Schooling, Wages, and the Role of Unobserved Ability in the Philippines

Yaraslau Zayats[§]

Department of Economics
University of North Carolina at Chapel Hill

Abstract

The paper analyzes the impacts of an individual's unobserved ability on schooling and wages in the context of a developing country using rich data from the Cebu (Philippines) Longitudinal Health and Nutrition Survey. Unlike any previous study, my model allows for grade repetition and school reentry after dropping out of school. Both phenomena are common in developing countries in general, and in the Philippines in particular. Semiparametric approach is used to control for an individual's unobserved ability. The results strongly indicate that children with lower innate ability enter school at a later age and complete fewer years of school. They are also more likely to drop out of school at all levels of education, but the effect of lower ability diminishes at higher levels of education. While a standard Mincer equation yields a 4.5 percentage point return to an additional year of schooling, my model estimates this return to be only 2.7 percentage points. An omitted ability bias appears substantial. While completing additional years of schooling can compensate for lower innate ability, such substitution would be costly. It would take about three additional years of education to compensate for one standard deviation lower innate ability in terms of labor market returns. Improving school quality appears to increase achievement test scores only a bit, and lower pupil-teacher ratios yield only slightly higher rates of elementary school completion. Higher family income appears to benefit both attendance and completion of elementary school, but these effects are quite small despite being very precisely estimated.

[§] I thank Thomas Mroz, David Blau, and Wilbert Van Der Klaauw for their insightful comments and suggestions, and to Linda Adair and Connie Gultiano for kindly answering my questions regarding the data. I also thank Paul Schultz, John Tyler and other participants of the NEUDC 2005 for valuable comments. All errors are mine.

1. Introduction

The primary challenge in studying the effect of education on wages is the fact that more able individuals choose more education. If an individual's ability is poorly controlled for by the measured variables, it is possible that the more educated individuals would have received higher wages even without their additional schooling. The measured effects of schooling on wages, therefore, potentially incorporate the effects of ability on wages, giving rise to what is called ability bias in the returns to schooling. Economists have used multiple approaches to resolve it. However, it remains one of the most challenging identification problems in empirical research. Recently, concerns have been raised in the literature regarding the magnitude of the bias, pointing toward the need for more flexible estimation techniques and better controls for unobserved ability. While the accumulated evidence on the significance of ability bias in the estimated returns to schooling in the United States is quite impressive, few studies for developing countries have addressed this issue directly.

My analysis aims to fill that void in the literature. I analyze the impacts of an individual's unobserved ability on schooling and wages in the context of a developing country. Using data from the Cebu Longitudinal Health and Nutrition Survey, from the Philippines, I try to answer the following questions that are crucial for public policy in developing countries. Are low-ability individuals more likely to drop out of school than people with higher ability? If so, what can be done to keep the low-ability dropouts in school longer? More importantly, would this additional schooling benefit individuals in the labor market? In other words, do we see a significant return to schooling when we look at their wages? Does this return differ by an individual's ability?

Numerous questions to which this study seeks to find answers are potentially relevant for many other developing countries. The Philippines, and the Cebu region in particular, have been undergoing a rapid transition from agriculture and low-skill manufacturing to a service and technology oriented economy during the last twenty years. This is the type of transition that one can expect many other developing countries to go through in the next few decades.

I use an economic model of schooling, test scores, and wages. I model both school attendance and school completion for each school year. I allow for grade repeats and school reentry after dropping out of school. Both phenomena are common in the Philippines in particular. None of the previous studies addressed the problem of grade repeats and school reentry at the individual level. I model cognitive achievement test scores similar to the analysis by Hansen, Heckman, and Mullen (2004). The relationships among these sets of outcomes provide a semiparametric identification of the unobservable “ability.” The inclusion of a key unobserved factor as a determinant of cognitive achievement test scores and IQ test scores provides a reason to label the unobserved factor as “ability.” It is important to note, however, that “ability” as used in the paper only refers to those unobserved characteristics that impact each of the modeled outcomes.

The results strongly indicate that children with lower innate ability enter school at a later age and complete fewer years of school. They are also more likely to drop out of school at all levels of education, but the effect of lower ability diminishes at higher levels of education. While a standard Mincer equation yields a 4.5 percentage point return to an additional year of schooling, my model estimates this return to be only 2.7 percentage points. The omitted ability bias appears substantial. While completing additional years of schooling can compensate for lower innate ability, such substitution would be costly. It would take about three additional years of education to compensate for one standard deviation lower innate ability in terms of labor market returns.

The next section briefly discusses the existing literature. Section 3 describes the data. Section 4 poses major questions of interest. Section 5 presents the model. Section 6 discusses the results. Section 7 concludes.

2. Literature

Ability bias represents one of the oldest problems in labor economics. The literature dedicated to this issue is voluminous, especially as it affects estimates of the returns to schooling. The approaches used in the literature to remove the ability bias can be classified into several groups. One approach dates back to Griliches and Mason (1972) and involves the use of available measures of ability as proxies for unobserved ability that is rewarded in the labor market. Including such measures in the regression should mitigate the endogeneity of schooling, but not completely eliminate it as long as the measures of ability are not perfect proxies. Empirically, the estimated return to schooling is generally reduced when unobserved ability is proxied. For example, Blackburn and Neumark (1995) find that the usual OLS estimates of the return to schooling, with proxies for ability omitted, are upward biased by roughly 40%.

A second approach uses the differences across siblings in levels of schooling and wages, relying on the assumption that much of the unobserved ability is common across siblings and is consequently differenced out.[1] The relevant studies include Behrman and Taubman (1976), Ashenfelter and Krueger (1994), Ashenfelter and Rouse (1998), and Behrman and Rosenzweig (1999), to name a few. The within-twins estimators generally indicate an upward bias in the OLS estimates if ability is ignored, but differ significantly in the magnitude of the bias. Bound and Solon (1999) and Neumark (1999), however, argue that between-twins differences in schooling are

[1] Based on this assumption, comparing monozygotic twins is even better since they share identical genetic endowments.

not random, but are chosen endogenously. As a result, between-twins estimates of the return to schooling might suffer from the endogeneity biases similar to conventional cross-sectional estimates. Moreover, differencing between twins wipes out much of the exogenous variation and inevitably exacerbates the measurement error problem (Griliches 1979).

A third approach exploits natural variation in determinants of schooling decisions, such as the interactions between quarter of birth and compulsory schooling laws, to create valid instruments for schooling as in Angrist and Krueger (1991, 1992). This approach tends to find at best no omitted-ability bias in the estimated returns to schooling.[2] Bound, Jaeger, and Baker (1995) show, however, that Angrist and Krueger's estimates may suffer from finite-sample bias that arises from weak correlation between quarter of birth and schooling. Rosenzweig and Wolpin (2000) discuss natural experiments in great detail and analyze a variety of recently used instruments that are based on natural experiments. The authors point out an extraordinary range of estimates across the studies that use instruments based on natural experiments. They argue that, in the presence of heterogeneity in returns to schooling, instruments identify local average treatment effects (Imbens and Angrist 1994), that is, the effects for the group or groups whose behavior is influenced by intervention, and different instruments affect different groups of people. Using a very simple model of schooling choice, they show that the date-of-birth (as in Angrist and Krueger 1991) and child-gender (as in Butcher and Case 1994) instruments identify the returns to schooling for different ability groups in the population. A similar concern but from a different perspective is expressed by Card (2001).

Another group of methods involves semiparametric and nonparametric estimation techniques for tackling the problem of ability bias. For instance, Belzil and Hansen (2002) use a

[2] They either find no significant changes in the estimates or a *negative* bias in the OLS estimates. A negative bias in the OLS estimate would indicate a presence of a measurement error in schooling rather than omitted-ability bias (omitted-ability bias is expected to have a *positive* bias in the OLS estimates).

panel of white males from the NLSY and estimate a structural dynamic programming model of schooling decisions with unobserved heterogeneity in both school ability and market ability, in which the wage regression is estimated using splines. The results cast doubt on the validity of the high returns to education reported in the literature. Contrary to conventional wisdom (Card 1999), the log wage regression is found to be convex in schooling.[3] A linear wage regression appears to be severely misspecified. The analysis strongly rejects the hypothesis of orthogonality between market ability and realized schooling and indicates the existence of a positive ability bias.

Essential to our analysis is the study by Hansen, Heckman and Mullen (2004). One dimension of the study is a semiparametric model that the authors develop for estimating the effect of schooling on achievement test scores. Assuming that a person's *latent* ability cannot be affected by schooling, the authors test whether *manifest* ability, as measured by ASVAB achievement tests, is affected by schooling when both schooling and manifest ability are allowed to be affected by latent ability. The schooling and test score equations are related only via what they call unobserved innate ability. The authors prove nonparametric identification of the distribution of latent ability. The model provides a flexible way of estimating an individual's unobserved ability.

The literature on the returns to schooling and ability bias in the context of developing countries deserves a separate discussion. If in the United States a private return to education is in the range of 5-12 percent (Burtless 1996), in developing countries this return is found to be generally much higher. Psacharopoulos (1994) reports the average private return to education in developing countries to be 29 percent for primary education, 18 percent for secondary education, and 20 percent for higher education. Even though there has been an enormous number of studies that estimate Mincer-type wage equations using data from developing countries (see the reviews in

[3] Namely, the marginal returns to schooling are 1 percent per year or less until grade 11, then increase to 3.7 percent in grade 12, and exceed 10 percent only between grade 14 and 16. The average return, measured from grade 7, increases smoothly from 0.4 percent (grade 7) to 4.6 percent (grade 16).

Schultz 1988, Strauss and Thomas 1995), very few studies have a measure of ability available in the data. Boissiere, Knight, and Sabot (1985), Psacharopoulos and Velez (1992), Alderman et al. (1996a), and Glewwe (1996) are the notable exceptions. In two of these studies, sample sizes are either less than or barely exceed two hundred. All the authors use Raven's test score (Raven's Progressive Matrices) as a measure of innate ability. Raven's test scores tend to have little direct effect on wages, but considerably affect achievement scores, which in turn significantly affect wages. The effects of completed schooling are similar to those of Raven's tests: schooling's effect on wages is mostly indirect, operating through the cognitive skills as measured by achievement tests.

It is worth noting, however, that the use of Raven's tests as a measure of innate ability is controversial. The major concern is well expressed by Glewwe and Jacoby (1994, 851), who point out that: "This test [the Raven's abstract thinking test] was never intended as such [as an indicator of "innate" ability, independent of schooling]". In the data they use, there is, conditional on age, a strong positive association between Raven's scores and years of acquired schooling. Their data set is not the only example – in the Pakistani data that Alderman et al. (1996a) use, Raven's test scores are significantly higher for men than for women. This difference in Raven's test scores is potentially related to the fact that men acquire more schooling than women in Pakistan, which would imply that Raven's test scores are influenced by schooling. This point is reinforced by the fact that the difference in the amount of completed schooling appears to be unrelated to possible differences in innate ability between Pakistani men and women – single-sex schools are predominant in Pakistan and the girls are disadvantaged in terms of school availability (Alderman et al. 1996b).

The literature review would be incomplete if I did not mention the research that has been done on returns to schooling in the Philippines. Schady (2003) uses data from a recent nationwide household survey, the 1998 Annual Poverty Indicator Survey, to estimate returns to schooling for Filipino men. The results suggest convexity – the returns to both primary and secondary education are lower than those for tertiary education. As a result, the returns to primary and secondary education are considerably smaller than the conventional rates in the literature. Depending on the specification, the mean rate of return ranges from 6.2 to 9.4 percentage points for primary education and 6.9-10.0 percent for secondary education (based on Schady 2003, Table 2). Schady also finds sheepskin effects in the returns, i.e., within a given level of education, the returns to completing the last year of primary school, high school, or college are higher than the returns to any year below the last one. Both of these results can be driven by ability bias. Data limitations preclude the author from fully exploring such a possibility.[4]

In summary, for the last forty years the literature has recognized ability bias as a serious econometric problem. Economists used multiple approaches to resolve it. None of them provides a universal fix. Recently, concerns have been raised regarding the magnitude of the bias, pointing toward the need for more flexible estimation techniques and better controls for unobserved ability. While the accumulated evidence on ability bias in the United States is quite impressive, few studies for developing countries have addressed this issue directly.

3. Data

The data come from the Cebu Longitudinal Health and Nutrition Survey (CLHNS). The CLHNS follows a representative cohort of Filipino children born between May 1, 1983 and April 30, 1984 in 33 randomly chosen barangays[5] (17 urban and 16 rural) of the Metropolitan Cebu

[4] His analysis partially controls for ability by including measures of parental education and by using within-sibling estimates. He finds no significant changes in the results. It is unclear, however, to what extent these measures can control for innate ability.
[5] “Barangay” is the smallest administrative unit in the Philippines; it can be thought of as a community or district.

region.[6] Metro Cebu is the second largest metro area in the Philippines, with a population of 1.4 million (as of the 1990 census). Contrary to the commonly held view that a "metro area" is urban by definition, Metro Cebu encompasses vast agricultural areas reaching deep into Cebu Island.[7]

Multiple follow-up surveys have been made for the last twenty years, tracking the children from their birth up to the present day. The latest surveys are 1991-1992, 1994-1995, 1998-1999, 2002-2003, and 2004-2005 follow-up surveys, with the latter survey being finished this fall. The CLHNS data sets provide detailed, up-to-date information on each child, including early childhood development, family background, household, and community characteristics, as well as information on the characteristics of schools children attended. As with any longitudinal data, the sample attrition across the surveys is of potential concern. My analysis hinges on surveys starting from the 1991-1992 survey (the first that provides information on schooling). During the 1991-1992 survey 2,260 children were surveyed, and the 2002-2003 survey (the latest survey with available data) contains information about schooling decisions for 2,040 individuals. The attrition appears to be fairly low.[8] The actual sample that I use includes only those for whom it was feasible to construct complete schooling trajectories from four panels. The sample consists of 1982 individuals. Appendix A provides details on the construction of the variables. Descriptive statistics of the variables are in Appendix B.

[6] First, a single-stage cluster sampling procedure was used to randomly select 33 barangays from the Metro Cebu area. Then the barangays, which contained about 28,000 households, were completely surveyed in late 1982 and again in early 1983 to locate all pregnant women. Women of the selected communities who gave birth between May 1, 1983 and April 30, 1984 were included in the sample.

[7] At the time of the 1980 census, Metro Cebu included 155 urban and 88 rural barangays based on the Census Bureau classification (148 urban and 95 rural barangays based on the reclassification made by the CLHNS researchers).

[8] Looking across all the surveys, most of the attrition happened during early childhood. Out of 3,080 nontwin live births, only 2,600 households were surveyed during the first two years of children's lives. The attrition was mostly due to death or migration out of Metro Cebu.

4. Questions of Interest

For the last few decades, the Philippines have gone through a rapid economic development. The Cebu region exemplifies that transition. This region has been undergoing a transition from agriculture and low-skill manufacturing to a service and technology oriented economy, with substantial population growth as well as rapid economic growth. Six of the top ten products produced in Cebu are high technology (e.g., semiconductors, electronic watches, etc.). This is the type of transition that one can expect many other developing countries to experience in the next few decades.

In the Philippines, basic education consists of six years of primary school and four years of secondary school; obtaining a university degree normally takes an additional four to five years. Under the Philippine Constitution, both primary and secondary education are free in public schools. However, the proportion of secondary schools that are public has been considerably smaller, especially in rural areas.[9] Also, while primary education is mandatory, secondary education is voluntary in the Philippines.

Accelerated economic development in the Philippines has been associated with educational expansion. As a result, the Philippines have achieved one of the highest school enrollment rates, especially in primary schools, among less developed countries. For example, during school year 1990/1991, when most of our sample entered school, the net enrollment rate in primary schools was 95.3 percent (1991 Philippine Development Report 1992). These gains, however, have been offset by low school completion rates. The proportion of students enrolled at the beginning grade who reached the final grade of primary school at the end of the required number of years of study in year 1990/1991, for instance, was 68.2 percent (1991 Philippine

[9] In 1997/1998, for instance, public primary schools accounted for 92.3 percent of total primary enrollments, while public secondary schools accounted for only 72.0 percent of total secondary school enrollments (Behrman, Deolalikar, and Soon 2002).

Development Report 1992). Dropping out of school and grade repetition account for this low rate. About 40 percent of our sample repeated a grade at least once. Despite the fact that almost all of the individuals in our sample enrolled in school at some point, the cohort survival rate for primary education was only 69.5 percent. Seventeen percent of the sample never made it to high school. Of those who went to high school, 26.3 percent did not finish by age nineteen.

This naturally raises several questions. What factors affect youths' decision to drop out? Are individuals with lower innate ability more likely to drop out of school than to people with higher ability? If so, what can be done to keep the low-ability dropouts in school longer? More importantly, would this additional schooling benefit individuals in the labor market? In other words, do we see a significant return to schooling when we look at their wages? Does this return differ by an individual's ability? These are the questions I seek to answer in this paper. Knowing these answers should provide important lessons for policymakers in many developing countries that will experience similar economic changes over the coming decades.

5. Model

Overview

The model is developed to answer the questions posed in the previous section. It can be divided into three parts, corresponding to school grade progression, test scores, and labor market outcomes. All of the outcomes are modeled as functions of unobserved ability. The intuition behind modeling innate ability is simple. An individual's innate ability is never observed. Any cognitive test (either achievement or intelligence) is only a proxy for innate ability. It is always unclear how good such a proxy is. Generally, test scores are affected by, among other factors, the amount of acquired schooling at the time the tests are taken.[10] The semiparametric approach that I use to control for an individual's innate ability allows me to avoid such problems. This approach is based on the methodology developed by Hansen, Heckman, and Mullen (2004). I specify a one-factor model, where an unobserved factor enters all outcomes of interest. The inclusion of the unobserved factor as a determinant of cognitive achievement test scores and IQ test scores provides a reason to label the unobserved factor as "ability." It is important to note, however, that "unobserved ability" as used in the paper only refers to the collection of unobserved characteristics that impact each of the modeled outcomes.

The only dependence among all outcomes comes from a common unobserved ability. All of the equations are estimated simultaneously using full-information maximum likelihood (FIML) with Gauss-Hermite quadrature approximation for the unobserved ability, which is assumed to follow a standard normal distribution. Below, the model is outlined in greater detail.

School Grade Progression

The school grade progression part of the model serves two purposes. First, it helps to identify factors that affect an individual's decision to attend school and to successfully complete each year. I model both attendance and successful completion since, despite high enrollment rates, as previously noted, we observe substantial dropping out in the Philippines, as well as subsequent school reentry, and grade repetition. These phenomena are common in developing countries in general; to the best of my knowledge, however, none of the previous studies addressed the problem of grade repeats and school reentry at the individual level.

[10] Hansen, Heckman, and Mullen (2004), for example, estimate that one year of schooling increases the AFQT score, on average, between 2.79 and 4.2 percentage points.

The second purpose of the school grade progression part is to control for the endogeneity of schooling – all of the schooling outcomes are modeled as functions of unobserved ability, which reflects the fact that more able individuals, generally, choose to acquire more schooling.

Within each educational level (primary school, secondary school, and tertiary education), progression through school grades is modeled by two binary outcomes. They represent the decisions and behavior of each individual and his/her family with respect to schooling every year.[11] First, a person must decide whether to enroll in school (variable *ATTND*) and then each individual has an opportunity to successfully finish a grade (variable *SUCSS)*.[12] *SUCSS* captures dropping out as well as failing to advance to the next grade.

In terms of economic behavior, each individual maximizes his/her utility subject to the budget constraint. The resulting subsequent lifetime indirect utility from attending school during school year t is:

$$V_t(ATTND_t = 1) = U_t(ATTND_t = 1) + \beta E\left[V_{t+1}(ATTND_t = 1)\right] + \varepsilon_{t,1}$$

Lifetime indirect utility from not attending school during school year t is:

$$V_t(ATTND_t = 0) = U_t(ATTND_t = 0) + \beta E\left[V_{t+1}(ATTND_t = 0)\right] + \varepsilon_{t,0},$$

where $\varepsilon_{t,1}$ and $\varepsilon_{t,0}$ represent preference shocks and are assumed to be independently and identically distributed as Type I extreme value distribution. It follows that an individual decides to enroll in school if and only if the difference in the indirect utilities is greater than zero. The latent variable $ATTND_t^*$ measures this difference in utilities:

$$ATTND_t^* \equiv U_t(ATTND_t = 1) + \beta E\left[V_{t+1}(ATTND_t = 1)\right] + \varepsilon_{t,1} - U_t(ATTND_t = 0) - \beta E\left[V_{t+1}(ATTND_t = 0)\right] - \varepsilon_{t,0}$$

---

[11] Since the attendance and completion rates across the three groups are different, there is no need for modeled effects to be constant across these groups. I allow the schooling outcome parameters to differ across the three educational groups (grades 1-6, grades 7-10, grade 11 and above).

[12] The variable *SUCSS* is modeled if and only if the person attended school that school year, i.e., if the variable *ATTND* is equal to one.

Similar logic applies to the successful completion of the grade, $SUCSS_t^*$, and all other discrete outcomes in this model. For primary school, I model *ATTND* for each person starting with the year after the first school entry, conditional on completed schooling as of the time of that decision.[13]

I approximate the latent indexes $ATTND^*$ and $SUCSS^*$ as:

$$ATTND_{it}^* = \gamma_{ATD}(1-ATTND_{i,t-1}) + \phi_{ATD} ATTND_{i,t-1}(1-SUCSS_{i,t-1}) + \alpha'_{ATD} X_i + \beta'_{ATD} Z_{it} + \varphi'_{ATD} C_{it} + \gamma_{ATD} S_{it} + \delta_{ATD} f_i + \xi_{it}$$

$$SUCSS_{it}^* = \gamma_{SUC}(1-ATTND_{i,t-1}) + \phi_{SUC} ATTND_{i,t-1}(1-SUCSS_{i,t-1}) + \alpha'_{SUC} X_i + \beta'_{SUC} Z_{it} + \varphi'_{SUC} C_{it} + \gamma_{SUC} S_{it} + \delta_{SUC} f_i + \zeta_{it}$$

The terms $\gamma_{ATD}(1-ATTND_{i,t-1})$ and $\phi_{ATD} ATTND_{i,t-1}(1-SUCSS_{i,t-1})$ are included to capture costs associated with the decisions to repeat a grade and to reenter school, respectively. γ_{ATD} represents the effect of not attending school the previous school year and ϕ_{ATD} represents the effect of failing the grade attended during the previous school year. Two similar terms are included in $SUCSS_{it}^*$ to reflect the fact that successfully completing a grade might be easier if the person repeats the grade and that successfully completing a grade might be harder if the person was out of school for some time. The vector X_i represents individual characteristics including age, sex, and a low birth weight dummy as a health measure. The vector Z_{it} consists of family background variables including household income, household size, family business dummy, parental education, and caretaker's household dummy. The vector C_{it} includes community characteristics including urban/rural dummy, population density, food prices, and school quality characteristics. The variable S_{it} represents the amount of successfully completed schooling by the beginning of school year *t*. The variable f_i stands for unobserved ability. The error terms (ξ_{it} and ζ_{it}) are independent of each other and logistically distributed.

[13] The first school entry is modeled as a separate outcome.

Entering School on Time

Initial school entry is modeled as a separate outcome. For simplicity, it is chosen to be a binary outcome – on time vs. late entry to school, with "on time" meaning "entered school by age 7.5." Note that "on time" entry controls for the attendance of the first year in school. The latent index specification is:

$$N_i^* = \alpha_N' X_i + \beta_N' Z_{it-1} + \varphi_N' C_{it-1} + \theta_N f_i + \omega_i ,$$

where subscript "*t*-1" stands for using lagged values (from the time the child was 2 years old) of the variables. Community variables are constructed as the averages of community characteristics from the time the child was 2 years old and 1991-1992 survey. Lagged and averaged characteristics are used for two reasons. One is the fact that sending the child to school is a complex decision, likely to be affected by the past as well as the present. The second reason is to provide additional identification: the variation in the exogenous characteristics at the time of the child's 2nd birthday is different from the present. This is crucial since "on time" entry is at the very beginning of school grade progression and acquired schooling enters in all subsequent outcomes. Aggregate primary school quality characteristics from the 1994-1995 survey are used as a proxy for primary school quality in the area at the time the decision is made to send the child to school. Aggregate school quality characteristics are constructed by computing averages of school quality characteristics across local schools within a certain area using geographical coordinates of schools (for more details, see Zayats 2004).

Test Scores

Three cognitive achievement tests (Math, English, Cebuano) were administered during the 1994-1995 follow-up survey. For the purpose of our analysis, Math and English test scores are used. All children who were surveyed took the tests independent of schooling status. Additionally, the Philippines Non-Verbal Intelligence Test ("IQ test" for simplicity) developed by Guthrie, Tayag, and Jimenez (1977), was administered in the 1991-1992 and 1994-1995 surveys. The IQ test is comparable to Raven's Coloured Progressive Matrices, which are heavily used in empirical research on developing countries as a measure of innate ability. I use the IQ test scores from the 1991-1992 survey, since at that time only a fraction of the sample was already in school. Test scores are modeled similar to Hansen, Heckman, and Mullen (2004), who in their turn extend the factor analysis model used in psychometrics. The k^{th} test score is modeled as

$$T_{k,i} = \beta_k' X_{k,i} + \mu_k(s_i) + \lambda_k(s_i) f_i + \varepsilon_{k,i}(s_i) \qquad k = 1 \text{ (Math)}, 2 \text{ (English)}, 3 \text{ (IQ)}$$

$X_{k,i}$ includes all exogenous regressors (individual, parental, community, and school characteristics) and s_i measures completed education as of the time of the test. $\mu_k(s_i)$ is a level effect of schooling that is uniform across unobserved ability levels. The effect of unobserved ability on test scores can vary by completed schooling at the time of the test, and it is given by $\lambda_k(s_i)$. Both f and $\varepsilon(s)$ are assumed to be independent and have zero means. $\mu_k(s_i)$ is further parameterized as $\mu_k(s_i) = \alpha_k S_i$. A more flexible specification would be a second- or third-degree polynomial, e.g., $\alpha_{1,k} S_i + \alpha_{2,k} S_i^2 + \alpha_{3,k} S_i^3$, but linearity is not very restrictive given that all schooling variation at the time of testing is within primary school only. $\lambda_k(s_i)$ is similarly specified as $\lambda_k(s_i) = \rho_{0,k} + \rho_{1,k} S_i$.

Earnings

Modeling returns to schooling involves two outcomes. One is the selection into work for pay after leaving school. It resolves the endogeneity of the experience in the wage equation.[14]

[14] I assume that working for pay contributes to human capital accumulation only if an individual is out of school. Therefore, for each individual, the experience in the wage regression is the number of years she/he worked for pay while not attending school.

The second outcome is a wage equation destined to provide the estimates of the return to schooling.

Wages are modeled as of the time of the 2002-2003 survey, when approximately thirty percent of the sample were still in school. Our analysis of wages is limited to those individuals who are not in school by the time of the 2002-2003 survey, so selection into work is modeled explicitly only for those who are out of school. I model the work decisions, for those not in school, starting from the school year 1997/1998,[15] when most of the sample was thirteen years of age:

$$R_{it}^{*} = \alpha_R' X_{it} + \beta_R' Z_{it} + \varphi_R' C_{it} + \psi_R' L_{it} + \delta_R f_i + \xi_{it}$$

where R_{it}^{*} is a latent index for whether person *i* is a wage worker during school year *t*; L_{it} includes local labor market variables such as the average wage in the area.

Wages are modeled by specifying the equation for the logarithm of hourly wage rate. Several specifications are used. I start with a separate Mincer-type equation, which is routinely used in the literature on returns to schooling, $\ln W_i = \alpha_0 + \alpha_1 S_i + \alpha_2 exper_i + \xi_i$. I do not include the quadratic in experience due to very young age of the workers. In this specification, the assumption that the only cost of additional schooling is forgone wages will yield α_1 as the private rate of return to schooling.

The preferred specification allows i) the rate of return to education to vary across individuals by unobserved ability and ii) unobserved ability to affect the wages directly. This specification is:

$$\ln W_i = \alpha_w' X_i + \beta_w' Z_{i,2002} + \varphi_w' C_{i,2002} + \gamma_w S_i + \eta_w S_i \cdot f_i + \delta_w f_i + \xi_{it}$$

[15] For the school year 1996/1997, only nineteen people reported working for pay while being out of school.

Other variables in the equation are used to capture the formation of human capital besides schooling and ability, as well as to control for observed heterogeneity in, for instance, local labor markets.

Likelihood Function

The individual likelihood after integrating out unobserved ability is the following:

$$
\begin{aligned}
& L_i(N=n, T_1=t_1, \ldots, T_3=t_3, S_T=s_T, S=s, R_m=r_m, \ldots, R_M=r_M, \ln W=w)= \\
& =\sum_{k=1}^{K} \pi_k \Big\{ \; \Pr(N=0 \,|\, f_{i,k})^{1-N} \Pr(N=1 \,|\, f_{i,k})^{N} \cdot \\
& \cdot \prod_{j=year\_after\_1st\_entry}^{sch.year\_2002} \left[\begin{array}{l} \left(\Pr(ATTND_j=1 \,|\, f_{i,k}) \Pr(SUCSS_j=1 \,|\, f_{i,k})^{SUCSS_j} \Pr(SUCSS_j=0 \,|\, f_{i,k})^{(1-SUCSS_j)} \right)^{ATTND_j} \cdot \\ \cdot \Pr(ATTND_j=0 \,|\, f_{i,k})^{1-ATTND_j} \end{array} \right] \cdot \\
& \cdot f_1(t_1 \,|\, f_{i,k}) \cdot f_2(t_2 \,|\, f_{i,k}) \cdot f_3(t_3 \,|\, f_{i,k}) \cdot \prod_{m=start\_work}^{M} \left[\Pr(R_m=0 \,|\, f_{i,k})^{1-R_m} \Pr(R_m=1 \,|\, f_{i,k})^{R_m} \right] \cdot f_w(\ln W=w \,|\, f_{i,k})^{R_M} \; \Big\},
\end{aligned}
$$

where K is the number of points of support chosen for the Gauss-Hermite quadrature, π_k is the probability weight that the unobserved ability f takes on the mass point f_k. The sample likelihood is given by the product of the individual likelihoods.

Identification

Hansen, Heckman, and Mullen (2004) prove nonparametric identification of unobserved ability and the identification of the model in a static version of this model. The factor structure assumption for the unobserved ability and the concept of "measurable separability" are key to the identification. The latter, in their model, boils down to having individuals with different amounts of schooling at the time tests are taken. Heckman and Navarro (2005) provide a detailed proof of semiparametric identification for more general dynamic discrete choice models in which agents sequentially update the information on which they act. The outcomes are allowed to be mixed discrete-continuous.

Additionally, Bhargava (1991), Mroz and Surette (1998), and Mroz and Savage (forthcoming) show that panel-data relationships like those in our model[16] provide many more identification conditions than one might achieve by simply counting the number of contemporaneous exogenous variables excluded from an equation of interest. The intuition behind this identification is simple. As an example, consider the amount of completed schooling as of the previous year that enters as an explanatory variable, for instance, all of the grade progression outcomes in our model. At any point in time lagged completed schooling is implicitly a function of all past exogenous characteristics; in a "reduced form equation" it would be a function of community characteristics from all previous time periods. Hence, for a given amount of previous schooling at any point in time, there will be a time variation in the past determinants of schooling that do not affect contemporaneous schooling decisions. This provides implicit exclusion restrictions, i.e., additional multiple identifications, to our model.

6. Results

The model is estimated using FORTRAN with analytic first derivatives, in conjunction with the GQOPT optimization library. The number of mass points used for Gauss-Hermite quadrature is 15 (further increase in the number of quadrature points did not improve the likelihood function). The estimates are reported in Appendix C.

In each of our outcomes, impact of the unobserved factor operates in the direction one would expect unobserved ability to operate. The estimates suggest that boys enter school later than

[16] Households' places of residence are treated as exogenous in this model. For that reason, the analysis contains numerous randomly varying and time-specific exogenous variables. These include community characteristics like urban dummy, food prices, school characteristics and local wage rate.

girls. Conditional on gender, children with lower ability enter school at a later age (Table C5). The same applies to the children with poor health as measured by the child's height at the time of his/her second birthday. The latter is in agreement with findings of Glewwe, Jacoby, and King (2001), even though I do not control for the endogeneity of a child's health in the model.

As can be seen from Tables C6-C8, children with lower ability face lower probabilities of attending school. They are also much more likely to drop out of school at all three levels of education (Tables C9-C11), with the effect of lower ability diminishing at higher levels of education. For example, one standard deviation decrease in unobserved ability implies a 7.5 percentage point higher probability of dropping out of elementary school, a 6.7 percentage point higher probability of dropping out of high school, and a 4.7 percentage point higher probability of dropping out of college.[17]

A key question is whether we can keep the low-ability dropouts in school longer. More importantly, would this additional schooling benefit individuals in the labor market? The answers to these questions lie in the wage equation: if the return to schooling is large in absolute terms, then the counterfactual additional schooling would certainly, on average, pay off for school dropouts. However, if the return is small, then additional resources spent on making this subgroup of population stay longer in school might be wasteful, at least for the low-ability subgroup. In this respect, our estimates from the wage equations are informative. While standard Mincer-type wage regression (Table C13) yields a 4.5 percentage point return per additional year of schooling (which is in broad agreement with Schady 2003), our model reveals that the introduction of unobserved ability and controlling for the endogeneity of acquired schooling reduces the estimated return by almost 2 percentage points, down to 2.7 (I allow returns to schooling to vary by ability by

[17] These numbers, as well as all other estimates for discrete outcomes (Tables 9-16), represent average marginal effects (i.e., marginal effects are computed for each individual and then averaged across the sample).

introducing the ability-schooling interaction, but the corresponding estimate is essentially zero). In other words, I find evidence of a considerable omitted ability bias in the conventional estimates of the return to schooling, the notion introduced almost thirty years ago by Griliches (1977). At the same time, the estimated coefficient on unobserved ability is 8 (although it is statistically insignificant). This implies that one would have to acquire about three additional years of education to compensate for one standard deviation lower innate ability in terms of labor market returns, ceteris paribus.

When discussing the results from the wage equations a word of caution is in order. Our sample represents very young wage workers (about nineteen years old at the time of the 2002-2003 survey). So early in their careers some of them may exhibit unusual behavior, confounding the effects of schooling and ability. For example, some high-ability individuals might choose to stay out of school and take low-paying jobs to get more experience. Also, as was previously pointed out, a significant fraction of our sample was still in school at the time of the 2002-2003 survey. We, potentially, do not observe the entire range of completed schooling and, perhaps, ability. While my results might have limitations with respect to generalizing to the whole population, they are quite meaningful for the subgroup, that is, primary and high school dropouts, as most of school reentry happens within a year or two in our sample. Note that the most recent round of the CLHNS (2004-2005, when the youth are twenty-one or twenty-two years old) is almost finished and the data are expected to be available by December. I intend to incorporate these new data into the analysis promptly, which should resolve most of the current limitations.

Looking at average marginal effects, improving school quality appears to increase achievement test scores. These effects, however, are quite small. Decreasing the local pupil-teacher ratio, for example, by one standard deviation, 5.19, is expected to increase Math test scores by only .32 score points, or less than one tenth of the standard deviation. Lower pupil-teacher ratio yields higher rates of elementary school completion, but the effect is similarly small. A one standard deviation decrease in the pupil-teacher ratio is expected to increase the elementary school completion rate by .8 percentage points. Surprisingly, the fraction of women with primary education and the fraction of women with more than primary education in each community (proxies for high school quality) have only small effects on the outcomes of interest. Looking at the effects of low birth weight, it is worth noting that low birth weight seems to hurt children at early stages of education, as reflected by lower test scores and lower probability of completing primary school. However, this effect virtually disappears later on. Higher family income appears to benefit both attendance and completion of elementary school, and it strongly affects high school and post-secondary school attendance.

The above discussion is based on the analysis of average marginal effects, and these do not reflect all of the complex relationships among our outcomes. To provide a more comprehensive assessment, I make a series of policy simulations by: 1) doubling household income in all time periods; 2) increasing the mother's education by one standard deviation, i.e., by 3.29 years of education; 3) assigning low birth weight to everyone in the sample; 4) decreasing local pupil-teacher ratio by one standard deviation, i.e., by 5.19; or 5) increasing the fraction of women with more than primary education in each barangay by .14 (i.e., decreasing the fraction of women with less than primary education to zero). The approach to implementing simulations is standard: a whole life-cycle to age at the time of the 2002-2003 survey is generated for each individual using estimated structural parameters of the model based on the specified policy change. The standard errors on the effects are estimated using a parametric bootstrap with 50 iterations. The resulting effects of policy changes on major outcomes of interest are reported in Table 1.

The effects are qualitatively similar to the previously discussed average marginal effects, with the increase in the mother's education producing the largest effect on the outcomes. For instance, while increasing the mother's education raises the probability of successfully completing elementary school by 2.1 percentage points, doubling household income in all time periods leads to only a .6 percentage point increase in the rate of successful elementary school completion.

Three key extensions I plan to make in the near future are i) relaxing the normality assumption for the unobserved ability; ii) allowing schooling inputs to affect low-ability individuals differently from high-ability people; and iii) allowing returns to schooling to vary across educational levels (primary school, high school, and college).

Table 1. Policy simulation results

Outcome	Doubling household income	Increasing mother's education	Assigning low birth weight to everyone	Decreasing local pupil-teacher ratio	Increasing fraction of women with more than primary education
Entered school on time	0.0186 (0.0172)*	0.0552 (0.0081)	0.0079 (0.0059)	-0.0214 (0.0135)	0.0000 (0.0000)
Attended *elementary* school	0.0107 (0.0036)	0.0357 (0.0045)	0.0020 (0.0021)	0.0061 (0.0043)	0.0000 (0.0000)
Attended *high* school	0.0093 (0.0058)	0.0248 (0.0052)	-0.0009 (0.0026)	-0.0065 (0.0075)	-0.0040 (0.0116)
Attended *college*	0.0172 (0.0108)	0.0247 (0.0109)	-0.0014 (0.0050)	0.0020 (0.0129)	-0.0645 (0.0729)
Successfully completed the grade, *elementary* school	0.0063 (0.0029)	0.0213 (0.0029)	0.0017 (0.0012)	0.0049 (0.0030)	0.0000 (0.0000)
Successfully completed the grade, *high* school	-0.0029 (0.0046)	0.0158 (0.0030)	-0.0011 (0.0019)	-0.0011 (0.0049)	0.0072 (0.0118)
Successfully completed the grade, *college*	0.0053 (0.0167)	0.0324 (0.0178)	-0.0033 (0.0054)	-0.0044 (0.0221)	0.0171 (0.0458)
Math test scores	0.3539 (0.1872)	2.2841 (0.1659)	0.2042 (0.0923)	0.3145 (0.2832)	0.0000 (0.0000)
English test scores	0.6955 (0.1685)	2.4074 (0.1857)	0.1440 (0.0830)	0.4168 (0.2580)	0.0000 (0.0000)
Completed schooling as of 2002	0.1339 (0.0324)	0.5213 (0.0405)	0.0189 (0.0233)	0.0187 (0.0539)	0.0055 (0.0620)
Log of the hourly wage rate	0.0089 (0.0091)	0.0156 (0.0161)	0.0025 (0.0073)	0.0003 (0.0073)	0.0188 (0.0136)

* Standard errors are in parentheses. The standard errors are estimated using parametric bootstrap with 50 iterations.

7. Conclusion

Using rich data from the Cebu Longitudinal Health and Nutrition Survey, the paper analyzes the role of an individual's unobserved innate ability in explaining school attendance and completion, and early labor market outcomes of young Filipino adults.

I find that children with lower innate ability enter school at a later age, complete fewer years of school, and are more likely to drop out of school at all levels of education. From a policy making perspective, I find that enhanced conventional school inputs, such as pupil-teacher ratios, do little to keep young children in school. My results suggest that in a country with relatively high primary school enrollment and completion rates, like the Philippines, policies oriented toward the achievement of universal primary education might need to be more refined than just increasing educational expenditures. A one-size-fits-all approach is unlikely to be successful, while more experimenting at the pilot project level could suggest important new policies and approaches.

With respect to labor market outcomes of school dropouts in the Philippines, I find that the returns to education, after controlling for ability, are small. This implies that a simple focus on increasing completed schooling may yield inefficient policies. It is important to note, however, that the sample of workers is quite young; no one was older than twenty at the time of the 2002-2003 survey. Data for 2004-2005, however, will be available within the next month. I also find that individuals with lower unobserved ability are considerably more disadvantaged in terms of labor market returns. While completing additional years of schooling can compensate for lower innate ability, such substitution would be costly. It would take about three additional years of education to compensate for one standard deviation lower innate ability in terms of labor market returns. If an individual's unobserved ability can be affected during early childhood, the payoff for improving it would be substantial. At the same time, policies should be geared toward the creation of more flexible labor markets so that lower-ability individuals can find appropriate jobs where they can be productive.

Appendix A: Constructed Variables

The variable “*Entered school on time*” is equal to one if a child entered school at age less than 7.5 years old, it is zero otherwise. “*Low birth weigh”* is equal to one if the weight of a child at birth was 2.5 kilograms or less, zero otherwise. “*Age as of*” represents the age of a person at the beginning of the school year. The school year starts in June in the Philippines. “*Completed schooling at t”* represents the number of successfully completed grades by school year *t*.

School quality characteristics that I use are measures aggregated from individual-level school measures. The reason for doing this is the fact that individual school quality measures cannot be constructed for everyone in the sample, but only for those who attended a “primary only” type of school (as opposed to “primary and high school in one” or “high school only”). Although a “primary only” type of school is predominant in Cebu (around 87-90 percent of all schools), I did not want to lose a portion of the sample. Two measures are used for primary school: pupil-teacher ratio and public school dummy. They are constructed based on the school questionnaires administered during the 1994-1995 survey and on a supplemental 1996 survey.

None of the CLHNS data contain high school characteristics. To resolve that issue, I have merged 2000 census data from the Philippines at the barangay (community) level with my sample by barangay of residence. Such measures as “Fraction of women with primary education in barangay” and “Fraction of women with more than primary education in barangay” were constructed to proxy for the quality of high schools in the areas of residence.

The *Household income* variable represents the average household income per week. It is calculated as the sum of three sources of income: 1) resources generated within and by the household (home gardening, income in kind, remittances, pensions, rent savings, etc.); 2) individual earnings (wages, piecework, fishing, self-employment); and 3) group earnings (livestock and farming).

All the pecuniary measures (like *household income* and *food prices*) were deflated to January 1983 pesos.

For all dynamic variables, like *household and community characteristics*, the data are assigned in the following way: years 1990-1993 use the data from the 1991-1992 survey, years 1994-1996 use the data from the 1994-1995 survey, years 1997-1999 use the data from the 1998-1999 survey, years 2000-2002 use the data from the 2002-2003 survey. The year sequence starts from 1990 because only twenty-two people attended school in year 1989 (once again, all references to years are references to school years, e.g., “year 1990” means “school year 1990/1991”).

School grade progression

The variables ATTND and SUCSS are created for each educational subgroup. Modeling of ATTND_elementary starts with the year after the first school entry, conditional on completed schooling as of the time of that decision; ATTND_high has a nonmissing value starting with the year right after the year when the last grade of primary school was completed; ATTND_college is modeled starting with the year right after the year when the last grade of high school was completed.

Earnings

In the final sample, 1,781 reported working, of whom 1,333 were working for pay. Only 1,234 were out of school at the time of the 2002-2003 survey. In the analysis of earnings, we limit the sample to only those who reported both working and being out of school by the time of the 2002-2003 survey, that is 1,179 people. Out of these 1,179, wage workers comprise 931. Five people are dropped as outliers in the hourly wage rate distribution (these five reported hourly wages above 400 pesos, while the 99th percentile had 250 pesos per hour). That leaves us with 926 wage workers (509 men and 417 women). Hourly wage rate was computed using available information on: 1) reported earnings per day, 2) reported number of days working per week, and 3) reported number of hours working per week. For those who reported “no regular workday” as their number of working days per week, it is assumed they worked five days a week (48 individuals).

Appendix B: Descriptive Statistics

Table B1. Summary Statistics of Time-Invariant Variables

Variable	Mean	Std. Dev.	Min	Max
Male	.5292	.4993	0	1
Low birth weight	.1231	.3262	0	1
Entered school on time	.7674	.4226	0	1
Math test	30.5621	11.0814	0	58
English test	27.4187	10.4548	0	59
IQ test	32.8548	6.6368	5	47
Working experience (in years)	1.1771	1.1950	0	6
Log of the hourly wage rate	2.5798	.9477	-1.5404	5.9915
Mother's education (log)	1.9795	.4885	0	2.9444
Father's education (log)	1.8554	.5598	0	2.8904
Local pupil-teacher ratio	39.1714	5.1930	22.5	55.6
Fraction of public schools in the area	.9547	.0639	.6988	1

Table B2. Summary Statistics of Time-Specific Variables

Variable	Mean	Std. Dev.	Min	Max
Height at 2nd birthday (log)	-.2339	.0440	-.4155	-.1109
Age as of IQ test date	8.6600	.2756	8.1667	9.0833
Age as of achievement test date	11.7402	.4066	10.8333	12.8333
Completed schooling as of IQ test date	1.3094	.7036	0	3
Completed schooling as of achievement test date	4.0940	1.0400	0	6
Household income (lagged)	5.4759	.5189	4.5511	9.8669
Urban (averaged, time of child's 2nd birthday and 1991-92 survey)	.7356	.4268	0	1
Population density (log, averaged)	8.6589	1.5952	4.5642	11.1956
Price of kerosene (log, averaged)	.8725	.3209	-.0594	1.5009
Price of bananas (log, averaged)	-1.5713	.1975	-2.4487	-1.0186
Price of corn (log, averaged)	.9167	.1545	.4322	1.1842

Table B3. Summary Statistics of Time-Variant Variables

Variable	1990 Mean (Std. Dev.)	1996 Mean (Std. Dev.)	2002 Mean (Std. Dev.)
Household size	6.9511 (2.3427)	7.1302 (2.4768)	6.9119 (2.7924)
Family business	0.3468 (0.4506)	0.4425 (0.4968)	0.5094 (0.5000)
Household income (log)	5.9486 (0.5531)	5.9382 (0.7642)	6.2006 (0.8188)
Household income net of individual's (log)	–[*]	–	6.1087 (0.8732)
Caretaker's household	0.9344 (0.2476)	0.9173 (0.2756)	0.8476 (0.3595)
Age (by the beginning of school year *t*)	6.6609 (0.2759)	12.6609 (0.2759)	18.6609 (0.2759)
Completed schooling (by the beginning of school year *t*)	0.0096 (0.0975)	4.9723 (1.2350)	8.9945 (2.5119)
Attended *elementary* school during the year *t*	0.9545 (0.2132)	0.9066 (0.2912)	–[†]
Attended *high* school during the year *t*	–	1.0000[‡] (0.0000)	0.2175 (0.4129)
Attended *college* during the year *t*	–	–	0.7690 (0.4222)
Successfully completed the grade, if in *elementary* school that year	0.8338 (0.3724)	0.9394 (0.2387)	0.4167[§] (0.5149)
Successfully completed the grade, if in *high* school that year	–	0.8918 (0.3108)	0.8644[§] (0.3429)
Successfully completed the grade, if in *college* that year	–	–	0.8891[§] (0.3143)
Missed school last year	0.9889 (0.1048)	0.0439 (0.2049)	0.5585 (0.4967)
Failed last grade	0.0015 (0.0389)	0.0469 (0.2115)	0.0691 (0.2537)
First year of high school	–	0.3875 (0.4873)	0.0747 (0.2629)
First year of college	–	–	0.3734 (0.4838)
Working for pay	–	–	0.7486 (0.4340)
Local wage rate for unskilled labor	–	–	15.6800 (6.9304)
Fraction of women with primary education	–	0.1712 (0.0689)	0.1727 (0.0679)
Fraction of women with more than primary	–	0.6874 (0.1660)	0.6852 (0.1617)
Urban	0.7356 (0.4411)	0.7306 (0.4437)	0.7184 (0.4499)

[*] This variable (as well as some variables below) is used in modeling "working for pay," which is modeled starting from 1997, and therefore does not have nonmissing observations prior to 1997.
[†] Unless otherwise noted, here and below the variable is missing if it is irrelevant for the year *t*, e.g., no one was in high school in 1990, etc.
[‡] 1.0 means that <u>all of those</u> who were eligible to go to high school that year (i.e., all who completed elementary school by 1997) did go to school during the school year 1997.
[§] The number represents the value for the school year 2001, not 2002. For the school year 2002 no "successful completion" was modeled since our sample was surveyed during that school year and, for that reason, there is no information on whether that school year was successfully completed.

Population density (log)	8.6148 (1.6982)	8.5382 (1.5489)	8.8054 (1.4091)
Price of bananas	0.2712 (0.0447)	0.2216 (0.0226)	0.1697 (0.0387)
Price of corn	2.8529 (0.2991)	3.0597 (0.1735)	2.2843 (0.1357)
Price of kerosene	1.2554 (0.7488)	1.9332 (0.1127)	2.5301 (0.2557)

Table B4. Summary Statistics of Some Time-Variant Variables (all years)

Variable	**Obs**	**Mean**	**Std. Dev.**
Attended *elementary* school	11489	0.8922	0.3101
Attended *high* school	8638	0.7609	0.4265
Attended *college*	1178	0.8820	0.3227
Successfully completed the grade, *elementary* school	12176	0.9212	0.2694
Successfully completed the grade, *high* school	6451	0.8859	0.3179
Successfully completed the grade, *college*	811	0.8520	0.3553

Appendix C: Estimates

Table C1. Math Test Scores

Variable	Estimate	Standard error	t-statistic
Completed schooling as of test date	2.1125	0.2749	7.6830
Age as of test date	2.8786	0.4892	5.8850
Male	-3.2162	0.4313	-7.4570
Low birth weight	-1.5214	0.6314	-2.4090
Caretaker's household	0.7527	0.6759	1.1140
Mother's education (log)	5.1670	0.4991	10.3530
Family business	0.3655	0.3603	1.0140
Household size	-0.1839	0.0781	-2.3560
Household income (log)	0.4162	0.2479	1.6790
Urban	3.3400	0.7728	4.3220
Price of bananas	-22.9137	8.5506	-2.6800
Price of corn	-0.7767	1.1067	-0.7020
Price of kerosene	2.1595	1.7458	1.2370
Population density (log)	-1.0676	0.2455	-4.3490
Local pupil-teacher ratio	-0.0621	0.0468	-1.3290
Fraction of public schools in the area	-1.1960	4.2673	-0.2800
Constant	-9.3178	10.4155	-0.895
f (unobserved ability)	2.9237	0.7358	3.973
*f*S* (schooling-ability interaction)	1.2277	0.1708	7.186

N=1,953, $\sigma_{\varepsilon} = 4.80$

Table C2. English Test Scores

Variable	Estimate	Standard error	t-statistic
Completed schooling as of test date	0.9808	0.2647	3.705
Age as of test date	3.3837	0.4735	7.146
Male	-3.8373	0.4105	-9.348
Low birth weight	-1.1679	0.6023	-1.939
Caretaker's household	0.3564	0.6466	0.551
Mother's education (log)	6.0434	0.4502	13.425
Family business	-0.6126	0.346	-1.771
Household size	-0.2962	0.0731	-4.052
Household income (log)	1.0106	0.2441	4.14
Urban	2.4438	0.7637	3.2
Price of bananas	-5.8483	8.8165	-0.663
Price of corn	-0.211	1.0005	-0.211
Price of kerosene	4.6883	1.6753	2.799
Population density (log)	-0.5747	0.2261	-2.542
Local pupil-teacher ratio	-0.0841	0.0505	-1.664
Fraction of public schools in the area	-2.5295	4.1817	-0.605
Constant	-29.0479	9.3420	-3.109
f (unobserved ability)	1.9474	0.6806	2.861
*f*S* (schooling-ability interaction)	1.4139	0.1526	9.265

N=1,953, $\sigma_\varepsilon = 4.42$

Table C3. IQ Test Scores

Variable	Estimate	Standard error	t-statistic
Completed schooling as of test date	1.3248	0.2721	4.869
Age as of test date	-4.0188	0.5861	-6.857
Male	-0.348	0.2926	-1.189
Low birth weight	-0.7279	0.4124	-1.765
Caretaker's household	0.4923	0.5217	0.944
Mother's education (log)	2.8393	0.3501	8.11
Family business	-0.0525	0.2974	-0.176
Household size	-0.2305	0.0622	-3.707
Household income (log)	0.7222	0.2983	2.421
Urban	0.5586	0.5388	1.037
Price of bananas	-6.7527	3.7945	-1.78
Price of corn	0.2456	0.4849	0.506
Price of kerosene	0.13	0.1982	0.656
Population density (log)	0.1329	0.1667	0.798
Local pupil-teacher ratio	-0.0028	0.0334	-0.084
Fraction of public schools in the area	1.326	2.915	0.455
Constant	55.8985	6.6202	8.444
f (unobserved ability)	3.804	0.314	12.114
*f*S* (schooling-ability interaction)	-0.8788	0.2072	-4.242

N=1,949, $\sigma_{\varepsilon} = 5.17$

Table C4. Log of Hourly Wage Rate

Variable	Estimate	Standard error	t-statistic
Male	0.3877	0.07	5.537
Age	0.1345	0.1157	1.163
Experience	0.0186	0.0356	0.524
Completed schooling	0.0267	0.0232	1.153
Urban	0.0071	0.1196	0.06
Population density (log)	0.007	0.0371	0.189
Local wage rate for unskilled labor	-0.0084	0.0051	-1.653
Constant	-0.3009	2.1626	-0.139
f (unobserved ability)	0.0805	0.1235	0.652
*f*S* (schooling-ability interaction)	-0.0008	0.0135	-0.06

N=918, $\sigma_{\varepsilon} = 0.92$

Table C5. Entered School on Time

Variable	Av. Marg. Effect	t-statistic
Male	-0.0478	-2.41
Caretaker’s household	0.0004	0.01
Low birth weight	-0.0482	-1.638
Height of the child	1.0761	4.368
Household income (lagged)	0.0276	1.268
Mother’s education (log)	0.1476	6.841
Family business	-0.0402	-1.846
Urban (averaged across time)	-0.1140	-3.007
Population density (log, averaged)	0.0311	2.763
Price of kerosene (log, averaged)	-0.0855	-2.139
Price of bananas (log, averaged)	0.0653	1.178
Price of corn (log, averaged)	-0.1392	-1.72
Local pupil-teacher ratio	0.0037	1.681
Fraction of public schools in the area	0.2703	1.41
f (unobserved ability)	0.0586	5.026

N=1,963

Table C6. "Did individual *i* attend ELEMENTARY school during school year *t*?"

Variable	Av. Marg. Effect	t-statistic
Missed school last year	-0.0929	-23.126
Failed last grade	-0.0477	-11.715
Completed schooling as of t	-0.0069	-4.27
Age as of t	-0.0113	-9.12
Male	-0.0073	-2.039
Low birth weight	-0.0042	-0.988
Caretaker's household	0.0050	1.155
Mother's education (log)	0.0213	5.68
Family business	0.0028	0.892
Household size	-0.0019	-3.311
Household income (log)	0.0050	2.024
Urban	0.0028	0.499
Price of bananas	-0.0433	-1.31
Price of corn	0.0185	2.943
Price of kerosene	-0.0013	-0.284
Population density (log)	0.0008	0.446
Local pupil-teacher ratio	-0.0004	-1.311
Fraction of public schools in the area	-0.0184	-0.563
f (unobserved ability)	0.0206	8.584

N=11,489, N of individuals=1,953

Table C7. "Did individual *i* attend HIGH school during school year *t*?"

Variable	Av. Marg. Effect	t-statistic
First year of high school	-0.1467	-13.535
Missed school last year	-0.2777	-25.284
Failed last grade	-0.2235	-20.06
Completed schooling as of t	-0.1008	-19.882
Age as of t	-0.0351	-7.272
Male	-0.0113	-1.535
Low birth weight	-0.0077	-0.746
Caretaker's household	0.0276	2.792
Mother's education (log)	0.0412	4.707
Family business	0.0041	0.603
Household size	-0.0034	-2.648
Household income (log)	0.0142	2.542
Urban	-0.0041	-0.285
Price of bananas	0.0741	1.059
Price of corn	-0.0009	-0.038
Price of kerosene	-0.0300	-2.302
Population density (log)	-0.0064	-1.284
Local pupil-teacher ratio	0.0002	0.305
Fraction of public schools in the area	-0.0876	-1.359
Fraction of women with primary education	-0.3306	-3.394
Fraction of women with more than primary	-0.0341	-0.664
f (unobserved ability)	0.0204	4.01

N=8,638, N of individuals=1,736

Table C8. “Did individual *i* attend COLLEGE during school year *t*?”

Variable	Av. Marg. Effect	t-statistic
First year of college	0.2268	4.484
Missed school last year	-0.2173	-4.993
Failed last grade	-0.3461	-7.97
Completed schooling as of t	0.0610	2.17
Age as of t	-0.0402	-2.496
Male	-0.0121	-0.743
Low birth weight	0.0163	0.583
Caretaker’s household	0.0312	1.349
Mother’s education (log)	0.0401	1.872
Family business	-0.0138	-0.867
Household size	-0.0035	-1.176
Household income (log)	0.0264	2.27
Urban	0.0302	1.105
Price of bananas	0.0169	0.077
Price of corn	0.0143	0.158
Price of kerosene	-0.0242	-0.556
Population density (log)	0.0081	0.553
Local pupil-teacher ratio	-0.0005	-0.28
Fraction of public schools in the area	0.1575	1.114
Fraction of women with primary education	-0.3131	-0.682
Fraction of women with more than primary	-0.3555	-1.403
f (unobserved ability)	0.0089	0.847

N=1,178, N of individuals=586

Table C9. "Did individual *i* successfully complete the grade during school year *t*, ELEMENTARY school?"

Variable	Av. Marg. Effect	t-statistic
Missed school last year	-0.0602	-9.221
Failed last grade	-0.0113	-1.542
Completed schooling as of t	0.0031	0.718
Age as of t	0.0010	0.368
Male	-0.0507	-7.923
Low birth weight	-0.0192	-2.204
Caretaker's household	0.0161	1.853
Mother's education (log)	0.0630	8.438
Family business	0.0026	0.49
Household size	-0.0044	-3.865
Household income (log)	0.0127	2.845
Urban	0.0127	1.305
Price of bananas	-0.0819	-1.347
Price of corn	0.0220	2.765
Price of kerosene	0.0001	0.033
Population density (log)	0.0008	0.26
Local pupil-teacher ratio	-0.0012	-1.826
Fraction of public schools in the area	0.0931	1.485
f (unobserved ability)	0.0748	14.971

N=12,176, N of individuals=1,957

Table C10. "Did individual *i* successfully complete the grade during school year *t*, HIGH school?"

Variable	Av. Marg. Effect	t-statistic
First year of high school	-0.0234	-1.218
Missed school last year	0.0075	0.346
Failed last grade	-0.0866	-7.048
Completed schooling as of t	0.0247	2.512
Age as of t	-0.0082	-1.215
Male	-0.1158	-10.637
Low birth weight	0.0056	0.383
Caretaker's household	0.0573	4.259
Mother's education (log)	0.0691	5.491
Family business	0.0186	2.103
Household size	-0.0001	-0.055
Household income (log)	0.00005	0.006
Urban	-0.0364	-1.862
Price of bananas	-0.0248	-0.288
Price of corn	-0.0115	-0.386
Price of kerosene	-0.0084	-0.528
Population density (log)	-0.0006	-0.09
Local pupil-teacher ratio	0.0001	0.104
Fraction of public schools in the area	0.2408	2.717
Fraction of women with primary education	0.0281	0.178
Fraction of women with more than primary	0.0523	0.624
f (unobserved ability)	0.0671	10.147

N=6,451, N of individuals=1,731

Table C11. "Did individual *i* successfully complete the grade during school year *t*, COLLEGE?"

Variable	Av. Marg. Effect	t-statistic
First year of college	-0.0650	-0.131
Missed school last year	-0.1576	-2.02
Failed last grade	0.0588	0.47
Completed schooling as of t	-0.0090	-0.018
Age as of t	0.0857	1.829
Male	-0.0795	-2.345
Low birth weight	0.0464	0.689
Caretaker's household	0.0877	1.559
Mother's education (log)	0.1061	2.212
Family business	0.0133	0.405
Household size	-0.0035	-0.537
Household income (log)	0.0070	0.29
Urban	-0.0719	-0.99
Price of bananas	0.0902	0.213
Price of corn	-0.0632	-0.433
Price of kerosene	0.0096	0.133
Population density (log)	0.0254	0.915
Local pupil-teacher ratio	0.0011	0.255
Fraction of public schools in the area	0.1293	0.446
Fraction of women with primary education	0.0907	0.132
Fraction of women with more than primary	0.1467	0.454
f (unobserved ability)	0.0467	1.972

N=811, N of individuals=545

Table C12. Working for Pay During the Year *t*

Variable	Av. Marg. Effect	t-statistic
Age as of t	0.1169	13.426
Male	0.0291	2.144
Low birth weight	0.0465	2.411
Mother's education (log)	-0.0712	-5.211
Family business	-0.0831	-6.122
Household size	0.0037	1.614
Household income net of individual's (log)	-0.0286	-3.539
Urban	-0.0060	-0.226
Price of bananas	0.6305	3.747
Price of corn	0.1605	3.182
Price of kerosene	0.0455	1.939
Population density (log)	-0.0007	-0.074
Local wage rate for unskilled labor as of t	0.0015	1.452
Local pupil-teacher ratio	0.0007	0.547
Fraction of public schools in the area	0.3655	2.882
Fraction of women with primary education	-0.0994	-0.577
Fraction of women with more than primary	0.0583	0.669
f (unobserved ability)	-0.0096	-1.324

N=3,898, N of individuals=1,454

Table C13. Mincer-type Log Wage Regression

Variable	Estimate	Standard error	t-statistic
Male	.4057	.0628	6.46
Experience	.0258	.0301	0.86
Completed schooling	.0447	.0143	3.13
Constant	1.9696	.1511	13.03

N=918, σ_{ε} = 0.92, R^2 =0.04

References

Alderman, H., J. R. Behrman, D. R. Ross, and R. Sabot. 1996a. The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market. *Oxford Bulletin of Economics and Statistics* 58(1): 29-55.

Alderman, H., J. R. Behrman, D. R. Ross, and R. Sabot. 1996b. Decomposing the Gender Gap in Cognitive Skills in a Poor Rural Economy. *Journal of Human Resources* 31(1): 229-254.

Angrist, J. D. and A. B. Krueger. 1991. Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics* 106(4): 979-1014.

Angrist, J. D. and A. B. Krueger. 1992. The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples. *Journal of the American Statistical Association* 87(418): 328-336.

Ashenfelter, O. and A. Krueger. 1994. Estimates of the Economic Return to Schooling from a New Sample of Twins. *American Economic Review* 84(5): 1157-1173.

Ashenfelter, O. and C. Rouse. 1998. Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins. *Quarterly Journal of Economics* 113(1): 253-284.

Behrman, J. R., A. B. Deolalikar, and L. Y. Soon. 2002. Promoting Effective Schooling through Education Decentralization in Bangladesh, Indonesia, and Philippines. *ERD Working Paper Series No. 23.* Asian Development Bank.

Behrman, J. R. and M. R. Rosenzweig. 1999. "Ability" Biases in Schooling Returns and Twins: A Test and New Estimates. *Economics of Education Review* 18(2): 159-167.

Behrman, J. R. and P. Taubman. 1976. Intergenerational Transmission of Income and Wealth. *American Economic Review* 66(2): 436-440.

Belzil, C. and J. Hansen. 2002. Unobserved Ability and the Return to Schooling. *Econometrica* 70(5): 2075-2091.

Bhargava, A. 1991. Identification and Panel Data Models with Endogenous Regressors. *Review of Economics Studies* 58(1): 129-140.

Blackburn, M. L. and D. Neumark. 1995. Are OLS Estimates of the Return to Schooling Biased Downward? Another Look. *Review of Economics and Statistics* 77(2): 217-230.

Boissiere, M., J. B. Knight, and R. H. Sabot. 1985. Earnings, Schooling, Ability, and Cognitive Skills. *American Economic Review* 75(5): 1016-1030.

Bound, J. and D. A. Jaeger. 1997. On the Validity of Season of Birth as an Instrument in Wage Equations: A Comment on Angrist and Krueger's "Does Compulsory School Attendance Affect Schooling and Earnings?" *NBER Working Paper # 5835*.

Bound, J., D. A. Jaeger, and R. M. Baker. 1995. Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association* 90(430): 443-450.

Bound, J. and G. Solon. 1999. Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling. *Economics of Education Review* 18(2): 169-182.

Burtless, G. 1996. Introduction and Summary. In *Does Money Matter? The Effects of School Resources on Student Achievement and Adult Success*, ed. Gary Burtless, 1-42. Washington, D.C.: Brookings Institution.

Butcher, K. F. and A. Case. 1994. The Effect of Sibling Sex Composition on Women's Education and Earnings. *Quarterly Journal of Economics* 109(3): 531-563.

Card, D. 1999. The Causal Effect of Education on Earnings. In *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card, 1801-1863. Amsterdam: Elsevier Science, North-Holland Publishers.

Card, D. 2001. Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica* 69(5): 1127-1160.

Glewwe, P. 1996. The Relevance of Standard Estimates of Rates of Return to Schooling for Education Policy: A Critical Assessment. *Journal of Development Economics* 51(2): 267-290.

Glewwe, P. and H. Jacoby. 1994. Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana. *Journal of Human Resources* 29(3): 843-864.

Glewwe, P., H. G. Jacoby, and E. M. King. 2001. Early Childhood Nutrition and Academic Achievement: A Longitudinal Analysis. *Journal of Public Economics* 81(3): 345-368.

Griliches, Z. 1977. Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica* 45(1): 1-22.

Griliches, Z. 1979. Sibling Models and Data in Economics: Beginnings of a Survey. *Journal of Political Economy* 87(2): S37-S64.

Griliches, Z. and W. M. Mason. 1972. Education, Income and Ability. *Journal of Political Economy* 80(2): S74-S103.

Guthrie, G. M., A. Tayag, and J. P. Jimenez. 1977. The Philippine Non-Verbal Intelligence Test. *Journal of Social Psychology* 102: 3-11.

Hansen, K. T., J. Heckman, and K. J. Mullen. 2004. The Effect of Schooling and Ability on Achievement Test Scores. *Journal of Econometrics* 121(1-2): 39-98.

Heckman, J. and S. Navarro. 2005. Dynamic Discrete Choice and Dynamic Treatment Effects. Working Paper, Department of Economics, University of Chicago.

Imbens, G. W. and J. D. Angrist. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62(2): 467-475.

Mincer, J. 1958. Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy* 66(4): 281-302.

Mroz, T. and T. Savage. Forthcoming. The Long-Term Effects of Youth Unemployment. *Journal of Human Resources*.

Mroz, T. and B. Surette. 1998. Post-Secondary Schooling and Training Effects on Wages and Employment. Mimeo, University of North Carolina at Chapel Hill.

National Economic and Development Authority. 1992. *1991 Philippine Development Report*. Republic of the Philippines.

Neumark, D. 1999. Biases in Twin Estimates of the Return to Schooling. *Economics of Education Review* 18(2): 143-148.

Psacharopoulos, G. 1994. Returns to Investment in Education: A Global Update. *World Development* 22(9): 1325-1343.

Psacharopoulos, G. and E. Velez. 1992. Schooling, Ability, and Earnings in Colombia, 1988. *Economic Development and Cultural Change* 40(3): 629-643.

Rosenzweig, M. R. and K. I. Wolpin. 2000. Natural "Natural Experiments" in Economics. *Journal of Economic Literature* 38(4): 827-874.

Schady, N. R. 2003. Convexity and Sheepskin Effects in the Human Capital Earnings Function: Recent Evidence for Filipino Men. *Oxford Bulletin of Economics and Statistics* 65(2): 171-196.

Schultz, T. P. 1988. Education Investments and Returns. In *Handbook of Development Economics*, Vol. 1, ed. H. Chenery and T. N. Srinivasan, 543-630. Amsterdam: North-Holland.

Strauss, J. and D. Thomas. 1995. Human Resources: Empirical Modeling of Household and Family Decisions. In *Handbook of Development Economics,* Vol. 3A, ed. J. Behrman and T. N. Srinivasan, 1883-2023. Amsterdam: North-Holland.

Zayats, Y. 2004. School Quality, Family Background, and Student Achievement: The Case of Cebu. Working Paper, Department of Economics, University of North Carolina at Chapel Hill.