Returns to education in India

Scott Fulford[*]

December 2012

Abstract

Despite the evidence for high returns to education at an individual level, large increases in education across the developing world have brought disappointing returns in aggregate. This paper shows that the same pattern holds in India by building aggregates from micro-data so that the comparability and quality issues that plague cross-country analyses are not a problem. In India both men and women with more education live in households with greater consumption per capita. Yet in aggregate, comparing across age cohorts and states, better educated male cohorts consume only about 4% more than less well educated ones. Better educated female cohorts do not live in households with higher consumption. This result is robust to: (1) using econometric models that account for survey measurement error, (2) different measures of household consumption and composition, (3) allowing returns to differ by state, and (4) age mismeasurement. Comparing state returns to a measure of school quality, it does not seem that poor quality is responsible for the low returns.

---

[*]Boston College Department of Economics, 140 Commonwealth Ave, Chestnut Hill, MA 02467; email: scott.fulford@bc.edu. This work has benefited from the advice and comments of Angus Deaton, Anne Case, and Chris Paxson.

1 Introduction

Education is often given the throne in the pantheon of development (Case, 2006) because of the extensive evidence for high returns at the individual level (Psacharopoulos and Patrinos, 2004). The empirical case for strong returns beyond the individual, when general equilibrium effects and externalities may be important, is not so clear. Despite huge increases in education in developing countries after 1960 the accompanying increases in production have been disappointing, prompting Pritchett (2001) to ask "where has all the education gone?" Since it is the returns to additional education for society, not the individual, that matter for policy, good estimates of the social returns are crucial.

One reason for the lack of relationship as suggested by Temple (2001) and Cohen and Soto (2007) is that the quality of the data comparing across countries is poor. Another reason highlighted by Pritchett (2001) may be that education exhibits strong externalities. For example, if education acts largely as a signal of quality then there may be high returns for individuals, but low returns to increases in education on average. Alternatively peer effects or innovation spillovers may result in positive education externalities, leaving the individual returns too low.[1]

To overcome these difficulties this paper focuses on India. Focusing on a single large country removes the problem of comparability and data quality which plague cross-country comparison. Further, since this paper builds estimates of aggregates directly from micro data across 25 years of national sample surveys it is possible to compare the individual and aggregate returns.

The basic approach of this paper and the conclusions it reaches can be summarized in two figures. Figure 1 shows the educational attainment of successive birth cohorts. There has been a tremendous increase in educational attainment in India accompanied by domestic and international efforts to build schools and train teachers.[2] From a low base of less than a year on average for women, and about two years for men for those born around 1920, the number of years of education

---

[1]In the growth literature there seems to be a relationship between the level of education and subsequent growth (Barro, 1991) although Bils and Klenow (2000) suggest that the causal relationship need not run from education to growth.

[2]For example, the District Primary Education Programme by the Government of India with World Bank support during the 1990s.

has increased steadily to around six years for women and nearly eight for men in the last cohort to have finished its education. Have these gains in education brought similar gains in material well-being?

Figure 2 shows the per capita household consumption of birth cohorts from select years between 1983 to 2005. Although younger cohorts are on average better educated, and sometimes substantially better educated, they do not appear to have systematically higher consumption. That suggests quite low returns to education. The bulk of the paper will be devoted to setting up the estimation to make the comparison between figures 1 and 2 systematic. Adding in all available household surveys, comparing across states, allowing consumption to vary systematically with age and year across cohorts, accounting for measurement error, and using several different approaches to account for family composition, all confirm the basic observation: Between cohorts, the returns to education for males are around 3-4% and appear to have *decreased* after 1991 following a major liberalization of the economy. Women in better educated cohorts do not appear to live in households with higher per capita household consumption.

On an individual level, there do appear to have been substantial returns to education, so these results are not driven by using consumption or some other peculiarity of the data. Across the population and within each age cohort, there is a strong positive relationship between education and consumption: individuals with more education live in households with greater per capita household consumption, with an extra year associated with an increase in consumption of between 5 and 8% for both men and women. The relationship is approximately the same pooling the entire population: better educated individuals live in households with more consumption. The estimated slopes appear to be larger for women. Such estimates have well known biases and cannot capture general equilibrium effects. Under reasonable assumptions, I show aggregating avoids these problems.

Aggregating over cohorts further allows me to correct for measurement errors both from sampling using the approach of Fuller (1987, Section 3.1.2) and because of age misreporting. Age misreporting is extensive in India: around one quarter of Indians have their ages incorrectly reported in the National Sample Surveys (see table 1), and the least educated are the most likely to

misreport their ages. At first glance that appears to create a large measurement error problem. Yet I show that unlike the standard measurement error which will tend to create attenuation to zero, even education specific age misreporting does not lead to bias when estimating using appropriately defined cohorts.

Low consumption returns for men and women does not mean that there are low returns to education in all dimensions. Not all externalities are within cohorts, and so there may be benefits even if the cohort is better of in terms of consumption. A more educated population may cause higher growth for everyone, even if the better educated cohorts do not benefit more. Education may contribute to a more desirable civic society, may improve health, and may be valuable in its own right (Sen, 1999). Education, particularly for women, may also lead to a more equitable distribution across society and within the family.

With a population of more than a billion people and large increases in education, the overall returns to education in India are important by themselves, but India's diverse states provide an additional way to examine what drives the returns to education. For example, Foster and Rosenzweig (1996) find that returns increased in areas where the Green Revolution allowed the use of new technologies. Yet I shot that the returns at the cohort level do not seem to vary substantially at the state level. One reason the returns may be so low is that the quality of education may be poor (see, for example, PROBE Team (1999)). I use a test administered to school age children as part of the nationally representative India Human Development Survey in 2005 (Desai, Vanneman, and National Council of Applied Economic Research, 2008) to get a sense of the quality of the education systems across states. There is a weak positive relationship between the returns across states and the added value that an additional year of education in that state brings. Yet going from the worst state to the best in the benefit of an extra year of education on reading raises the returns by less than 2%, and moving from the best to worst in mathematics by less than 1%. It is not clear that quality by itself is responsible for the low returns.

Most studies of the returns to education in India (and other developing countries) consider only wage rates, which means that they apply only to a very small, almost entirely male and urban,

section of the population. Duraisamy (2002) uses the employment rounds of the National Sample Survey (NSS) in 1983 and 1993-1994 to estimate a Mincerian regression of log wages on education for individuals. Dutta (2004) goes one step further, separating out the regular wage earners from the casual wage earners, arguing that there is a dual labor market. Education and inequality are also closely linked. Decomposing by quantiles, Azam (2012) examines changes in the wage structure in urban India over the three employment rounds of the NSS and suggests the returns are increasing. Kijima (2006) also finds increasing returns at the higher education levels after 1991, and uses this increase to explain rising inequality in India.

2 Returns to education

2.1 Consumption returns to education

This section lays out a simple log linear relationship between individual consumption and education. Such an approach can be motivated through a life-cycle model with production, saving, and a human capital decision or simply taken as a reduced form. I model the human capital decision which would give rise to such a reduced form in appendix A. Education is related to log consumption at time t:

$$\ln c_{it} = z_i + b_i + \theta_{it} + \beta S_i + \gamma S_\tau + h(a_{it}) + u_{it}, \tag{1}$$

where $\ln c_{it}$ is log consumption, z_i is ability, b_i is family background such as caste or family wealth or land, S_i is the level of education, S_τ is the average education of the cohort τ to which i belongs, $h(a_{it})$ is a function of age at time t that allows for experience and degeneration in old age, θ_{it} allows consumption to vary over time in ways unrelated to education, and u_{it} is unobserved variation specific to that individual.

Equation (1) differs from the standard individual approach for considering returns to education in two ways. The first is that it allows general equilibrium effects of cohort education S_τ. These

could be either positive or negative. While there may also be general equilibrium effects from education beyond the cohort, these are not directly measurable while still allowing for time varying effects. How important the education of other cohorts is depends on the degree of substitutability in production among cohorts (see, for example, Card and Lemieux (2001)).

The second way equation (1) differs from the standard approach is that it measures returns to education using consumption. There are several reasons for focusing on consumption returns rather than earnings. The first is philosophical: if we are interested in policy and its affects on welfare, then consumption is the relevant variable, although education may affect consumption through labor income. The consumption and income returns to education need not be the same, although the two are intimately linked. For example, if the permanent income hypothesis holds exactly, then consumption in each year is an individual's permanent income, and so the consumption returns are the correctly measured income returns to education.

The second reason to focus on consumption is more concrete: everyone consumes, and so has potential consumption returns to education, while not everyone has labor income. Less than half of men, and only a very small fraction of women work directly for wages in India, and it is still a largely rural society.[3] This paper is therefore one of the very few that estimates any returns for women or for the large majority of Indians who live in rural areas but do not get most of their income from wages. That is especially important since the selection into wage work may be one of the reasons to obtain additional education. A third reason to focus on consumption is that it is less sensitive to income shocks since these are likely to be at least partially smoothed. Consumption is also measured consistently across a wide population, while income, especially self-employment income in agriculture, is not.

---

[3]While the fraction working for a wage has increased over time, survey calculations from the NSS 50th round in 1993-1994 (a large round in the middle of the analysis) show that in the rural sector 20% of men and 10% of women worked for any kind of wage (formal, informal, paid in-kind) on any day in the week preceding the survey. In urban areas 29% of men worked for wages, and only 7.5% of women. More work for wages at some point in the year, but wage work is not the norm.

2.2 Estimating cohort returns to education

The consumption equation (1) allows for ability, family background and age to have level effects. It is well known that such an equation will give estimates of the true returns β only if the education choice is unrelated to the unobservables (Card, 1999). The family background bias is likely to be a particularly large problem in India. Some castes which traditionally placed great value on education may also be at the top of the consumption distribution, and so an individual regression such as equation (1) will mistake high consumption accompanied by high education for high returns to education. Credit constraints may also be strongly related to family background. On the other hand, if schooling is measured with error, attenuation bias may tend to push $E[\hat{\beta}]$ downwards, so that the estimates could be too high. More importantly, the parameter of policy interest $\beta + \gamma$ is not identified at the individual level. While the individual returns are useful for an individual making investment decisions, policy should focus on the social returns.

Aggregation solves both problems under some reasonable assumptions. Along with the consumption and education of each individual suppose we also know the state where the individual resides and her age. Then it is possible to divide observations into age cohorts within geographic areas. Discarding all individuals not in these cohorts, take the mean over individuals of the reduced form consumption equation. Then in state s, for cohort τ which is the year of birth of the oldest member, measured in year t:

$$(\ln c)_{\tau st} = z_{\tau s} + b_{\tau s} + \theta_{\tau t} + (\beta + \gamma)S_{\tau s} + h_{t-\tau} + u_{\tau st}. \tag{2}$$

Here $h_{t-\tau}$ is the average of the age function over the cohort, with $t-\tau$ being the age, and a subscript of τ indicates the average value for that cohort. The average is taken over the log consumption, not the log of average consumption. Equation (2) is not identified, since there are cohort, age, and year effects included, so some additional assumptions are necessary. It seems reasonable to assume that the average ability of adjacent cohorts in the same state is the same. Similarly, the family background on average in the state is not likely to have changed between cohorts. Finally,

if shocks on an aggregate level affect cohorts the same way, then $\theta_{\tau t}$ becomes θ_t. Then it is possible to combine cohort effects and create a single state fixed effect θ_s. The basic estimating equation becomes:

$$\ln c_{\tau st} = \theta_s + \theta_t + (\beta + \gamma)S_{\tau st} + h_{t-\tau} + u_{\tau st}. \tag{3}$$

By comparing the consumption of cohorts, it is possible to to estimate the social returns without the biases of the individual estimates.

As shown in Moffitt (1993) and discussed in Verbeek (2005), using pseudo-panels is econometrically the same as using age indicators as instruments. Using regions and age cohorts is equivalent to using the interaction of age and region indicators as instruments. Duflo (2001), for example, uses a school building program in Indonesia which increased the number of years of education for some cohorts in some district more than others as an instrument to estimate returns to education. The identification in Duflo (2001) comes from the variation in education by district and age, and the assumption that there was no underlying variation between cohorts in the unobservables associated with the variation (see also Lleras-Muney (2005) for a similar approach using state drop out age laws in the United States). The identification in this paper rests on a similar assumption, but does not rely on a single program, so it is not as easy to rule out other contemporaneous changes. If other programs within states affected cohorts differently and at the same time as increases in education, these may bias results. Such programs must be cohort specific, however. For example, a program to build health clinics that affects all cohorts is not a problem. To the extent such cohort specific changes are an issue. one might expect them to be positively correlated with education. That will tend to produce an upward bias suggesting that the estimated returns are too high.

Measurement error is a problem both at the individual and aggregate level, but aggregation provides a way to correct the estimates. By the time the surveys are divided by sex, state, and cohort, there may not be a large number of observations per cohort, leading to error prone estimates of the true population cohort means. Substituting survey cohort means for the true population cohort mean will then tend to create a bias, even if the variables on the individual level are measured without error. As proposed by Deaton (1985) and Fuller (1987), it is possible to use measurement

error models to correct for this bias. The correction is possible because although the sample mean is an error ridden estimate of the population mean, with an estimate of the sample variance it is possible to remove the added variation from the measurement error and so estimate without bias. Moreover, this approach also deals with any effects from individual measurement error, since this will be included in the estimates of the cohort errors. I implement the two stage method of moments estimator in Fuller (1987, Section 3.1.2). The major complication is that because the surveys differ in size each cohort-year observation must have its own variance and covariance with all of the other variables. That substantially complicates the data management, programming, and estimation.

Two additional problems arise using using household consumption to consider education and its effect on individual welfare across age groups. The first problem is that the size and composition of the household are not fixed. Households with many children may consume substantially less than a household with the same number of adults, and there may be economies of scale for larger households.

As is clear from figure 1 there are fluctuations over the life cycle in per capita expenditure with a dip in the late 20s for men and earlier for women. The dip come from changes in household composition that are not fully captured by per capita expenditure. The life-cycle variation in household composition can be seen in figures 3 and 4 which show how the composition of the household changes with the age of an individual.

To deal with the household composition and other effects of age, I show results from four different methods that deal with household composition in slightly different ways. First, I always use per capita expenditure and include a flexible function of age. The age function captures the problem that the average household with a 30 year old female in has a different composition than one with a 50 year old female in it and so will deal with systematic changes over ages. An age function is also necessary from equation (1) to account for other systematic changes with age such as experience. Second, I can include a direct measure of household composition such as the fraction of the household that is children in each cohort. That allows the household composition to differ by cohort and time and so capture differences in household composition over time. Third,

9

I show consumption renormalized with equivalence scales to allow children to consume less than adults. I construct a measure of log per capita equivalent expenditure using the equivalence scale that an adult woman has 90% of the consumption of an adult male, a male child under 17 has 70% and a female child need 60%. In the regressions I call this $\log PCE^a$. While somewhat arbitrary, these equivalence scales are in the mid-range of the scales used in Drèze and Srinivasan (1997) to examine widowhood and poverty in India. Fourth, in addition to equivalence scales, larger households may have economies of scale. I construct $\log PCE^b$ using the same equivalence scales but also including an economies of scale parameter of 0.8. Lanjouw and Ravallion (1995) find a parameter of 0.6 for Pakistan although this estimate is probably not precise. Drèze and Srinivasan (1997) suggest that a parameter around 0.85 may be reasonable. Varying the equivalence and scale parameters within reasonable ranges does not affect the estimates appreciably.

The second problem with using household consumption is that schooling may affect the distribution within the household as in, for example, Kooreman and Wunderink (1996, Chapter 3) who consider how higher earnings may affect bargaining power within the household. Even in there is no net change in household consumption from eduction, redistribution may shift consumption to groups with higher marginal utility within the household. The estimates will miss redistributive effects within the household—but so would estimates using income if there is some household sharing so this problem is only made more evident by using consumption.

3   The data and cohort definitions

The Indian National Sample Surveys (NSS) are nationally representative surveys which examine consumer expenditure, employment and unemployment, migration and other facets of life in India. The surveys vary in size; the large rounds, which occur approximately every five years, contain approximately 120,000 households and 600,000 individuals and the smaller rounds including around 50,000 households. The first micro level data available is from 1983.[4]

---

[4]Although most rounds span multiple years, for simplicity in figures and the text I refer to them by a single year. This paper uses the following consumer expenditure surveys: 38 (Jan. 1983 - Dec. 1983); 42 (July 1986 - June 1987);

There has been a great deal of controversy about changes in the survey methodology, particularly focused on the 55th round.[5] What is not generally recognized is that rounds other the the 55th show substantial changes as well. To deal with these changes, I use the standardized consumption data constructed by Fulford (2011) which recalculates the household consumption so that it is comparable across rounds. I adjust prices across India using Deaton (2003), and use the all India price CPI for industrial workers and agricultural laborers from the Labour Bureau to update for urban and rural areas in between and after the major rounds.

The NSS does not collect years of education except in 1983, but instead records the level of education reached: literate, primary, middle, secondary (sometimes broken into higher secondary) and post-secondary. The years of school necessary to attain each level vary somewhat by state. Appendix B details the construction of years of school from the education information in the surveys. While there do appear to be differences in the returns to different levels of education, most of the changes at the cohort level are moving from literate to having finished primary school, so at an aggregate level the cohort estimates are based on a local average mostly weighted around primary school.[6]

In order to compare survey estimates of cohorts across time, the underlying cohort population needs to be the same. Movement into and out of states must be low, so that a cohort does not change from the addition or loss of members. Migration can be an especially large problem if the best educated and most able from rural areas all leave for the city. There is very low migration between states and internationally in India so such migration does not affect the estimates.[7] While

43 (July 1987 - June 1988); 50 (July 1993 - June 1994); 51 (July 1994 - June 1995); 52 (July 1995 - June 1996); 53 (Jan. 1997 - Dec. 1997); 54 (Jan. 1998 - June 1998); 55 (July 1999 - June 2000); 56 (July 2000 - June 2001); 57 (July 2001 - June 2002); 58 (July 2002 - Dec. 2002); 59 (Jan. 2003 - Dec. 2003); 60 (Jan. 2004 - June 2004); 61 (July 2004 - June 2005); 62 (July 2005 - June 2006); 63 (July 2006 - June 2007); and 64 (July 2007 - June 2008), as well as the employment/unemployment surveys from the 38th, 50th, and 61st rounds.

[5]In particular changes in the recall period for consumption were introduced after the 50th round and two side by side questions were asked in the 55th round. These differences, and the differences in overall growth between surveys and National Accounts, have been the subject of much debate (Deaton and Kozel, 2005). In the analysis I include the 55th round, but I find that excluding it does not change the results.

[6]Allowing different education levels to have different returns in the cohort estimation does not change the results but the returns at different education levels are not always well estimated so these regressions are not shown to conserve space.

[7]The employment and unemployment survey of round 55 of the NSS, which took place from July 1999-June 2000, considers migration. 93% of men in rural areas have never moved, while 75% of men in urban areas have never moved,

urban-rural within a state migration is surprisingly low, it is not so low that it can be ignored. For this reason, I define cohorts to include both urban and rural areas within a state. Doing so limits the analysis since it would be useful to know how returns differ between urban and rural areas, but given the free movement in India and that the most educated are the most likely to move to urban areas, examining rural and urban areas separately could produce misleading results.

Like migration, if a person misreports her age, then she is classified in the wrong cohort. Age heaping at ages ending in 0 or 5 is extensive in India with approximately 25% of the population misreporting assuming that ages are distributed uniformly. This problem can be examined systematically using Myers' Blended Index (MBI) which is a useful measure of the total amount of age heaping, although certainly not the only one (Siegel and Swanson, 2004, 138). MBI is approximately the percentage of people whose ages would have to be changed to achieve an even distribution. A higher number means more age heaping, with a lower bound of 0 if each digit appears 10% of the time and an upper bound of 90 if all ages ended in the same digit.

Table 1 shows the MBI calculated for men and women in urban and rural areas and by different education groups. It is clear that age misreporting is extensive particularly in rural areas. It is also decreasing with education suggesting that there may be a problem using cohorts defined by ages that are often misreported.

When creating age cohorts, however, I average over five years which means that it is only cross-cohort measurement error that matters in this context since everyone else will be in the correct cohort, although possibly with the wrong age. Five year cohorts always contain either a 5 or a 0. Since that form of rounding is the predominant in the surveys, it should remove most of the problem.

Yet one might still be worried that there is a bias from age misreporting that is related to education. If the least educated get their ages wrong the most often, then there may be a bias in

---

the vast majority of both come from within the same state. Many more females have moved: 43% in rural and 43% in urban, but that is almost entirely on marriage and only vary rarely crosses state borders. In total, in rural area only 7% of migrants come from another state or country, and 20% from another state or country in urban areas. Since a small portion of the population has migrated in any case and most of the population is rural, only about 3% of a cohort based on a state will be incorrectly classified. This proportion varies by state and is higher in popular destination states such as Maharashtra or Delhi.

estimating the returns to education using cohorts. I show in appendix A, however, that there is no bias even from education specific age misreporting. The intuition is that since those with incorrectly recorded ages bring their education with them, the differences between cohorts decreases both in terms of consumption and education. The slope is unchanged, although it becomes less well estimated.

4 Results

4.1 Individual returns to education

Within each cohort there is a strong positive relationship that those with higher education live in households with higher consumption. Table 2 shows regressions at an individual level of log per capita household expenditure on education and state indicators for males and females. The tables only include estimates from the large rounds so the slope is estimated quite precisely; there are between ten and thirty thousand observations per cohort. The regressions use survey weights and so are population estimates. Each cell represents the slope from a separate regression, either for an individual cohort or all cohorts together. I refer to the estimates as "returns" even though they are likely to be estimated with substantial bias.

The returns show variation by cohort and year, but are generally around 4-6% for men and 5-8% for women. Pooling all cohorts in a survey year and including age indicators, the returns are approximately 5% for men and 6% for women. For both men and women the estimated return increases both with age in a year and also with the same cohort as it ages, although the gradient seems to be larger for women.

Next to the consumption returns, the table shows the slope of log daily wages with education for comparison from the three large employment/unemployment rounds. These are generally higher than the consumption returns, ranging around 4% for cohorts in their 20s to around 13% for cohorts in their 50s. These results are comparable to Duraisamy (2002) and Dutta (2004), although I do not break the sample into casual and regular wage work as suggested by Dutta (2004), and the

sample includes both urban and rural. Since so few people work for wages in India, it is not clear that the wage returns are comparable to the consumption returns which include the entire sample. Indeed, since the more educated are much more likely to work for wages in India it is not clear that regressions conditioning on working for a wage tell much about the returns to education. Both consumption and wage estimates have family background and ability biases which may be even stronger for the wage returns, since the selection into the highly paid regular wage sector may be particularly related to family background and connections.

Finally, the last rows for both men and women allow education levels to have different effects from each other. The values shown are the annual returns based on the average additional years it takes to obtain the next level of education as calculated in appendix B. The returns to literacy and secondary and post-secondary education are much higher than the marginal return to primary over literacy and middle over primary.

4.2   Cohort returns to education

Aggregating the individuals into cohorts, I can compare the cohort returns to education as they vary across states, cohorts, and time. Table 3 shows the results of estimating the returns for five year age cohorts defined by major states.[8] I exclude cohorts from smaller states because the sample sizes become quite small for these after dividing by sex and cohort. The regressions include state and year indicators, as well as an age quintic.

The returns are somewhat smaller for men than in the individual regressions: between 2.5 and 3% depending on the particular adjustment for household composition. The returns for women are very small: cohorts with better educated women do not live in households with higher per capita consumption. Moreover, the estimate is precise so it is possible to reject any returns for women. The results are nearly the same no matter which measure of consumption is used or whether the

---

[8]Major states are the 19 states with populations over 10 million in the 2001 census which excludes Himachal Pradesh and Uttaranchal (which would be kept at 6 million threshold) but includes includes Delhi and Jammu and Kashmir which a 20 million threshold used by the NSS excludes. The population covered by the major states is greater than 95%.

regressions include the fraction children.

Allowing the returns to vary after 1991 provides a slightly altered picture. For both men and women the returns appear lower after 1991. This temporal variation allows higher returns for women before 1991, with the average return over both periods still close to zero. The change after 1991 is not a result of cohort aging. Older cohorts who move past 60 years old are replaced by younger cohorts as they get old enough to be included, and so the population distribution by age is approximately constant in each year.

To check that these results are not a fluke of using measurement error models, I show the same regressions using OLS in the lower section of table 3. I do not include any measurement error corrections or heteroskedasticity corrections. The male returns are slightly higher, there is still a decline in returns after 1991, and the female returns are still insignificantly different from zero. The typical assumption that measurement error leads to attenuation bias is not applicable here. With a full covariance matrix that allows for correlated errors no single variable is necessarily biased to zero. The results are also not sensitive to the exclusion of the 55th round.

One important question is how is it possible for women and men to have different levels of per capita household consumption and different returns since they live together. The average woman in a given cohort lives with a different composition of the population than the average man in the same cohort, and better educated women need not live in the same households, on average, as less well educated women. Figures 3 and 4, for example, show that the composition of households by age of the woman differs from that of men. Across the population there are household with very different consumption and this analysis concludes that women from better educated cohorts are not systematically in higher consumption households.

The low returns are also not driven by households that mimic society by encompassing all age groups. Figure 3 shows the average age of the members of the household compared to the ages of the individuals who compose it. It is clear that the age of an individual is strongly related to the average age in a household: older people live in older households.

Aggregating one level further up, the last column of table 3 shows the results for age cohorts

defined across all of India. The cohorts include all individuals of the correct age, not just those living in major states. Defining cohorts across India removes the variation by state, and places more importance on the assumption that, after dealing with age and household composition, cohorts are the same except for their education. Controlling for measurement error is less important since the sample from each cohort is large. The returns for men are similar in size to the aggregate returns for cohorts defined by state. The returns tend to be negative for women. The pattern suggests that with all of the variation coming between cohorts of different ages, it may be difficult to differentiate between differences in education and differences in household composition and age, especially with a possible fall in returns after 1991. Nonetheless, the evidence supports the observation that men in better educated cohorts have only slight returns to education, and women from better educated cohorts do not live in households with more consumption.

4.3   Returns by state: Does quality matter?

It is also possible to allow the returns to vary by state. Table 4 shows measurement error regressions allowing different states to have their own returns to education, but maintaining the same age function and year effects across states. These results are broadly similar to the results constraining the returns to be the same: the returns are generally low for men, and zero for women.

The states generally have quite similar returns, except for Delhi which stands out with negative returns. Delhi is actually a special union territory, not a state, but in size it is comparable to a state, and it is administered separately. This special status as city, state, and national capital coupled with a higher degree of migration from the rest of the country may explain its low returns. The assumption that the cohorts measured in each survey are from the same population is the least plausible for Delhi.

Perhaps the reason returns are so low is that the quality of education is poor. Using the India Human Development Survey (IHDS, see Desai, Vanneman, and National Council of Applied Economic Research (2008)) it is possible to roughly compare schooling quality across India in 2005. Quality in 2005 may be only loosely related to education quality when the oldest cohorts

were in school, so these results can be only suggestive. The IHDS administers a simple test of reading, mathematics, and writing to all children age eight to eleven in its nationally representative survey of 41,500 households. The tests are graded into categories based on whether the child can do basic mathematics or reading tasks. Children in higher grades should have higher test scores so I measure the quality of schooling in each state by the increase that an extra year of school brings in the odds of being able to accomplish the next most complicated task using an ordered logit with state indicators and a state–grade interaction. The coefficients reported in table 4 are from the state–grade interaction and so represent for each state the increase in the log odds of going from one reading or mathematics category to the next associated with an extra year of school.

Although there is substantial variation in the value added of an additional year of education by state that variation is only weakly related to the returns. Figure 5 plots the returns for men and school quality measure against each other and shows a linear best fit. It does not appear the measured quality and the returns to education are strongly related. Along the best fit line going from the best to the worst state in terms of quality raises returns by less than 2% for reading and by less than 1% for mathematics (I exclude Delhi from both calculations). Doing the same order logit but dropping the state fixed effects and including only the state–grade interaction does not change these results—the best fit is actually slightly negative in that case.

5   Conclusion

Despite large apparent consumption returns to education estimated at the individual level, returns at the five year age cohort level show that men with an extra year of education live in households with only 3-4% more per capita consumption, and the returns appear to have decreased after 1991. The aggregate consumption returns for women are close to zero.

Do these results make sense? Aggregate returns may be below the estimates of individual returns because of bias in the individual estimates and because returns at the cohort level include general equilibrium effects at the cohort level. I also deal with possible attenuation from measure-

ment error and so remove this possible bias. Some of the low returns may be explained by low quality of education, but moving from the best to worst state in India does not produce significantly higher returns.

For both men and women, some of the difference may be explained by low returns at the individual level to levels of education between attaining literacy and finishing secondary school. Since most of the population weighted gains in education have been at the primary and middle school level (see figure 1) where even the biased individual estimates suggest low returns (see table 2), it is perhaps no surprise that the aggregate returns are quite low. While highly educated Indians have enjoyed many of the gains from India's recent growth and so may have high returns to completing a post-secondary education, most Indians have not finished middle school and so are still facing low returns.

Low returns at the cohort level may be from general equilibrium effects, but it seems likely that an important reason for the low returns may be that the Indian economy cannot make effective use of better educated cohorts in production. There may be low returns to education in the traditional agriculture or casual labor which employs most Indian men. Women in India still primarily work within the household, and there may be only limited gains in household production from education. That has the promising possibility that changes in the Indian economy, particularly in its employment of women, may yield large benefits. There may still be returns in other dimensions that are difficult to capture such as health, the education of the next generation, or redistribution within the household. Yet it seems that in India the positive benefits to consumption measured at the level that actually matters for policy are disappointingly small.

References

Azam, Mehtabul. 2012. "Changes in Wage Structure in Urban India, 1983-2004: A Quantile Regression Decomposition." *World Development* 40 (6):1135 – 1150.

Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *The Quarterly Journal of Economics* 106 (2):pp. 407–443.

Bils, Mark and Peter J. Klenow. 2000. "Does Schooling Cause Growth?" *The American Economic Review* 90 (5):pp. 1160–1183.

Card, David. 1999. "The causal effect of education on earnings." In *Handbook of Labor Economics*, edited by Orley C. Ashenfelter and David Card. Amsterdam: Elsevier Science, 1801–1863.

Card, David and Thomas Lemieux. 2001. "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis." *The Quarterly Journal of Economics* 116 (2):pp. 705–746.

Case, Anne. 2006. *Understanding Poverty*, chap. The Primacy of Education. Oxford: Oxford University Press, 269–285.

Cohen, Daniel and Marcelo Soto. 2007. "Growth and human capital: good data, good results." *Journal of Economic Growth* 12:51–76.

Deaton, Angus. 1985. "Panel data from time series of cross-sections." *Journal of Econometrics* 30.

———. 2003. "Prices and Poverty in India, 1987-2000." *Economic and Political Weekly* 38 (4):362–368.

Deaton, Angus and Valerie Kozel. 2005. "Data and Dogma: The Great Indian Poverty Debate." *The World Bank Research Observer* 20 (2):177–199.

Desai, Sonalde, Reeve Vanneman, and National Council of Applied Economic Research. 2008. "India Human Development Survey (IHDS) 2005." Inter-university Consortium for Political and Social Research (ICPSR). Accessed 12 November 2011.

Drèze, Jean and P. V. Srinivasan. 1997. "Widowhood and poverty in rural India: Some inferences from household survey data." *Journal of Development Economics* 54 (2):217–234.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *The American Economic Review* 91 (4).

Duraisamy, P. 2002. "Changes in returns to education in India, 1983-1993: by gender, age-cohort and location." *Economics of Education Review* 21:609–622.

Dutta, Puja Vaudeva. 2004. "The structure of wages in India, 1983-1999." Poverty Research Unit at Sussex: Working paper 25.

Foster, Andrew D. and Mark R. Rosenzweig. 1996. "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution." *The American Economic Review* 86 (4):pp. 931–953.

Fulford, Scott. 2011. "The effects of financial development in the short and long run." Boston College Working Paper 741.

Fuller, Wayne A. 1987. *Measurement Error Models*. New York: John Wiley and Sons.

Kijima, Yoko. 2006. "Why did wage inequality increase? Evidence from urban India 1983-99." *Journal of Development Economics* 81 (1):97–117.

Kooreman, Peter and Sophia Wunderink. 1996. *The Economics of Household Behavior*. New York: St. Martin's Press.

Lanjouw, Peter and Martin Ravallion. 1995. "Poverty and Household Size." *The Economic Journal* 105 (433):1415–1434.

Lleras-Muney, Adriana. 2005. "The Relationship Between Education and Adult Mortality in the United States." *The Review of Economic Studies* 72 (1):189–221.

Moffitt, Robert. 1993. "Identification and estimation of dynamic models with a time series of repeated cross-sections." *Journal of Econometrics* 59 (1-2).

Pritchett, Lant. 2001. "Where Has All the Education Gone?" *The World Bank Economic Review* 15 (3):pp. 367–391.

PROBE Team, The. 1999. *Public Report on Basic Education in India*. New Delhi: Oxford University Press.

Psacharopoulos, George and Harry Anthony Patrinos. 2004. "Returns to investment in education: a further update." *Education Economics* 12 (2):111–134.

Sen, Amartya. 1999. *Development as Freedom*. New York: Alfred A. Knopf.

Siegel, Jacob S. and David A. Swanson. 2004. *The Methods and Materials of Demography*. San Diego: Elsevier Academic Press.

Temple, Jonathan R.W. 2001. "Generalizations that aren't? Evidence on education and growth." *European Economic Review* 45 (4 -6):905 – 918. 15th Annual Congress of the European Economic Association.

Verbeek, Marno. 2005. "Psuedo panels and repeated cross-sections." Prepared for: Matyas and Sevestre, The Econometrics of Panel Data: Handbook of the Theory with Applications, 3rd edition, Kluwer Academic, 2006.

A Age misreporting and bias

This section shows that both age misreporting will not cause bias in estimates of the returns to schooling measured across cohorts as long as all cohorts suffer from the same kinds of problems. In passing, it also demonstrates that the standard ability or family background biases of so much concern are not a problem when estimating returns using cohorts.

To calculate returns from cohorts, different age groups must have different levels of education. The simple model which is a version of the approach in Card (1999) assumes this occurs because of changes in costs to acquiring education between cohorts such as from school building programs. Suppose that consumption is a simplified version of equation 1 given by $\ln c_i = z_i + \beta S_i$ and each individual has an observed age $\hat{a}_i$ and a true age a_i. Obtaining education is costly and reduces utility. Students choose the level of their education to maximize $\ln c - (h(a_i) - z_i)S - S^2/2$ where $h(\cdot)$ is a function of the true age, which I will assume is constant within each cohort. We could think of a h as a school building program that only affects younger cohorts: more able children have lower costs in both both cohorts, and children with more schools built for their cohort have lower costs. Then each student chooses $S_i = \beta + b_i - h(a_i)$, and $\ln c_i = (1 + \beta)b_i + \beta^2 - \beta h(a_i)$. There is some distribution of z_i which is the same across cohorts and has mean z, and each cohort has a unit mass of individuals. Since $h(a_i)$ changes with age, this simple model explains why different cohorts have different levels of education on average. Note that the standard problems of estimating the returns to schooling apply here: Since S_i is a function of z_i, an estimate of β which does not account for z_i will be biased.

First, assume that the observed ages are the true ages and consider an older cohort A and a younger cohort B. Then the difference in schooling is $S_A - S_B = h_B - h_A$ where h_A is the value of $h(a_i)|i \in$ cohort A. The difference in consumption, after taking the cohort mean and subtracting, is $\ln c_A - \ln c_B = \beta(h_B - h_A)$, and so estimating the return on schooling using the change in earnings over the change in schooling gives:

$$\hat{\beta} = \frac{\beta(h_B - h_A)}{h_B - h_A} = \beta$$

so that the true returns to schooling are captured by the average cohort returns to schooling. Using cohorts removes the ability bias or family background bias.

Now suppose that there is some error in age reporting: some fraction λ of the A cohort underreport their ages, but that who underreports is random and unrelated to b_i. Then observed schooling in the B cohort is $\hat{S}_B = (1/(1+\lambda))((\beta - h_B + b) + \lambda(\beta - h_A + b))$ and $\hat{S}_A = S_A$. Similarly, $\ln \hat{c}_B = (1/(1+\lambda))(\ln c_B + \lambda \ln c_A)$ and $ln\hat{c}_A = lnc_A$. Then calculating the returns to schooling and combining similar terms:

$$\hat{\beta} = \frac{ln\hat{c}_A - \ln \hat{c}_B}{\hat{S}_A - \hat{S}_B} = \frac{c_A - c_B}{S_A - S_B} = \beta$$

since $(\ln \hat{c}_A - \ln \hat{c}_B) = \ln c_A - (1/(1+\lambda))(\ln c_B + \lambda \ln c_A) = \ln c_A - \ln c_B$. So random reassignment does not change the estimated returns to schooling.

This same argument applies to overestimated age errors, and twoway misreporting in which people randomly report their ages as too old or too young. With flow both ways, however, it is possible to equalize the observed values of the two cohorts, leaving the estimated $\hat{\beta}$ undefined. Even if they are not exactly the same, the closer $\hat{S}_A - \hat{S}_B$ gets to zero, the weaker the estimation will be. The major problem from age misreporting is not bias, but that it reduces variability by pulling the two populations closer together.

It is unlikely that there is more than a small fraction of people who cross cohorts defined over five years. What is of greater concern is that those who misreport may be different in some way than those who do not. Suppose that in each cohort there are some good $z_i = b_h$ and some bad $z_i = b_l$ abilities where $z_h > z_l$ by assumption, and they are divided such that $b = (1-\nu)b_h + \nu b_l$ and this division is the same across cohorts. Ability does not change on average. But who misreports is dependent on ability: some fraction λ of the bad z_l's from each cohort incorrectly under-report their ages and so are included in the following cohort, but that none of the good b_h misreport. The situation is then analogous to table 1: since the b_h also have more education, age misreporting and education are linked.

This situation of education dependent misreporting does not lead to bias since each cohort will gain and lose exactly the same type of members. The people misreporting will tend to have different levels of h, but there will be no bias since the types cancel. To see this, suppose that there exists an older cohort C which loses members to A, and a younger cohort D to which B loses members. D and C are excluded from the analysis, say because they are too young or too old. Then the type and h of each observed age cohort is $\hat{h}_A = (1 - \nu)h_A + \nu(1 - \lambda)h_A + \lambda\nu h_C$, $\hat{b}_A = (1 - \nu)b_h + \nu(1 - \lambda)b_l + \lambda\nu b_l$, $\hat{h}_B = (1 - \nu)h_B + \nu(1 - \lambda)h_B + \lambda\nu h_A$, and $\hat{b}_B = (1 - \nu)b_h + \nu(1 - \lambda)b_l + \lambda\nu b_l$. The important point is that $\hat{b}_B = \hat{b}_A$, since each cohort has the same mix from the other cohort above it and loses the same to the cohort below. Calculating the returns to education as:

$$\hat{\beta} = \frac{\ln \hat{c}_A - \ln \hat{c}_B}{\hat{S}_A - \hat{S}_B} = \beta$$

and there is no bias. The same argument works for upward age misreporting and for more complicated distributions.

The model and evidence in this section suggest that bias from age misreporting is not likely to be a problem. First, by taking five year age cohorts which always contain a 0 or 5, most age misreporting takes place within a cohort and so is irrelevant. Second, even when there is age heaping, it will not cause a bias even when it is systematically linked to education as long as the relationship holds for all cohorts.

B Calculating years of education

The NSS only asks directly about the number of years of education in 1983. In all other years it instead asks about educational attainment. The exact survey options have changed slightly over the years as the prevalence of certain adult literacy programs has come and gone and as new graduate programs have been added or subtracted. These make almost no difference to the totals since they affect very few people and I subsume them into larger categories. One change that is important is the differentiation between higher secondary and secondary or lower secondary school. This

differentiation appears in the data for the first time in the 50th round. Before that if a person had completed either secondary or higher secondary school they were counted as having had secondary education. To keep the comparison, I continue this definition, so that secondary school throughout refers to having completed either secondary or higher secondary school.

According to the survey instructions the criteria for finishing each level depends on the particular state. To calculate the years of education, I assign each category the number of years it takes to reach that level of education in the state. In most states, for example, someone finishing middle school will have had at least 8 years of formal education. They may have even have had a year of secondary school, but have not had two otherwise they would have finished (lower) secondary. The 38th round is the only round to ask about years of education as well as level (see Dutta (2004) for a table showing the years calculated from the 38th). According to the Government of India's Department of School Education and Literacy, in most states primary school is five years and middle is another three, while all states follow the standard 10 years for secondary plus two for higher secondary.[9] These numbers are close to the estimated results from the 38th round, with secondary a mixture of secondary and higher secondary. However, the NSS instructions are to ask about the level of education as established by the states, and there is some diversity in how the states define primary and secondary. Some states have primary lasting four years, three for middle, and another three for secondary (this is the structure for for Assam, Goa, Gujarat, Karnataka, Kerala, Maharashtra, Meghalaya, Mizoram, Dadra & Nagar Haveli, Daman & Diu, Lakshadweep according to the Department of School Education and Literacy). In Andhra Pradesh and Orissa, middle school is for only two years. According to the Department, all states define secondary as being the 10th year, and higher secondary the 12th year. These figures are mostly consistent with the survey evidence in 1983 by state, but not entirely.

To deal with these differences by state, I use the survey evidence from the 38th round to give the number of years corresponding to primary and secondary education by state for all rounds. This approach has the advantage of also allowing reasonable estimates of the number of years over

---

[9]See http://education.nic.in/secedu.asp, (accessed 5 September 2007)

a particular level the average person has. For example, some people who finished primary had an extra year of education which will be counted in the average. 1983 is approximately in the middle of the cohorts considered, although it is the earliest survey. Since no other surveys ask for years it is not possible to compare the changes in years and levels over time.

For illiterate, literate but below primary, and post secondary, I assign the years of education from 1983 calculated using the full survey over all states. These do not change much based between states, and are not affected by definitional changes, and so it is better to estimate them more precisely. For rounds which split secondary and higher secondary I assign these categories 10 and 12 years, while for rounds which do not I use the estimated secondary years from the 38th round.
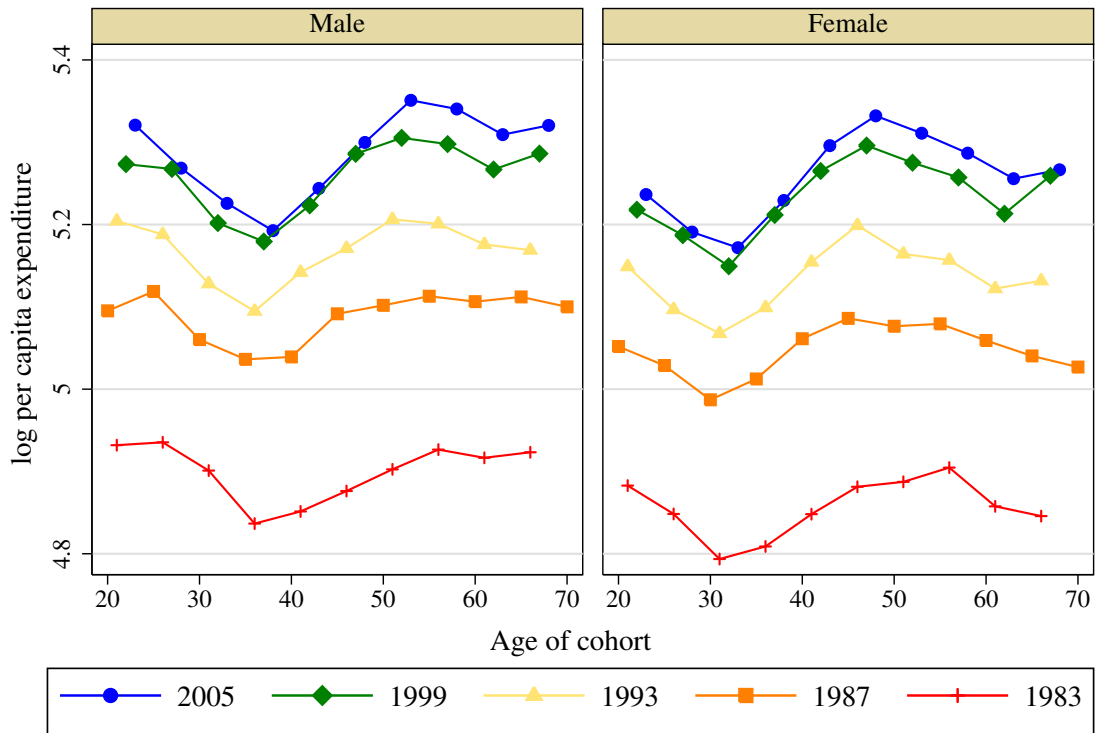
Figure 1: Education by cohort birth year
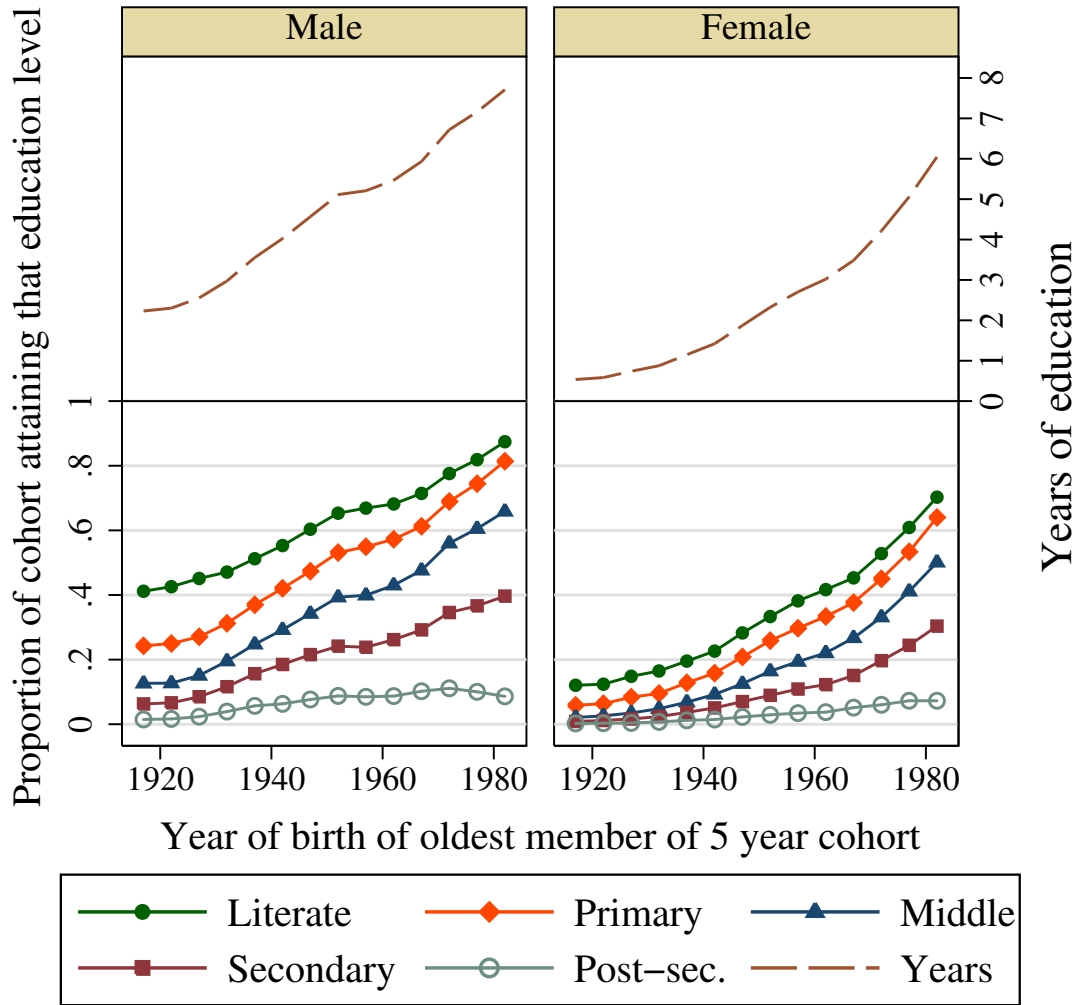
Notes: Age cohorts are five years wide and are labeled by the oldest member. The years of education is calculated from the level of education, see the appendix. The cohort education is the mean for each cohort calculated from all surveys in which the cohort would have finished most of its education (the oldest member of the cohort was 21 or older), without accounting for survey size. The oldest and youngest cohorts are thus less well estimated, and the youngest cohort may not have completed all of its post-secondary education. Secondary includes both secondary and higher secondary. Source: National Sample Surveys, various years.

Figure 2: Log per capita household consumption by cohort age and year

Figure 3: Mean household age and individual age
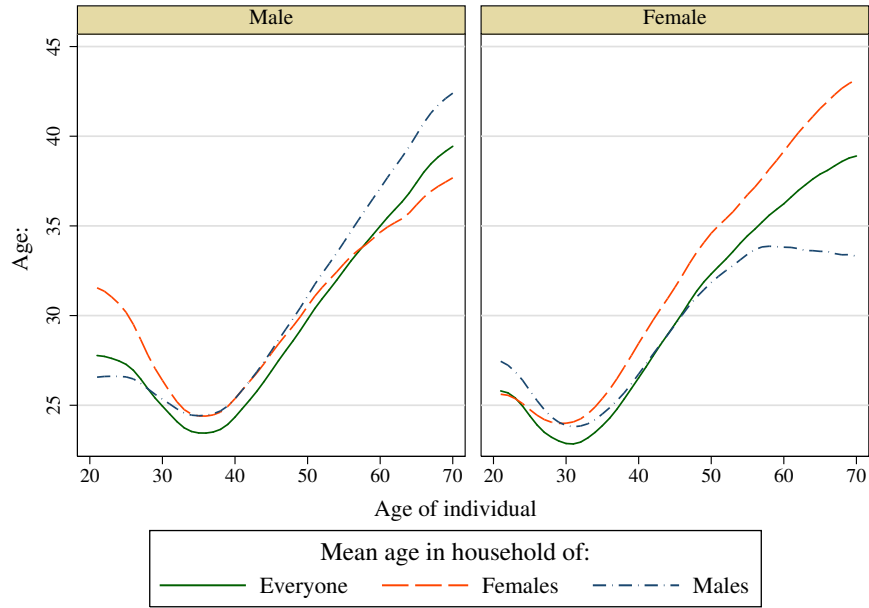
Notes: From all rounds of the NSS with a polynomial smoother. The mean age is the mean age of members in the household that a man or woman is member of including that individual where appropriate.

Figure 4: Household composition and individual age
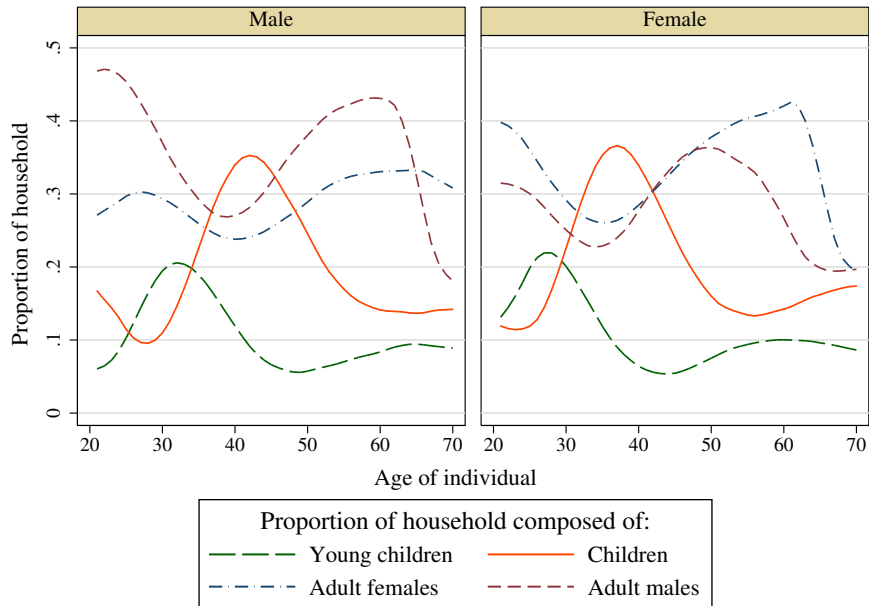
Notes: From all rounds of the NSS with a polynomial smoother. The vertical axis is the fraction of that demographic group in the household that a man or woman is member of including that individual where appropriate. Young children are age six and under. Children are ages 7-16. Everyone else is an adult. Uses a polynomial smoother.

Figure 5: Returns to education and education quality

Reading

State education quality (reading)

Mathematics

State education quality (mathematics)
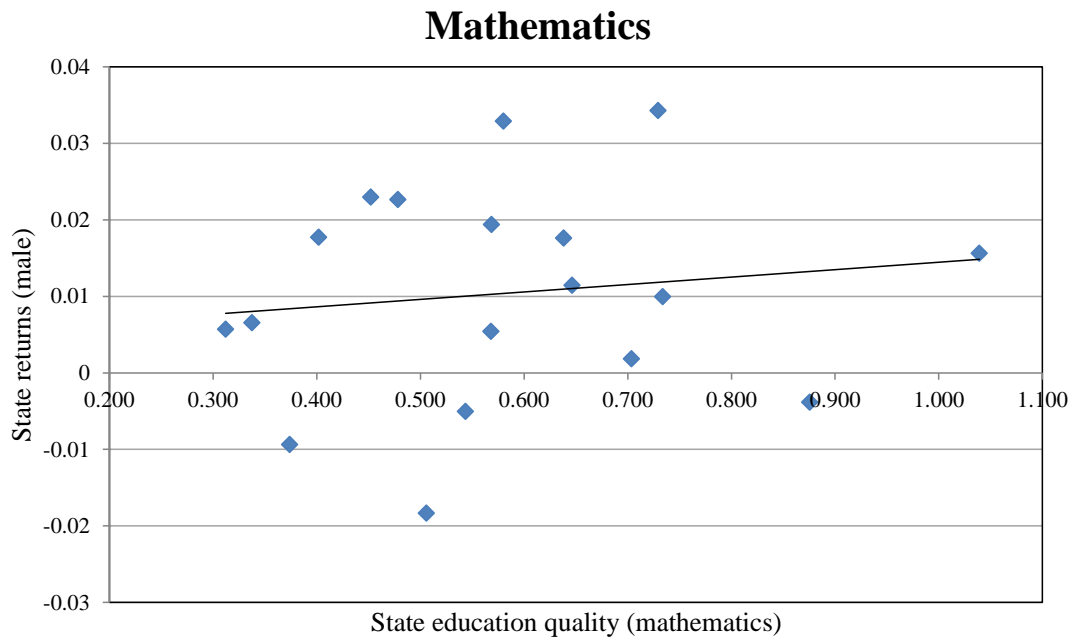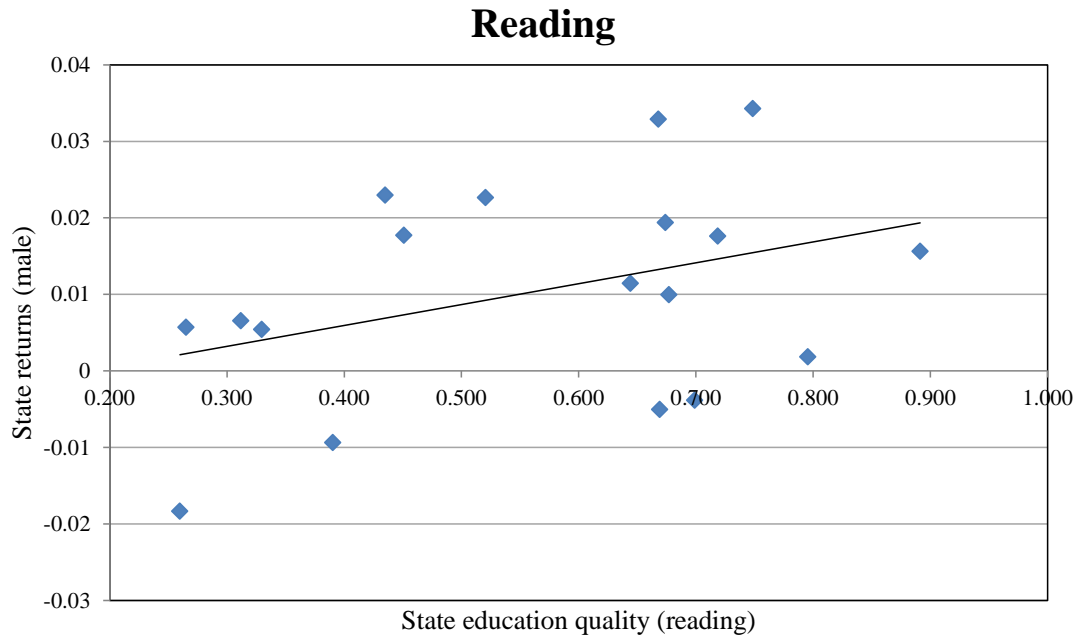
Notes: Shows the returns to education by state from table 4 and the estimated quality of education as measured by the increase that a year of education brings to reading and math tests in an ordered logit using the IHDS (Desai, Vanneman, and National Council of Applied Economic Research, 2008). The returns and best fit line exclude Delhi.

Table 1: Age heaping in India: Myers' Blended Index

	Rural Male	Rural Female	Urban Male	Urban Female
Overall	24.4	23.3	19.1	19.3
By Education				
Not literate	33.2	28.2	33.0	28.9
Literate	27.6	23.3	27.7	24.8
Primary	26.1	22.0	23.6	19.6
Middle	23.4	19.6	20.4	18.2
Secondary	19.7	15.2	16.6	13.4

Notes: From the 60th round of the NSS by education and the 55th overall, calculated with the appropriate multiplier. The informant is the person interviewed in the survey. The age range is 16-65. Myers' blended index is the half the sum of absolute deviations from 10% of the proportion of final digits summed from a rolling starting point. The index is thus, approximately, the percentage of observations that are miscoded from age heaping. Uses code written by G. Rodríguez and T. Pullum., (http://data.princeton.edu/eco572/digitpref.html).

Table 2: Individual relationship of education and consumption by cohort

Cohort	Age in 1983	Log(Per Capita Consumption)					Log(Day Wage)		
		1983	1987	1993	1999	2005	1983	1993	2005
Male									
1977 - 1981	6				0.040	0.046			0.043
1972 - 1976	11			0.040	0.040	0.049		0.012	0.063
1967 - 1971	16			0.037	0.042	0.048		0.036	0.074
1962 - 1966	21	0.044	0.042	0.046	0.047	0.052	0.027	0.064	0.094
1957 - 1961	26	0.045	0.048	0.047	0.050	0.059	0.063	0.078	0.105
1952 - 1956	31	0.051	0.051	0.051	0.054	0.062	0.081	0.088	0.124
1947 - 1951	36	0.053	0.053	0.052	0.054	0.061	0.095	0.099	0.131
1942 - 1946	41	0.056	0.056	0.054	0.055		0.108	0.117	
1937 - 1941	46	0.058	0.059	0.059			0.112	0.114	
1932 - 1936	51	0.065	0.063				0.123		
All cohorts		0.051	0.051	0.046	0.046	0.052	0.088	0.074	0.076
All cohorts education level									
Literate		0.059	0.063	0.057	0.049	0.043			
Primary		0.026	0.025	0.014	0.017	0.020			
Middle		0.044	0.036	0.033	0.026	0.034			
Secondary		0.077	0.076	0.071	0.068	0.072			
Post Sec.		0.070	0.077	0.077	0.087	0.098			
Female									
1977 - 1981	6				0.044	0.054			0.061
1972 - 1976	11			0.045	0.051	0.056		0.024	0.068
1967 - 1971	16			0.051	0.055	0.063		0.061	0.080
1962 - 1966	21	0.051	0.054	0.058	0.061	0.069	0.039	0.072	0.101
1957 - 1961	26	0.060	0.061	0.061	0.060	0.072	0.089	0.092	0.109
1952 - 1956	31	0.065	0.064	0.063	0.069	0.078	0.108	0.102	0.139
1947 - 1951	36	0.072	0.070	0.069	0.075	0.079	0.114	0.116	0.149
1942 - 1946	41	0.077	0.077	0.073	0.080		0.125	0.135	
1937 - 1941	46	0.084	0.077	0.079			0.127	0.138	
1932 - 1936	51	0.084	0.085				0.142		
All cohorts		0.064	0.063	0.056	0.055	0.060	0.102	0.081	0.079
All cohorts education level									
Literate		0.103	0.107	0.093	0.080	0.077			
Primary		0.035	0.033	0.021	0.024	0.025			
Middle		0.048	0.043	0.041	0.037	0.040			
Secondary		0.082	0.077	0.071	0.074	0.086			
Post Sec.		0.058	0.070	0.084	0.090	0.100			

Note: Each value is the result of estimating $log(c_{it}) = \alpha + I_s + \beta$Years of education $+ u_i$ where I_s is an indicator for the state. Each cohort regression uses only individuals from that particular cohort in that year. The all cohorts regressions use members of all cohorts with maximum age between 22 and 60 and includes an indicator for cohort age. The years listed are the major rounds of the NSS, and so cohort sample sizes are approximately 19,000 in each cell for consumption and an average 6,900 per cell for men and 2,300 for women for the daily wage. All estimates are statistically significant at a high level of confidence (not reported). The education returns at each education level are adjusted for the number of additional years required to attain that level.

Table 3: Cohort returns to education

Cohort definition:			Major States			All India
Dependent Variable:	$\ln(PCE)$	$\ln(PCE)$	$\ln(PCE)$	$\ln(PCE^a)$	$\ln(PCE^b)$	$\ln(PCE)$
Estimator: Measurement error two stage method of moments estimator						
			Male			
Years Education	0.025**	0.031**	0.032**	0.026**	0.028**	0.040**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)
Yrs Ed * Post 1991		−0.009**				
		(0.002)				
			Female			
Years Education	−0.0014	0.0076	0.0004	−0.0030	−0.0041	−0.0288**
	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.006)
Yrs Ed * Post 1991		−0.0061*				
		(0.003)				
	(0.088)	(0.088)				
Estimator: Least squares						
			Male			
Years Education	0.035**	0.041**	0.041**	0.036**	0.041**	0.043**
	(0.036)	(0.005)	(0.005)	(0.005)	(0.005)	(0.012)
Yrs Ed * Post 1991		−0.010**				
		(0.003)				
			Female			
Years Education	0.0026	0.0208**	0.0058	0.0021	0.0016	−0.0243**
	(0.002)	(0.006)	(0.004)	(0.004)	(0.004)	(0.009)
Yrs Ed * Post 1991		−0.0130**				
		(0.003)				
Frac. young children	Yes	Yes	No	No	No	No
Frac. children	Yes	Yes	No	No	No	No
State Indicators	Yes	Yes	Yes	Yes	Yes	No
Age quintic	Yes	Yes	Yes	Yes	Yes	Yes
N	1216	1216	1216	1216	1216	80

Note: Standard errors are in parentheses. * indicates t-stat ≥ 1.96 and ** t-stat ≥ 2.33. The cohorts defined on major states include an age quintic, and state and time dummies. The last column does not divide cohorts across states and so only uses age variation. $\ln(PCE)$ is the cohort mean (adjusting for survey methodology) log per capita household consumption measured in constant rupees; $\ln(PCE^a)$ adjusts for household composition using equivalence scales; $\ln(PCE^b)$ adjusts for equivalence and economies of scale. The Frac. young children is the fraction of household members who are less than 6 years old; the Frac. children are those aged 7-16. The measurement error model is from Fuller (1987, Section 3.1.2).

Table 4: Cohort returns to education by state

| Dependent Variable: | log(per capita expenditure) | | | | Test Scores in 2005 | |
	Male		Female		Reading	Math
Cohort Years Ed *						
Jammu & Kashmir	0.002	(0.0059)	−0.001	(0.0066)	0.312	0.338
Punjab	−0.001	(0.0083)	−0.006	(0.0063)	0.669	0.544
Haryana	0.009	(0.0086)	−0.012	(0.0069)	0.677	0.734
Delhi	−0.123**	(0.0299)	−0.067**	(0.0095)	0.758	0.707
Rajasthan	0.020**	(0.0066)	0.003	(0.0135)	0.674	0.569
Uttar Pradesh	0.032**	(0.0073)	0.009	(0.0098)	0.668	0.580
Bihar	0.033**	(0.0088)	0.005	(0.0138)	0.749	0.729
Assam	−0.017*	(0.0087)	−0.007	(0.0077)	0.260	0.506
West Bengal	−0.010	(0.0150)	−0.009	(0.0113)	0.699	0.875
Jharkhand	0.017*	(0.0083)	0.006	(0.0118)	0.891	1.039
Orissa	−0.014*	(0.0072)	−0.025**	(0.0088)	0.390	0.374
Chhattisgarh	0.000	(0.0071)	−0.013	(0.0089)	0.796	0.704
Madhya Pradesh	0.009	(0.0078)	−0.018	(0.0111)	0.644	0.646
Gujarat	0.018**	(0.0074)	0.003	(0.0076)	0.719	0.638
Maharashtra	0.024**	(0.0064)	0.011	(0.0064)	0.435	0.452
Andhra Pradesh	0.020**	(0.0070)	0.011	(0.0094)	0.451	0.402
Karnataka	0.001	(0.0075)	−0.002	(0.0077)	0.265	0.312
Kerala	0.006	(0.0073)	−0.003	(0.0057)	0.330	0.568
Tamil Nadu	0.022**	(0.0072)	0.010	(0.0064)	0.521	0.478
Frac. children	Yes		Yes			
Year indicators	Yes		Yes			
State indicators	Yes		Yes		Yes	Yes
Age quintic	Yes		Yes			
Obs.	1216		1216		11625	11584

Note: Standard errors are in parentheses. * indicates t-stat $\geq$ 1.96 and ** t-stat $\geq$ 2.33. $\log(PCE)$ is the cohort mean (adjusting for survey methodology) log per capita household consumption measured in constant rupees. Each regression includes the fraction young children, fraction children, an age quintic, and state indicators (coefficients not shown). Education is cohort education in years as constructed in the appendix. The included states are the states with populations greater than 10 million in the 2001 census. The log per capita estimation is done using the measurement error model of Fuller (1987, Section 3.1.2) and includes all of the covariances of the interactions with $\log(PCE)$ and the fraction children and young children. The benefits that an extra year of education in each state brings to the test is based on an ordered logit with state indicators and a state–years of school interaction using survey weighting from the IHDS (Desai, Vanneman, and National Council of Applied Economic Research, 2008). The coefficients reported are from the state–years of education interaction. The Frac. young children is the fraction of household members who are less than 6 years old; the Frac. children are those aged 7-16.