# Educational Enrollment and Attainment in India:

# Household Wealth, Gender, Village, and State Effects

Deon Filmer

Lant Pritchett

September 22, 1998

*Abstract*: This paper uses the National Family Health Survey (NFHS) data collected in 1992-93 to estimate the determinants of child (aged 6 to 14) enrollment and educational attainment of a recent cohort (aged 15 to 19) in India. The analysis produces five major results. First, using an index of assets as a proxy for household wealth shows enormous gaps between the enrollment and attainment of children from rich and poor households. While 82 percent of the children from the richest 20 percent complete grade 8 only 20 percent of children from the poorest 40 percent of households do. Second, the wealth gaps vary widely across states of India. Third, gender differences exacerbate these differences, so while 80 percent of girls from households in the top 20 percent complete grade 8 only 9.5 percent of girls from the poorest 40 percent do so. Fourth, the physical presence or absence of school facilities in the rural villages explains only a very small part of enrollment differences. Fifth, there are huge gaps in the enrollment rates of observationally equivalent households across states, especially among the poor. For instance, enrollment rates are 44 percentage points higher in Kerala than for an observationally equivalently poor household in Bihar. We conclude with an examination of the state specific policies that could account for such differences.

# Educational Enrollment and Attainment in India: Household Wealth, Gender, Village, and State Effects[1]

*Introduction*

We used NFHS household data from 1992/93 from each state of India to establish five principal empirical findings.

- First, in India on average there are large gaps in enrollment rates between the rich and poor, which persist even after controlling for a range of child, household, village, and state effects.
- Second, the "wealth gap" in enrollment varies widely within India across states, from 9 percentage points in Kerala to 56 percentage points in Uttar Pradesh.
- Third, gender plays a large, but also variable, role across states.
- Fourth, the physical presence or absence of schools within a village has a relatively small effect on enrollment rates.
- Fifth, even after controlling for household and village characteristics there are large, state specific differences on enrollment rates, which are concentrated in differences in enrollments of the poor.

After presenting these findings, we speculate on what produces the large variations across Indian states. We present evidence that suggests the quality of basic schooling is a plausible candidate to explain a large part of the state specific effect, but that improving quality is not merely a matter of raising spending.

*I) The NFHS data: Education and Household Wealth*

The data available for this analysis present both a unique opportunity, as well as a challenge. The opportunity is the fact that the data collected by the National Family Health Surveys (NFHS) in 1992 and 1993 used nearly identical questionnaires for each state, with large samples designed to be representative at the state level[2]. Sample sizes for each state ranged from 1000 in the small northeastern states to almost 10,000 in Uttar Pradesh. Overall the survey covered over 88,000 households and about 500,000 individuals.

Although the primary goal of the NFHS was the collection of data on fertility and child health from a primary female respondent from each household, questions on educational history were asked about all household members. Most of the analysis presented here comes from the following three questions:

- Has (name) ever been to school?
- If attended school: What is the highest grade (name) completed?
- If attended school and is aged less than 15 years: Is (name) still in school?

Overall only 68 percent of those 6 to 10 and 66 percent of those aged 11 to 14 are reported as still in school. While this is substantially less than is suggested from official government enrollment data, this direct measure is likely to be reliable (the discrepancy is

[1] We would like to thank Zoubida Alloua, Gunnar Eskeland, Keith Hincliffe, Valerie Kozel, Peter Lanjouw, Marlaine Lockheed, Martin Ravallion, and Adrian Verspoor for comments.

[2] The NFHS surveys were modeled closely on the Demographic and Health Surveys (DHS) which have been carried out in nearly a hundred countries over the past two decades.

explored extensively in a recent book on basic education in India, World Bank, 1997)[3].

As is well known, educational enrollments and attainments vary widely across Indian states. The percentage of 6 to 10 year in school ranges from only 50 percent in Bihar to 96 percent in Kerala and the percentage of those 11 to 14 who are in school ranges around the national average of 66 percent from 54 percent in Bihar to 94 percent in Kerala and Mizoram. The percentage of adults 15 to 65 who have ever attended school ranges from 42 percent in Bihar to 93 percent in Mizoram, and has a national average of 55 percent. Average years of attainment of those who even attended to school, ranges much less than enrollments and is close to 8 years of schooling in all states (except Delhi)[4].

*B) Constructing a proxy for long-run household wealth*

The goal is to estimate the determinants of enrollment with characteristics at four levels: child, household, village, and state. A major challenge is the fact that a household's economic status is an important factor but the NFHS did not collect information on household consumption expenditures or household income. The NFHS did however inquire about household ownership of various assets and characteristics of the household's dwelling. We use these variables to create an index to proxy for household "wealth."

We avoid the problem of assigning the appropriate weights to each asset by using use the statistical procedure of principal components. Principal components is a

[3] The present analysis focuses on differences in enrollments across groups (wealth, gender) and therefore these differences in absolute levels are less relevant.

[4] For more on the comparison between these basic results from the NFHS to results from other sources see World Bank (1997).

technique for extracting a small number of variables that best represent the common information in a larger set of related variables by creating a series of linear combinations of the original variables. The first principal component is created by choosing the weights on each of the variables such that the linear combination captures the greatest amount of information common to all the variables. We simply assume that what causes the most common co-movement of the asset variables is a household's wealth. This methodological approach is defended in a separate paper (Filmer and Pritchett, 1990).

In forming the index we used twenty one of the NFHS asset questions, which can be grouped into three types. First, eight questions about household ownership of consumer durables (clock/watch, bicycle, radio, television, bicycle, sewing machine, refrigerator, car). Second, twelve questions about characteristics of the household's dwelling (three about toilet facilities, three about the source of drinking water, two about rooms in the dwelling, two about the building materials used, and one if the main source of lighting is electric and one if the main fuel used for cooking is biomass). Third, a single question about whether the household owned more than 6 hectares of land.

An asset index A is constructed for each household *j* based on the means and standard deviations of the asset variables and their associated "scoring factors" reported in Table 1 according to the formula:

$$A_j = f_1 \times (a_{j1} - a_1) / (s_1) + .... + f_N \times (a_{jN} - a_N) / (s_N)$$

where $f_1$ is the "scoring factor" for the first asset, $a_{j1}$ is the household's value for the first asset and $a_1$ and $s_1$ are the mean and standard deviation of the values of the first asset over all households. The mean value of the index is by construction, zero. The standard deviation is 2.3.

Since all the variables (except "number of rooms") take only the values of zero or one, the weights have an easy interpretation. A move from 0 to 1 changes the index by $f_i/s_i$. Therefore, a household that owns a clock has an asset index higher by .54, owning a car raises a household's asset index by 1.21 units, and using biomass for cooking lowers the index by .67.

Each household is assigned to an "economic status" group depending on whether their value of the index places them in either the bottom 40 percent, the middle 40 percent, or top 20 percent of households in India[5]. From now on, purely for expository convenience, we will refer to these are the poor, the middle and the rich, asking the reader to keep firmly in mind that this is not following the any of the usual definitions of "poverty" and that we are *not* proposing the asset index for use in poverty analysis[6].

The difference in the average index between the poor and the middle is around 2. An example of a combination of assets that would produce this difference is owning a radio (.54), having a kitchen as a separate room (.37), having electricity for lighting (.57), and having a dwelling not of all low quality materials (-.55). The richest 20 percent have a wealth index almost four units higher than the middle 40 and this additional difference is equivalent to owning a motor scooter (.91), a television (.83), having a flush toilet (.75), a house of all high quality materials (.73) and not using biomass as a cooking fuel (.67) (again, this is merely an example of one possible combination of assets).

[5] Cutoff points for these quantiles were based on a ranking of individuals, that is the bottom 40 percent refers to the households in which the bottom 40 percent of people live.

[6] Likewise, we are not claiming to measure current "living standards" by using an asset index.

Table 1: Scoring factors and summary statistics for variables entering the computation of the first principal component

| | All India | | | | Poorest 40 percent | Middle 40 percent | Richest 20 percent |
|---|---|---|---|---|---|---|---|
| | Scoring factors | Mean | Std. Dev | Scoring factor* (1/std. dev.) | Mean | Mean | Mean |
| 1=own clock/watch | 0.270 | 0.533 | 0.499 | 0.54 | 0.164 | 0.739 | 0.985 |
| 1=own bicycle | 0.130 | 0.423 | 0.494 | 0.26 | 0.264 | 0.510 | 0.621 |
| 1=own radio | 0.248 | 0.396 | 0.489 | 0.51 | 0.101 | 0.522 | 0.838 |
| 1=own television | 0.339 | 0.209 | 0.407 | 0.83 | 0.000 | 0.127 | 0.866 |
| 1=own sewing machine | 0.253 | 0.182 | 0.385 | 0.66 | 0.015 | 0.179 | 0.580 |
| 1=own motorcycle/scooter | 0.249 | 0.082 | 0.274 | 0.91 | 0.001 | 0.031 | 0.375 |
| 1=own refrigerator | 0.261 | 0.068 | 0.252 | 1.04 | 0.000 | 0.006 | 0.353 |
| 1=own car | 0.129 | 0.012 | 0.107 | 1.21 | 0.000 | 0.001 | 0.059 |
| 1=drink.water from pump/well | -0.192 | 0.609 | 0.488 | -0.39 | 0.800 | 0.569 | 0.242 |
| 1=drink.water from open source | -0.041 | 0.040 | 0.195 | -0.21 | 0.057 | 0.036 | 0.005 |
| 1=drink.watr frm non-piped srce | -0.002 | 0.019 | 0.138 | -0.01 | 0.016 | 0.027 | 0.012 |
| 1=flush toilet | 0.308 | 0.217 | 0.412 | 0.75 | 0.005 | 0.175 | 0.797 |
| 1=pit toilet/latrine | 0.040 | 0.086 | 0.280 | 0.14 | 0.040 | 0.127 | 0.111 |
| 1=none/other toilet | 0.001 | 0.001 | 0.029 | 0.03 | 0.001 | 0.001 | 0.001 |
| 1=main source of lighting elect | 0.284 | 0.510 | 0.500 | 0.57 | 0.143 | 0.700 | 0.989 |
| Number of rooms in dwelling | 0.159 | 2.676 | 1.957 | 0.08 | 1.975 | 2.965 | 3.739 |
| 1=kitchen is a separate room | 0.183 | 0.536 | 0.499 | 0.37 | 0.312 | 0.643 | 0.848 |
| 1=cookg fuel is wood/dung/coal | -0.281 | 0.776 | 0.417 | -0.67 | 0.956 | 0.841 | 0.224 |
| 1=dwelling all high quality mat. | 0.309 | 0.237 | 0.425 | 0.73 | 0.005 | 0.218 | 0.821 |
| 1=dwelling all low quality mat. | -0.273 | 0.483 | 0.500 | -0.55 | 0.832 | 0.308 | 0.017 |
| 1=own >6 acres land | 0.031 | 0.115 | 0.319 | 0.10 | 0.075 | 0.155 | 0.126 |
| Economic status index | | 0.000 | 2.32 | | -2.00 | 0.071 | 3.857 |

Note: Scoring factor is the "weight' assigned to each variable (divided by its standard deviation) in the linear combination of the variables that constitute the first principal component.
Source: NFHS 1992/93

The first principal component works reasonably well as it explains 25.6 percent of the variation in the twenty-one asset variables[7]. The last three columns of table 1 compare the average asset ownership across the poor, middle and rich households. The asset index produces sharp differences in nearly every asset: clock ownership is 16 percent for the poor versus 98 percent for the rich, while the poor use biomass (wood/dung/coal) almost exclusively (96 percent) the rich only do so 20 percent of the

[7] Filmer, Pritchett, and Tan (1998) report the variation explained for 27 countries and the average is around 25 percent.

time. An important question is whether the asset index loads excessively on variables that are dependent on infrastructure (electricity, piped water) rather than household specific variables. On this score the clean separation on items not related to infrastructure like quality of materials in the household (only .5 percent of the poor versus 82.1 percent of the rich) and having a kitchen as a separate room (31 percent of the poor versus 85 percent of the rich) is reassuring.

The levels of the index that define the groups are calculated on an all India basis so states differ in the number of households in each group. Table 2 presents, for each state, the distribution of individuals across the different economic status groups. For instance, a richer state like Punjab only has 8.4 percent of its households in the bottom 40 percent and 44 percent in the top 20 percent while poorer states like Uttar Pradesh have 48 percent in the bottom 40 percent and only 15 percent in the top 20.

The last two columns of Table 2 compare these numbers with the most recent state rankings by poverty or State Domestic Product (SDP) per capita. Nationwide the poverty rate was 36 percent and hence is roughly comparable to the bottom forty percent cutoff which we use. The classifications agree that Punjab, Haryana, and Kerala have better than average economic status and that Bihar, Orissa and Uttar Pradesh have worse than average status. The rank correlation of the poverty rate and the fraction in the bottom 40 percent is .794 (p-value<.001). That said, there are differences, like Maharashtra, which looks richer (27 percent in bottom 40 versus 37 percent poverty rate) and Andrah Pradesh which looks poorer (39 in bottom 40, but poverty rate of only 22 percent).

Table 2: Distribution of individuals across groups and state level poverty and net domestic product (states sorted from smallest to largest proportion in the all-India bottom 40 percent).

| | Proportion of people in each group based with groups derived from economic status index | | | State poverty rate (headcount index) | Per capita net state domestic product |
|---|---|---|---|---|---|
| | Bottom 40 pct. | Middle 40 pct | Top 20 pct. | | |
| Delhi | 1.3 | 21.9 | 76.8 | | |
| Goa | 5.6 | 45.5 | 48.9 | | 10128 |
| Himachal Pradesh | 6.8 | 71.3 | 21.9 | 28.58 | |
| Punjab | 8.4 | 47.4 | 44.3 | 11.46 | 10857 |
| Haryana | 10.5 | 58.1 | 31.5 | 25.22 | 9609 |
| Jammu | 14.5 | 55.1 | 30.4 | | |
| Kerala | 15.1 | 63.9 | 21.1 | 25.12 | 5065 |
| Mizoram | 18.1 | 61.2 | 20.8 | | |
| Nagaland | 20.3 | 65.4 | 14.3 | | |
| Gujarat | 26.8 | 43.2 | 30.0 | 24.15 | 7586 |
| Maharashtra | 26.9 | 41.2 | 31.9 | 36.82 | 9270 |
| Karnataka | 27.6 | 52.1 | 20.3 | 32.91 | 6313 |
| Manipur | 27.6 | 54.0 | 18.4 | | |
| Tamil Nadu | 32.5 | 45.7 | 21.8 | 35.40 | 6205 |
| Meghalaya | 37.9 | 49.1 | 13.0 | | 5769 |
| Arunachal Pradesh | 38.1 | 51.4 | 10.5 | | 6359 |
| Andrah Pradesh | 39.0 | 40.6 | 20.4 | 21.87 | 5802 |
| Rajasthan | 39.7 | 42.7 | 17.6 | 27.46 | 5035 |
| Tripura | 41.8 | 50.1 | 8.0 | | |
| West Bengal | 44.3 | 38.6 | 17.2 | 36.94 | 5901 |
| Uttar Pradesh | 48.6 | 36.3 | 15.1 | 41.55 | 4280 |
| Madhya Pradesh | 49.4 | 34.3 | 16.3 | 42.46 | 4725 |
| Orissa | 54.4 | 36.5 | 9.1 | 48.64 | 3963 |
| Assam | 58.3 | 32.1 | 9.6 | 41.09 | 5056 |
| Bihar | 61.5 | 27.5 | 11.0 | 55.15 | 3280 |
| All India | 40.0 | 40.0 | 20.0 | 36.16 | 6380 |

Notes: The rank correlation coefficient between the percent in the bottom 40 percent and the poverty rate is 0.794 (p-value <.001), the rank correlation between the percent in the bottom 40 percent and per capita state product is -0.864 (p-value <.001). Sources: NFHS, 1992/93 and Haque, Lanjouw and Ravallion, 1998, and Agrawal and Varma, 1996. Data on the Headcount Index are for 1993/94.

There are similarly high correlations of the ranking by the percent in the bottom 40 percent group and Net State Domestic Product per capita, where the rank correlation is -.864 (p-value<.001)[8]. Again certain states look different by the two rankings. For example, Kerala looks richer by the index (15 percent in bottom 40 percent for a per capita SDP of 5,065) while Assam looks poorer (58 percent in the bottom 40 percent for a

per capita SDP of 5,056). However, the poverty rate is substantially higher in Assam than in Kerala (41 versus 25 percent) and the share in the bottom 40 percent is consistent with this ranking which is perhaps reassuring.

While the first principal component of assets might well serve as a reasonable overall index, there are two concerns. First, the second principal component appears to be capturing an additional dimension of the data, rich rural households without access to modern infrastructure. Second, the urban-rural differences using the asset index produce a much greater disadvantage for rural areas than do poverty rates. Although this is of some concern for urban/rural comparisons, it will not greatly affect out results much as we include dummy variables for rural/urban status or carry out the analysis for rural households only.

*C) Descriptive statistics.*

Figure 1 shows the attainment profiles for those aged 15 to 19 in each state for the three economic groups: each panel shows the proportion who have completed each grade among children who live in the poorest 40 percent, the middle 40 percent, and the wealthiest 20 percent of households[9]. Table 3 shows the fraction of children enrolled by wealth group. Also, Table 3 shows the probability a child aged 15 to 19 completed grade 8 classified by the household's asset index.

---

[8] The rank correlation between the poverty rate and per-capita state domestic product is -.7286 (p-value =.002).

[9] Because the economic groups are based on the all India sample, there are sometimes very few observations from which to derive the numbers displayed here. When the number of observations for any subgroup drops below 40 the attainment profile is not shown. For example in Andrah Pradesh there are only thirty-two urban males in the lowest economic group and therefore the attainment profile is not shown for that group.

Both Figure 1 and Table 3 show clearly how well the children from richer households do in all states. On average 94 percent of children aged 6 to 14 from the upper 20 percent are in school. This high enrollment rate of the rich is remarkably consistent across states, above 90 percent in all but three states (Arunachal Pradesh, Assam, and Tripura). Over 70 percent of 15 to 19 year olds from the richest economic group have completed grade 8 in all but two states (Meghalaya, Arunachal Pradesh). Nearly all children from rich households begin school, finish at least primary school and the vast majority finish through basic education of grade 8.

In sharp contrast, among the poorer part of the population educational attainment is dismal. Only half of the poor children aged 6 to 14 are in school. Only two out of five children aged 15 to 19 from poor households finished grade 5. Only one in five finished the eight years of basic education.

The gap in educational enrollment and attainment between the rich and poor is enormous on average in India, but varies a great deal across states. The gap in enrollment varies from a minor 0.08 in Kerala to a substantial 0.56 in Bihar. The gap in the attainment of grade 8 varies from (a non-negligible) 0.39 in Kerala to 0.72 in Orissa.

**Figure 1: Attainment profiles for ages 15 to 19, by economic group**

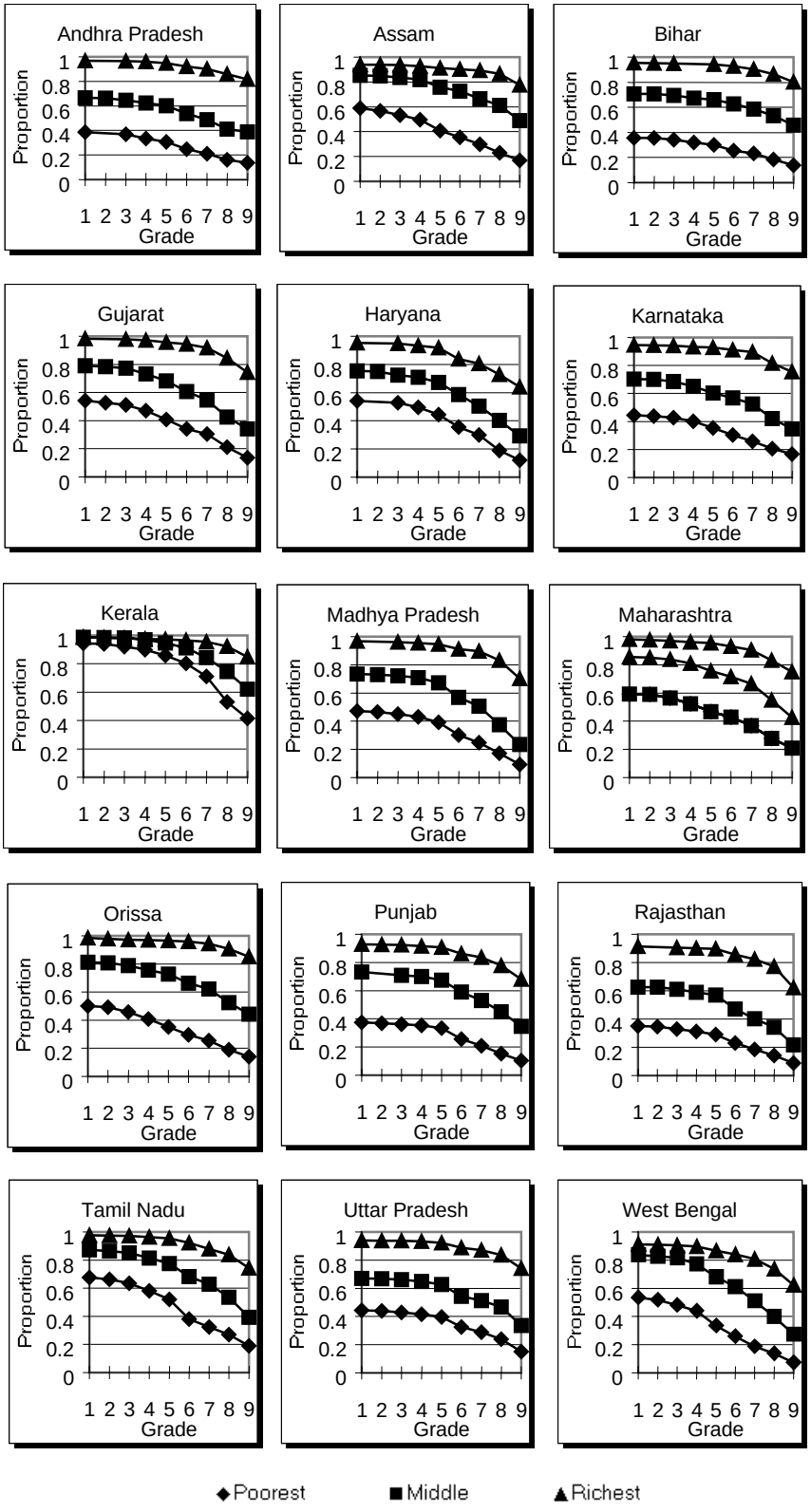


Table 3: Basic statistics on education status by wealth group

| State | Proportion of 6 to 14 year olds who are currently "in school" | | | | Proportion of 15 to 19 year olds who have completed at least grade 8 | | | |
|---|---|---|---|---|---|---|---|---|
| | Average | Bottom 40 percent | Top 20 percent | Wealth gap (top - bottom) | All | Bottom 40 percent | Top 20 percent | Wealth gap (top-bottom) |
| Kerala | 0.949 | 0.887 | 0.975 | 0.088 | 0.749 | 0.531 | 0.923 | 0.392 |
| Goa | 0.937 | 0.774 | 0.973 | 0.200 | 0.703 | 0.344 | 0.848 | 0.504 |
| Himachal Pradesh | 0.908 | 0.724 | 0.970 | 0.246 | 0.565 | 0.233 | 0.818 | 0.585 |
| Mizoram | 0.907 | 0.768 | 0.974 | 0.205 | 0.567 | 0.190 | 0.844 | 0.654 |
| Manipur | 0.902 | 0.804 | 0.991 | 0.186 | 0.610 | 0.359 | 0.927 | 0.568 |
| Nagaland | 0.896 | 0.824 | 0.980 | 0.157 | 0.572 | 0.354 | 0.865 | 0.511 |
| Delhi | 0.872 | 0.477 | 0.924 | 0.448 | 0.685 | . | 0.766 | . |
| Jammu | 0.857 | 0.666 | 0.979 | 0.313 | 0.541 | 0.195 | 0.833 | 0.638 |
| Tamil Nadu | 0.825 | 0.717 | 0.950 | 0.232 | 0.518 | 0.269 | 0.838 | 0.570 |
| Maharashtra | 0.820 | 0.671 | 0.962 | 0.290 | 0.579 | 0.279 | 0.832 | 0.554 |
| Haryana | 0.813 | 0.605 | 0.957 | 0.352 | 0.480 | 0.189 | 0.728 | 0.539 |
| Punjab | 0.808 | 0.427 | 0.957 | 0.531 | 0.571 | 0.153 | 0.777 | 0.624 |
| Tripura | 0.795 | 0.710 | 0.873 | 0.163 | 0.395 | 0.187 | 0.789 | 0.603 |
| Gujarat | 0.757 | 0.552 | 0.962 | 0.410 | 0.504 | 0.212 | 0.845 | 0.633 |
| Meghalaya | 0.749 | 0.601 | 0.959 | 0.358 | 0.326 | 0.150 | 0.667 | 0.516 |
| Arunachal Pradesh | 0.711 | 0.585 | 0.865 | 0.279 | 0.340 | 0.184 | 0.585 | 0.400 |
| Karnataka | 0.708 | 0.507 | 0.943 | 0.437 | 0.447 | 0.205 | 0.816 | 0.611 |
| Assam | 0.703 | 0.615 | 0.846 | 0.231 | 0.422 | 0.229 | 0.866 | 0.637 |
| Orissa | 0.697 | 0.552 | 0.969 | 0.416 | 0.395 | 0.189 | 0.908 | 0.719 |
| West Bengal | 0.678 | 0.527 | 0.902 | 0.375 | 0.338 | 0.137 | 0.734 | 0.597 |
| Andrah Pradesh | 0.639 | 0.457 | 0.917 | 0.460 | 0.419 | 0.160 | 0.859 | 0.698 |
| Madhya Pradesh | 0.626 | 0.461 | 0.937 | 0.476 | 0.367 | 0.172 | 0.832 | 0.661 |
| Uttar Pradesh | 0.614 | 0.484 | 0.939 | 0.455 | 0.424 | 0.239 | 0.836 | 0.598 |
| Rajasthan | 0.593 | 0.414 | 0.91 | 0.496 | 0.345 | 0.141 | 0.773 | 0.632 |
| Bihar | 0.514 | 0.378 | 0.942 | 0.564 | 0.381 | 0.183 | 0.864 | 0.681 |
| All India | 0.677 | 0.500 | 0.942 | 0.442 | 0.447 | 0.204 | 0.824 | 0.620 |

Source: Calculated from NFHS data, 1992-93, Haque, Lanjouw and Ravallion, 1998

An implication of the small differences among the rich and huge differences in the enrollment rates of the poor is that that differences in attainment across states are largely driven by the extent to which states have been able to reach the bottom part of the economic distribution and bring them into the educational system. For instance, Tamil Nadu and Rajasthan are not that different in the percent of the households falling into the India-wide bottom 40 percent: 37 percent in Tamil Nadu and 43 percent in Rajasthan. Their average educational attainments are quite different however: only 52 percent

completed grade 5 in Rajasthan as compared to 74 percent in Tamil Nadu. What causes this large difference? In both states the attainment of grade 5 of the rich is high, 96 percent in Tamil Nadu versus 90 percent in Rajasthan. What differs is how likely the poor are to reach grade 5 while in Tamil Nadu 52 percent of the poor population reached grade 5 this was only true of 29 percent of the poor in Rajasthan, a gap between the two states of 23 percentage points.

Income differences not only affect the enrollment and attainment of children, but they exacerbate gender differences: the gender gap is much larger for the poor than the rich. Table 4 shows that, overall, the gender gap in current enrollment is 24 percentage points for the poorest group and is close to zero at 3 percentage points for the richest group. Over 93 percent of both males and females from the richest quintile are in school whereas less than 40 percent of girls from the poorest group are in school. The percentage that complete grade 8 is 31 percent for boys but only 9.5 percent for girls from poor households, a gender gap of 22 percentage points. In contrast among the top 20 percent of households 85 percent of boys and 80 percent of girls complete grade 8.

Again, the difference across states is striking with the gender gap in enrollments in the poorest group being close to zero in Meghalaya, Nagaland, and Kerala, and reaching as high as 32 percentage points in Uttar Pradesh and 44 percentage points in Rajasthan. By contrast, the gap is consistently close to zero in the richest group except for Arunachal Pradesh. The results for attainment are similar. In the poorest group the results vary from a female advantage in Kerala of 11 percentage points to a female deficit of 29 percentage points in Uttar Pradesh where only 8 percent of female children aged 15 to 19 have completed grade 8. In the richest group the gap is less than 10 percentage

points in all states except for Rajasthan where it is 12 percentage points and West Bengal where it is 19 percentage points.

Table 4: Gender gaps enrollment and attainment by economic group

| | Proportion of 6-14 year olds who are currently enrolled | | | | Proportion of 15-19 year olds who have completed grade 8 | | | |
|---|---|---|---|---|---|---|---|---|
| | Bottom 40 percent | | Top 20 percent | | Bottom 40 percent | | Top 20 percent | |
| State | Female | Gap (Male-Female) | Female | Gap (Male-Female) | Female | Gap (Male-Female) | Female | Gap (Male-Female) |
| Meghalaya | 0.605 | -0.007 | 0.954 | 0.010 | 0.133 | 0.034 | . | . |
| Nagaland | 0.822 | 0.003 | 0.966 | 0.025 | 0.293 | 0.146 | 0.896 | -0.078 |
| Kerala | 0.882 | 0.011 | 0.972 | 0.006 | 0.579 | -0.105 | 0.937 | -0.028 |
| Goa | 0.743 | 0.067 | 0.972 | 0.003 | 0.229 | 0.229 | 0.825 | 0.045 |
| Mizoram | 0.715 | 0.099 | 0.963 | 0.021 | 0.191 | -0.002 | 0.843 | 0.003 |
| Assam | 0.561 | 0.106 | 0.803 | 0.085 | 0.183 | 0.097 | 0.819 | 0.085 |
| West Bengal | 0.475 | 0.107 | 0.852 | 0.108 | 0.094 | 0.090 | 0.648 | 0.190 |
| Tripura | 0.655 | 0.107 | 0.895 | -0.041 | 0.165 | 0.047 | . | . |
| Manipur | 0.728 | 0.140 | 0.990 | 0.001 | 0.317 | 0.075 | 0.959 | -0.081 |
| Arunachal Prad. | 0.504 | 0.158 | 0.776 | 0.179 | 0.150 | 0.072 | . | . |
| Punjab | 0.339 | 0.161 | 0.955 | 0.005 | 0.089 | 0.104 | 0.773 | 0.009 |
| Tamil Nadu | 0.636 | 0.161 | 0.950 | 0.000 | 0.177 | 0.202 | 0.821 | 0.038 |
| Maharashtra | 0.581 | 0.177 | 0.959 | 0.005 | 0.138 | 0.283 | 0.799 | 0.071 |
| Madhya Prad. | 0.355 | 0.200 | 0.923 | 0.027 | 0.053 | 0.221 | 0.788 | 0.085 |
| Orissa | 0.448 | 0.204 | 0.948 | 0.037 | 0.095 | 0.201 | 0.881 | 0.052 |
| Karnataka | 0.393 | 0.212 | 0.932 | 0.023 | 0.098 | 0.234 | 0.797 | 0.039 |
| Andrah Pradesh | 0.347 | 0.219 | 0.890 | 0.052 | 0.072 | 0.182 | 0.820 | 0.086 |
| Delhi | 0.364 | 0.230 | 0.928 | -0.007 | . | . | 0.784 | -0.033 |
| Gujarat | 0.430 | 0.232 | 0.948 | 0.027 | 0.100 | 0.209 | 0.804 | 0.083 |
| Haryana | 0.467 | 0.254 | 0.945 | 0.023 | 0.056 | 0.247 | 0.715 | 0.026 |
| Jammu | 0.513 | 0.278 | 0.983 | -0.008 | 0.089 | 0.201 | 0.819 | 0.025 |
| Bihar | 0.228 | 0.291 | 0.939 | 0.005 | 0.065 | 0.253 | 0.851 | 0.024 |
| Himachal Prad. | 0.564 | 0.298 | 0.965 | 0.010 | 0.136 | . | 0.808 | 0.020 |
| Uttar Pradesh | 0.313 | 0.315 | 0.932 | 0.013 | 0.081 | 0.293 | 0.826 | 0.021 |
| Rajasthan | 0.173 | 0.437 | 0.861 | 0.092 | 0.017 | 0.229 | 0.710 | 0.116 |
| All India | 0.375 | 0.239 | 0.929 | 0.025 | 0.095 | 0.218 | 0.796 | 0.056 |

Note: Cells are empty if there are fewer than 40 individuals from which to calculate the quantity.
Source: Calculated from NFHS data, 1992-93

The main purpose of the classification of households into groups by economic status is to examine what fraction children in each group of households are in school. While we cannot compare the results of the classification of assets and expenditures directly using an Indian data source with both, we can make the following comparison.

Table 5: Enrollment rates by quintile, household per capita consumption and asset index.

| Quintile | Enrollment of rural children aged 6-14 when household quintiles are constructed by | |
|---|---|---|
| | Per capita consumption expenditures | Asset index |
| 1 (poorest) | 49 | 42 |
| 2 | 61 | 58 |
| 3 | 70 | 71 |
| 4 | 76 | 84 |
| 5 (richest) | 82 | 94 |
| Difference between 5 and 1 | 33 | 52 |

Source: Calculated from NFHS data, 1992-93 and Haque, Lanjouw and Ravallion, 1998

Table 5 shows the fraction of children aged 6 to 14 in rural areas of India by quintile when children are classified by our asset index from the NFHS data or by per capita consumption expenditures (not accounting for household composition or economies of scale) using consumption expenditures from the NSS. While the enrollment rates of children from the middle quintile agree almost exactly (70 versus 71 percent), the enrollment rate profile based on quintiles from household consumption expenditures from NSS data has a flatter profile (from 49 to 82) than the profile based on an asset index[10]. Those classified as poor by the assets index have an enrollment rate that is 7 percentage points lower (49 versus 42) than those classified as poor by expenditures while the asset index rich have one which is 12 percentage points higher (94 versus 82) than the expenditure rich.

[10] This tendency of the classification by consumption expenditures to produce a flatter profile is also found in Pakistan (Filmer and Pritchett, 1998).

*II) The determinants of enrollment: household, child, village, and state effects*

Armed with the educational outcome data on the one hand and the indicator of economic status on the other we now address how enrollments differ within states according to the economic status of the household and, controlling for it, how they are affected by gender, location, and the presence of schools.

*A) Empirical specification*

To disentangle the determinants of school enrollment, we now turn to a multivariate model. The model, estimated as a probit regression, is specified as

$$E_i^* = \sum_{j=2,5} \beta_j \times Q_{ij} + \alpha \times X_i + \varepsilon_i$$

where $E_i^*$ is an unobserved variable whose observed counterpart, whether or not a child aged 6 to 14 is currently in school, is defined as

$E_i = 1$ if $E_i^* >= 0$
$= 0$ otherwise.

Wealth effects are specified by including the $Q_{ij}$ s which are dummy variables equal to one if child *i* is in a household in quintile *j* (the reference quintile is the poorest quintile).

In all of the samples the variables included besides wealth are the *child* variables of a dummy variable for gender, child's age and age squared, the *household* variables of age of the head of the household, whether the household head ever attended school, the highest grade completed of the household head, whether the household was Hindu, whether the household is from a scheduled caste or tribe[11].

The other variables included depend on whether the sample pools data from urban and rural areas or is limited to rural areas only.

As data on school availability and other village characteristics is limited to rural areas, the rural only sample the variables include three dummy variables for the presence of (1) a primary school (2) a primary and a "middle" school and (3) a primary, "middle" and a secondary school. In addition a large set of other village level variables capturing village infrastructure is included[12]. In the pooled samples a dummy variable is included for urban/rural location of the household.

The regressions are estimated separately for each state and then for India as a whole. The all-India regressions include dummy variables for each state (with Bihar as the reference state). Instead of presenting the complete set of equations for each of the 25 states plus all-India we have divided the results up into sections that report the results each of the levels: household, child, and village. Each set of results are however taken from the full regression specifications.

*B) Household effects: The impact of wealth*

Table 6 presents the marginal effects of quintiles of household wealth on the probability a child aged 6 to 14 is in school. The results show that there is a strong wealth effect in the probability of enrollment. All else equal, a child from a household in the highest quintile is about 31 percentage points more likely to be in school than a child from the poorest quintile. Moreover, the effects are starkly ordered across the quintiles: being in the second quintile increases the probability of being in school by 10 percentage points and each subsequent quintile increases the probability by roughly 7 percentage points (10.3 to 16.9 to 24.1 to 30.7).

[11] Also included as a set of variables so as to not lose observations when data about the head of household was missing.

Table 6: Marginal effects of wealth on the probability of being in school for ages 6 to 14, urban and rural (Probit regression results for selected variables). Regions sorted by the "quintile 5" coefficient in the rural sample.

| | Pooled urban and rural samples | | | | Rural sample only | | | |
|---|---|---|---|---|---|---|---|---|
| | Quintile 2 | Quintile 3 | Quintile 4 | Quintile 5 | Quintile 2 | Quintile 3 | Quintile 4 | Quintile 5 |
| Mizoram | 0.030 | 0.073 | 0.112 | 0.083 | -0.012 i | 0.026 i | 0.018 i | -0.096 i |
| Himachal Pr. | -0.035 i | 0.031 i | 0.045 i | 0.062 i | -0.086 i | 0.005 i | 0.013 i | 0.026 i |
| Kerala | 0.017 i | 0.038 | 0.059 | 0.046 | 0.014 i | 0.037 | 0.058 | 0.042 |
| Goa | 0.019 i | 0.042 | 0.064 | 0.098 | 0.024 i | 0.038 | 0.063 | 0.054 |
| Nagaland | -0.004 i | 0.027 i | 0.017 i | 0.064 | 0.001 i | 0.037 i | 0.007 i | 0.065 i |
| Manipur | 0.032 i | 0.055 | 0.085 | 0.073 | 0.037 | 0.049 | 0.095 | 0.095 |
| Jammu | 0.039 i | 0.079 | 0.146 | 0.160 | 0.028 i | 0.066 | 0.118 | 0.119 |
| Tamil Nadu | 0.006 i | 0.061 | 0.106 | 0.143 | -0.001 i | 0.078 | 0.119 | 0.142 |
| Tripura | 0.080 | 0.115 | 0.138 | 0.079 i | 0.066 | 0.136 | 0.137 | 0.155 |
| Delhi | 0.055 i | 0.072 i | 0.115 | 0.446 | | | 0.087 | 0.160 |
| Maharashtra | 0.048 | 0.084 | 0.124 | 0.199 | 0.049 | 0.093 | 0.163 | 0.164 |
| Assam | 0.131 | 0.202 | 0.212 | 0.133 | 0.139 | 0.212 | 0.187 | 0.172 |
| Haryana | 0.072 | 0.093 | 0.186 | 0.234 | 0.084 i | 0.107 | 0.229 | 0.196 |
| Arunachal Pr. | 0.137 | 0.215 | 0.239 | 0.242 | 0.121 | 0.217 | 0.226 | 0.212 |
| Orissa | 0.082 | 0.206 | 0.231 | 0.263 | 0.095 | 0.229 | 0.250 | 0.251 |
| Meghalaya | 0.011 i | 0.081 | 0.188 | 0.197 | 0.011 i | 0.083 i | 0.209 | 0.257 |
| Gujarat | 0.057 | 0.106 | 0.179 | 0.294 | 0.066 | 0.145 | 0.210 | 0.273 |
| West Bengal | 0.152 | 0.242 | 0.290 | 0.271 | 0.124 | 0.226 | 0.287 | 0.284 |
| Punjab | 0.035 i | 0.104 | 0.207 | 0.336 | 0.022 i | 0.110 | 0.246 | 0.286 |
| Karnataka | 0.088 | 0.185 | 0.253 | 0.296 | 0.074 | 0.191 | 0.267 | 0.303 |
| Madhya Prad. | 0.121 | 0.198 | 0.268 | 0.348 | 0.135 | 0.220 | 0.297 | 0.371 |
| Uttar Prad. | 0.135 | 0.188 | 0.271 | 0.382 | 0.152 | 0.196 | 0.282 | 0.372 |
| Andrah Prad. | 0.077 | 0.151 | 0.261 | 0.322 | 0.083 | 0.126 | 0.270 | 0.387 |
| Rajasthan | 0.082 | 0.158 | 0.296 | 0.388 | 0.065 | 0.180 | 0.339 | 0.406 |
| Bihar | 0.150 | 0.248 | 0.400 | 0.426 | 0.167 | 0.255 | 0.425 | 0.526 |
| All India | 0.103 | 0.169 | 0.241 | 0.307 | 0.111 | 0.185 | 0.269 | 0.315 |

Source: NFHS 1992-93

Notes: All underlying probit coefficients for displayed variables are significant except those indicated by "i". Marginal effects are evaluated at the means of the other variables. In addition to the displayed variables, the probit regression includes age, age squared; gender, age, and schooling of the head of the household; a dummy for Hindu. The regression for the rural sample includes dummy variables for village infrastructure (for example for the presence of a paved road, a PHC clinic, a post office, a marketshop). All India regression includes dummy variables for state (see below).

The results for the rural only sample are very similar even though a host of additional village level factors are included in the model are included. In particular, the rural sample includes information on school availability and village infrastructure so these wealth effects represent the effects of household wealth controlling for the fact that

[12] For example whether or not there is a post-office, a regular market, a health center.

the poor are more likely to live in less well developed villages. Even with these additional controls the magnitude of the wealth effects are nearly identical in the all India sample (11.1, 18.5, 26.9, and 31.5 percentage points respectively).

While the effects are large on average, there is a large amount of variation in the magnitude of the wealth effects across states. For example in rural areas, a child from the highest quintile in Kerala is about 4 percentage points more likely than one from the poorest quintile to be in school. In Himachal Pradesh the difference is estimated to be only 2.6 percentage points. In Bihar the difference is 53 percentage points[13]. Focusing on rural areas only exacerbates the differential with the Kerala-Bihar wealth gaps now being 4 to 53 percentage points respectively.

The results on household wealth here are consistent with those from other studies. For example, NCAER (1994) found that the difference in the percent of children aged 6 to 14 years old who had ever attended school between children from households with per capita incomes of less than Rs3,000 and children from households with per capita incomes of more than Rs10,000 was 25 percentage points, on average taken over 14 major states. The range in the difference was smallest in Kerala where there was no difference found, and largest in Punjab where it was 55 percentage points. When focusing on only rural areas, they find an average gap of 17 percentage points whereas the difference based on our results is remains about 30 percentage points.

Haque, Lanjouw and Ravallion (1998) find similar differences across the quintiles in the raw enrollment rates (see Table 4). Moreover, their estimated marginal effect of a

[13] Recall that the quintiles are based on the all India sample so that the highest quintile in each state refers to the same level of wealth.

1 percent change in household per capita consumption expenditures on enrollment of 0.178, applied to the percentage difference in average consumption between the highest and lowest per capita quintiles in India, is extremely close to what we get if we use the estimates in Table 6[14].

Behrman and Knowles (1998) summarize estimates on the income elasticity of educational attainment from many different countries. The elasticity for the poorer countries is consistent with an elasticity estimate of close to 0.18.

Table 7: Estimates of the elasticity of schooling outcomes with respect to incomes

| Country | Year | Outcome measure | Elasticity |
|---|---|---|---|
| Ghana | 1987/9 | school attainment | 0.18-0.56* |
| Nepal | 1980/1 | grade attainment | 0.38* |
| Bangladesh | 1980/1 | attendance | 0.20 |
| Pakistan | 1989 | numeracy and literacy | 0.05-0.23* |
| Cote d'Ivoire | 1985/7 | school attainment | 0.14-0.42* |
| Bolivia | 1989 | grade attained | 0.04* |
| Nicaragua | 1977/8 | grades completed | 0.02-0.07 |
| Brazil | 1970 | completed years | 0.09-0.16* |
| Brazil | 1982 | completed years | 0.06-0.22* |
| Venezuela | 1987 | years | 0.01* |
| Taiwan | 1989 | years of schooling | 0.12-0.33* |

Source: Adapted from Behrman and Knowles (1997)
Notes: * indicates that the underlying estimate was significant at the 10 percent level. Country/years are sorted by purchasing power parity adjusted per capita GDP.

*C) Child effects: The role of gender*

The effect of gender on schooling decisions, and its variability across states, in India is widely recognized (for recent examples see Murthi, Guio, Jean Drèze, 1995, Filmer, King, and Pritchett, 1997). As Table 8 shows there are large differences in the magnitude of the gender gap across states, effects which persist even after controlling for household and village characteristics. The male advantage in enrollment is slight, less

[14] For example, if the difference in average per-capita consumption between the richest and poorest quintiles is 139 percent, then their marginal effect estimate of 0.178 implies a 25 (139*0.178)

than 5 percent, in Kerala, Himachal Pradesh, Goa and the Northeastern states. Then there are a set of states in which the male advantage is substantial and always statistically significant, from Assam at 7.3 percentage points to Maharashtra at 13.5 percentage points. Then there is a jump and there are nine states where the male advantage exceeds 15 percentage points. In these states boys are from 18.6 percentage points more likely in Orissa to an dismaying 45.8 percentage points more likely to be enrolled in Rajasthan. As these states include several large states, such as Uttar Pradesh (34.5 percentage points), the all-India average gender gap is 23.7 percentage points.

The effects discussed here are estimated from a model that does not include interaction effects between gender and wealth. Table 8 suggests, however, that the effects may be even more severe for the poor. Even in this additive model the combination of wealth and gender effects paints a depressing picture for girls from poor households, but one that varies widely across India. In rural areas of India on average a girl from a poor (bottom 20 percent) household is 55.2 (31.5+23.7) percentage points less likely to be in school than a boy from a rich (top 20 percent) household. But in Kerala she is only 4.5 percentage points less likely to be in school, while in Uttar Pradesh she is 71.7 percentage points less likely. In Bihar and Rajasthan the combination of gender and wealth gaps produces a gap between the most and least socially favored groups of a staggering 86.3 and 86.4 percentage points.

---

percentage point difference in the enrollment rates. This compares to our all India estimate of 31 percentage points difference between the poorest and richest quintiles of the wealth index.

Table 8: Marginal effects on the probability of being in school for ages 6 to 14, rural only (Probit regression results for selected variables)

| | Male | | SCST | | Primary school | | Primary and "middle" school only | | Primary, middle, and secondary schools | |
|---|---|---|---|---|---|---|---|---|---|---|
| Meghalaya | -0.010 | i | 0.130 | i | -0.073 | i | -0.046 | i | -0.048 | i |
| Kerala | 0.003 | i | -0.036 | | | | | | 0.068 | i |
| Goa | 0.008 | i | -0.014 | i | | | -0.038 | | -0.012 | i |
| Nagaland | 0.018 | i | -0.042 | i | 0.037 | i | 0.021 | i | 0.011 | i |
| Mizoram | 0.030 | | | | | | -0.008 | i | 0.129 | i |
| Manipur | 0.043 | | -0.010 | i | -0.017 | i | -0.095 | i | -0.024 | i |
| Tripura | 0.046 | | -0.055 | i | | | 0.026 | i | -0.074 | i |
| Himachal Pradesh | 0.050 | | -0.045 | | -0.011 | i | 0.027 | i | -0.041 | i |
| Delhi | 0.058 | i | -0.145 | | | | -0.105 | i | -0.015 | i |
| Assam | 0.073 | | -0.013 | i | 0.002 | i | 0.044 | i | -0.084 | i |
| Punjab | 0.074 | | 0.004 | i | | | -0.051 | i | -0.081 | |
| Tamil Nadu | 0.104 | | 0.024 | i | | | 0.059 | | 0.046 | i |
| West Bengal | 0.104 | | -0.082 | | -0.002 | i | 0.100 | | 0.026 | i |
| Jammu | 0.126 | | -0.049 | | -0.055 | | -0.036 | i | -0.039 | i |
| Arunachal Pradesh | 0.133 | | 0.142 | | -0.114 | i | -0.121 | i | -0.108 | i |
| Maharashtra | 0.135 | | -0.098 | | | | 0.025 | i | 0.072 | i |
| Orissa | 0.186 | | -0.139 | | -0.079 | i | -0.019 | i | 0.041 | |
| Karnataka | 0.189 | | 0.008 | i | | | 0.044 | i | 0.160 | |
| Haryana | 0.192 | | -0.012 | i | | | 0.041 | i | 0.036 | i |
| Gujarat | 0.197 | | 0.029 | i | | | 0.047 | i | 0.109 | i |
| Madhya Pradesh | 0.225 | | -0.070 | | | | 0.042 | i | 0.145 | |
| Andrah Pradesh | 0.233 | | -0.026 | i | | | 0.062 | i | 0.095 | i |
| Bihar | 0.337 | | -0.066 | | 0.132 | | 0.223 | | 0.103 | i |
| Uttar Pradesh | 0.345 | | -0.101 | | 0.025 | i | 0.044 | i | 0.125 | |
| Rajasthan | 0.458 | | -0.051 | i | 0.039 | i | 0.148 | i | 0.094 | i |
| All India | 0.237 | | -0.053 | | 0.037 | | 0.073 | | 0.083 | |

Source: NFHS 1992-93

Notes: The displayed coefficients correspond to a change in the dummy variable from zero to one, evaluated at the means of all the other variables. All underlying probit coefficients for displayed variables are significant except those indicated by "i". In addition to the displayed variables, the probit regression includes age, age squared; gender, age, and schooling of the head of the household; a dummy for Hindu, dummy variables for village infrastructure (for example for the presence of a paved road, a PHC clinic, a post office, a marketshop). All India regression includes dummy variables for state (see below).

*D) Village effects: The impact of the presence or absence of school facilities*

The "availability gap" in rural areas, that is the difference in the proportion of 6 to 14 year olds who are in school between those villages with no school and those villages that have a primary, middle, and secondary school is only 8.3 percentage points[15]. This availability gap can be contrasted to the effect of being in the second versus the poorest quintile which is 11.1 percentage points, and in being from the richest versus the poorest quintile which is 31.5 percentage points.

Moreover, while the wealth terms reported in Table 6 were statistically significant for almost every state, Table 8 shows that the effect of even the largest difference in school availability (that is comparing a village that has all schools to one which has no schools) was positive and statistically significant in only 4 of the 24 states (Karnataka, Madhya Pradesh, Orissa and Uttar Pradesh). This lack of statistical significance is not simply the result of insufficient precision of estimation, as the impact was estimated to be negative in 10 of the 25 states.

These results suggest that the physical presence or absence of schools in a village is not an overwhelming factor in determining enrollment rates. Moreover, since only about 17 percent of villages in the NFHS sample were reported as lacking a school, and since the incremental enrollment was estimated at 4 percent when a primary school was available, the scope for increasing the enrollment rates of the poor by expansion in the number of villages with schools is limited.

[15] We are looking here at the relationship between school presence and outcomes setting aside the issue of the possibility of the non-random placement of schools.

However, this result should not be over-interpreted to mean that school availability does not matter, as there are many ways in which the presence of a school does not mean school availability. First, a school that is reported as "present" may be dysfunctional. In their review of conditions of schools in Uttar Pradesh, Drèze et al (1996) found that in many cases teachers are not present, or even that the school facility has been converted to other uses, such as a cattle shed or storage. Second, even though a school may be present in the village it could still be at a considerable distance from some children. We have no way of testing the impact of distance within a village even though it is often empirically found to be important in other countries. Third, even though a school may be present it may not be socially *available* to all students as there may be social exclusion. That is, parents in certain parts of India may not feel comfortable sending their female children to school if the school lacks appropriate facilities, or if it is unsafe for girls, or if there are not female teachers present. Moreover, social conditions of caste or income may play a role in the availability of schools to individual children, in the sense they are made to feel (directly or indirectly) unwelcome in the school.

*E) State Effects: The "unknown"*

We have seen, so far, the effects on enrollment of characteristics of the child (gender), of the household (wealth) and of the village (school facilities), and now we turn to effects at the level of the state. We first show that even after controlling for all the child, household, and village effects there are still large differences in enrollment rates. Moreover, these state specific effects are concentrated among the poor, for whom nearly all of the differences in enrollment probabilities are state specific

Table 9: Marginal effects on the probability of being in school for ages 6 to 14, (Probit regression results)

| | | All India | | All India Poorest 40 percent only | | Rural | |
|---|---|---|---|---|---|---|---|
| | Zero / one variable | Marginal Effect | T-ratio | Marginal Effect | T-ratio | Marginal Effect | T-ratio |
| Quintile 2[a] | * | 0.103 | 12.32 | 0.129 | 12.39 | 0.111 | 9.87 |
| Quintile 3 | * | 0.169 | 16.94 | | | 0.185 | 17.92 |
| Quintile 4 | * | 0.241 | 22.55 | | | 0.269 | 20.77 |
| Quintile 5 | * | 0.307 | 23.53 | | | 0.315 | 18.69 |
| Male | * | | | | | 0.237 | 8.42 |
| Rural male[b] | * | 0.070 | 3.85 | 0.048 | 1.36 | | |
| Urban Female | * | -0.107 | -6.19 | -0.197 | -6.31 | | |
| Rural Female | * | -0.149 | -6.70 | -0.232 | -5.06 | | |
| Scheduled caste / Scheduled tribe | * | -0.047 | -3.87 | -0.060 | -4.47 | -0.053 | -4.37 |
| Age | | 0.206 | 13.37 | 0.241 | 15.25 | 0.232 | 13.20 |
| Age squared | | -0.011 | -16.89 | -0.013 | -19.22 | -0.012 | -16.47 |
| Head is male | * | -0.092 | -5.64 | -0.116 | -5.47 | -0.119 | -5.90 |
| Head's age | | 0.001 | 4.29 | 0.002 | 5.74 | 0.002 | 5.41 |
| Head ever attended school | * | 0.072 | 6.73 | 0.093 | 7.97 | 0.071 | 6.88 |
| Head's highest grade completed | | 0.019 | 16.27 | 0.028 | 13.16 | 0.023 | 19.31 |
| Head information missing | * | 0.094 | 4.42 | 0.121 | 6.01 | 0.112 | 4.75 |
| Hindu | * | 0.109 | 5.11 | 0.111 | 4.28 | 0.119 | 5.38 |
| Primary school in village | * | | | | | 0.037 | 2.10 |
| Primary and middle school in vill. | * | | | | | 0.073 | 3.05 |
| Prim., mid., and secondary in vill. | * | | | | | 0.083 | 6.43 |
| Nearest town within 5 km | * | | | | | 0.018 | 1.31 |
| Nearest railroad within 5 km | * | | | | | -0.001 | -0.11 |
| Nearest bus within 5 km | * | | | | | 0.014 | 1.71 |
| Paved road in village | * | | | | | 0.006 | 0.42 |
| Electricity in village | * | | | | | 0.019 | 1.10 |
| PHC clinic in village | * | | | | | -0.006 | -0.27 |
| Health subcenter in village | * | | | | | -0.011 | -1.09 |
| Hospital in village | * | | | | | -0.015 | -1.00 |
| Dispensary in village | * | | | | | 0.001 | 0.11 |
| Health guide in village | * | | | | | 0.001 | 0.05 |
| Bank in village | * | | | | | 0.009 | 0.92 |
| Co-op in village | * | | | | | 0.007 | 0.55 |
| Post-office in village | * | | | | | -0.009 | -0.60 |
| Market in village | * | | | | | -0.021 | -2.95 |
| Cinema house in village | * | | | | | 0.003 | 0.31 |
| Pharmacy in village | * | | | | | 0.016 | 1.15 |
| Mahila Mandal | * | | | | | -0.022 | -1.01 |
| Flood within the last two years | * | | | | | -0.003 | -0.22 |
| Drought in the last two years | * | | | | | -0.007 | -0.56 |

Notes: The marginal effect for a zero/one variable is the effect of a change in the variable from zero to one on the probability of a child being in school. The specification includes dummy variables for each state. T-ratios refer to the underlying probit coefficient. [a/] Reference group is quintile 1. [b/] Reference group is urban male.

In all of the above regressions the regressions were run both state by state and at the national level where binary indicators for each of the 25 states were included. Table 9 reports the results of the "all India" regressions for three sample specifications. The first set of results are all of India including both urban and rural areas (for example, the final row "All India" estimates of the wealth quintiles in Table 6 were derived from this regression). The next set of results are for the same urban and rural coverage but limited to those households in the bottom 40 percent. The last set of results are from the regression limited to rural areas, which allows the introduction of the variables for school availability and other village infrastructure. In each of these regression, the coefficients on the state dummy variables can be interpreted as the differences in enrollments between households that are observationally "equivalent" except for living in different states.

Table 10 presents both state by state raw averages and the estimated state effects. Bihar is chosen as the reference state so in all cases the average or effect in Bihar is by definition zero and therefore states that do better than Bihar have positive values and those that do worse have negative values (of which there are none for average enrollments). Column 1 shows how much higher each state's enrollment is than Bihar's, a gap which is as high as 43 percent age points for Kerala.

Column 4 shows how much of the raw differences is explained away by variations across states in household and village characteristics. The first thing to notice is that the large differences across states present in the raw averages persist. Some states do only slightly better than Bihar (Andrah Pradesh 6.3 percentage points higher, Uttar Pradesh 5.3 percentage points higher, Rajasthan 1.8 percentage points higher, Madhya Pradesh 5.6 percentage points higher) and some states doing much better than Bihar (Kerala 25.3

percentage points higher, Himachal Pradesh 20.5 percentage points higher, Gujarat 22 percentage points higher). The differences across state in enrollment rates are not due only to differences in average state characteristics such as wealth or the education of adults.

Table 10: Enrollment probabilities, adjusted and unadjusted

| | Enrollment rate of 6-14 year olds (minus by the rate in Bihar) | | | Effect of state dummy variable in "All India" probit regression of "in school" of 6 to 14 year olds (reference state is Bihar) | | |
|---|---|---|---|---|---|---|
| | Full sample | Poorest 40 percent | Poorest 40 percent rural areas only | Full sample | Poorest 40 percent | Poorest 40 percent rural areas only |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Kerala | 0.435 | 0.509 | 0.515 | 0.253 | 0.442 | 0.450 |
| Nagaland | 0.382 | 0.446 | 0.444 | 0.241 | 0.441 | 0.445 |
| Manipur | 0.388 | 0.426 | 0.418 | 0.233 | 0.396 | 0.394 |
| Gujarat | 0.423 | 0.396 | 0.182 | 0.221 | 0.188 | 0.200 |
| Mizoram | 0.393 | 0.390 | 0.390 | 0.237 | 0.394 | 0.387 |
| Himachal Pradesh | 0.394 | 0.346 | 0.353 | 0.205 | 0.311 | 0.323 |
| Tamil Nadu | 0.311 | 0.339 | 0.346 | 0.183 | 0.325 | 0.322 |
| Tripura | 0.281 | 0.332 | 0.334 | 0.178 | 0.317 | 0.298 |
| Maharashtra | 0.306 | 0.293 | 0.308 | 0.175 | 0.285 | 0.300 |
| Jammu | 0.343 | 0.288 | 0.293 | 0.197 | 0.312 | 0.325 |
| Assam | 0.189 | 0.237 | 0.246 | 0.157 | 0.257 | 0.275 |
| Haryana | 0.299 | 0.227 | 0.240 | 0.128 | 0.228 | 0.237 |
| Meghalaya | 0.235 | 0.223 | 0.229 | 0.198 | 0.330 | 0.339 |
| Arunachal Pradesh | 0.197 | 0.207 | 0.215 | 0.176 | 0.309 | 0.348 |
| Goa | 0.243 | 0.174 | 0.458 | 0.115 | 0.375 | 0.414 |
| Orissa | 0.183 | 0.174 | 0.181 | 0.131 | 0.180 | 0.180 |
| West Bengal | 0.164 | 0.149 | 0.157 | 0.120 | 0.167 | 0.188 |
| Karnataka | 0.194 | 0.129 | 0.129 | 0.092 | 0.132 | 0.136 |
| Uttar Pradesh | 0.100 | 0.106 | 0.119 | 0.053 | 0.094 | 0.113 |
| Delhi | 0.358 | 0.099 | | 0.058 | 0.027 | |
| Madhya Pradesh | 0.112 | 0.083 | 0.086 | 0.056 | 0.083 | 0.087 |
| Andrah Pradesh | 0.125 | 0.079 | 0.084 | 0.063 | 0.123 | 0.114 |
| Punjab | 0.294 | 0.049 | 0.060 | 0.151 | 0.136 | 0.138 |
| Rajasthan | 0.079 | 0.036 | 0.036 | 0.018 | 0.065 | 0.076 |
| Bihar | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Unweighted average | 0.257 | 0.229 | 0.243 | 0.146 | 0.237 | 0.254 |
| Unweighted standard deviation | 0.120 | 0.142 | 0.144 | 0.073 | 0.131 | 0.128 |

Source: NFHS 1992-93.
Note: States have been ranked by column 2.

Second, although the differences in average wealth and adults' education do not explain all of the cross-state variation in enrollments, they do explain a large part of it.

Some states that do much better in absolute terms than Bihar do so because their incomes are high. For example, the enrollment rate of those 6 to 14 is 29.4 percentage points higher in Punjab than in Bihar, but Punjab is relatively wealthy and controlling for this the "Punjab" effect of enrollments of equivalent households is only 15.1 percentage points. Even more striking are the states of Delhi and Goa which have much higher absolute enrollment rates but because they are urban and relatively wealthy their "state" effect is low (5.8 and 11.5 percentage points respectively).

Columns 2 and 5 repeat the analysis restricting the sample to households that are in the bottom 40 percent. The differences in the "state effect" for the poor (column 5) are very much larger than the "state effect" for the full sample (column 4). So while living in Kerala versus Bihar made an *average* child 25.3 percentage points more likely to attend school (conditional on the observed characteristics) a *poor* child was 44.2 percentage points more likely to attend school in Kerala than Bihar. There are similarly large differences with Bihar in Tamil Nadu (32.5 percentage points), Maharashtra (28.5 percentage points), and Himachal Pradesh (31.1 percentage points). The difference is much smaller for states like Uttar Pradesh (9.4 percentage points), Rajasthan (6.5 percentage points), and Madhya Pradesh (8.3 percentage points).

*III) Speculation: Explaining wealth effects and the magnitude of the state effects*

We now examine the possible explanations of the wealth effects as well as the state specific effects. We argue that state differences in the quality of schooling is a potentially important explanation of differences in enrollment.

*A) Why are there wealth effects?*

Having established that there are large differences in children's enrollment rates across the wealth of the household, the question is why? And what might be the potential implications for policy? There are two facts to be explained. First is the large size, on average, of the wealth gap. Second is the variation in the wealth gap in enrollment across states. Any explanation of the wealth gaps has to be consistent with both large *and* variable effects.

It is useful to begin with a model in which it is surprising that there is a link between income and enrollment. Suppose education is a pure investment, household are perfectly inter-generationally linked, credit markets are perfect (in the strict sense that all intertemporal contracts are enforceable) and investment opportunities in education are equally distributed across households. In such a model schooling should not be influenced by the present financial wealth of the household. Such a model is useful, not as a guide to reality, but as an organizing framework of the ways in which reality differs from this model: credit constraints, differing investment opportunities, education as a consumption good.

*Credit constraints.* One explanation for observed wealth effects is that lower wealth is associated with unobserved higher credit constraints. While there are certainly some impacts of credit constraints on enrollments, we doubt they are very large for two

reasons[16].

First, unlike the acquisition of capital assets whose price is a substantial fraction of wealth and which need to be purchased with payments up front (like houses, or cars, or land, or factories) human capital is always purchased on a pay as you go basis. If a household wants 10 years schooling for a child as a capital investment they don't need to pay up front, but pay year by year. These year to year payments are, at the lower levels of basic education at least not only small as a fraction of household wealth, but small as a fraction of the flow of consumption.

Second, the variation across states is not well explained by credit constraints. The rich are at near universal enrollment in nearly all Indian states and hence the gap between the rich and poor is due to the shortfall of the poor from universal enrollment. The estimated gap between the poorest and richest varies from 2.6 and 4.2 percent in Himachal Pradesh and Kerala to 37 percent in Uttar Pradesh and 53 percent in Bihar. These large differences across India are certainly not accompanied by similar large differences in the perfection of capital markets as a means for financing education. Nor are there similarly large differences in the expressed or revealed demand for the use of credit for basic education.

*Distribution of investment opportunities*. The second source of income effects is that poorer households do not in fact face the same returns to education as richer households. There are three main explanations; academic performance and labor market effects, pure social exclusion.

[16] The argument that credit constraints are unlikely to affect schooling investments is hard to maintain when even among firms in the US most investment is financed through retained earnings.

First, poor children arrive at school with lower academic potential due to a variety of effects from growing up in a poorer household and hence do not face the same academic returns to schooling. Low levels of nutrition, for instance, has long term effects through a variety of pathways. Intellectual stimulation within the household through the presence of books, adult attention, and other media is also likely to be low. This implies that children from poorer household arrive at school (or would arrive at school) much lower in the distribution of performance. Even if a year of schooling brings equal increments of learning but those ranked low in the absolute performance are more likely to drop out, this could account for higher drop out of the poor.

Second is that the potential returns to education, or at least the perception or expectation of these returns, may be related to income or wealth. Investing in education might be considered a poor choice for certain groups if the (pecuniary) rewards cannot be realized. This differing "access to returns" has been shown to lead to much of the lagging in educational outcomes of Central American indigenous peoples (Patrinos, 1997). The extent to which the wealth quintiles in our model are correlated with these expectations may explain the relationship between wealth and enrollment.

The third possibility is that the poor are socially excluded from attending school, even if they wanted to do so. As seen above, the presence or absence of school facilities has small average effect on enrollments. This is consistent with large gaps in enrollments within villages between the rich and poor. In turn, this is consistent with either strong within village location effects due to distance as a deterrent or with social exclusion of poorer children.

We do include a control for whether the household was from a scheduled caste or

tribe (Table 9) in our specification but this variable is almost certainly too crude a filter to capture caste effects. Nonetheless, the estimated effects are typically substantial: even controlling for wealth, a child from a scheduled caste or tribe household is about 5 percent less likely to be in school.

*Education as consumption* Education has a large consumption component and therefore it is not surprising to see a households with higher incomes demanding more of it. Moreover, there is a substantial theoretical and some empirical literature on the relationship between the demand for the "quantity" and the "quality" of children by households. If wealthier households demand more "quality", that is more educated children, and less "quantity" then in the partial analysis of schooling and income one would expect to see a positive relationship. Finally, if children's time can be thought of as being devoted to work (inside or outside the household), schooling, and leisure then if work time decreases with income it is likely that both the demand for schooling and leisure will increase.

*B) Explaining the state effects*

*The potentially large role of quality* Improving the quality of basic education is perhaps the key priority in India today (World Bank, 1997). The role of quality is sometimes downplayed by pointing to the problem of increasing enrollment. Expenditures on access is at times simply portrayed as benefiting the poor who are not in school while improvements in quality are often claimed to benefit only the children of the rich who are already in school. However, this portrayal misses a crucial point, that the poor may not be in school precisely *because* of the low quality and that improvements in quality are a key instrument to increase the enrollments of the poor.

Children of the rich might expect to go on beyond the eight years of basic education to secondary school and perhaps university. Since having attended primary school is a prerequisite for these higher levels, children of the rich will not respond to low public school quality by dropping out, but rather by shifting to private schools getting tutoring and other private preparations for the necessary examinations *in addition* to primary school.

In contrast, children of the poor often have no expectation of receiving schooling beyond primary school, so primary school is undertaken for its own sake, for what might actually be learned in primary school, as opposed to its gatekeeper role for access to higher levels. In this case if schools are of low quality and children are learning very little then poor parents (who have other needs from their children's time) will withdraw their children from school. The result of this would be that the enrollment of the poor in basic education would be much more sensitive to the quality of basic level schools than the rich

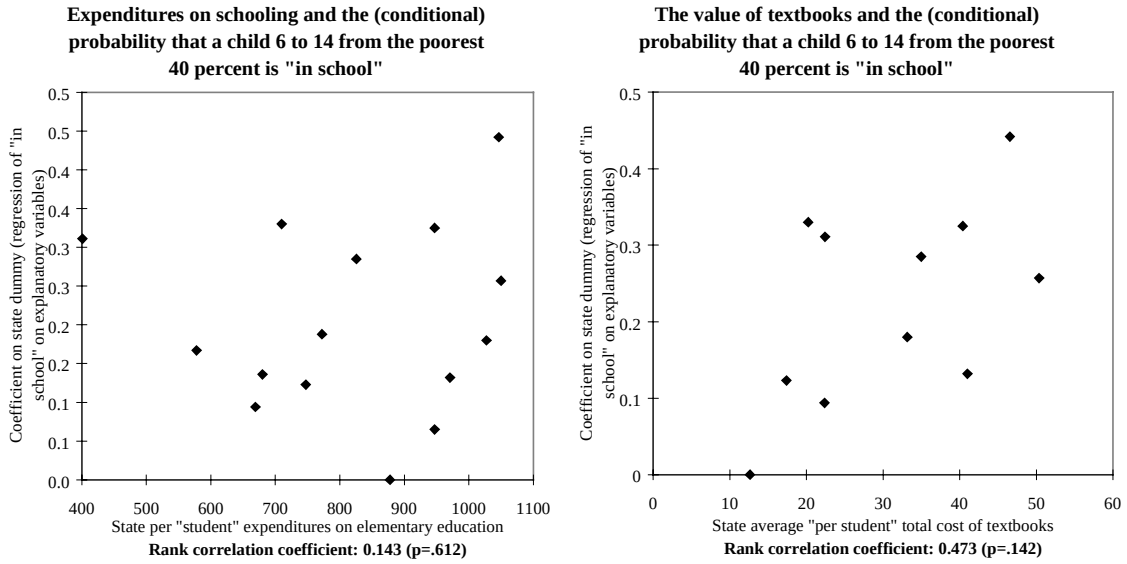
Therefore, the benefits of improving quality are not just to those already in school, but may be a key instrument in increasing the enrollments of the poor. A static analysis of the benefit incidence of quality improvements versus expansion would be misleading here, as the enrollment impacts of expansion on the poor at low quality would be overstated and the enrollment impacts of improving quality understated.[17]

*Quality is not just spending* Are there differences in policies that explain these large differences in enrollment rates of the poor? Figure 2 shows the relationship between state per student expenditures and the state effect on (conditional) enrollments of

[17] This difference of marginal and average benefit incidence in India is a key point of Lanjouw and Ravallion (1998).

the bottom 40 percent. Both visually and numerically there is very little evidence of such a relationship. This lack of a general effect is not surprising, as there is huge literature that supports the proposition that, while additional spending has the *potential* to raise school quality there is certainly no necessary connection between school quality and school spending (Filmer and Pritchett, 1998).

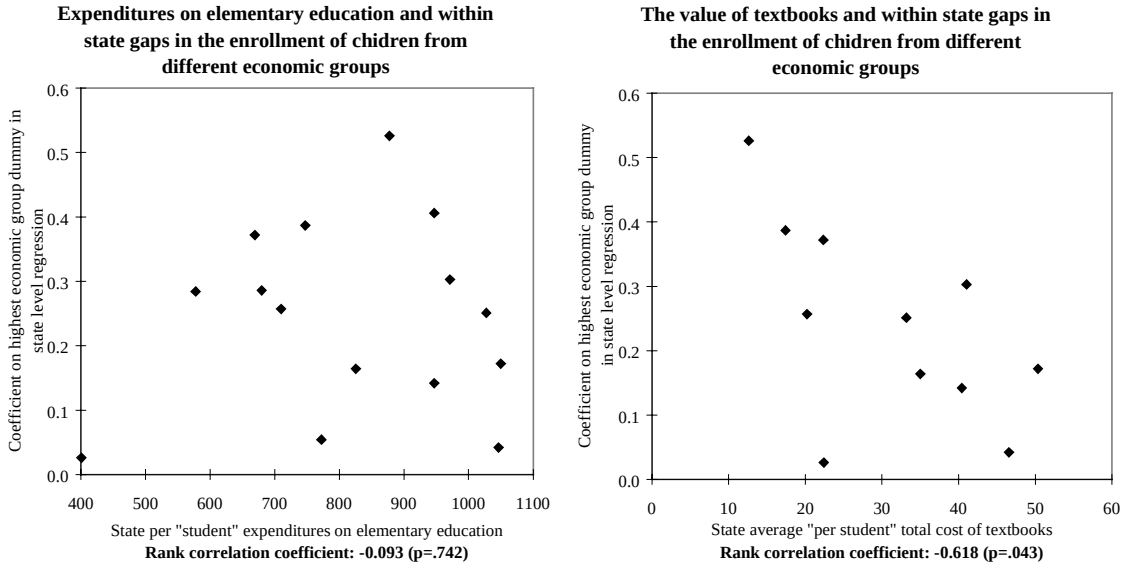
## Figure 2: "Quality" and the unexplained variation in enrollments




However, there is a possibility that other types of expenditures are a better proxy for school quality than total expenditures. The second panel of Figure 2 shows the relationship between the total retail cost of textbooks per student as reported in World Bank (1997) (which is not what the government spends, as parents absorb differing amounts of the cost across states and may also reflect cost differences due to inefficient production) and the enrollment of the poor. In this case there is a much stronger relationship both graphically and numerically (a rank correlation of .473 versus .143 for total expenditures).

Figure 3 shows the relationship between the average cost of textbooks per student and the gap in enrollment between the rich and the poor. In this figure the relationship is obvious and striking (a correlation of .62 and statistically significant, even with only 11 observations): states with higher levels of expenditures on textbooks have much lower gaps in the enrollment rates between the rich and poor. Again, this is in contrast to total expenditures which displays no such relationship (either visually or numerically).

## Figure 3: "Quality" and the wealth gaps in enrollments




These empirical results are indicative and are not to be taken literally as supporting a recommendation of spending more on textbooks per se, but rather are hueristic support for three well known points about state level educational policies. First, simply increasing overall expenditures is not the best way to either improve education or bring the poor in schools. Second, there is an important agenda to improve the *quality* of basic education in India and that those improvements in quality will be a necessary component of efforts to raise enrollment rates of the poor. The recent World Bank report on primary education in India is an excellent guide to many of these policies. Third, there

are specific actions that improve the quality of schooling and that focusing on those actions is the key priority.

*Conclusions*

We have not focused on the aggregate figures but recent work has estimated the number of 6 to 10 year olds not in school to be 32 million (out of 105 million) in 1993 (World Bank, 1997). This lack of completion of even minimum levels of schooling is concentrated among the poor and among females, and is concentrated in certain states. The example of Kerala, Himachal Pradesh, and Tamil Nadu illustrate that low enrollments are not an inevitable consequence of poverty. Universal enrollment is possible in India, but only through efforts to improve the quality of schools and through a concerted effort to reduce social exclusion ensure the poor and females are brought into the educational system.

*References*

Behrman, Jere R. and James C. Knowles, 1997. "How Strongly is Child Schooling Associated with Household Income?" University of Pennsylvania and Abt Associates. Mimeo.

Drèze et al, 1996. review of school conditions in UP

Drèze, Jean, and P.V. Srinivasan, 1997. "Widowhood and Poverty in Rural India: Some Inferences from Household Survey Data," *Journal of Development Economics* 54:217-234.

Filmer, Deon, Elizabeth M. King, and Lant Pritchett, 1997. "Gender Disparity in South Asia: Comparisons Between and Within States," World Bank Policy Research Working Paper No. 1867. Washington, DC.

Filmer, Deon and Lant Pritchett, 1998. "Estimating Wealth Effects without Income of Expenditure Data -- or Tears: Educational Enrollment in India," Policy Research Working Paper, forthcoming, The World Bank. Washington, DC.

Filmer, Deon, and Lant Pritchett, 1998. "The Effect of Household Wealth on Educational Attainment Around the World: Demographic and Health Survey Evidence," Policy Research Working Paper, forthcoming, The World Bank. Washington, DC.

Glewwe, Paul and Gillette Hall, 1995. "Who is Most Vulnerable to Macroeconomic Shocks? : Hypotheses Tests Using Panel Data from Peru," LSMS Working Paper No. 117. The World Bank. Washington, DC.

Hammer, Jeffrey, 1998. "Health Outcomes across wealth groups in Brazil and India," mimeo, DECRG, The World Bank. Washington, DC.

Lanjouw, Peter and Martin Ravallion, 1995. "Poverty and household size," *Economic Journal: The Journal of the Royal Economic Society*. 105(433):1415-34.

Montgomery, Mark, Kathleen Burke, Edmundo Paredes, and Salman Zaidi, 1997. "Measuring Living Standards with DHS Data: Any Reason to Worry?" mimeo, Research Division, The Population Council. New York, NY.

Murthi, Mamta, Anne-Catherine Guio, and Jean Drèze, 1995. "Mortality, Fertility, and Gender Bias in India; A District Level Analysis," *Population and Development Review* 21(4):745-782

NCAER (National Council of Applied Economic Research), 1994. "Non-enrollment, Drop-out, and Private Expenditures on Elementary Education: A Comparison across States and Population Groups." mimeo. New Delhi.

Patrinos, Harry Anthony. 1997. "Differences in Education and Earnings Across Ethnic Groups in Guatemala," *Quarterly Review of Economics and Finance* 37:809-821.

Haque, Peter Lanjouw and Martin Ravallion, 1998. "A Poverty Profile for India 1993-94," mimeo, DECRG, The World Bank. Washington, DC.

Jalan, Jyotsna and Martin Ravallion, forthcoming. "Transient Poverty in Postreform Rural China," *Journal of Comparative Economics.*

World Bank, 1997. *Primary Education in India*. The World Bank, Washington, DC.