086	-0.0057
	(0.0467)	(0.0425)	(0.0693)		(0.0316)	(0.0090)
1981 Max Population (000s) sq in PC	0.0118	-0.0027	-0.0154	xx	0.0024	0.0001
	(0.0057)	(0.0062)	(0.0078)		(0.0047)	(0.0014)
1998 Population (000s)		0.0437	0.0899	xx	0.062	0.0452
		(0.0476)	(0.0557)		(0.0446)	(0.0096)
1998 Population (000s) Sq		-0.0014	-0.0037	xx	-0.0017	-0.0006
		(0.0040)	(0.0043)		(0.0036)	(0.0003)
% Perm Houses		1.7332	1.1143	xx	2.0205	1.2917
		(0.3273)	(0.5953)		(0.3123)	(0.0954)
Observations	804	804	804	804	804	6070
Chi-sq/F-Test (GHS Rule = 0)	32.59					
Pseudo R-sq	0.2					
Number of QGH 1998			259			
Prob > chi2			0	0		
Prob > F	0	0			0	0
Adj R-sq	0.07	0.08			0.24	0.13

The first column in the table show the first-stage of the IV strategy in the reduced sample of "Program-PCs" only where a Program PC is defined as the PCs where at least one village received a GHS after 1981. Columns (2) and (3) are the second-stages for the linear IV estimator; Column (4) reports the estimated marginal impact of GHS and bootstrapped standard-errors for a bivariate probit specification (xx represents variables included in the regression, but whose marginal coefficients and standard errors we have not bootstrapped for computational convenience). Columns (5) and (6) presented the reduced form estimates for the Program PC and non-Program PC samples (the latter serves as a falsification test for our Instrument). Standard errors in parentheses.

Table VIII - Private School Existence - The Female Teacher Channel?

PANEL A

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: Private School Existence				Dependent Variable: Percentage of Women with Middle and Above Education			
	Probit	OLS - Controls & PC Location Dummies	First Difference	First diference & PC Dummies	OLS	OLS- Controls & PC Location Dummies	First Difference	First diference & PC Dummies
Years Exposure - GHS	0.0044	0.0059						
	(0.0010)	(0.0016)						
Years Exposure - GPS	0.0016	-0.0002						
	(0.0006)	(0.0007)						
Years Exposure - BHS	0.0013	0.002						
	(0.0002)	(0.0003)						
Years Exposure - BPS	0.0002	0.0004						
	(0.0002)	(0.0003)						
Treatment- Received GHS					0.0221	0.015	0.015	0.0183
					(0.0037)	(0.0042)	(0.0031)	(0.0039)
1998-1981 Population (000s)			0.0798	0.116			-0.0014	0.0039
			(0.0071)	(0.0081)			(0.0012)	(0.0013)
Change in Exposure - GHS			0.1515	0.16				
			(0.0255)	(0.0250)				
Change in Exposure - GPS			0.0103	-0.008				
			(0.0081)	(0.0107)				
Change in Exposure - BHS			-0.0645	-0.0314				
			(0.0438)	(0.0693)				
Change in Exposure - BPS			-0.0144	-0.0126				
			(0.0088)	(0.0114)				
Location Dummies	NO	YES	NO		NO	YES	NO	
Cluster-Specific Time-Trends	NO	NO	NO	YES	NO	NO	NO	YES
Observations	6854	6761	6854	6854	6967	6767	6964	6964
Pseudo R-sq	0.12							
Adj R-sq		0.34	0.07	0.3	0.01	0.5	0.003	0.38

PANEL B

	(9)	(10)	(11)	(12)
	Dependent Variable: Private School Existence			
	Probit	Controls & PC FEs	First Difference	First diference & PC Dummies
% middle & above females	0.4149	0.52		
	(0.0819)	(0.1217)		
% middle & above males	0.3506	0.0783		
	(0.0469)	(0.0738)		
Change in % Females middle+			1.0146	0.5801
			(0.1029)	(0.1153)
Change in % Males middle+			0.0498	-0.0118
			(0.0531)	(0.0716)
1998-1981 Population (000s)			0.0839	0.1186
			(0.0076)	(0.0080)
Observations	6967	6873	6964	6964
Pseudo R-sq	0.17			
Prob > chi2				
Adj R-sq		0.34	0.09	0.3

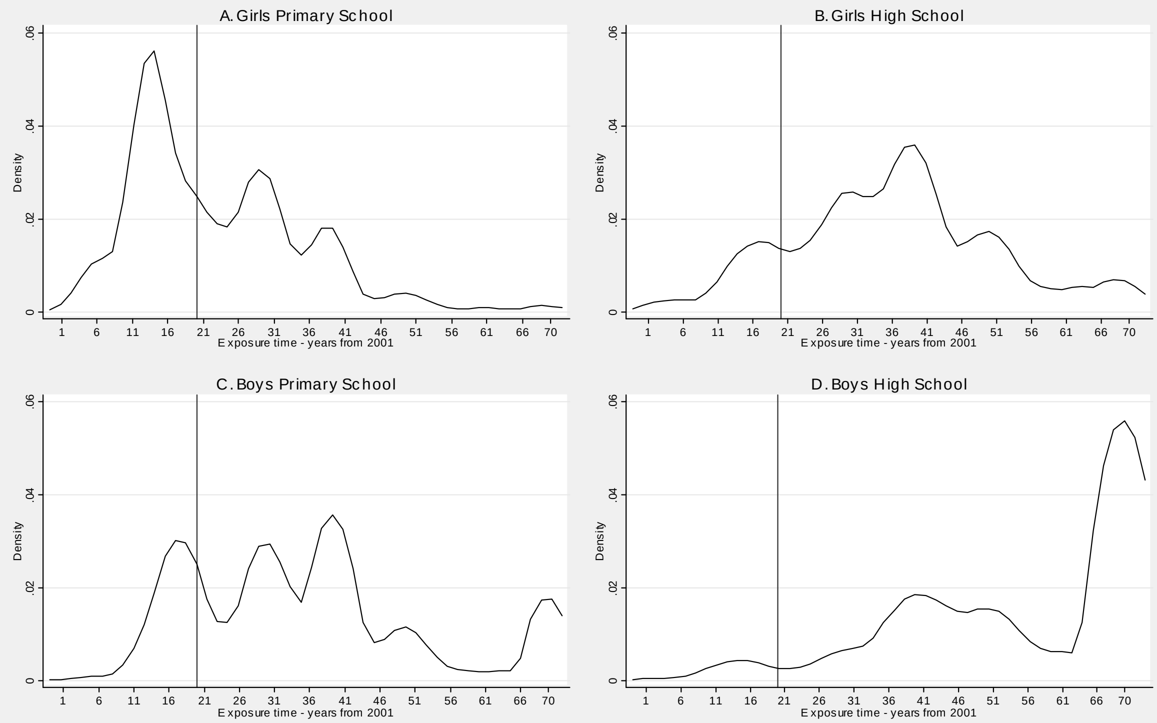
Columns (1) to (4) examines the relationship between different types of government schools and private school existence. Column (1) is a probit, Column (2) a linear specification with location dummies; Columns (3) and (4) are the village-level first-difference. Columns (5) to (8) look at the impact of GHS on female higher education. Columns (9) to (12) looks at private school existence and female/male education.

Table IX - Supply Side Impact - Teaching Costs

	(1)	(2)	(3)	(4)	(5)
	OLS - Controls & QH Dummies	Heckman- Controls & QH Dummies	Heckman - Controls & QH Dummies, BPS	Control Function Controls & QH Dummies	Control Function Controls & QH Dummies, BPS
Treatment- Received GHS	-0.1977	-0.2015	-0.2037	-0.2031	-0.2094
	(0.1078)	(0.0790)	(0.0794)	(0.1079)	(0.1083)
Years Exposure - BHS	0.0006	0.0004	0.0004	0.0002	0.0002
	(0.0010)	(0.0008)	(0.0008)	(0.0011)	(0.0011)
1998 Population (000s)	0.0329	0.0022	0.0127	-0.0161	-0.0047
	(0.0233)	(0.0320)	(0.0309)	(0.0451)	(0.0433)
1998 Population (000s) Sq	-0.0004	0.0004	0.0001	0.0010	0.0007
	(0.0010)	(0.0011)	(0.0010)	(0.0015)	(0.0014)
Observations	877	6967	6967	877	877
Pseudo R-sq					
Prob > chi2		0	0		
Adj R-sq	0.15			0.15	0.15

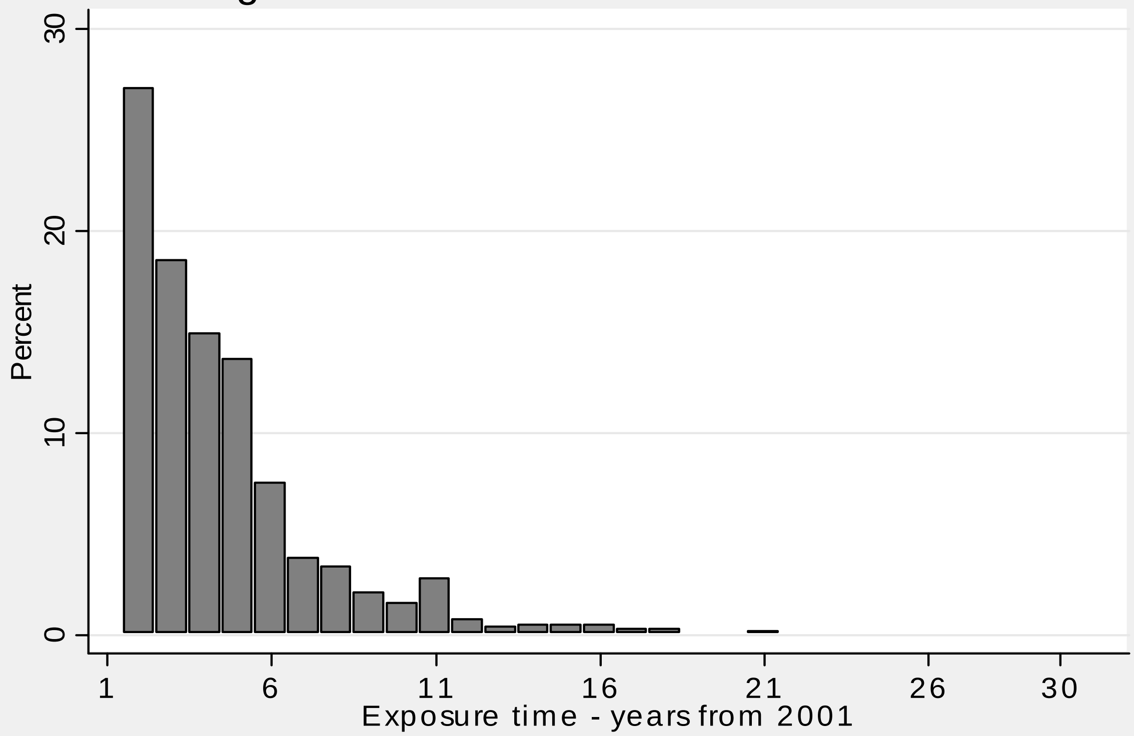
Columns (1) to (5) examines the relationship between average wage bill in private schools and government high schools. Column (1) runs an OLS specification. Columns (2)-(3) run a Heckman selection model to take into account the fact that the LHS variable is only observed in villages where private schools exists. Column (3) differs in that it includes an additional instrument for the selection stage - the number of government boys primary schools. Columns (4)-(5) present an alternate "control function" method to account for the selection issue by directly including polynomials in the predicted probability of observing a positive wage in the wage regression. Column (5) differs in that it includes an additional instrument for the selection stage - the number of government boys primary schools.

Figure I. Government School Exposure
conditional on receiving school



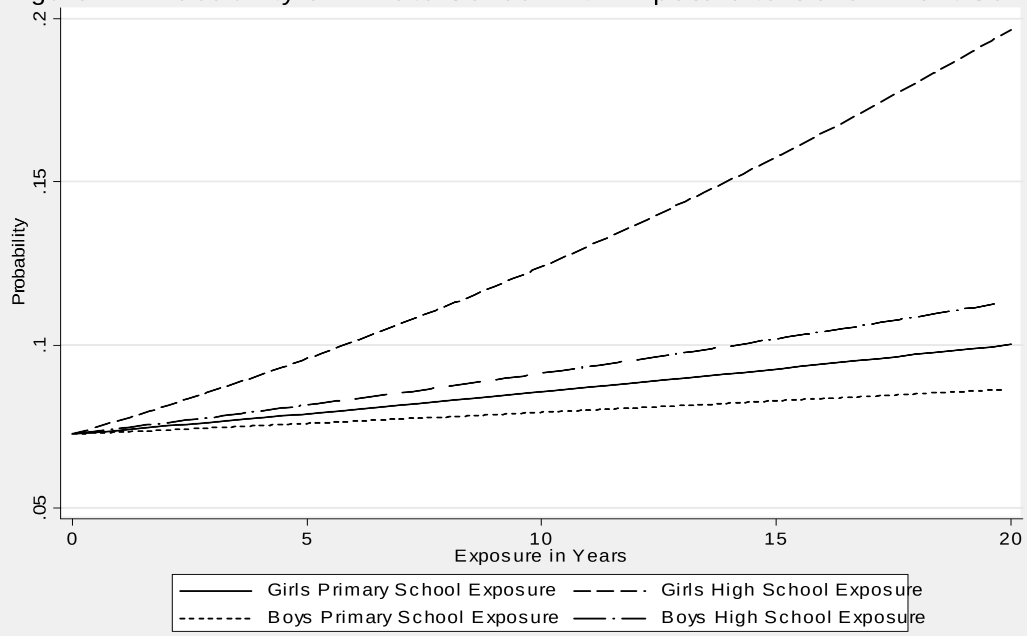
Note: The figure plots the number of years that schools have existed in different villages for four types of schools--Girl's Primary Schools (Panel A), Girl's High Schools (Panel B), Boy's Primary Schools (Panel C) and Boy's High Schools (Panel D). For instance, a large number of villages have had a girl's primary school for 11-16 years in Panel A. The data are from the Punjab Educational Management Information System. Schools more than 70 years old are coded as 70.

Figure II. The Growth of Private Schools

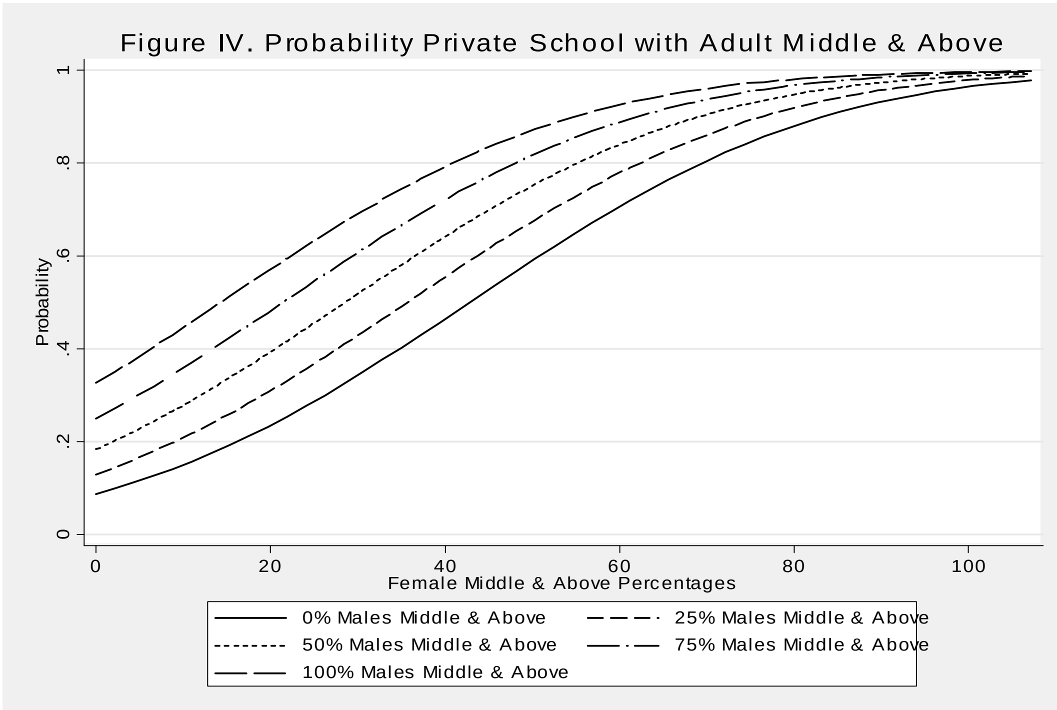


Note: The figure plots the number of years that schools have been active from a census of all private schools in the country in 2001. For instance, 25 percent of all private schools in the country had been active for only 1 year at the time of the survey. The data are based on the survey of private educational institutions carried out by the Federal Bureau of Statistics (2001).

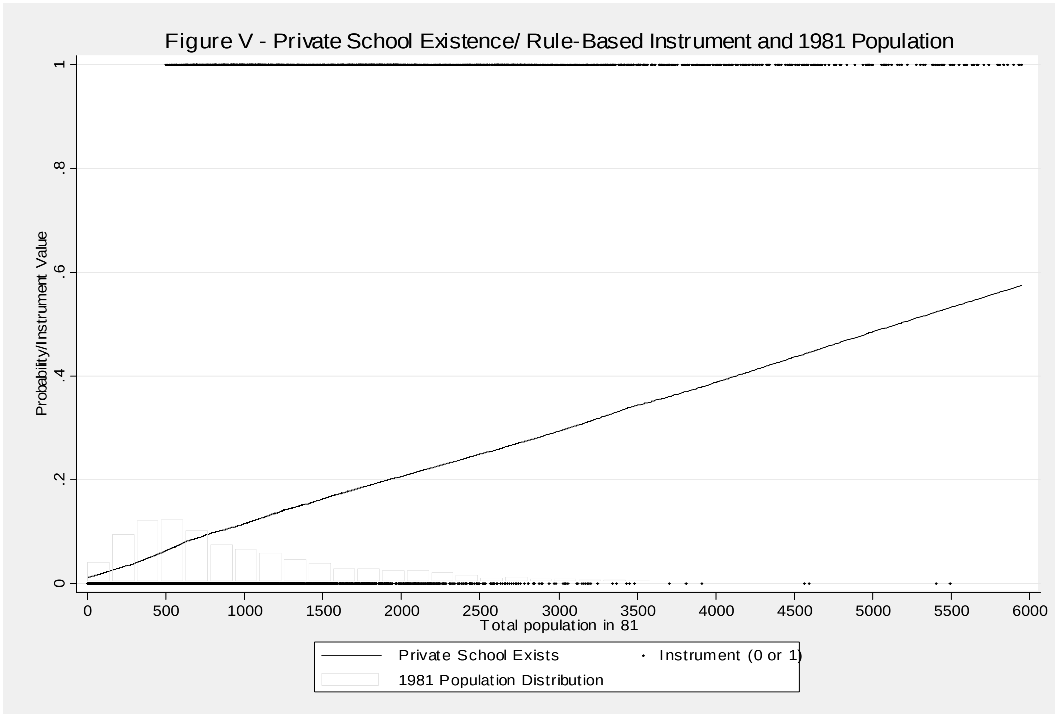
Figure III. Probability of Private School with Exposure to Government Schoc



Note: The figure plots the predicted probability of private school existence against the number of years that different types of schools have existed in the village. The predicted probability is based on a probit regression, with all schools included in a single regresion. The data area based on the census of private schools (Federal Bureau of Statistics) matched to public schools (Educational Management Information System for Punjab).

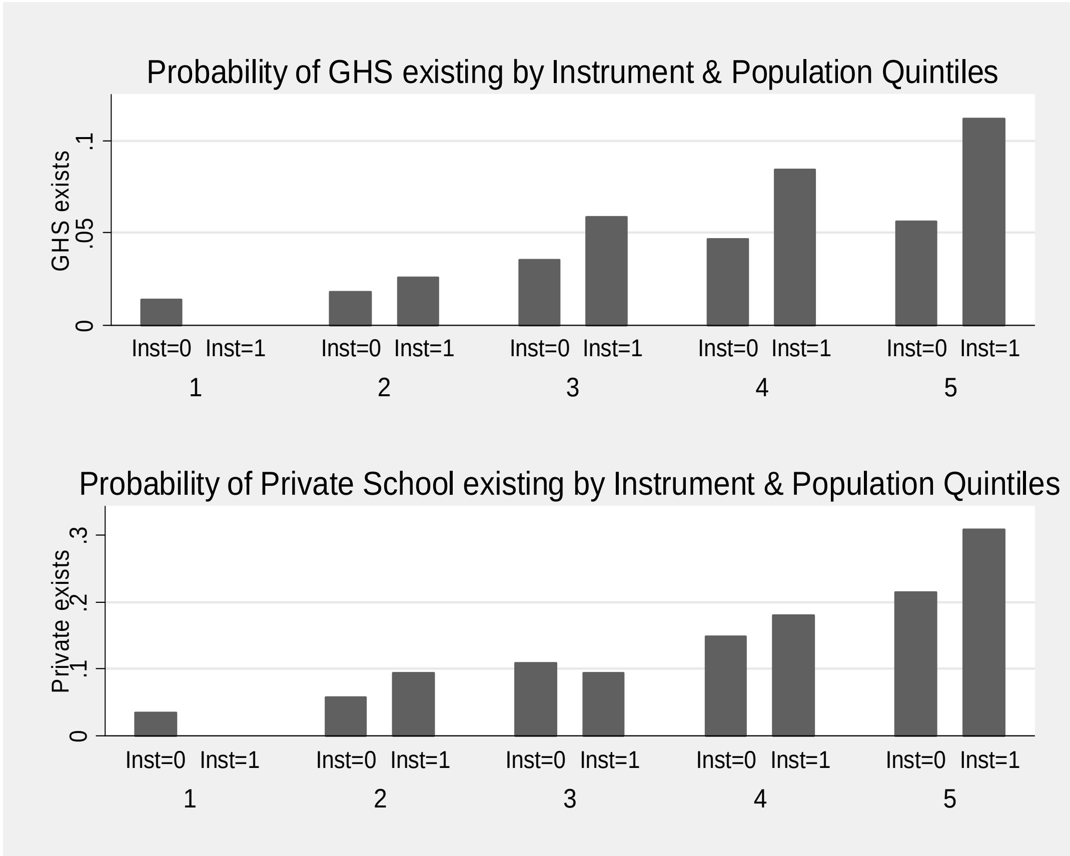


educated males and females in the village. Movements along the curve show the increase in probability with increases in the percentage of secondary-educated females; each different curve shows increases with increases in the percentage of secondary-educated men. The predictions are based on a probit regression. The data area based on the census of private schools (Federal Bureau of Statistics) matched to village-level census data for Punjab(Population Census Organization).



Note: The figure shows the probability of private school existence against village population (the line), the histogram of village populations, and the assignment of villages to eligible and non-eligible groups (which takes the value 0 or 1). All villages below 500 are ineligible, as are villages whose population is not the highest in their patwar-circle. Above villages of size 500, there is considerable variation in eligibility even at the same population. The data are based on the private school census and the population census.

Figure VI



Note: The top panel shows the percentage of villages with a girl's high school in villages that are assigned to the eligible and non-eligible groups; the panel below shows the percentage of villages with a private school in the same two groups. Villages are divided into 5 population quintiles to examine variation within similar population groups. The top-panel is equivalent to the first-stage of the IV regression, the bottom-panel represents the reduced form.

APPENDIX TABLE I

DIFFERENCES IN MEANS in 1998

	Treated	Not Treated	Difference
1998 Female/Male Ratio	0.941	0.946	-0.004
	(0.013)	(0.003)	(0.013)
1998 % Males w/ IC cards	0.710	0.693	0.017
	(0.025)	(0.006)	(0.026)
1998 % Females w/ IC cards	0.508	0.487	0.021
	(0.028)	(0.006)	(0.028)
1998 % of HHs with potable Water	0.064	0.064	0.001
	(0.014)	(0.003)	(0.014)
1998 % HHs w/ Perm. Housing	0.065	0.066	-0.001
	(0.014)	(0.003)	(0.014)
1998 % HHs w/ Electricity	0.083	0.084	0.003
	(0.016)	(0.004)	(0.016)
1998 % HHs w/ TV	0.033	0.029	0.004
	(0.010)	(0.002)	(0.010)
1998 % HHs w/ Radio	0.025	0.026	-0.001
	(0.009)	(0.002)	(0.009)
1998 % HHs w/ NewsPaper	0.025	0.026	-0.001
	(0.009)	(0.002)	(0.009)
1998 Village Land Area	2403.774	1609.910	793.8642***
	(158.671)	(28.484)	(132.106)
1998 per capita Land Area	0.957	1.812	-0.855*
	(0.060)	(0.108)	(0.483)