©The Pakistan Development Review
53:2 (Summer 2014) pp. 205–221

# Income Growth, School Enrolment and the Gender Gap in Schooling: Evidence from Rural Pakistan

HANAN G. JACOBY and GHAZALA MANSURI

Household panel data document a remarkable closing of the gender gap in school enrolment in rural Pakistan between 2001 and 2004. During this 3-year period, there was an 8 point increase in the percentage of girls entering school, while the corresponding increase for boys was less than 2 percentage points. More than half of the rise for girls can be explained by the substantial increase in household incomes, whereas comparatively little is accounted for by increased school availability. Unpacking these enrolment trends and their determinants requires solving the classic period-age-cohort identification problem. The paper shows how to do so using auxiliary information on the distribution of school entry ages.

*JEL Classification:* O15, O40, I 25, I21
*Keywords:* School Enrolment, Gender, Income Growth, Gender Gap

## 1. INTRODUCTION

Large gender gaps in schooling persist in much of South Asia and yet are still not well understood. How much of the lower female enrolment and attainment relative to males can be explained by differences in the gender-specific returns to education [e.g., Behrman, *et al.* (1999)], by poverty, or by other barriers to schooling that differentially affect girls is the central question in formulating and targeting policies to address the gender gap in educational outcomes.

Pakistan, historically, has had one of the largest education gender gaps in the world, being especially pronounced in rural areas [Alderman, *et al.* (1996)]. While this gap has been closing over time, it remains high. Moreover, there is substantial variation in the gender gap within the country, with the two largest provinces providing a dramatic contrast. Girls' enrolment has been substantially higher in Punjab than in Sindh, even though the difference in boy's enrolment across these two provinces has been slight. How much of this cross-sectional variation in the gender gap in schooling can be attributed to the greater poverty in Sindh relative to Punjab?

Recent economic trends in Pakistan can help answer this question. Rural incomes grew robustly from 2001-04, largely due to external factors, such as the easing of drought and increased remittances in the aftermath of 9/11. This income growth was thus not driven by

Hanan G. Jacoby <hjacoby@worldbank.org> and Ghazala Mansuri < gmansuri@worldbank.org> are at the Development Research Group, World Bank, Washington, DC, USA, respectively.

technical progress that might have also altered the relative returns to education or the shadow price of child (or adult) time. Furthermore, the percentage growth in rural incomes between 2001 and 2004 was of the same order of magnitude as the baseline cross-sectional income differential between rural Punjab and Sindh. Our principal objective is to estimate the extent to which household income explains gender-specific enrolment patterns in rural Pakistan.

To be sure, there were other salient developments over this same period, notably the continuing construction of rural schools. Alderman, *et al.* (1996), in their analysis of a cohort of rural Pakistanis born in the 1960's, find that lack of local schools for girls was the main source of the gender gap in cognitive skills. To assess the relevance of this conclusion for recent cohorts, we also consider the role of school availability. Of course, new school construction may reflect increasing local demand for education, which itself could be a function of income growth. Given the lag in school construction, however, the establishment of schools after 2001 should largely reflect income growth (or other trends) *prior* to 2001.

A large and expanding literature examines the impact of income *shocks* on transitory (year-to-year or season-to-season) changes in school enrolment or attendance [e.g., Duryea, *et al.* (2007); Jacoby and Soufias (1997)]. Less empirical attention has been paid, however, to longer-run processes underlying trends in school entry decisions (i.e., ever enrolled). Glewwe and Jacoby (2004) use household panel data to show that income growth led to a rise in school enrolment in Vietnam in the mid-1990s. Their paper does not focus on school entry, but rather conflates entry and dropout behaviour, nor does it consider gender differentials in enrolment trends. By contrast, our interest is in whether a child was ever enrolled in school, an important decision margin in a setting where a large fraction of children, particularly girls, never go to school.

In order to unpack enrolment trends and their determinants using data on cohorts of young children one must first solve the classic period-age-cohort effect identification problem. Because of the linear relationship between year, age, and cohort, it is generally impossible to separate their independent effects, even with panel data. Our approach uses auxiliary information on the distribution of the age of school entrants to back out the change in probability of ever enrolling in school during childhood. Once this age effect is 'purged', period and cohort effects can be separately estimated without having to make any *ad hoc* identifying assumptions.

Using this method, we find an 8 percentage point increase in the proportion of girls who ever enrolled in school between 2001 and 2004. This is an average increase *across all cohorts among children aged 5-12 in 2001 that could potentially have enrolled in school in response to changing economic conditions*; i.e., The corresponding figure for boys is between 1 and 2 percentage points and is not statistically significant. Important cohort effects are also found for girls, but not for boys. Practically all of the movement in girls' school enrolment over the sample period occurred in Sindh; the 2001-04 cross-cohort enrolment increase for girls is 13 percentage points there, but only 2 percentage points in Punjab. Thus, in rural Sindh, the gender gap in school entry fell by about 9 percentage points in just 3 years. Increases in household income explain around sixty percent of the overall increase in girls' school enrolment, whereas the establishment of new schools plays only a minor role. It is possible that policy efforts to increase enrolment among girls such as Tawana Pakistan or the middle school stipends program for girls may account for some of the increased enrolment. Male work migration rates from Sindh also rose in the post 2001 period. Mansuri (2006) has shown a substantial impact of migration on school enrolment, particularly for girls.

The paper presents in Section I a simple description of enrolment trends in rural Pakistan, followed in Section II with a more sophisticated decomposition into period, age, and cohort effects. Section III then analyses the underlying determinants of the observed enrolment trends.

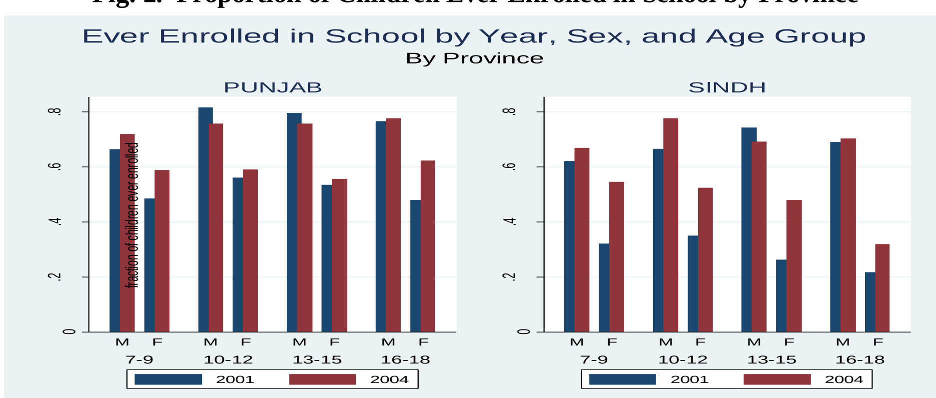
## 2. DATA AND TRENDS IN SCHOOL ENROLMENT

The data for this analysis is sourced from the Pakistan Rural Household Surveys (PRHS) of 2001 and 2004. PRHS-01 is a representative survey of rural Pakistan, consisting of around 2800 households in all four provinces (Punjab, Sindh, NWFP, and Balochistan). PRHS-04 follows up households in the two most populous provinces, Punjab and Sindh, to form a panel of about 1600 households.

For the purposes of obtaining descriptive statistics that are comparable across years, we treat the panel sample as a repeated cross-section, selecting all individuals aged between 7-18 years in each year. This leaves us with 1374 households contributing 3495 children in 2001 and 3734 children in 2004 (households need not contribute children in both years). Note that, for now, our sample is not restricted to children of household members. Doing so would exclude quite a few married women under the age of 19. For example, in 2001, 24 percent of 17 year old girls and 34 percent of 18 year-olds were already married; the corresponding figures in 2004 are 17 percent and 27 percent. Since girls who marry early are much less likely to have ever been enrolled in school, excluding them would overstate the proportion of 16-18 year-old girls ever enrolled. Selective marriage is not a concern in the subsequent econometric analysis where we focus on a sample of younger children.

Figure 1 shows the proportion of children by age-gender group ever enrolled in school (including pre-school) in 2001 and 2004. There appear to have been substantial gains for girls, both absolutely and relative to boys. A provincial breakdown of the same numbers in Figure 2 reveals that the biggest changes occurred in Sindh province, which also had far lower base (i.e., 2001) in girls' school enrolment than Punjab. As we discuss next, however, comparisons of proportions ever enrolled, even for a given age, confound year and cohort effects and hence must be interpreted carefully.

**Fig. 2. Proportion of Children Ever Enrolled in School by Province**



## 3. DECOMPOSING ENROLMENT TRENDS

In examining trends in proportions of children ever enrolled in school, one faces the classic period-cohort-age effect identification problem [see, e.g., Hall, *et al*. (2005) for a recent discussion]. The problem arises from the need to focus on children who are young enough to still be entering school over the relevant period. To fix ideas, we first describe the three effects in question:

**Period effect:** The change in enrolment of a given cohort over time captures shifts in the economic and policy environment. Period effects are only relevant for cohorts that could potentially have entered primary school in response to these shifts, which means for children no older than 12 in the base year.

**Cohort effect:** Differences in enrolment across cohorts in a given year may reflect longer-term secular trends in enrolment. For example, since we know that period effects and age effects (see below) are zero for children 13 and older, we can infer from figure 1 that there has been a sizeable cohort trend in girls' enrolment, which may have resulted in higher likelihood of later cohorts to be enrolled.

**Age effect:** As a child ages the odds of ever enrolling in school increase, or, at least, cannot decrease. In the context we study, most children enter school by age 9, with a very small percentage enrolling between age 10 and 12. Thus, age effects are only relevant for children up to age 12.

Consider, now, the unrestricted dummy variable regression (linear probability model), using two rounds of data from 2001 and 2004,

$$e_{it} = p \cdot I(year = 2004) + \sum_j c_j \cdot I(cohort = j) + \sum_k a_k \cdot I(age = k) + u_{it} \qquad \ldots \qquad (1)$$

where $e_{it}$ is an indicator for whether the child was ever enrolled in school. Since $cohort = age - 3 \cdot I(year = 2004)$, it is evident that the period effect ($p$), cohort effects ($c$), and age effects ($a$) are not separately identified. This identification problem cannot be avoided by selecting a single age group, since in this case it would be impossible to estimate the cohort effect (a given cohort consists of children at two different ages in the two rounds of the survey).

Generally speaking, without further *ad hoc* restrictions on the coefficients in (1), little can be said about the period effect [see Hall, *et al*. (2005)]. We propose an identification strategy that makes use of auxiliary information, possibly even from a different data set. The advantage of our strategy is that it eschews arbitrary parameter restrictions.

### 3.1. Purging the Age Effect

Consider a sample consisting of children age $k = 5, ..., K$. Given the innocuous normalisation $a_5 = 0$, we may write the age effects (the coefficients in Equation (1)) as

$$a_k = E(e_{it} \mid age = k) - E(e_{it} \mid age = 5) \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (2)$$

Suppose now that we have information on the age of school entrant $AE$ for a (possibly different) sample of children. Since $\Pr(e_{it}=1\,|\,age)=\Pr(AE\le age)$, age effects may be written as

$$a_k=\Pr(AE\le k)-\Pr(AE\le 5) \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (3)$$

Thus an estimator for $a_k$ is simply the difference in proportions of children who entered school at or before age $k$ and those who entered at or before age 5. This calculation is best performed on a sample of older children to avoid the censoring problem. In particular, for children younger than 10 there is still a nontrivial probability that those not yet enrolled in school may enter at a later date. We also estimate the $\hat{a}_k$ separately for boys and girls, but, with enough data, one could do so with respect to other characteristics, such as province.

Given the $\hat{a}_k$, one can calculate $\hat{\phi}_{it}=\sum_{k=5}^{K}\hat{a}_k I(age=k)$ and replace the dependent variable in (1) by $\tilde{e}_{it}=e_{it}-\hat{\phi}_{it}$, proceeding from there as though age effects were identically zero. In other words, the regression

$$\tilde{e}_{it}=p\cdot I(year=2004)+\sum c_j\cdot I(cohort_{it}=j)+u_{it} \qquad \ldots \qquad \ldots \qquad \ldots \qquad (4)$$

is equivalent to (1). Clearly, the parameters $p$ and $\boldsymbol{c}$ are now separately identified.

It may not be immediately obvious why the procedure just outlined 'works'? One might think that, if there are indeed cohort effects in $e_{it}$, these should be present in $\Pr(AE\le k)$ as well, and thus our age effects correction cannot be applied uniformly across cohorts. Note, however, that the estimate of $a_k$ essentially 'differences out' $\Pr(e_{it}=1)$. To see this, observe that, if no child ever enrols in school beyond age 12, then $\Pr(AE>12)=\Pr(e_{it}=0)$, implying $\Pr(AE\le k)=\Pr(e_{it}=1)-\Pr(12\ge AE>k)$. Consequently, $a_k=\Pr(12\ge AE>5)-\Pr(12\ge AE>k)$ contains no information on the overall probability of ever enrolling in school.

What *is* being assumed, however, is that the distribution of school entry ages *conditional* on eventual enrolment is stationary, or at least changes slowly. In other words, we are assuming that it is reasonable to impute the $a_k$ estimated retrospectively using a sample of 11-15 year olds in 2004 to 8 year-olds in 2004.[1] Though it seems unlikely that the distribution of $AE$ would change substantially within such a short time, cohort effects in $\hat{a}_k$ can be investigated formally.

In Table 1, we calculate the gender-specific $\hat{a}_k$ separately for 11–14 year-olds and for 15–18 year-olds and then test whether the differences are significantly different from zero across the two cohorts. The bootstrap t-tests reveal no significant differences at any enrolment age. Thus, there is no noticeable shift in the distribution of enrolment ages across cohorts. This is true even though (as Figure 1 shows), there is a very substantial cohort effect in enrolment for girls.

[1]There is no way to calculate the $\hat{a}_k$ directly for cohorts just entering school in 2004 precisely because many have yet to enrol. It is also worth noting that age of school entrants was only asked in PRHS-04, not in the 2001 survey. This, however, is of little relevance for our procedure.

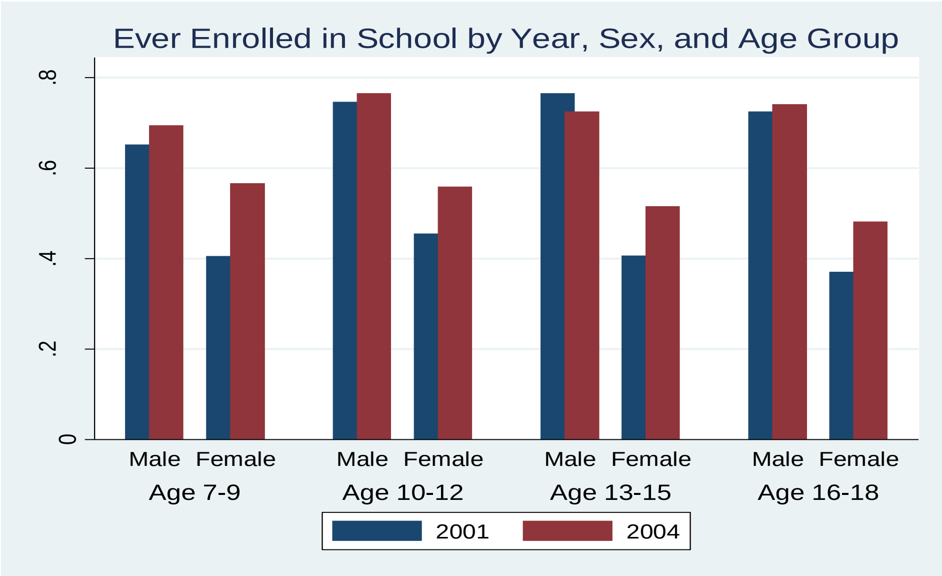
Table 1

*Changes in Age of Enrolment Distribution by Cohort*

| | $\hat{a}_k$ | | | Bootstrap |
|---|---|---|---|---|
| | 11-14 | 15-18 | | *t*-test |
| *k* | cohort | cohort | Difference | (p-value) |
| | | | ***Boys*** | |
| 6 | 0.177 | 0.206 | -0.029 | 0.219 |
| 7 | 0.279 | 0.314 | -0.035 | 0.206 |
| 8 | 0.325 | 0.372 | -0.047 | 0.104 |
| 9 | 0.355 | 0.390 | -0.034 | 0.238 |
| 10 | 0.374 | 0.407 | -0.033 | 0.260 |
| 11 | 0.391 | 0.409 | -0.018 | 0.544 |
| 12 | 0.395 | 0.411 | -0.016 | 0.583 |
| Sample Size | 588 | 567 | | |
| | | | ***Girls*** | |
| 6 | 0.114 | 0.101 | 0.013 | 0.525 |
| 7 | 0.183 | 0.190 | -0.006 | 0.800 |
| 8 | 0.231 | 0.228 | 0.003 | 0.916 |
| 9 | 0.250 | 0.236 | 0.014 | 0.611 |
| 10 | 0.268 | 0.248 | 0.020 | 0.469 |
| 11 | 0.279 | 0.248 | 0.031 | 0.266 |
| 12 | 0.292 | 0.250 | 0.042 | 0.142 |
| Sample Size | 545 | 416 | | |

*Notes:* See text for definition of $\hat{a}_k$ . Bootstraps use 1000 replications each.

**Fig. 1. Proportion of Children Ever Enrolled in School.**



### 3.2. Estimating the Period Effect

Having dealt with the age effect, we now consider the period effect in greater detail. In order to distinguish period and cohort effects, we must follow the same cohorts over calendar time. Our empirical analysis thus utilises the following sample structure:

| Year | Child Age in Survey Year | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2001 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| 2004 | | | | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

In principle, one can use data from such a sample to estimate a full set of interactions between cohort and year; i.e.,

$$\begin{aligned} \tilde{e}_{it} &= \sum_t \sum_j \alpha_{jt} \cdot I(cohort = j) \cdot I(year = t) + u_{it} \\ &= \sum_j (\alpha_{j04} - \alpha_{j01}) \cdot I(cohort = j) \cdot I(year = 2004) + \sum_j \alpha_{j01} \cdot I(cohort = j) + u_{it} \end{aligned} \qquad (5)$$

Equation (5) provides for a separate period effect for each cohort, $\alpha_{j04} - \alpha_{j01}$. Imposing the (testable) restriction of a common period effectdelivers the Equation (4).

Identification of period and cohort effects from Equations (4) (or (5)) does not require panel data. The decomposition could just as well be done using repeated cross sectional data and estimated using ordinary least squares. However, for comparability with the subsequent analysis (which does require household panel data), we estimate Equation (4) using household fixed effects. The choice between OLS and household fixed effects, at any rate, is of little consequence for the decomposition of year and cohort effects.

Note, finally, that, although we could do so in principle, we do not follow individual children over time; we only follow households. Thus, a given household might contribute a completely different set of children to the sample each round. Given our interest in the cumulative outcome "ever been enrolled", following individuals is not particularly useful. Since we do not, therefore, we remove *individual* fixed effects, the cohort effects do not drop out from Equation (4) as they otherwise would [see, e.g., Hall, *et al*. (2005)] and thus they can still be identified.

### 3.3. Results of the Decomposition

Given the imperative to maximise the number of cohorts followed over time, our sample for the decomposition differs from that underlying Figures 1 and 2. As already mentioned, we select only 5-12 year-olds in 2001 and 8-15 year-olds in 2004, giving a total estimation sample of 4705 (child-year) observations contributed by 1001 panel households. Two additional restrictions underlie this sample: First, we only choose children of household members, although for the age range we consider this is of little import since very few girls have yet married. Second, our sample excludes households that do not contribute at least one child in each survey round.

Table 2 reports the decomposition of enrolment trends into period and cohort effects, after netting out age effects. All coefficients are allowed to differ by sex. For purposes of comparison, specifications (1) and (2) use the 'raw' enrolment variable, $e_{it}$, and do not control for cohort, allowing only gender-specific period effects and intercepts. The difference between the two regressions is that the first is estimated by OLS, and the second by household fixed effects. As already indicated, including household fixed effects is of practically no consequence at this stage of the analysis.

Table 2

*Decomposition of Enrolment Trend into Period and Cohort Effects*

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Male x 2004 | **0.114** | **0.115** | **0.017** | **0.017** | **0.017** | **0.014** | **0.013** |
| | (0.015) | (0.015) | (0.014) | (0.014) | (0.014) | (0.014) | (0.014) |
| | [0.000] | [0.000] | [0.231] | [0.220] | [0.223] | [0.339] | [0.353] |
| Female x 2004 | **0.156** | **0.152** | **0.080** | **0.080** | **0.080** | **0.077** | **0.077** |
| | (0.016) | (0.016) | (0.016) | (0.016) | (0.016) | (0.016) | (0.016) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| Female | **–0.237** | **–0.234** | **–0.134** | **–0.055** | **0.002** | – | – |
| | (0.022) | (0.021) | (0.020) | (0.039) | (0.057) | | |
| | [0.000] | [0.000] | [0.000] | [0.158] | [0.965] | | |
| Male x Cohort | | | | | **-0.001** | | **-0.003** |
| | | | | | (0.005) | | (0.005) |
| | | | | | [0.860] | | [0.605] |
| Female x Cohort | | | | | **–0.017** | | **-0.019** |
| | | | | | (0.005) | | (0.005) |
| | | | | | [0.000] | | [0.000] |
| Fixed Effects | No | HH | HH | HH | HH | HH-sex | HH-sex |
| Age Effects (Adjusted) | No | No | Yes | Yes | Yes | Yes | Yes |
| Cohort Effects | No | No | No | Unres. | Linear | Unres. | Linear |
| *F*-test *p*-value * | | | | 0.779 | 0.218 | 0.375 | 0.405 |
| Total Sample (Households) | 4705 | 4705 | 4705 | 4705 | 4705 | 4587 | 4587 |
| | (1001) | (1001) | (1001) | (1001) | (1001) | (985) | (985) |

*Notes:* Standard errors adjusted for clustering on household in parentheses; *p*-values in square brackets.

* In specifications (4) and (6) the 14 restrictions tested are: all period effects are equal across cohorts for males and females; In specifications (5) and (7) the 12 restrictions tested are: gender-specific cohort dummies, which can be collapsed into gender-specific linear cohort trends.

Our adjustment for age effects, starting with specification (3), has a big impact on the estimated coefficients. This should be expected, given our sample structure. Children are 3 years younger on average in 2001 than in 2004 and for this reason alone are less likely to have enrolled in school. Not correcting for age effects thus greatly exaggerates the period effect. In specification (3) the period effect for boys essentially vanishes, while that for girls falls by about half as compared to specification (2).

Specifications (4) and (5) add cohort effects, in the first case by including an unrestricted set of cohort dummies, and in the second case, with a linear cohort trend. The linear trend cannot be rejected in favour of the unrestricted dummies. While there is no significant cohort trend for boys, there is a negative trend for girls. That is, the later a girl was born the more likely she was to have been enrolled in school. The inclusion of cohort trends, however, does not affect the estimated period effect for either boys or girls. The insignificant *F*-test in Table 2 also indicates that a completely unrestricted model (cohort-year interaction dummies) fits the data no better than a restricted (common) period effect.

Looking at the behaviour of the female dummy coefficient across specifications (3)-(5), there are signs of a co-linearity problem. The female dummy and its interaction with the cohort trend are highly correlated with each other. So, it might be difficult to distinguish the effect of being a girl *per se* versus the effect of being a girl of successively later vintage. One way to avoid this problem is not to estimate the female effect in the first place. This can be accomplished by replacing household fixed effects with household-sex fixed effects, which absorb the female dummy (effective sample size falls a bit because there are some households contributing only a single girl or boy that must be dropped). Specifications (6) and (7) thus include household-sex fixed effects, while allowing for unrestricted and linear cohort trends. Once again, the linear trend cannot be rejected, while the remaining coefficients are virtually unaffected by the inclusion of household-sex fixed effects. Later we use specification (7) to deal with a similar, but even more severe, co-linearity problem.

## 4. EXPLAINING THE PERIOD EFFECT

The next step is to quantitatively assess the contribution of different economic factors to the period effects in enrolment. Our empirical approach is to re-estimate Equation (4), replacing the year dummy with a vector of time-varying regressors. Specifically, we focus on income growth and school construction.

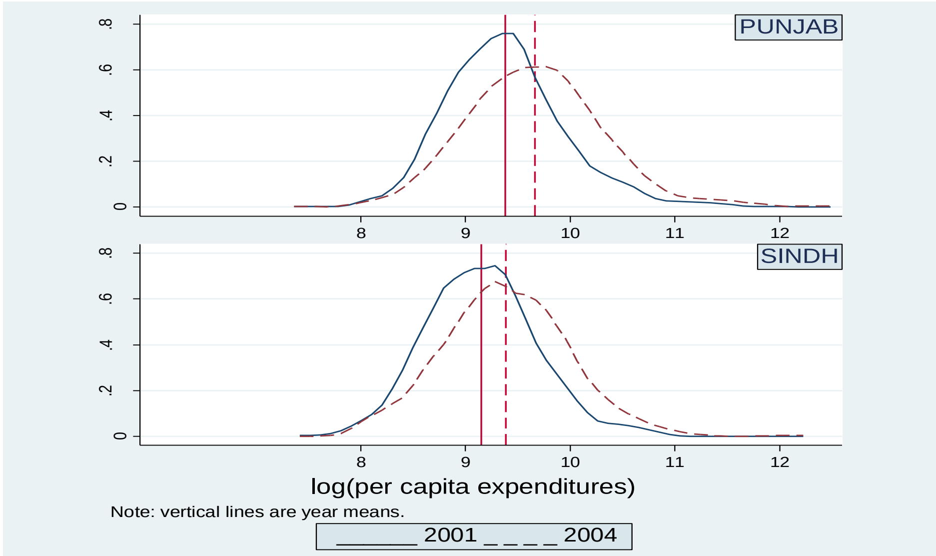
### 4.1. Income Growth

Our measure of income is per capita household expenditures.[2] The 2001 and 2004 PRHS surveys have essentially identical household expenditure modules, so the resulting expenditure aggregates are perfectly comparable across years after controlling for inflation. Figure 3 displays the distributions of log per capita expenditures by year and province based on the panel sample.[3] Household consumption grew substantially in both provinces; by around 28 percent on average in Punjab and by 23 percent in Sindh. As of 2004, the average household in Sindh had achieved almost the same income level as the average household in Punjab in 2001.

[2]Glewwe and Jacoby (2004) rationalise the use of household expenditures as a measure of the shadow value of wealth in the context of a dynamic model of human capital accumulation wherein child school enrolment and consumption are household decision variables. Thus, after properly accounting for endogeneity, the partial correlation between enrolment and consumption reflects a well-defined wealth effect on the demand for schooling.

[3]The panel sample may not be adequately representative of the rural population of the two provinces. In particular, 260 households from whom expenditure data were gathered in 2001 were not followed up in 2004. This sample loss was mainly due to administrative problems. A regression of 2001 log per capita expenditures on a province dummy and a dummy for whether the household does not appear in PRHS-04 reveals that, on average, base-year expenditures are 10 percent lower for households lost in 2004. However, this entire effect is due to 71 households from 4 villages (3 in Punjab, 1 in Sindh) that could not be revisited due to security concerns. Otherwise, the lost households are no different in terms of baseline wealth than those that were followed-up.

**Fig. 3. Household Expenditure Distributions by Province**




In a cross-section, household expenditures and child school enrolment are likely to be jointly determined and may thus be positively correlated for reasons having little to do with the increased affordability of schooling as income rises. Specifically, given positive schooling costs, any change (shock) in school enrolment independent of changes in wealth will be associated with a change in consumption. Having no direct way to handle such feedback,[4] we argue next that it should not cause significant bias.

Consider the stripped down regression model

$$e_{it} = \beta \log(C_{it}) + u_{it} \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \qquad (6)$$

with cohort effects suppressed and the period effect captured only by $C_{it}$, per-capita expenditures on all goods other than schooling. Conceptually, we would like $C_{it}$ to represent *ex-ante* consumption; i.e., to reflect the resources available to the household *prior* to any change in enrolment. However, what we observe is *ex-post* consumption (or changes therein), which we denote by $C'_{it}$. It is reasonable to suppose that $C'_{it} = C_{it} - \gamma e_{it}$, where $\gamma > 0$, since enrolling a child in school reduces the resources that could otherwise be spent on the consumption of other goods, either because of direct education costs or the forgone income from the child's labour.

Assume now that total annualised per-child enrolment costs are proportional to ex-ante consumption; i.e. $\gamma = \delta C_{it}$. Thus, wealthier households pay proportionally more for tuition, books, uniforms, etc. and/or their children's time has a higher opportunity cost.

[4]In principle, it might be possible to instrument consumption changes with household characteristics that predict whether income grew over the relevant time period. For example, households with relatively more un-irrigated land would have been more affected by the 2001 drought, or households with more migrants in 2001 would have benefited more from the post-9/11 increase in foreign remittances. In practice, however, such instruments performed poorly in our data. This approach would also require a first-difference specification in household means of the enrolment variable as in Glewwe and Jacoby (2004).

Given this, the relationship between ex-post (observed) consumption and ex-ante consumption is $\ln(C'_{it}) = \ln(C_{it}) + \ln(1-\delta e_{it})$. Substituting into (6) gives

$$e_{it} = \beta \ln(C'_{it}) - \beta \ln(1-\delta) e_{it} + u_{it} \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (7)$$

The least squares estimate of β thus converges in probability to $\frac{\beta_0}{1+\beta_0 \ln(1-\delta)}$, which for δ not too large and $0 < \beta_0 < 1$ is approximately $\beta_0(1 + \beta_0\delta)$. So,

$$\frac{\hat{\beta}_{OLS} - \beta_0}{\beta_0} \cong \beta_0 \delta \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (8)$$

which indicates that the bias in the least squares estimate is positive and, in percentage terms, is roughly equal to the true value of β times δ. In all of the empirical specifications below, none of which correct for feedback, the (over) estimates of β never exceed 0.3. We can be assured, therefore, that $\beta_0 < 0.3$. Thus, in order for the feedback bias in these estimates to exceed 10 percent, the value of δ would have to be greater than 0.33; in other words, enrolment costs per child would have to account for at least a third of ex-ante consumption! More realistic values of δ imply a negligible bias in $\hat{\beta}_{OLS}$.

Household expenditures may also be endogenous with respect to school enrolment decisions due to measurement error in expenditures. Noise in household expenditure data will result in the usual attenuation bias, which, in contrast to the case of feedback bias just discussed, can be quite substantial. To correct for this, we need an instrument correlated with household consumption expenditures, but not with the measurement error in this variable. A natural candidate is the village-year (leave-one-out) mean of expenditures as calculated from the full sample (i.e., including households that do not contribute children to the panel sample). As we will see, this instrument performs extremely well in terms of first-stage explanatory power.

### 4.2. School Construction

The 2004 PRHS includes a census of schools within each village. In addition to knowing the type of school (primary, middle or boys only, girls only, or mixed), we also have the date the school was established. Using this information, we can construct indicators for whether a girls' (boys') primary (middle) school was present in the village at the time of each survey. The same can be done for schools of given type within the *settlement* where the household resides, since most villages have multiple settlements. Due to mobility constraints, especially for girls, it may matter more that the appropriate school is located in the same settlement rather than merely in the same village.[5] On the other hand, establishing the very first girls' school in an entire village may have a greater effect on enrolment than adding the tenth school, even though that school happens to be in the same settlement.

[5]On mobility constraints for girls see Khan (1998), Jacoby and Mansuri (2010) and Jacoby and Mansuri (2013).

Because we include household fixed effects, identification of the impact of school availability on enrolment comes from schools that were established since 2001. Given that the panel sample covers only 93 villages with 274 settlements, there may not be enough new schools in the data to estimate the effects of interest. Indeed, this is a particular problem for boys' schools, as Table 3 indicates. For example, not a single one of our sample villages that did not have a boys' primary school prior to 2001 received one in the subsequent 3 years, although two settlements within these villages did get a new school. Likewise, there was a paucity of new middle school construction in these villages. Thus, the percentage of boy observations in our sample for which there is a change in school availability between 2001 and 2004 never exceeds one. For girls' schools the situation is somewhat better, so there may be hope of identifying school availability effects for girls.[6]

Table 3

*Changes in School Availability 2001-2004*

| | Primary Schools | Middle Schools |
|---|---|---|
| **Boys** | | |
| No. of Villages | 0 (0.0) | 1 (1.0) |
| No. of Settlements | 2 (0.7) | 1 (1.0) |
| **Girls** | | |
| No. of Villages | 2 (1.6) | 2 (3.6) |
| No. of Settlements | 6 (2.6) | 3 (3.2) |

*Note:* Percent of sample observations for that gender residing in relevant village or settlement in parentheses. There are a total of 93 villages and 274 settlements.

### 4.3. Main Results

Table 4 displays the determinants of the period effects. In other words, the gender-specific period effects in specification (5) of Table 2 are replaced here with log per capita expenditures interacted with a male and female dummy, as well as with girls' primary school availability in the village interacted with a female dummy. Given the lack of variation (see Table 3), we do not attempt to estimate school availability effects for boys.

The first specification is estimated using household fixed effects; the second deals with measurement error in expenditures using as instruments village leave-out means interacted with the gender dummies. Shea partial $R^2$s for the two first-stage regressions are quite high; 0.19 for the boy-expenditure interaction and 0.16 for the girl interaction. The second-stage expenditure coefficients behave exactly as one would expect with measurement error. The female coefficient, already positive and significant in the OLS, increases substantially in magnitude. The male income effect, meanwhile, remains insignificant across specifications. There is also some evidence that, for girls, the addition of a girls' primary school in the village increases enrolment.

[6]Primary school availability changes little in part because by 2001 nearly every village, and indeed many settlements, already had one. Specifically, in 2001, 99 (75) percent of the boys and 94 (69) percent of the girls in our sample had a primary school for their respective gender in their village (settlement).

Table 4

*Determinants of Period Effects in School Enrolment*

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Male x log(pcexp) | **0.017** | **–0.003** | **0.013** | **0.042** | **0.042** | **0.042** |
| | (0.019) | (0.046) | (0.021) | (0.050) | (0.050) | (0.050) |
| | [0.356] | [0.954] | [0.527] | [0.406] | [0.406] | [0.406] |
| Female x log(pcexp) | **0.072** | **0.262** | **0.068** | **0.185** | **0.170** | **0.186** |
| | (0.021) | (0.061) | (0.023) | (0.068) | (0.068) | (0.067) |
| | [0.001] | [0.000] | [0.003] | [0.006] | [0.012] | [0.006] |
| Female x girl's primary school | **0.183** | **0.158** | **0.385** | **0.366** | **0.368** | **0.289** |
| | (0.077) | (0.081) | (0.135) | (0.132) | (0.132) | (0.097) |
| | [0.018] | [0.052] | [0.004] | [0.005] | [0.005] | [0.003] |
| Female x girl's middle school | – | – | – | – | **0.132** | – |
| | | | | | (0.093) | |
| | | | | | [0.157] | |
| Female | **–0.633** | **–2.538** | – | – | – | – |
| | (0.244) | (0.637) | | | | |
| | [0.010] | [0.000] | | | | |
| Male x cohort | **0.000** | **–0.001** | **–0.002** | **–0.002** | **–0.002** | **–0.002** |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| | [0.953] | [0.869] | [0.685] | [0.685] | [0.685] | [0.685] |
| Female x cohort | **–0.018** | **–0.019** | **–0.019** | **–0.019** | **–0.019** | **–0.019** |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| Fixed effects | HH | HH | HH-sex | HH-sex | HH-sex | HH-sex |
| | 4628 | 4628 | 4508 | 4508 | 4508 | 4508 |
| Total sample (households) | (987) | (987) | (970) | (970) | (970) | (970) |
| % period effect (female) explained by growth in | | | | | | |
| Income + schools | 26 | 84 | 30 | 67 | 68 | 68 |
| Income only | 22 | 81 | 22 | 59 | 55 | 60 |

*Notes:* Standard errors adjusted for clustering on household in parentheses; *p*-values in square brackets. Specifications (1) and (3) are estimated by fixed effects. Specifications (2), (4)-(6) are estimated by fixed effects-IV using interactions with the village-year leave-one-out mean of log (pcexp) as instruments. Specifications (1)-(5) define school availability at the village level, whereas specification (6) does so at the settlement level.

One worry, however, is the alarming increase in the female dummy variable coefficient, becoming unrealistically large in specification (2). The problem, again, is colinearity; this time between the female dummy and the female dummy interaction with log per capita expenditures. Mechanically, these two variables must be highly correlated, which they are, and instrumenting expenditures only exacerbates the problem. As before, our solution is to purge the female dummy altogether by including household-sex fixed effects. Comparing specifications (3) and (4), then, yields a similar conclusion about the influence of measurement error in the expenditure coefficient, except that the estimated income effect is only about two-thirds as large as before (0.185 vs. 0.262). The primary school availability effect for girls more than doubles in magnitude, however.

The last two columns in Table 4 explore alternative specifications of the school availability effect for girls. In specification (5), we control for the presence of a middle school for girls in the village. While greater middle school availability does increase the likelihood of ever enrolling a girl, the effect is not significant and the coefficient is far less than the corresponding one for girls' primary schools. Specification (6) replicates specification (4) using girls' primary school availability at the settlement level. Recall from Table 3 that under this second definition of primary school availability somewhat more girls in our sample experience a change in availability over the 2001-04 period (2.6 percent versus 1.6 percent). The resulting primary school coefficient, however, is little changed.

By way of summary, we calculate the fraction of the period effect explained by changes in the time-varying covariates. This exercise is only relevant for girls, since period effects are negligible for boys. The second to last row in Table 4 shows that, for the preferred specifications (household-sex fixed effects with correction for measurement error), we can explain more than two-thirds of the period effect in girls' enrolment. The figures in the last row show that almost all of this is due to income growth; very little of the period effect is explained by growth in school availability, which is not surprising given that there is hardly any change in school availability in our sample.

### 4.4. Provincial Differences

We now turn to the question raised at the beginning: As we have seen, between 2001 and 2004, average income in Sindh rose to about the level of Punjab in 2001. Girls' school enrolment in Sindh followed a similar pattern, also rising to about the level observed in Punjab in 2001. Of course, this may just be coincidence; the fact that these trends line up by no means implies that the income rise in Sindh was entirely responsible for the increase in girls' enrolment.

To investigate the question, we present a province-level analysis in Table 5. By far the largest period effect for girls is in Sindh: In 2004, the proportion of girls who had ever enrolled in school was 13 percentage points higher than in 2001. Looking at province-specific results in specification (2), we see that the income effect is also by far the largest (and only significant) for girls in Sindh. The primary school availability effect, by contrast, is important only in Punjab; more precisely, it is only *estimable* in Punjab, because there was no change in school presence in any of the Sindh villages in our sample.

Even though girls' school enrolment in Sindh appears much more responsive to income changes than in Punjab, income growth explains less than half of the period effect in Sindh (see bottom of Table 5). This suggests that factors other than income must be responsible for at least half the convergence in girls' enrolment between Pakistan's two largest provinces. By the same token, it is unlikely that the 2001 gap in girls' enrolment between Punjab and Sindh can be mostly explained by Punjab's greater wealth.

Table 5

*Province-level Decomposition of Enrolment Trends*

| | Punjab | | Sindh | | |
|---|---|---|---|---|---|
| | (1) | (2) | (1) | (2) | (3) |
| Male x 2004 | **–0.009** | **–** | **0.040** | – | – |
| | (0.019) | | (0.021) | | |
| | [0.624] | | [0.057] | | |
| Female x 2004 | **0.020** | **–** | **0.133** | – | – |
| | (0.019) | | (0.024) | | |
| | [0.306] | | [0.000] | | |
| Male x log(pcexp) | – | **0.005** | **–** | **0.071** | **0.071** |
| | | (0.070) | | (0.071) | (0.071) |
| | | [0.942] | | [0.314] | [0.314] |
| Female x log(pcexp) | – | **0.061** | **–** | **0.261** | **0.287** |
| | | (0.082) | | (0.098) | (0.104) |
| | | [0.457] | | [0.008] | [0.006] |
| Female x log(pcexp) x | | | | | **–0.521** |
| No girl's primary school in 2001 | | | | | (0.194) |
| | | | | | [0.007] |
| Female x girl's primary | – | **0.386** | – | – | – |
| school (in village) | | (0.136) | | | |
| | | [0.004] | | | |
| Male x cohort | **0.003** | **0.004** | **–0.010** | **–0.010** | **–0.010** |
| | (0.007) | (0.007) | (0.008) | (0.008) | (0.008) |
| | [0.603] | [0.550] | [0.206] | [0.230] | [0.230] |
| Female x cohort | **–0.016** | **–0.016** | **–0.021** | **–0.022** | **–0.022** |
| | (0.007) | (0.007) | (0.008) | (0.008) | (0.008) |
| | [0.022] | [0.024] | [0.008] | [0.005] | [0.004] |
| Total sample | 2374 | 2329 | 2213 | 2179 | 2179 |
| (households) | (524) | (514) | (461) | (456) | (456) |
| % period effect (female) explained by growth in | | | | | |
| Income + schools | | 147 | | 44 | 40 |
| Income only | | 77 | | 44 | 40 |

*Notes:* Standard errors adjusted for clustering on household in parentheses; *p*-values in square brackets. All specifications include household-gender fixed effects. Specifications with log (pcexp) interactions, are estimated by IV to correct for measurement error.

A final question to consider, as far as Sindh is concerned, is whether the response of girls' enrolment to income growth depends on school availability? In particular, for around 6 percent of girls in the Sindh subsample, no (girls') primary school existed in their village in 2001. Since they would have had no school to go to, it would be very surprising if the enrolment of these girls rose with household income. That this indeed did not happen is confirmed by the results in the final column of Table 5. The response of enrolment to income growth for girls without a primary school in 2001 is not significantly different from zero ($p$-value = 0.15), whereas it remains significantly positive for girls who did have access to a village primary school.[7]

## 5. CONCLUSIONS

Recent years have seen a marked closing of the gender gap in school enrolment in rural Pakistan. This paper has shown how to use panel data to isolate changes in school entry attributable to shifting economic conditions. Using this approach, we have established that income growth has played an important role in drawing an increasing number of girls into school. Meanwhile, very little of the observed enrolment changes can be explained by new school construction.

Despite the enrolment gains observed in the 2001-2004 period, the overall gender gap in schooling remained significant and the findings of this paper suggest that the much lower girls' school enrolment observed in Sindh as compared to Punjab cannot be attributed entirely to the large income differences between the two provinces. A recent paper that focuses specifically on this residual gender gap [Jacoby and Mansuri (2013)], finds that much of the residual gender gap can be explained by social constraints. In particular, it finds that social stigma greatly discourages school enrolment among low-caste children, with low-caste girls, the most educationally disadvantaged group, being the worst affected. However, it also shows that low-caste households who can escape stigma invest at least as much in schooling as high caste households, indicating similar returns to schooling across caste groups. These results suggest that, from a policy perspective, it may be important to deliberately target gender specific social barriers to schooling in addition to any policies that target schooling demand through transfers.

## REFERENCES

Alderman, H., J. Behrman, D. Ross, and R. Sabot (1996) Decomposing the Gender Gap in Cognitive Skills in a Poor Rural Economy. *Journal of Human Resources* 31:1, 229–54.

Behrman, J., A. Foster, M. Rosenzweig, and P. Vashishtha (1999) Women's Schooling, Home Teaching, and Economic Growth. *Journal of Political Economy* 107:4, 682–714.

Duryea, S., D. Lam, and D. Levinson (2007) Effects of Economic Shocks on Children's Employment and Schooling in Brazil. *Journal of Development Economics* (Forthcoming).

Glewwe, P. and H. G. Jacoby (2004) Economic Growth and the Demand for Education: Is there a Wealth Effect? *Journal of Development Economics* 74:1, 33–51.

Hall, B., J. Mairesse L. Turner (2005) Identifying Age, Cohort, and Period Effects in Scientific Research Productivity: Discussion and Illustration using Simulated and Actual Data on French Physicists. Cambridge, MA. (NBER Working Paper 11739).

Jacoby, H. G. and G. Mansuri (2010) Watta-Satta: Exchange Marriage and Women's Welfare in Rural Pakistan. *American Economic Review* 100, 1804–1825.

[7]The same exercise for primary schools within the *settlement* in 2001 yields similar, but statistically weaker, results. In this case, 26 percent of girls in the sample had no primary school in their settlement in 2001.

Jacoby, H. G. and G. Mansuri (2013) Crossing Boundaries: How Social Hierarchy Impedes Economic Mobility. The World Bank. (Under Review).

Jacoby, H. G. and E. Skoufias (1997) Risk, Financial Markets, and Human Capital in a Developing Country. *Review of Economic Studies* 64:3, 311–336.

Mansuri, Ghazala (2006) Migration, School Attainment and Child Labour. The World Bank. Washington, DC. (Policy Research Working Paper # 3945).

Khan, A. (1998) *Female Mobility and Social Barriers to Accessing Health and Family Planning Services: A Qualitative Research Study in Three Punjabi Villages*. Islamabad: Ministry for Population Welfare, London School of Tropical Hygiene and Medicine and Department for International Development, British Government.