CHILD SCHOOLING IN PERU: EVIDENCE FROM A SEQUENTIAL ANALYSIS OF SCHOOL PROGRESSION

Sarmistha Pal*
Cardiff Business School

Abstract: Primary enrolment rates are very high in Peru, but so are the failure and drop-out rates, especially beyond the primary level. Thus an analysis of child schooling should take account of the conditional sequence with the previous level and self-selection into the next higher level of schooling. This cannot be done using standard univariate or ordered logit/probit models of school enrolment/grade attainment. This paper applies a unique correlated sequential probit model with unobserved individual specific heterogeneity to determine the nature of school progression at primary, secondary and post-secondary levels in Peru. This entails richer results, argued to be better than the standard static estimates. In particular, parental education, household expenditure, sibling composition and local adult market participation rates are found to affect different levels of schooling differently. While parental education is crucial for child school enrolment at the primary level, sibling composition and household expenditure turn out to be significant for attainment at the secondary level. However, grade repetition at primary and secondary levels and market participation rates are important for a child to move on to the post-secondary levels.

Keywords: Child schooling, School progression, Resource constraint, Sibling composition, Sequential probit model, Limited dependent variable.

JEL Classification: I21, J13, O15

*Address for correspondence: Cardiff Business School, Colum Drive, Cardiff CF10 3EU, UK. Tel. 00-44-29-20876164; fax: 00-44-29-20874419; e-mail: pal@cf.ac.uk. The author is much grateful to the Managing Editor and also the Journal referees for their helpful and constructive comments. She also wishes to thank Cardiff Business School for the research grant and Gerry Makepeace for comments on an earlier draft of the paper. The usual disclaimer applies.

CHILD SCHOOLING IN PERU: EVIDENCE FROM A SEQUENTIAL ANALYSIS OF SCHOOL PROGRESSION

(2nd revision)

1. INTRODUCTION

There have been a large and growing number of empirical studies examining the determinants of child schooling in developing countries of Asia, Africa and Latin America. Different indicators of schooling have been used in this literature, including completed years (Birdsall, 1985), current enrolment (Singh, 1992), ever attended (Cochrane, Mehra and Osheba, 1986), grades attained or grades failed (Drèze and Kingdon, 2001) and delayed enrolment (Glewwe and Jacoby, 1994). The estimation n is often separated for boys and girls and for rural and urban areas. Both demand (e.g., household income and parental education) and supply (e.g., variables reflecting quantity and quality of schools; see Drèze and Kingdon, 2001) factors have been identified as explanations for low educational attendance and attainments in these low-income regions. Most of these studies are essentially static in nature: usually ordinary least squares or two-stage least squares methods are used to determine completed years of schooling; determination of school enrolment/ever attended is often based on some univariate probit/logit method while ordered probit/logit models are used to determine grades attained.

Using 1994 Peru Living Standards Measurement Study (LSMS) data, we examine the nature school progression among 10 to 20 years old Peruvian children. Unlike many Asian or African countries a very high proportion (about 97% in our sample) of Peruvian children gets enrolled in primary schools while failures and dropout rates too are significantly higher beyond the primary levels. Existing studies on Peru however tend to overlook these important features of schooling. For example, Ray (2000a) uses univariate logit estimates to analyse the key determinants of child labour[1] in Peru[2] and Pakistan. In a subsequent attempt Ray (2000b) uses separate univariate logit estimates of child schooling and child labour in these two countries and their interaction with adult labour. Patrinos and Psacharapoulos (1997; P&P hereafter) made a more direct attempt to capture late entry and high failure rates as they use a univariate logit model to determine an aggregated index of age-grade distortion called the over-age indicator. This indicator measures if the child under consideration had a normal progress (in terms of schooling years) relative to his/her current age. Although this measure serves as a proxy for school repetition and late entry (because presence of overaged children in the classroom reduces overall quality), there are some difficulties. First, the over-age indicator does not distinguish between enrolment/attainment at primary, secondary or post secondary levels and as such cannot take account of the sequential nature of school progression where there is self-selection from one level to the next higher level. More seriously, the over-age indicator is determined using child employment as an explanatory variable, which raises serious endogeneity issue in these household decision models.

We depart from this tradition and develop a dynamic sequential framework of school progression conditional on attainment at the previous level and self-selection into the next higher level. This is based on a unique correlated sequential probit model[3], which allows us to identify the children who have progressed much less than others and also to locate at what level of schooling this has happened. This is

[1] Child's participation in wage employment constitutes important opportunity costs of schooling. There is now a growing literature on child labour in many low-income countries of Asia, Africa and Latin America. A useful collection of articles in this respect can be found in Grootaert and Patrinos (2000). Child labour is however endogeneous to child schooling and very few studies determine child schooling after taking account of child labour.

[2] This study too is based on 1994 Peru LSMS data as ours.

[3] The only use of this model in a schooling context that I am aware of is a study of intergenerational educational mobility in Malaysia (Lillard and Willis, 1994) where the effect of parental education on child education is studied.

important for any assessment of policies geared to boost child schooling because it is based on a full understanding of the nature of the selection process across different levels of schooling. We shall, in particular, focus on the following three levels of transition: (a) considering all sample children, whether a child gets enrolled in a primary school; (b) among those enrolled in primary schools, whether a child moves to the secondary level and (c) among those enrolled in secondary schools, whether a child moves on to the post-secondary level. While decision (a) relates to school enrolment, decisions (b) and (c) relate to school attainment, i.e., school progression from primary to secondary level and that from secondary to post-secondary levels respectively. Thus our analysis of child school progression combines both indicators of enrolment and attainment in a sequential framework. Movement from the primary to the secondary level is conditional on the successful completion of the final year of the primary school; similarly moving from the secondary to the post-secondary level requires one to pass the final year at the secondary level. We also take account of the process of self-selection at each higher level as only a fraction of children successfully completing primary (or secondary) schools will move on to the secondary (or post-secondary) schools[4]. In addition to child's ability, we control for sibling composition, household resource constraint, parental preferences and some community characteristics and obtain selectivity corrected correlated sequential probit estimates of school progression.

The paper is novel in a number of ways. Standard modelling techniques used in most existing studies on child schooling (including those specific to Peru) fail to capture the specific characteristics of child schooling in Peru, where primary enrolment rates are high along with high failures and drop out rates. Thus school

---

[4] For example, decision (b) selects those who successfully complete primary level and move on to the secondary level while decision (c) selects those completing secondary schools and move on to the post-

progression is a better indicator of child schooling than school enrolment/attainment. In this respect, sequential probit estimates are argued to be better than the corresponding ordered probit estimates. For someone at the secondary level, for example, the sequential probit model takes account of the fact that the person has completed the primary level to reach the secondary level while the ordered probit model neither takes account of the achievement at the previous level nor does it correct for any self-selection into the next higher level of schooling. We have also calculated a likelihood ratio statistic to compare the sequential and ordered probit estimates. Secondly, we have taken account of the individual-level unobserved heterogeneity that has not been considered by most existing studies. For example, individual health problem, ability/disability, if any, may significantly affect school enrolment/attainment though we do not observe this in our data set. Ignoring the unobserved heterogeneity may however seriously bias estimates of schooling. We compare estimates with and without individual specific unobserved heterogeneity and find that unobserved heterogeneity is significant in our sample. Thirdly, though a child's labour market participation is considered to be an important indicator of the opportunity cost of schooling, it is endogenous to his/her schooling decisions and cannot be entered as an explanatory variable in the schooling equation as in P&P, for example. Given the problems of measuring these costs in the LSMS data[5], we instead include local (segment-specific) participation rates in home production and market jobs to obtain indirect evidence of the effect of job market participation on schooling at primary, secondary and post-secondary levels. Fourthly, household resource constraint plays an important role in determining child's school progression effect in

secondary level.

[5] For example, we only observe child's labour market participation in the past month or in the past year that may not represent child's labour market participation at the point when the a particular schooling decision was made.

societies characterised by capital market imperfections. Although household income is generally used in this respect, there are fluctuations in income from period to period especially in rural areas; hence, we consider household per capita expenditure to be a long-term indicator of household income. Since household expenditure is endogenous to decisions regarding child schooling, we use expenditure per capita *predicted* by characteristics of the household head (e.g., if male, if married, if educated), his/her spouse (e.g., if educated), household ownership of various assets and variables reflecting household demographic composition. Finally, we transform the continuous age variable into piecewise linear age splines, which allows for discrete non-linearity in analysing the effects of different age ranges on different levels of schooling (primary/secondary/post-secondary). The latter would also reflect the incidence of over-aged students at *different* levels of schooling. In other words, these age splines offer indirect evidence of late entry, if any, at a given level of schooling.

Most existing studies suggest how individual/household characteristics affect school enrolment/attainment. Our results are richer than most of these existing studies, allowing us to explicitly establish how *different* individual, household and community characteristics may affect school progression at *different* levels *differently*. In general, child's ability, sibling composition, parental education and some community characteristics turn out to be significant determinants of school progression at any level though these factors may affect schooling at different levels differently. For example, P&P suggest that having a greater number of younger siblings means more age-grade distortion. We however find that age composition of siblings is important only at the secondary level though not at the primary or post-secondary level. Secondly, father's schooling is significant for age-grade distortion in P&P's study though we find parental education is significant for primary and secondary levels, but not for post-secondary levels. Thirdly, existing studies generally highlight the adverse effects of job market participation on child schooling. However, assuming a child's participation in wage employment is a fraction of the local adult market participation rate, we find that local market participation rate has adverse effects on school performance only at the post-secondary levels. The latter perhaps reflects the flexibility of part-time occupations in Peru so that there are no adverse effects on schooling at lower levels, but only at a much higher level. Finally the likelihood ratio test rejects the equality of parameters of sequential and ordered probit models at 1% or lower level of significance.

The paper is developed as follows. Section 2 discusses the analytical framework while section 3 describes the data, analyses the sequential probit estimates and compares these estimates with the corresponding ordered probit estimates. Section 4 concludes.

2. AN ANALYTICAL FRAMEWORK

Traditionally, demand for schooling is derived from the Beckerian 'common preference' model of household behaviour. Here a household maximises the joint utility function of all its members to determine the quantity and quality of children, consumption of leisure and other market goods as well as adult/child labour market participation decisions.

We rationalise the sequential analysis of school progression in terms of a simplified framework with only two levels of schooling, namely, primary and secondary[6]. Assuming that primary school enrolment is compulsory (since most children in Peru are enrolled in primary schools), parents only choose whether to send

[6] This could easily be extended to include a third level of schooling, namely, post-secondary level.

a child to the secondary level. Provided a child is able to move from the primary to the secondary level, parents would send a child to a secondary school if and only if the discounted value of returns from additional schooling tomorrow is higher than the discounted value of the additional cost today. Assuming children born to same parents are of similar ability, parental investment in secondary schooling would depend on costs of schooling (net of returns), parental attitude towards higher schooling, available household resources, especially in societies with credit market imperfections. Composition of siblings may also be important to capture household resource constraint or implicit opportunity costs of schooling, if any.

Peruvian children play an important role in production and other domestic work within the household. Both the Peruvian educational system and flexibility in part-time occupations often accommodate dual school-work activities especially at lower levels of schooling. Though education is primarily free in Peru (in terms of school fees) schooling costs can be significant in terms of costs on transportation, school uniform, utensils etc., especially for the poor. There are also issues relating to the opportunity cost of lost child earnings, which are also more important for children from poorer background. Given these (explicit and opportunity) costs of schooling, resource constrained parents may decide to send some children to schools while reserving others for home production and/or wage employment supplementing family incomes; the latter may in turn give rise to an inequality among siblings with respect to school progression. More educated parents, even if resource constrained, may still perceive higher returns to higher schooling and thus are more likely to encourage children to move beyond the primary level.

3. DATA AND EMPIRICAL RESULTS

Our analysis of sequential school progression among Peruvian children is based on the most recent 1994 Peru Living Standards Measurement Survey (LSMS) data. In the Peruvian system, there are six grades within primary schools and 5 grades within secondary schools. Assuming six to be the normal school entry age and no failures in any grade, a child would start secondary school at around age 12 and post-secondary school around age 17. However, late enrolments are common in South America (Psacharopoulos, 1997), especially in rural areas (Ilon and Moock, 1991. Hence we focus on children aged 10-20 years for our analysis.

There are a total of 3873 male and female children belonging to 1840 households in our sample. Only 3% of the sample children have zero level of education either because they have not been enrolled in a primary school or have left school soon after being enrolled. As high as 97% of the sample children were currently enrolled in primary/secondary/post secondary schools and all these children thus have some level of primary education. Students currently enrolled in secondary and post secondary schools have some level of secondary schooling while only those enrolled in post-secondary schools have completed secondary schooling. Considering the highest level of educational achievement, we consider three sequentially related transition decisions. First, whether a child has some level of primary education. Among the total 3873 children, only 120 (3%) have no schooling. Thus, 3753 children have some positive level of primary education. The second decision is whether a child has moved from a primary to a secondary school. Among 3753 children enrolled in a primary school, a total of 2030 children have gone to the secondary level. The rest (1723 children) have either not been able to complete the primary level or have not gone to a secondary school even after successfully completing the primary level. Finally the third decision is whether to move on to a post-secondary school. Among 2030 children aged 10-20 years, only 186 have gone beyond the secondary level while others have either not been able to complete the secondary level or have decided not to go to post-secondary level, even after completion of the secondary schools.

Adding these numbers, there is a cumulative total of (3753 + 2030 + 186) = 5969 children in the age group 10-20 years in our sample who have some positive level of schooling[7]. If we calculate the proportion of the children in each category out of the cumulative total of 5969 children with some schooling in this sequential framework, there are an alarmingly decreasing proportion of children progressing beyond the primary level of schooling. In particular, 63% have some level of primary education while only 34% (about half) have some level of secondary schooling. Only about 3% of these children have some post-secondary level of schooling.

3.1. A Sequential Probit Model of Child Schooling

Suppose an index of child quality is the level of schooling attained. Here we consider three schooling decisions ranked in ascending order: (a) whether to attend primary school; (b) whether to attend secondary schools or stop after the primary level and (c) whether to move on to post-secondary levels or stop after the secondary level. Accordingly, we can define S as indicating the levels of schooling as follows:

S = 0 if no formal education

= 1 if some level of primary education

= 2 if some level of secondary education

[7] Although we observe a cross-section of children 10-20 years old surveyed in 1994, by virtue of the conditional sequence of the schooling decisions, we can convert the data on schooling choices into a sequential one. Thus the total number of observations included in the sequential analysis of school progression turns out to be 6089, which is 5969 plus 120 children never enrolled.

= 3 if some level of post-secondary education

The conventional approach would be to apply an ordered probit/logit model (e.g., see Kingdon, 1998) of school progression that allows one to estimate the probability that S = 0, 1, 2, 3 among the sample children. However, these decisions are not only ordered, but also sequential in nature and as such decision (b) is conditional on decision (a) and decision (c) is conditional on decision (b). Thus ordered probit estimates would be biased since it does not take account of the conditional sequence or aspect of self-selection involved in these decisions[8]. Hence, following Lillard and Willis (1994), we model this problem as a correlated sequential probit model where the probit index functions at each decision point s = 0, 1, 2, 3 for a child i belonging to household j is given by:

$$I_{sij} = \beta_s' X_{sij} + \delta_{ij} + u_{sij}$$

where the decision to move from one level to the next higher level depends on a set of covariates X_{sij} which vary by the schooling decision (s), individual child (i) and also the household (j). We also include an individual-level unobserved heterogeneity term δ_{ij} that is constant across schooling decision s, which will account for the unobserved ability/disability, other health problems, if any, of a child that may affect school progression. The remaining residual variation is assumed to be captured by u_{sij}. Both δ_{ij} and U_{sij} are assumed to be normally distributed as follows:

$$u_{sij} \sim N(0,1)$$
$$d_{ij} \sim N(0, s^2_{d_{ij}})$$

[8] Similar concerns may also be raised against the use of univariate probit regressions to individually determine primary, secondary or post-secondary scholing levels, thus strenthening the justification for the case of a correlated sequential probit model.

The residual structure is basically a random effect variance component model with up to three replications. The number of replications is determined within the model and higher order decisions are made only for positive outcomes at earlier decisions. We incorporate residual heterogeneity to obtain consistent estimates as well as consistent standard errors (see Lillard and Willis, 1994 p. 1136-37).

Thus individual i will move from level S to S+1 if $I_{sij} > 0$ and drop out otherwise.

$$P(s = S) = P[I_{0ij} \leq 0] \text{ if } S = 0$$

$$= P [I_{1ij} > 0, I_{2ij} \leq 0] \text{ if } S = 1$$

$$= P [I_{1ij} > 0, I_{2ij} > 0, I_{3ij} \leq 0] \text{ if } S = 2$$

$$= P [I_{1ij} > 0, I_{2ij} > 0, I_{3ij} > 0] \text{ if } S = 3$$

In other words, decision to move to the next higher level is correlated with the previous decision and also the subsequent decisions are subject to selectivity[9] with respect to earlier decisions.

3.2. Model Specification

For a given level of schooling, the probability of success to move from one level to the next higher level depends not only on the ability of the child, but also on factors affecting parental investment in child schooling, including opportunity costs of schooling, parental resources and preferences, community characteristics defining local health/education/employment environment . It is, however, difficult to measure or observe many of these factors in household survey data of a given year. Hence, we apply great care to choose the most relevant available indicators.

[9] For example, we do not select children who complete primary schools but do not go to secondary schools.

Age and past grade performance of the child are used as instruments of the child's ability[10]. We consider the quartile age distribution of the children in the age group10-20 years and found that the first quartile is 12 years (which is also the normal entry year at the secondary level) while the third quartile is 17 years (which is the normal entry year for the post-secondary level). Using these two quartile age figures as two nodes, we generate a piecewise-linear spline transformation of the age variable. The latter gave rise to 3 new variables AGE12 (between 10-12 years), AGE1217 (between 12-17 years) or AGEGT17 (above 17 years). Each new variable represents the original age variable on a specific segment of its range so that the estimated effect of the splines is no longer linear, but piece-wise linear. These spline coefficients may directly be interpreted as slope coefficients (Panis, 1994). Secondly, we include the number of years failed at the primary level (FAILPRIM) and the total number of years failed at primary and secondary levels (FAILPS) as indicators of a child's ability at secondary and post-secondary levels respectively.[11]

Individual child's participation in wage employment, if any, constitutes an important measure of opportunity costs of schooling. In our original sample of 3873 children, only about 10% children actively participated in some wage employment in the last month of the survey date. While only 4 children without any education worked during this reference period, 175 (about 10%) of 1723 children with primary education worked. Among the rest, 192 (10.4%) of 1844 children with secondary education and only 17 (9%) of 186 children with post-secondary education worked. Also this information does not correspond to the decision points of moving from primary to secondary or from secondary to higher schools. It is therefore difficult to

[10] The unobserved component of ability is taken care of by the individual-level unobserved heterogeneity (see discussion in section 3.1 and also later in section 3.2).

[11] One may also argue that grade repetition at primary and secondary levels too are endogeneous since in part it may be caused by the decision regarding number of days attended in a school. Hence we have also estimated a

use this information in a sequential analysis of child school progression. Moreover child's participation in wage employment is endogenous to his/her participation in schooling. Hence to avoid the possible problem of endogeneity, we include local (segment specific) adult participation rates in home production and job markets instead in the schooling equations. Though imperfect, these participation rates would be exogenous and would account for its effect on school enrolment/attainment at different levels.[12] In particular, we include two variables, namely, OWN and MKT to denote local (specific to the particular segment of the rural location) adult participation rates in home production and market jobs respectively in the last one week. There are about 51% male in all the three samples corresponding to these three sequential schooling decisions. Similar proportions of male (44%) and female (45%) children had attained primary schools while a slightly higher proportion of boys (49% as against 46% for girls) attained secondary schools in our sample. Thus the apparent evidence of gender difference in schooling achievements is rather limited in our sample, as has been noted by Patrinos and Psacharapoulos (1997) as well for 1991 Peru LSMS data-set. Nevertheless we create two gender interaction terms MALEOWN and MALEMKT to account for the possible gender difference, if any.

Household resource constraint, if any, may exert an important influence on child schooling in societies with imperfect credit markets. Household income or expenditure is usually considered to be a measure of resource constraint[13]. But in a

---

parsimonius model by dropping these possibly endogenous variables. These estimates are shown in Appendix table A2, which retain our basic results.

[12] We have also considered including the wages of unskilled adults as an alternative to participation rates. LSMS provides information about the last payment in the past 7 days or past 12 months. This however would not be representative of the wages at the time when these schooling decisions were made. Though similar arguments would also apply to using participation rates, adult participation rates are subject to less fluctuations than adult unskilled wage rates over time. Hence we shall only use adult participation rates to account for opportunity costs of schooling in our analysis.

[13] We consider per capita expenditure as an indicator of household long-term income since we do not observe the household income at each point of child's schooling decisions (e.g., whether to move to primary, secondary or post-secondary schools).

household model, household income/expenditure is endogenous. Hence, we first predict per capita expenditure in terms of characteristics of the household head, his/her spouse, household assets and other family composition variables (assumed to be given for the cross-section sample for the survey year) and use this predicted value as a measure of household resources.

Age distribution of siblings is also included to account for the implicit opportunity costs of schooling, if any. Following P&P (1997), we classify siblings into three age categories: (a) pre-school group aged 0-6 years, (b) school-going group aged 7-18 years and (c) working group aged above 18 years. These age composition variables would, therefore, takes account of the competition among siblings, especially younger ones (who are not in working age category) for limited household resources. Thus larger number of dependent (non-earner) siblings would impose a more stringent resource constraint as the available family resources will be divided among more consumption units. Faced with this resource constraint, households may use different children to specialise in different activities including school participation, home production or wage employment[14].

Parental education is expected to account for parental attitude towards sending children to secondary or higher schools. It may also affect a child's motivation to do well in a school. In this respect, we include three levels of parental education namely, primary, secondary and post-secondary levels of schooling, for father (HEADEDN1, HEADEDN2, HEADEDN3) and mother (HSPEDN1, HSPEDN2, HSPEDN3). Among other household characteristics, we include the predicted value of per capita current expenditure on food, health and education (PREDEXP) as an instrument. As a

[14] These sibling composition variables may also be regarded as endogenous, though in our single crosssection framework we assume away the dynamics of family composition. However, in order to check the robustness of our estimates, we also estimate a model by excluding these sibling composition

direct measure of borrowing constraint, we include if the household has an existing long-term borrowing (DEBT). We also include if the household comes from an indigenous Peruvian community (INDIG); the omitted category here is the Spanish-speaking community. An overwhelming majority (about 85%) of children belonged to Spanish-speaking households as compared to indigenous Peruvian population[15]. We also consider if the household is headed by a male member and in as high as 88% cases the child came from a male-headed household[16]. These household characteristics may also account for household resource constraint, parental preferences as well as an interaction between these factors.

One also needs to control for the supply-side factors, if possible. In the absence of other factors signifying supply of schooling, we include type of school (public/private/ parochial) attended. Choice of a school, e.g., a private (fee paying) school, may reflect parental motivation to produce children of better quality (i.e., with higher schooling). However, for given level of parental education, it may also account for the quality of services provided by public/private schools An overwhelming majority (87%) of the sample children attended public school as opposed to only 7% attending private schools; the rest went to parochial schools. We include the variable PUBLIC to denote whether the child has gone to a public school. We also include a variable RURAL to indicate if the child lives in a rural location. Residential location

---

variables. These estimates are shown in Table A1 in the Appendix, which essentially retains the basic results as shown in Table 2. See further discussion in section 3.3.

[15] In the Peruvian Society Spanish-speaking European descendants are at the top of the Peruvian social ladder while the monolingual indigenous language speakers are at the bottom. Other Spanish-speaking indigenous people (e.g., mestizos, cholos) are in the intermediate ranks. Psacharopoulos and Patrinos (1994) suggest that almost half the monolingual indigenous population is concentrated in the bottom income decile of the Peruvian income distribution.

[16] We started our estimation by including all different characteristics of the household head (e.g., if the head is male, married, occupation of the head etc.) among other variables. However, except HEADMALE all other characteristics turned out to be insignificant. Hence we have dropped other characteristics of the head of the household from the final specification.

may signify the distance from the local schools or the quality of local public services, e.g., health/education/employment, in relation to an urban area.

Finally we need to take account of the unobserved individual-level heterogeneity as Lillard and Willis (1994) suggested that individual specific heterogeneity played a significant role in explaining inter-generational educational mobility in Malaysia. This may capture the unobserved individual specific ability/disability or other health problems, if any, that are not observable in our data-set, but that may significantly affect school performance of these children.

There are some identifying variables in these three equations corresponding to three levels of schooling and these relate to past performance (i.e., grade repetition) at primary and secondary levels and also the marital status of the individual. In particular, past performance and marital status are not important for children's enrolment at the primary level. However, we include number of failed years at primary levels (FAILPRIM) for the progression primary to secondary level. But the question of being married does not seem to be important for these children in our sample. Finally, we include the number of failed years in primary as well as secondary levels (FAILPS) and also the marital status (MARRIED) in determining whether a student moves from secondary to post-secondary level. We, however, retain very similar parental and household characteristics for all three schooling decisions (including the age splines) [17].

Endogeneity is a common problem in the choice of regression variables in this respect. This is because in life cycle models of household decisions quantity (i.e., family size determining sibling composition) and quality (i.e., child schooling, child labour) of children, marital status, household earnings and expenditure are all jointly determined. It is however difficult to drop all these variables and yet estimate a meaningful model of child schooling. Since our analysis is based on single cross-section information (for 1994), we may however assume number of children, their birth order, marital status etc. to be given for this year and thus ignore the dynamics of fertility, consumption choices and their implications for child schooling.[18] There are also some measurement problems as all our observations relate to the year 1994 and not when the schooling decisions were made. Hence, we consider the values of the regression variables observed in 1994 as instruments for the values of these variables when schooling decisions were made. This would allow us to focus on the hypotheses of our interest regarding child school progression.

Summary statistics of the selected arguments of school progression are shown in Table 1 for the sample children classified by three levels of schooling.

3.3 Results

Maximum likelihood sequential probit estimates with and without individual specific heterogeneity are summarised in Table 2 for three levels of schooling. A comparison of these two sets of estimates suggests some differences. Most often there are differences in t-statistics (though often not very significant); more importantly, one can identify significant differences: for example, mother's education is significant for primary schooling if we do not consider heterogeneity, but turns out to be insignificant when heterogeneity is included. In contrast, father's highest level of education is significant for secondary schooling when we include heterogeneity, but not when we do not. These differences are strengthened by the highly significant

---

[17] We include all three age variables in all three schooling equations to account for the non-linear effects of age, if any, on the particular level of schooling. Inclusion of these age variables also indicate evidence of late entry, if any, especially at primary/secondary level.

value of individual level heterogeneity terms for all three levels of schooling. We thus conclude that ignoring unobserved individual level heterogeneity would give rise to biased estimates and proceed to interpret the sequential probit estimates with heterogeneity for different levels of schooling. We also compare these estimates with those of the restricted models presented in Table A1 and Table A2 of the Appendix. In general, these restricted estimates yield very similar results with respect to the common variables as shown in Table 2. One interesting difference is that the indicator of expenditure turn out to be more significant when we exclude the sibling age composition variables (see Table A1), thus strengthening our hypothesis that these sibling composition variables account for the implicit opportunity costs of schooling among resource constrained households in societies with imperfect credit markets.

From non-enrolment to enrolment at primary level: Among various child characteristics, there is some indirect evidence of late entry in that children aged between 12-17 years may still be at the primary level while children above 17 years are significantly less likely to be in this level. Neither sibling characteristics nor household expenditure are very important for the primary school enrolment. However, parental education plays an important role for child's primary schooling. Primary and secondary education of the father as well as primary education of the mother significantly enhances the likelihood of primary education. Finally higher local market participation rate significantly encourages (in stead of lowering) achievement at the primary level.

Progression from primary to secondary schools: At the secondary level, however, more factors turn out to be significant for explaining child's school attainment. First, the probability of secondary schooling increases between age 10-17

[18] In order to check the robustness of our results, we also estimate some parsimonium models by excluding these possibly endogenous variables (see Appendix).

years, but decreases thereafter. Child's performance at the primary level is also important in that higher number of failures at the primary level lowers secondary schooling attainment. Unlike primary education, instrument of household expenditure is significant such that higher expenditure raises the probability of secondary attainment. Among various sibling composition variables, larger number of siblings in the working age group (SIBGT18) significantly enhances secondary school attainment. While more infant siblings lower the attainment at the secondary level, the variable SIB06 is not statistically significant. In other words, individuals from poorer household and/or with more dependent siblings are less likely to move from primary to secondary schools. Taken together, the latter reflects significant effects of household resource constraint and competition for limited resources among the siblings on secondary schooling. As before, parental education plays an important role – all levels of paternal education and primary level of maternal education variables significantly enhance the likelihood of moving to secondary schools. School quality is also statistically significant: children from public schools are less likely to do well at the secondary level.

Progression from secondary to post-secondary schools: In this case child's age and indicator of ability turn out to be more important than household resource constraint or parental education. Whether children attain some post-secondary level of education significantly depends on age and grade repetition at the primary and secondary levels. In particular, the probability of moving to a post-secondary school increases between age 12-17 years. Children who do not perform well at primary and secondary levels of education are also less likely to move up the post-secondary levels. As before, children from public schools are less likely to have higher level of post-secondary education. Also important is the local market participation rate; the likelihood of post-secondary attainment is lower if the local market participation rates are higher, thus providing some indirect evidence of an adverse effect of wage employment on child schooling (despite flexibility of part-time occupations in Peru). However, the instrument of household expenditure or sibling age-composition variables is not significant. We also do not find any evidence of significant gender differences in this respect.

Finally, we compare these sequential probit estimates with the corresponding ordered probit estimates (with and without individual specific unobserved heterogeneity). These ordered probit estimates (see Table 3) generally yield standard results (not corrected for self-selection at different stages of schooling). Unlike the sequential probit estimates, order probit ones cannot however show how different characteristics would affect different levels of schooling differently. We also compute likelihood ratio statistics to test if sequential and ordered probit estimates are equal. These statistics are 903.50 and 882.38 respectively for estimates with and without heterogeneity, thus both rejecting the null hypothesis of any equality between these two sets of estimates at 1% or lower level of significance.

4. CONCLUDING COMMENTS

Boosting education for children from all background is essential for economic development, which not only entails a process of human capital accumulation but also generates useful externalities (e.g., lower fertility and/or better child health). Thus a better understanding of the process of school progression with late entry, conditional sequence with the previous level and self-selection into the next higher level is useful for any reassessment of the present education and/or employment policy in Peru.

This paper analyses the factors determining school progression among Peruvian children aged 10-20 years. While primary enrolment is high, drop-out and failure rates are high too in Peru. This justifies the use of sequential analysis of school progression, which offers a finer understanding of the factors determining school progression. A correlated sequential probit model is used to do so and it is argued that these selectivity corrected estimates that allows for conditional sequence/self selection and late entry are an improvement over the univariate logit or the ordered probit estimates commonly used in the literature.

Our results based on sequential probit estimates of 1994 Peru LSMS data are richer than most existing ones in that it shows how *different* individual, household and community characteristics affect school progression at *different* levels *differently*. An improvement of school attainment, especially at primary/secondary level crucially depends on whether children have educated parents and siblings in the working age category; while the former indicates preferences of educated parents to invest in child schooling, the latter reflects the lower opportunity costs of schooling among children with more working siblings supplementing family income. More interestingly, labour market participation does not necessarily adversely affect school attainment at primary/secondary level, but only at the post-secondary level, perhaps reflecting the flexibility of part-time employment in Peru. Finally, choice of school type, i.e., public/private is important in school progression in Peru even among children born to same parents (with similar motivation for investment in child schooling) beyond the primary level, emphasizing the need for improving the quality of service provided by the public schools in the country; this is especially important at a higher level of schooling. These results would highlight the importance of two sets of policies geared especially to lower alarming rates of drop-out at the primary and secondary levels of schooling. First, in stead of discouraging participation in wage employment, government or school authority may encourage children, especially poorer ones, to participate in part-time employment available locally. This can be accompanied by a second set of policies to improve the quality of services provided by the schools, especially public schools where poorer children go to, including the quality of teaching and school environment as well as introducing more vocational programmes to help students to find part-time employment locally.

TABLE 1. MEANS AND STANDARD DEVIATIONS OF REGRESSION VARIABLES IN THE SEQUENTIAL SCHOOLING DECISIONS

	PRIMARY VS NO SCHOOLING			SECONDARY VS PRIMARY SCHOOLING			POST-SECONDARY VS SECONDARY SCHOOLING		
VARIABLE	NOBS	Mean	Std Dev	NOBS	Mean	Std Dev	NOBS	Mean	Std Dev
AGELE12	3873	11.71314	0.624546	3753	11.71276	0.625156	2030	11.99163	0.101389
AGE1217	3873	2.543248	2.085653	3753	2.529976	2.08168	2030	3.764532	1.528995
AGEGT17	3873	0.441518	0.902104	3753	0.432188	0.892547	2030	0.693103	1.049256
MALE	3873	0.507875	0.500003	3753	0.50786	0.500005	2030	0.514286	0.499919
FAILPRIM				3753	0.267786	0.557436			
FAILPS							2030	0.23399	0.540091
MARRIED							2030	0.023153	0.150425
SIB06	3873	0.906274	1.075528	3753	0.896616	1.069031	2030	0.675862	0.956603
SIB718	3873	2.904209	1.378279	3753	2.904077	1.374597	2030	2.666995	1.379797
SIBGT18	3873	1.181513	1.404078	3753	1.180123	1.404402	2030	1.473892	1.527813
PREDEXP	3873	3.705996	0.3588748	3753	3.711018	0.3569642	2030	3.825926	0.335092
DEBT	3873	0.1727343	0 .3780662	3753	0.172129	0.3775428	2030	0.1817734	0.3857527
HEADMALE	3873	0.87968	0.325378	3753	0.879297	0.325826	2030	0.868966	0.337521
HEADEDN1	3873	0.46708	0.49898	3753	0.465494	0.498874	2030	0.405911	0.491189
HEADEDN2	3873	0.316034	0.464986	3753	0.320011	0.466542	2030	0.357143	0.479276
HEADEDN3	3873	0.125742	0.331602	3753	0.127898	0.33402	2030	0.174877	0.379955
HSPEDN1	3873	0.408211	0.491566	3753	0.411937	0.492249	2030	0.38867	0.487568
HSPEDN2	3873	0.198812	0.399158	3753	0.199307	0.399533	2030	0.227094	0.419057
HSPEDN3	3873	0.068939	0.253383	3753	0.07061	0.256207	2030	0.097044	0.296091
PUBLIC				3753	0.898481	0.302055	2030	0.850246	0.356918
INDIG	3873	0.142009	0.349104	3753	0.140155	0.347194	2030	0.097044	0.296091
MALEMKT	3873	0.4301575	08253095	3753	0.4332534	0.8317286	2030	0.4990148	0.8994544
MKT	3873	0.8399174	1.006275	3753	.8457234	1.011089	2030	0.9615764	1.063532
MALEOWN	3873	0.8670281	1.313129	3753	.8681055	1.318205	2030	0.844335	1.335125
OWN	3873	1.682159	1.37192	3753	1.68372	1.379639	2030	1.622167	1.450889
RURAL	3873	0.615544	0.486529	3753	0.621903	0.484977	2030	0.747783	0.434392

Note: AGELE12: if between 10-12 years; AGE1217: if between 12-17 years; AGEGT17: if between 18-20 years. MALE: 1 if male; FAILPRIM: number of years failed at the primary level; FAILPS: number of years failed at the primary and secondary levels; MARRIED : 1 if married; SIB06: number of siblings between 0-6 years; SIB718 : number of siblings between 7-18 years; SIBGT18: number of siblings above 18 years; PREDEXP : predicted value of per capita household expenditure on food, health and education; HEADMALE: 1 if the household head is a male; HEADEDN1: 1 if the household head has primary education; HEADEDN2 : 1 if the household head has secondary education; HEADEDN3: 1 if the household head has post-secondary education ; HSPEDN1 : 1 if the spouse of the household head has primary education; HSPEDN2: 1if the spouse of the household head has secondary education; HSPEDN3 : 1 if the spouse of the household head has post secondary education; PUBLIC: 1 if the child goes to public school; INDIG: 1 if the household belongs to indigenous community; RURAL: 1 if the household lives in rural location; MKT: local adult market participation rate; OWN: local adult home participation rate; MALEMKT : interaction between MALE and MKT; MALEOWN: interaction between MALE and OWN.

TABLE 2. SEQUENTIAL PROBIT ESTIMATES WITH AND WITHOUT INDIVIDUAL HETEROGENEITY

	Primary vs. no schooling		Secondary vs primary schooling		Post-secondary vs. secondary schooling	
	NO HET♣	HET♣♣	NO HET♣	HET♣♣	NO HET♣	HET♣♣
CONS_1	-1.1939	-1.1356	-19.268 ***	22.6186 ***	19.0667 **	16.8057 *
	2.5222	3.1162	2.2131	2.7805	8.5263	9.7366
AGELE12	0.067	0.0633	1.2436 ***	1.4971 ***	-1.9986 ***	-2.0387 **
	0.097	0.1131	0.1227	0.1441	0.7353	0.8323
AGE1217	-0.0487	-0.0503	0.4495 ***	0.5309 ***	0.6354 ***	0.7316 ***
	0.0325	0.0374	0.021	0.0271	0.1705	0.191
AGEGT17	-0.1280 **	-0.1633 **	-0.1926 ***	-0.2240 ***	0.3949 ***	0.4724 ***
	0.0626	0.0712	0.0446	0.0505	0.0643	0.0774
MALE	-0.0424	-0.0331	0.2018	0.2395	-0.2722	-0.3646
	0.2152	0.2594	0.1239	0.1474	0.2086	0.2512
FAILPRIM			-0.5656 ***	-0.6520 ***		
			0.054	0.0665		
FAILPS					-0.8413 ***	-0.9939 ***
					0.2614	0.3226
MARRIED					-0.6441 *	-0.689
					0.375	0.4435
SIB06	-0.0402	-0.0531	-0.0443	-0.0806	-0.054	-0.0055
	0.0991	0.1246	0.066	0.0879	0.1295	0.1556
SIB718	0.0354	0.0334	0.0274	0.024	-0.035	-0.0074
	0.0695	0.087	0.0479	0.063	0.0864	0.1059
SIBGT18	-0.0384	-0.0429	0.1013 ***	0.1270 ***	0.0344	0.0472
	0.0451	0.0553	0.0246	0.0337	0.0414	0.0531
PREDEXP	0.4989	0.5729	0.9288 **	0.9580 *	0.1742	0.6418
	0.6002	0.7439	0.4414	0.5785	0.7681	0.9432
DEBT	-0.1521	-0.1591	-0.0358	-0.058	0.1466	0.1733
	0.1256	0.1632	0.0738	0.0991	0.125	0.1557
HEADMALE	-0.3081 *	-0.4076 *	-0.0402	-0.1446	-0.0734	-0.1786
	0.1668	0.2133	0.0961	0.1316	0.2081	0.2601
HEADEDN1	0.3172 **	0.4239 **	0.2454 **	0.3998 ***	0.1144	0.2441
	0.1503	0.1928	0.1067	0.1456	0.2459	0.3145
HEADEDN2	0.4982 **	0.6835 **	0.5465 ***	0.7804 ***	0.1741	0.3449
	0.2201	0.2842	0.1582	0.2132	0.3153	0.3967
HEADEDN3	0.4526	0.5716	0.4001	0.6218 *	0.5095	0.6217
	0.3067	0.3829	0.2538	0.3297	0.4129	0.5118
HSPEDN1	0.2408 *	0.2534	0.1313 *	0.2224 **	0.1601	0.2767
	0.1316	0.1652	0.0745	0.1019	0.1652	0.2131
HSPEDN2	-0.1471	-0.2074	0.1484	0.2427	0.1317	0.1846
	0.1905	0.2419	0.1311	0.174	0.2254	0.2822
HSPEDN3	0.308	0.3239	0.3091	0.4351	0.3093	0.3639
	0.4196	0.5026	0.2295	0.298	0.3329	0.4195
PUBLIC			-0.4894 ***	-0.6123 ***	-0.4444 ***	-0.5279 ***
			0.1487	0.1771	0.138	0.1685
INDIG	0.1207	0.0858	0.2892 **	0.2815	0.1176	0.2168

	0.1796	0.2276	0.13	0.1742	0.278	0.3438
RURAL	0.2872 **	0.2772 *	0.3441 ***	0.4178 ***	0.0006	-0.0059
	0.1296	0.16	0.0715	0.0984	0.1665	0.2063
MALEMKT	0.0099	0.0037	-0.0293	-0.0281	0.0653	0.0871
	0.1271	0.15	0.0622	0.0746	0.093	0.1135
MKT	0.1463 *	0.1972 **	0.0442	-0.0583	-0.1375 *	-0.1703 *
	0.0807	0.1006	-0.0462	0.0583	0.0718	0.088
MALEOWN	0.0197	0.0266	-0.0525	-0.048	-0.008	0.0174
	0.0912	0.1111	0.0452	0.0545	0.0703	0.0839
OWN	0.1555 **	0.1974 **	-0.0288	-0.0314	-0.0304	-0.0604
	0.0701	0.0865	0.0337	0.0427	0.0512	0.0634
	Individual-level unobserved heterogeneity					
SIGIND		0.6631 ***				
		0.0648				
	Log-likelihood function					
ln-L	-2169.36	-2136.26				

Note: ♣ : Estimates without heteorogeneity; ♣♣: Estimates with individual heterogeneity. Asymptotic standard errors are shown below each estimate. * denotes that a variable is significant at 10%, ** denotes that at 5% while *** denotes the same at 1% level of significance.

TABLE 3. ORDERED PROBIT ESTIMATES

	No HET	HET
Constant	-0.4926	-0.5531 **
	0.3007	0.2759
AGELE12	0.1330 ***	0.1334 ***
	0.0257	0.0233
AGE1217	0.1228 ***	0.1212 ***
	0.0062	0.0058
AGEGT17	0.0258 **	0.0284 ***
	0.0103	0.0097
MALE	-0.0235	-0.0193
	0.0158	0.0155
MARRIED	-0.1847 ***	-0.1785 ***
	0.0489	0.05
SIB06	-0.0365 ***	-0.0365 ***
	0.0077	0.0101
SIB718	-0.0172 ***	-0.0155 **
	0.0055	0.0073
SIBGT18	0.0237 ***	0.0241 ***
	0.0057	0.0071
EXPQRTL	0.0499 ***	0.0542 ***
	0.0078	0.01
HEADMALE	-0.0424 *	-0.0489
	0.0257	0.0317
HEADEDN1	0.1502 ***	0.1534 ***
	0.0223	0.0298
HEADEDN2	0.2499 ***	0.2569 ***
	0.027	0.0356
HEADEDN3	0.4063 ***	0.4143 ***

	0.0344	0.0428
HSPEDN1	0.0644 ***	0.0634 **
	0.0187	0.0247
HSPEDN2	0.0568 **	0.0564 *
	0.0259	0.0326
HSPEDN3	0.2010 ***	0.2095 ***
	0.0392	0.0462
FAILPRIM	-0.2258 ***	-0.2170 ***
	0.0505	0.049
FAILPS	0.1239 ***	0.1251 ***
	0.0451	0.0438
PUBLIC	0.3935 ***	0.4313 ***
	0.0171	0.0175
SPANISH	-0.0366	-0.0258
	0.0228	0.0298
RURAL	0.1389 ***	0.1337 ***
	0.0174	0.0228
SIGU	0.3928 ***	0.3345 ***
	0.005	0.0068
	Individual-level Unobserved heterogeneity	
SIGDELTA		0.2090 ***
		0.0111
ln-L	-2621.11	-2577.45

Note: Asymptotic standard errors are shown below each estimate. * denotes that a variable is significant at 10%, ** denotes that at 5% while *** denotes the same at 1% level of significance.

REFERENCES

Behrman, J.R. and P. Taubman. 1986. 'Birth Order, Schooling and Earnings', *Journal of Labour Economics*, S121-S145.

Birdsall, N. 1985. 'Public Inputs and Child Schooling in Brazil', *Journal of Development Economics* 18(1), pp. 67-86.

Cochrane, S.H., K. Mehra and I.T. Osheba. 1986. 'The Educational Participation of Egyptian Children'. Education and Training Series Report EDT45. Washington, D.C.: World Bank.

Drèze, J., and Kingdon, G. G., 2001, 'School Participation in Rural India,' *Review of Development Economics*, 5(1), pp. 1-24.

Glewwe, P. and H.Jacoby. 1993. 'Student Achievement and Schooling Choice in Low-Income Countries: Evidence from Ghana', *Journal of Human Resources*, pp. 843-64.

Grootaert, C and H.A. Patrinos. 2000. '*The Policy Analysis of Child Labour: A Comparative Study*', Macmillan, London.

Ilon, L. and P. Moock. 1991. 'School Attributes, Household Characteristics and Demand for Schooling: A Case Study of Rural Peru', *International Review of Education*, 37(4), pp. 429-51.

Kingdon, G. 1998. 'Does the Labour Market Explain Lower Female Schooling in India?' *Journal of Development Studies*, 35(1), pp. 39-65.

Lillard, L. and R.J. Willis. 1994. 'Intergenerational Educational Mobility: Effects of Family and State in Malaysia', *Journal of Human Resources*, pp. 1126-67.

Panis, C.W.A. 1994. 'The Piecewise Linear Spline Transformation with an Application to Age at First Marriage', Stata Technical Bulletin 18, pp. 27-29.

Patrinos, H. A. and G. Psacharopoulos. 1994. 'Indigenous People and Poverty in Latin America: An Empirical Analysis', World Bank, Washington DC.

Patrinos, H. A. and G. Psacharopoulos. 1997. 'Family Size, Schooling and Child Labour in Peru – An Empirical Analysis', *Journal of Population Economics*, pp. 387-405.

Ray, R. 2000a. 'Analysis of Child Labour in Peru and Pakistan: A Comparative Study', *Journal of Population Economics* 13(1), pp. 3-19.

Ray, R. 2000b. 'Child Labour, Child Schooling and Their Interaction with Adult Labour: Empirical Evidence for Peru and Pakistan', *The World Bank Economic Review*, 14(2), pp. 347-67.

Rosenzweig, M. 1986. 'Birth Spacing and Sibling Inequality: Asymmetric Information within the Family', International Economic Review, pp. 55-76.

Singh, R.D., 1992, 'Underinvestment, Low Economic Returns to Education, and the Schooling of Rural Children: Some Evidence from Brazil,' *Economic Development and Cultural Change*, Vol.40, No.3., pp.645-64.

Wolpin, K. 1984. 'An Estimable Dynamic Stochastic Model of Fertility and Child Mortality', *Journal of Political Economy*, pp. 852-74.

APPENDIX

Table A1. Sequential probit estimates excluding sibling composition variables

	Primary vs. no schooling		Secondary vs. primary schooling		Post-secondary vs. secondary schooling	
	NO HET	HET	NO HET	HET	NO HET	HET
CONS_1	-1.1899	-1.389	19.61 ***	23.678 ***	17.8495 **	16.9724 *
	1.2192	1.4575	1.5128	1.825	7.8477	8.6605
AGELE12	0.0692	0.0685	1.2506 ***	1.5117 ***	-2.007 ***	-2.083 ***
	0.0954	0.1111	0.1211	0.1423	0.7139	0.7876
AGE1217	-0.0473	-0.0509	0.4521 ***	0.5320 ***	0.6381 ***	0.7289 ***
	0.0314	0.0359	0.0208	0.0267	0.1676	0.1851
AGEGT17	-0.157 ***	-0.194 ***	-0.174 ***	-0.204 ***	0.4060 ***	0.4829 ***
	0.0554	0.064	0.0433	0.0493	0.0601	0.073
MALE	-0.0473	-0.0421	0.2106 *	0.2504 *	-0.2694	-0.3569
	0.2129	0.2562	0.1215	0.1441	0.2065	0.2488
MARRIED					-0.6386 *	-0.6711
					0.368	0.4374
FAILPRIM			-0.572 ***	-0.657 ***		
			0.0538	0.0662		
FAILPS					-0.847 ***	-0.995 ***
					0.2602	0.3219
PREDEXP	0.4999 ***	0.6306 ***	1.0038 ***	1.2004 ***	0.5128 **	0.7453 **
	0.1828	0.2319	0.1143	0.1539	0.2392	0.2906
DEBT	-0.1485	-0.1627	-0.0568	-0.085	0.1386	0.1607
	0.1211	0.1575	0.0725	0.097	0.1237	0.1539
HEADMALE	-0.2673 *	-0.3545 *	-0.0454	-0.1391	-0.0737	-0.1978
	0.1582	0.2012	0.0937	0.1273	0.1909	0.2422
HEADEDN1	0.3149 **	0.4126 **	0.2328 ***	0.3595 ***	0.0634	0.2304
	0.1333	0.1715	0.0874	0.1197	0.2169	0.2854
HEADEDN2	0.4993 ***	0.6687 ***	0.4964 ***	0.6714 ***	0.0677	0.2995
	0.1673	0.2173	0.1066	0.1461	0.2381	0.3102
HEADEDN3	0.4501 *	0.5459 *	0.3671 **	0.5126 **	0.3583	0.5683
	0.2489	0.3095	0.1872	0.2405	0.2772	0.3584
HSPEDN1	0.2397 **	0.2485 *	0.1179 *	0.1899 **	0.1442	0.2739
	0.1179	0.149	0.0651	0.0891	0.1478	0.1933
HSPEDN2	-0.1388	-0.2118	0.1044	0.162	0.084	0.1654
	0.1679	0.2145	0.1076	0.1416	0.1888	0.2392
HSPEDN3	0.3061	0.3076	0.2208	0.2879	0.2288	0.336
	0.3787	0.4521	0.2059	0.264	0.2548	0.3268
RURAL	0.2747 **	0.2645 *	0.4042 ***	0.4761 ***	-0.007	0.0167
	0.1121	0.1422	0.062	0.0856	0.1563	0.1931
INDIG	0.1252	0.1042	0.2792 ***	0.3009 ***	0.2003	0.2328
	0.1222	0.1548	0.085	0.1154	0.2104	0.2656
PUBLIC			-0.469 ***	-0.581 ***	-0.439 ***	-0.527***
			0.1426	0.1698	0.1369	0.1673
MALEMKT	0.0177	0.0145	-0.0386	-0.0365	0.0648	0.0834
	0.1259	0.1479	0.061	0.0731	0.0927	0.1126
MKT	0.1251 *	0.1711 *	0.0107	0.0088	-0.1228 *	-0.1456 *
	0.0758	0.0947	0.043	0.0537	0.0696	0.0846
MALEOWN	0.0233	0.0296	-0.0522	-0.0548	-0.0518	-0.0069
	0.0894	0.109	0..0602	0.0441	0.053	0.0702
OWN	0.1412 **	0.1821 **	0.0174	0.0262	-0.0213	-0.04

	0.0676	0.0836	0.031	0.0389	0.0487	0.0598
	Individual-level unobserved heterogeneity					
SIGIND		0.6104***				
		0.0870				
ln-L	-2217.03	-2163.24				

NOTE: Asymptotic standard errors are shown below each estimate. Significance: '*'=10%; '**'=5%; '***'=1%.

Table A2. Sequential probit estimates excluding grade repetition variables

	Primary vs. no schooling		Seondary vs. primary schooling		Post-secondary vs. secondary schooling	
	NO HET	HET	NO HET	HET	NO HET	HET
CONS_1	-1.0468	-1.0269	19.5218 ***	23.3173 ***	19.2144 **	17.3638 *
	2.5285	3.1716	2.1802	2.7919	8.2413	9.5455
AGELE12	0.0648	0.0608	1.2213 ***	1.4936 ***	-2.024 ***	-2.143***
	0.097	0.1147	0.1228	0.1459	0.7114	0.8161
AGE1217	-0.0483	-0.0519	0.4390 ***	0.5238 ***	0.6224 ***	0.7324 ***
	0.0325	0.0377	0.0204	0.0267	0.1658	0.1878
AGEGT17	-0.1292 **	-0.1631 **	-0.14***	-0.165 ***	0.4190 ***	0.5114 ***
	0.0626	0.0716	0.0442	0.0501	0.0628	0.0771
MALE	-0.0471	-0.0314	0.186	0.2324	-0.2838	-0.3852
	0.215	0.2625	0.1185	0.143	0.2064	0.2513
MARRIED					-0.5775	0.5495
					0.3671	-0.446
SIB06	-0.0455	-0.0542	-0.0266	-0.0559	-0.0494	0.0142
	0.0993	0.1268	0.0636	0.0867	0.1271	-0.1554
SIB718	0.0346	0.0299	0.0483	0.0451	-0.0247	0.0144
	0.0695	0.088	0.0467	0.0631	0.084	-0.1049
SIBGT18	-0.0387	-0.0445	0.1068 ***	0.1340 ***	0.0416	0.0584
	0.0451	0.0561	0.0242	0.0338	0.0406	-0.0534
PREDEXP	0.4651	0.5596	1.0364 **	1.1187 *	0.2122	0.7755
	0.602	0.7564	0.4272	0.5738	0.7492	-0.9367
DEBT	-0.1538	-0.1606	-0.0033	-0.0233	0.1403	0.1712
	0.1255	0.1662	0.0724	0.0996	0.1219	0.1542
HEADMAL	-0.3098 *	-0.4167 *	-0.0406	-0.1399	-0.1007	-0.2137
	0.167	0.2179	0.0931	0.1314	0.2066	0.2623
HEADEDN1	0.3211 **	0.4292 **	0.1713	0.3260 **	0.0483	0.1722
	0.1503	0.1958	0.1049	0.1472	0.2381	0.3091
HEADEDN2	0.5128 **	0.7061 **	0.4513 ***	0.6773 ***	0.102	0.2698
	0.22	0.2883	0.1541	0.213	0.3072	0.393
HEADEDN3	0.4701	0.5899	0.3608	0.5744 *	0.4647	0.5719
	0.3069	0.3887	0.2485	0.3289	0.4022	0.5077
HSPEDN1	0.2436 *	0.2541	0.1756 **	0.2688 ***	0.2045	0.3316
	0.1315	0.1674	0.0728	0.1023	0.1613	0.2119
HSPEDN2	-0.1465	-0.2139	0.2262 *	0.3314 *	0.1907	0.2539
	0.1906	0.2457	0.1271	0.1733	0.2188	0.2777
HSPEDN3	0.3156	0.3315	0.3806 *	0.5219 *	0.3917	0.4529
	0.4199	0.5091	0.2245	0.2966	0.3259	0.4171
PUBLIC			-0.539 ***	-0.688 ***	-0.469 ***	-0.557 ***
			0.1413	0.172	0.1375	0.1697
RURAL	0.2946 **	0.2861 *	0.3381 ***	0.4122 ***	-0.0182	-0.0213
	0.1296	0.1619	0.069	0.0969	0.1595	0.2009

INDIG	0.1103	0.0763	0.2590 **	0.2523	0.0738	0.1838
	0.1798	0.2311	0.1271	0.1745	0.2642	0.3309
MALEMKT	0.0103	-0.0001	-0.0215	-0.0216	0.0743	0.1027
	0.1271	0.1523	0.0612	0.0744	0.0912	0.1123
MKT	0.1479 *	0.2073 **	-0.0601	-0.0744	-0.1396 **	-0.1763 **
	0.0807	0.1025	0.0444	0.0574	0.0706	0.0875
MALEOWN	0.024	0.0253	-0.0648	-0.0705	-0.0023	0.0251
	0.0909	0.1122	0.0435	0.0533	0.0683	0.0824
OWN	0.1555 **	0.2002 **	-0.022	-0.0216	-0.039	-0.0717
	0.0701	0.0875	0.0322	0.0418	0.0498	0.0628
	Individual-level unobserved heterogeneity					
SIGIND		0.6639***				
		0.064				
Ln-L	-2244.8	-2206.08				

NOTE: Asymptotic standard errors are shown below each estimate. Significance: '*'=10%; '**'=5%; '***'=1%.