
**Lancaster University Management School**
**Working Paper**
**2001/004**

# High School Dropouts: A Longitudinal Analysis

Anh Ngoc Nguyen, Jim Taylor and Steve Bradley

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

# HIGH SCHOOL DROPOUTS: A LONGITUDINAL ANALYSIS

Anh Ngoc Nguyen
Jim Taylor
Steve Bradley

Department of Economics
Lancaster University
Lancaster
United Kingdom

August 2001

# HIGH SCHOOL DROPOUTS: A LONGITUDINAL ANALYSIS


## ABSTRACT

The two aims of this paper are (i) to identify the factors influencing the decision to drop out of high school and (ii) to estimate how the risk of dropping out varies during a student's high school years. The study is based on data from the *National Educational Longitudinal Study 1988-94* and uses event history techniques to investigate the dropout decision. A flexible piecewise-linear baseline hazard is estimated, controlling for unobserved heterogeneity both parametrically and non-parametrically. Factors such as previous educational attainment, ethnicity and a range of family background variables are found to influence the dropout decision. The risk of dropping out is estimated to vary substantially during the student's time in secondary school.


## I. INTRODUCTION

The long-term downward trend in high school non-completion rates experienced during the 1970s and 1980s came to a halt in the 1990s. The percent of those aged 18 to 24 who have not completed high school has remained at around 13% over the past decade. Dropping out of high school is therefore a serious problem in the USA. There is also a significant ethnic dimension to this problem since the proportion who fail to complete high school is higher for blacks and Hispanics than for whites.[1]

The consequences of not completing a high school diploma can be severe. Dropouts are at considerably higher risk of being unemployed and on low income than those who complete high school. They are seriously disadvantaged in the labour market not only by their dropout status but also because they have cut themselves off from further educational opportunities. The

[1] The proportion of 16 to 24 year olds who had not completed high school in 1999, for example, was 12.6% for blacks, 28.6% for Hispanics compared to 7.3% for whites. The dropout rate is high for Hispanic youths because no distinction is made in calculating this rate between those Hispanics born in the USA and those who are immigrants. More than half of the foreign-born Hispanic youths who were classified as dropouts had never been enrolled in a US school (NCES 2000, p.14). See also *Statistical Abstract of the United States 1999*, Table 302: High School Dropouts, by Race and Hispanic Origin.

unemployment rate for persons aged 25 and over in 1998, for example, was 7.1% for those who did not complete high school compared to 4.0% for those whose highest educational attainment was completion of high school (National Centre for Education Statistics 2000). The relative income gap between these two groups was similarly substantial.[2] The costs of high school dropouts are borne not only by the dropouts themselves but also by society as a whole. High school dropouts, for example, are far more likely to receive public assistance and dropouts make up a disproportionate percentage of the nation's prison population (NCES 2000, p.1). Reducing the dropout rate could therefore lead to lower welfare payments and lower crime rates. An additional benefit could be higher productivity growth since raising educational attainment levels will enhance the stock of human capital and facilitate the diffusion and transmission of knowledge.

This paper has two main aims. The first is to identify the factors influencing the decision to drop out of high school. The second is to investigate how the risk (or hazard) of dropping out varies during a person's high school years. The study is based on data obtained from the *National Educational Longitudinal Study 1988-94* (NELS).[3] This longitudinal data set followed a nationally representative sample of students who were 8$^{th}$ graders in 1988 through to 1994. The present study differs from previous work on high school dropouts in so far as we exploit the longitudinal nature of the NELS data set by using event history techniques to investigate the dropout decision.[4] In the present paper, we estimate a flexible piecewise-linear baseline hazard model, as suggested by Meyer (1990, 1995), and simultaneously control for unobserved

[2] In 1997, the median income for males who did not complete high school was $17,000 compared to $26,000 for those whose highest level of attainment was completion of high school. Source: *Money Income in the United States*, Department of Commerce, Bureau of the Census, Current Population Reports. See also Zemensky *et al.* (1998).
[3] Our study is based on the public-use data set since the restricted-use data set is available only within the United States.

heterogeneity both parametrically and non-parametrically (Lancaster 1979; Heckman and Singer 1984a, 1984b; Stewart 1996). By using duration modelling techniques, we are therefore able to investigate not only the determinants of dropping out but also its timing. In comparison to previous cross-sectional studies of the dropout decision, our approach enables us to control for unobserved differences between individuals and hence address the issue of state dependence.

The remainder of the paper is in six parts. Section II provides a discussion of theoretical models of the dropout decision and a review of previous studies. This is followed in section III by a description of the data and the variables used in the econometric analysis. Section IV discusses the econometric modelling procedures that we adopt, while section V discusses the results. Section VI concludes.

## II. WHY DO STUDENTS DROP OUT OF HIGH SCHOOL?

### *Theoretical considerations*

Human capital theory explains the decision to invest in education (Becker 1964, 1967, 1993; Becker and Tomes 1986). Since the decision to drop out of high school is simply the corollary of the decision to stay on, it follows that human capital theory is also a useful vehicle for explaining why some high school students become dropouts. In deciding whether to stay on or terminate schooling, a person is assumed to compare the present value of the expected benefits, such as increased lifetime earnings, with the present value of expected costs of tuition and foregone earnings. An individual will terminate formal schooling if the expected discounted costs exceed the expected discounted benefits. The optimal amount of time spent in education will also differ

[4] Duration models are still not used extensively in education economics. Recent examples of their use include Booth and Satchell (1995), Ehrenberg and Mavros (1995), van Ours and Ridder (2000), who investigate PhD completions, and DesJardins, Ahlburg and McCall (1999), who investigate college dropouts.

between individuals because they use a subjective discount rate when calculating the expected benefits and expected costs of schooling.

The amount invested in education will depend on two primary factors. First, a person's innate ability, which has a positive effect on the demand for education since the expected returns from investment in education will be higher for those with greater ability. More able students are consequently less likely to drop out. Second, since parents are concerned about the future welfare of their offspring, family income will influence the length of time their child continues in education. Poor families will have greater difficulty financing the education of their children not only because they have other competing expenditures but also because the cost of borrowing may be prohibitively high. Youths from low-income families are therefore more likely to terminate their education earlier than are youths from high-income families.

Leibowitz (1973) and Becker and Tomes (1986) provide a more detailed discussion of the major potential routes through which a child's family background may influence the dropout decision. First, genetic endowments are passed from parents to children via heredity, which partly determines a child's ability to benefit from schooling. Second, family income and wealth will determine the quantity of educational inputs purchased by parents for their children. Third, the education level reached by parents may affect the quality of home tuition and the amount of time they devote to home tuition.

*Previous literature*

The literature on high school dropouts and graduation goes well beyond the boundaries of economics. Rosenthal (1998), for example, reviews a wide range of empirical studies that aim to identify the non-school correlates of high school dropouts. According to Rosenthal, however, this research "is not at a very high technical level" with the result that "it is not possible to

predict dropout accurately nor to conceptualize adequately the dynamics of dropping out" (Rosenthal 1998, p. 429). Research by economists into high school dropouts can be grouped into three broad categories. The first category includes a wide range of family background variables, such as parental education, parents' occupation, working status of parents, family size and family structure, as suggested by human capital theory.

Parental education is often found to be the most important family related factor in determining high school graduation and dropping out. As the education level of parents increases the probability of becoming a dropout falls (Hill 1979, Mare 1980, Rumberger 1983, Manski et al 1992, Haveman, Wolfe and Spauling 1991, Evans et al 1992, Evans et al 1995, Sander et al 1995 Neal 1997). A potential drawback of these earlier studies is that they fail to allow for the possible non-linearity in the effect of parental education. In addition, many studies ignore the qualitative aspect of education by using years of schooling as a proxy for parental education. The effect of parental education on dropping out may be expected to vary, however, according to the level of education attained.

Parental occupation, parents' employment status and family income, which capture the effect of financial resources, also influence the probability of dropping out. Previous studies provide convincing evidence that youths from higher income families or whose parents are in higher occupational groups are more likely to graduate from high school and are less likely to drop out (Hill 1979; Marc 1980; Rumberger 1983; Neal 1997; Jonsson *et al.* 1997).

Family size and structure are also important determinants of dropping out. Several researchers have argued that there is a trade-off between the quantity and quality of children and this may influence decisions over family size (Hanushek 1992). The amount of resources, including parental time, available within the family to educate each child will fall as family size increases. Several studies have found that the number of siblings is negatively related to the

probability of graduating from high school (Mare 1980; Haveman, Wolfe and Spaulding.1991; Kane 1994; Jonsson *et al.* 1997) and positively related to the probability of dropping out (Hill 1979; Rumberger 1983; Evans *et al.* 1992). The influence of family structure has been investigated in various ways, the primary objective being to estimate the effect of being in a non-intact family (i.e. not living with both parents). Being in a single-parent family, for example, may reduce the resources available to invest in education (McLanahan 1985; Manski *et al.* 1992). It has also been argued that children from non-intact families are more likely to have suffered from emotional upheaval and disturbed social relations during family break-up (Jonsson *et al.* 1997).

The second category of determinants, which has received the least attention, relates to the personal characteristics of dropouts. The most consistent result of previous studies is the highly significant negative relationship between dropping out and measures of ability, such as test scores (Mare 1980; McElroy 1996; Chuang 1997). A common problem with earlier studies is that they do not take account of the endogeneity of test scores since students who regard education as a poor investment have less incentive to perform well in tests. Students intending to drop out are likely to do badly in tests. Bishop and Mane (2001) use prior test score in $8^{th}$ grade to overcome this problem.

The common perception is that youths from black and Hispanic ethnic minorities are more likely to drop out than whites. This perception stems directly from the raw dropout data, which indicates that (unconditional) dropout rates are higher for blacks and Hispanics than for whites. After conditioning for family related factors, however, the picture is less clear cut. Several studies indicate that blacks are more likely to graduate from high school and are less likely to drop out than whites, whereas the opposite has been found for Hispanic youths (Manski *et al.* 1991; Sander *et al.* 1995; Chuang 1997).

The third set of factors influencing the decision to drop out relates to wider environmental factors such as schooling and the effect of neighbourhood. In comparison with the voluminous research into the effect of schooling on educational attainment (Hanushek 1986, 1996; Haveman and Wolfe 1995; Wössman 2000), there is a paucity of research on the effect of school-related factors on high school graduation and dropping out. Attendance at a Catholic school, however, is highly negatively correlated with the probability of dropping out (Evans and Schultz 1995; Sander and Krautmann 1995; Neal 1997). Other studies attempt to control for school quality by including variables such as the dropout rate of the student's school (Astone *et al.* 1991), student / teacher ratio, percent of students eligible for free lunch, socio-economic status of each school's students and ethnic mix (Bishop and Mane 2001).

Controls have also been included to capture peer group effects. These include the proportion of a school's pupils who are economically disadvantaged due to low family income. A serious problem with such variables is that they are endogenous in models of educational outcomes (Evans, Oates and Schwab 1992). This is because the residential location decision of families with children is likely to be influenced by the quality of a neighbourhood's schools, thereby making peer group jointly determined with choice of neighbourhood and school. Controlling for the endogeneity of the explanatory variables, however, is often difficult due to data constraints.

## III. DATA AND VARIABLES

Our analysis of high school dropouts is based on data from the *National Educational Longitudinal Study 1988-94* (NELS). The NELS is a nationally representative sample of youths taken from over 1000 schools with more than 20,000 youths participating. Respondents were

first interviewed in 1988, the base year and then at two-yearly intervals (1990, 1992 and 1994). By 1994, 14915 participants were included in the third follow-up.[5] The base-year survey collected information from the youths themselves, their parents, school administrators and teachers. The first and second follow-ups were undertaken when most students were in their 10th and 12th grades respectively. The third follow-up survey was undertaken two years after most youths would normally have graduated from high school.

The NELS contains a detailed history of the high school experience of respondents, including whether or not they dropped out, whether they had permanently or temporarily dropped out, and when they dropped out for the first time. The survey provides information for 14527 students who proceeded to 9th grade, 1906 of whom (13.1%) were classified as dropouts by the time of the third sweep in the 12th grade. Most students graduated during May to June 1992. After the end of June 1992, very few dropouts were identified because the number of students at risk of dropping out fell from 12307 to 754 at this time.

An advantage of the NELS is that separate surveys were conducted for dropouts, which limits the problems arising from missing data. As the initial sweep of the NELS was undertaken when students were in 8th grade, the survey provides information (such as test scores) relating to each student's situation in the year preceding high school. The availability of prior information helps to overcome the problem of the endogeneity of explanatory variables. The main characteristics of high school dropouts are shown in Table 1.[6] Figure 1 shows that the frequency of dropping out varies over time in two ways. First, dropping out is more likely to occur in the first term of the school year, September/December, and less likely to occur during May/August, than during the period January/April. Second, and of greater interest here, there is a clear upward

[5] All follow-ups were designed to be representative of the national population through a process of refreshing the database.

trend in the dropout rate as students progress through the 9th to the 12th grade. Dropping out is therefore more likely to occur in higher grades than in lower grades.

*Dependent variable*

A drawback of the NELS is that it only allows first spell analysis of dropouts. We can identify the time at which a student first dropped out but not the time those who temporarily dropped out subsequently returned to school. It is therefore necessary to restrict our analysis of the dropout event to the initial continuous spell in high school. The dependent variable is the time elapsed between the beginning of the 9th grade and the first time a student exits. Specifically, duration in school is measured as the number of months elapsing between the start date in 9th grade (1st September 1988) and the exit date, or alternatively between the start date and the last date a student was observed in the survey (30th August 1994). This definition of dropouts is broader than the definition used by the National Centre for Education Statistics (NCES), which includes only those who do not achieve any high school qualification. Under the NCES definition, those with the General Education Diploma (GED) are not considered as high school dropouts. Under the definition of dropouts used in this paper, students are considered as dropouts even if they subsequently return to school or obtain the GED.[7] The inclusion of those who dropped out of high school but later obtained the GED is justified on the grounds that these particular dropouts are "not identical to high school graduates" and are "statistically indistinguishable from high school dropouts" (Cameron and Heckman 1993, p.2). Moreover, the question of ever having dropped out has its own importance, since even if dropouts return to the education system to

[6] The dropout rate is measured here as the proportion of students who had ever dropped out by the 12th grade (regardless of whether or not they later returned to complete their secondary education).
[7] The proportion of the sample obtaining the GED by 1994 was 5.2%.

obtain the GED (or any other qualification), their life course might still be adversely affected by having been a dropout.

*Explanatory variables*

The selection of variables is guided by the discussion in section II. Data availability also imposes constraints on the selection of variables.[8] Table 1 provides some descriptive statistics.

Variables reflecting the individual's personal characteristics, such as age, gender and ethnic origin are included in the model. Unlike previous studies, ethnic origin is observed for five separate groups, namely non-Hispanic white, non-Hispanic black, Hispanic, Asian, and American Indians.[9] A variable of particular interest is the academic attainment of students. Although separate test scores are available for reading, math and science, the composite test score achieved in 8$^{th}$ grade is used since it averages out the heterogeneity that may be present in individual test scores. It can also be genuinely regarded as an exogenous determinant of the dropout decision in high school.

Parental and family background are proxied by variables on family size, family structure, family income, parental education, parental working status and parental occupation. Of these the only variable that differs substantially from previous research is parental education. Several binary variables are constructed to capture the qualitative and possibly non-linear effect of parental educational attainment on the drop out decision of their children. The quality of schooling is reflected by variables for the type of school attended (public versus non-public

[8] A number of additional variables describing the behaviour of the respondents have been used in previous studies of dropouts based on the NELS:88 data set (Bishop and Mane 2001). These include, for example: TV and homework hours, reading for pleasure, number of books in the home, hours working for pay and an index for smoking. Since many of these variables are highly collinear and are highly correlated with an individual's test score, we have taken a more parsimonious approach in this paper in order to focus on what we believe to be the major potential determinants of the dropout decision.

school) and school size (measured by the number of pupils). The effect of neighbourhood and wider environmental factors are indicated by the type of area in which a student was living while at school (i.e. urban, suburban or rural).[10]

## IV. ECONOMETRIC METHODOLOGY

This section describes the econometric methods for estimating the determinants of the dropout decision (Andrews, Bradley and Stott 2001; Andrews, Bradley and Upward, 2001; Stewart 1996). Since we are interested in the timing of dropping out, a single risk hazard model is estimated. Recall that individuals can exit from school by either dropping out or by graduating. A small number of individuals (149) are also right-censored by the termination date of the survey.[11]The NELS records the number of months that the individual is observed in school. Our approach to analysing this data is to employ discrete-time duration methods.

A panel of individuals is constructed with the $i$-th individual contributing $j = 1, 2, \ldots, t_i$ observations. More precisely, the data is organised into sequential binary response form (Prentice and Gloecker 1978, Meyer 1990, 1995) where all observations $y_{ij}$ are zero except the period in which a dropout occurs. We therefore create a person-period data set in which each individual contributes as many 'pseudo observations' as there are periods that each individual is at risk of dropping out. The maximum number of periods (months) that an individual can be at risk is 72. However, because of 'data thinning' we are forced to collapse the observations into

[9] It should be noted that the sample size for American Indians is small (n = 212) and so the statistical findings for this ethnic group need to be interpreted cautiously.
[10] Several other variables were initially included in the statistical analysis, such as the proportion of pupils on free lunch, but were found to be statistically insignificant and were therefore dropped.
[11] We are not interested in estimating the parameters of the censored distribution and hence the censored are grouped with those who graduate. Since all individuals have a common start date and begin in the same state (i.e. 9th grade) the problems of left censoring and 'initial conditions' do not arise.

two-month intervals. Moreover, since there were very few dropouts during the final two-year period following the end of the 12th grade (May to June 1992), the intervals from 24 to 36 are also grouped into a single interval, giving 24 intervals in total.

The likelihood for the *i*-th individual is then given as:

$$L_i = \prod_{j=1}^{t_i} h_j(\mathbf{x}_i')^{y_{it}} [1 - h_j(\mathbf{x}_i')]^{1-y_{it}} \quad (1)$$

where $h_j$ is the hazard of exiting from school and $\mathbf{x}_i'$ is a vector of explanatory variables. We assume proportional hazards:

$$h_j(\mathbf{x}_i') = \bar{h}_j \exp(\mathbf{x}_i'\boldsymbol{\beta}) \quad (2)$$

where $\bar{h}_j$ is the baseline hazard. This is estimated non-parametrically to avoid the possibility of mis-specification. The explanatory variables affect the hazard via the complementary log-log link function:

$$h_j(\mathbf{x}_i') = 1 - \exp(-\exp(\mathbf{x}_i'\boldsymbol{\beta}) + \gamma_j) \quad (3)$$

Note that $\gamma_j \approx \log \bar{h}_j$ when the parameter estimates of the explanatory variables are set to zero, which means that the $\gamma_j$ can be interpreted as the log of a non-parametric piecewise linear baseline hazard.

It is also important to control for the possible effects of unobserved heterogeneity. Failure to control for unobserved differences between individuals may cause severe bias in the

estimation of the baseline hazard (Heckman and Singer 1984a, 1984b, Lancaster 1990). Standard practice suggests that it is possible to control for unobserved heterogeneity by including a positive-valued random variable ($v$), or mixture, into our model as follows:

$$h_j(\mathbf{x}_i', v_i) = \bar{h}_j \exp(\mathbf{x}_i'\boldsymbol{\beta})v_i \tag{5}$$

where $v$ represents the unobserved heterogeneity. The above model can be written as follows:

$$h_j(\mathbf{x}_i') = \bar{h}_j \exp(\mathbf{x}_i'\boldsymbol{\beta} + u) \tag{6}$$

where $u = \log v$ with density $f_u(u)$. The amended likelihood can be written as

$$L(\boldsymbol{\beta}, \gamma) = \prod_{i=1}^{n} \int_{-\infty}^{+\infty} \left[ \prod_{j=1}^{t} h_j(\mathbf{x}_i', u)^{y_{it}} [1 - h_j(\mathbf{x}_i', u)]^{1-y_{it}} \right] f_u(u)du \tag{7}$$

where $h_j(\mathbf{x}_i', u_i) = 1 - \exp[-\exp(\mathbf{x}_i'\boldsymbol{\beta} + \gamma_j + u_i)]$.

Two approaches have been used to model the unobserved heterogeneity. The first is to assume a particular parametric distribution for the heterogeneity term. In this case $f_u(u_i)$ in Equation (7) is replaced by parameters from, for instance, the Gamma or the Gaussian distribution. The advantage of the Gamma distribution is its analytical tractability since it yields a closed form solution (Lancaster 1979; Meyer 1990). The Gaussian distribution is justified on the grounds that the heterogeneity term might capture the large number of unobserved

characteristics (Narendrenathan and Stewart 1993; Stewart 1996). The problem with specifying a parametric distribution for the heterogeneity term is that the estimated parameters may be sensitive to the particular distribution adopted, especially where the baseline hazard is not sufficiently flexible.[12] An alternative approach suggested by Heckman and Singer (1984b) is to use the mass point technique, which approximates a continuous distribution by a finite discrete distribution of unrestricted form. The $u_i$ and $f_u(u_i)$ are replaced by a discrete mass point approximation. We adopt both the parametric and the mass point method in this paper so that we can check the sensitivity of the results to different methods (Andrews, Bradley and Stott 2001).

## V. RESULTS

Four models have been estimated. Model 1 assumes there is no unobserved heterogeneity and is included for comparative purposes only. The three remaining models assume the presence of unobserved heterogeneity but differ in the specification of the underlying distribution. Models 2 and 3 assume that the unobserved heterogeneity have a Gamma and a Gaussian distribution respectively, whereas model 4 assumes a discrete distribution with two mass points.[13] All four models are reported for the whole sample (see Table 2), whereas only model 4 is reported for the separate female and male analyses (see Table 3). Females and males were investigated separately to explore gender differences in the impact of the explanatory variables on the dropout decision. Since the heterogeneity parameters indicate the presence of unobserved heterogeneity, the

[12] Meyer (1990) suggests that the sensitivity of the parameters to the choice of distribution for the heterogeneity term may be a result of mis-specification of the duration distribution and that if the baseline hazard is sufficiently flexible the sensitivity problem may be less severe.

[13] The estimated mass points (-1.13 and 0.46) are asymmetric around the expected value of zero, indicating that the population does possess unobserved heterogeneity. Experimentation indicated that two mass points are sufficient to approximate the underlying distribution for the unobserved heterogeneity. The results are virtually unchanged when the number of mass points is increased to three.

discussion of the results is confined to models 2, 3 and 4. There is little difference between the three models in terms of the value of the log likelihood, and in view of its advantages over parametric approaches, our the discussion of the results and baseline hazards is restricted to the Heckman-Singer model.[14] To make it easier to discuss and interpret the results, the odds ratio (which is the risk of dropping out compared to the base group) has been calculated and included in the tables.

In terms of the baseline hazards, the main results are as follows. First, the estimated baseline hazard for the full sample is very similar for all models. The risk of dropping out increases steadily during the first three years in high school and then jumps onto a higher plane in the fourth and final year (see Figure 2). There is also a distinct seasonal pattern to dropping out, with the risk of dropping out being twice as high in September to December as in January to April. This is not surprising since the risk of dropping out is likely to be highest after the summer break due to students failing to return to school. Second, controlling for unobserved heterogeneity shifts the hazard upward, particularly in later periods. The fact that it remains positively sloped indicates that there is positive duration dependence in the hazard of dropping out. Third, although the shape of the baseline hazard is very similar for females and males, Figure 3 shows that it is markedly higher for females.

*Personal characteristics*

The most striking result is that the estimated coefficients on the ethnic minority variables are all negative and significant, indicating that all of the ethnic minority groups have a lower risk of dropping out than whites after controlling for the other variables in the model. The actual

[14] The odds ratio (or relative risk) is obtained by exponentiating the estimated regression coefficient, which is the log of the odds ratio.

dropout rates for black and Hispanic students, for example, are over 20% compared to 13% for whites and 5% for Asians. After controlling for all other variables, however, the odds ratio indicates that the risk of black students dropping out is 52% compared to whites (i.e. 48% lower than the white base group). For Hispanics and Asians, the risk of dropping out is 82% and 65% respectively compared to whites. American Indians also have a lower risk of dropping out than whites but the estimated coefficient is not statistically significant. Once controls are added for other determinants of dropping out, it is clear that students from ethnic minorities have a lower risk of dropping out than whites. The fact that actual dropout rates are nearly twice as high for blacks and Hispanics as for whites must therefore be a result of the adverse effects of other characteristics. Blacks, for example, are far more likely to be in the lowest quintile of test scores (41%) than whites (15%); and Hispanic students are far more likely to have parents who did not complete high school (33%) than whites (7%). Blacks and Hispanics are therefore far more likely than whites to possess characteristics that increase their probability of dropping out.

A variable with a very large influence on dropping out is a student's test score in the year before entering high school. The result that high achievers are much less likely to drop out than low achievers is closely in line with previous studies (Bishop and Mane 2001). The estimated coefficient on the composite test score indicates that an increase of one percentage point in the test score reduces the hazard rate by 4%. Hence, an increase in the test score by one standard deviation above the mean (10 percentage points) reduces the risk of dropping out to 60% (i.e. 40% less than those with a mean test score). This backs up the claim that high academic achievers have more to gain from schooling than low achievers and are consequently far less likely to drop out.

Older students have a much higher risk of dropping out compared to younger students. The effects are large. Those students who are one year behind their own age cohort, for example,

have a risk of dropping out that is over twice that for those students in the appropriate grade for their age. The risk for those held back two years is over six times greater. The opposite result holds for those students who progress to a higher grade more quickly than their peers in the same age cohort. The risk of dropping out for this elite group of students is 70% of the risk for the base group. The year cohort indicator may therefore be standing as a proxy for ability. The results also indicate that females have a 33% higher risk of dropping out than males.

*Family background*

Several family background factors are significantly related to the risk of dropping out. Family structure is one factor of major significance. Being in a single parent family, or in a family in which one of the parents is living with a partner, substantially increases the hazard of dropping out. The risk of dropping out is 50% higher for those in a single parent family and 37% higher for those living with a parent-cum-partner compared to students who live in an intact family. The other aspect of family background that has a strong effect on dropping out is parental education. Moreover, the results indicate that parental education has a non-linear effect on dropping out. The risk of dropping out for individuals with parents who are high school graduates or have some college education is 75% of that for those whose parents did not complete high school. The risk falls to 55% for individuals who have graduated from college and to 41% for those whose parents have a higher degree.

Several other family background variables, such as parental occupation, have less effect on the hazard of dropping out. The parental occupation variables are generally statistically insignificant, apart from the result that students with a father in either a professional or

managerial occupation have a lower risk of dropping out.[15] The risk for students whose father has a professional occupation, for example, is 56% of the risk for those whose father is in the base group (unskilled or semi-skilled). The lack of significance of the parental occupation variables is due to the fact that the effect of a student's socio-economic background is already taken into account by the inclusion in the model of parental education. When the parental education variables are excluded from the model, the occupational variables become highly significant.

*School and neighbourhood effects*

By far the largest school-related factor influencing the hazard of dropping out is attendance at a non-public school. Compared to attendance at a public school, the risk of dropping out for students attending a non-public school is 39% (i.e. 61% lower than those attending a public school). School size, however, is not found to be a significant factor in affecting the risk of dropping out.[16] There is also some evidence that students living in the South are at higher risk of dropping out than those in other parts of the country, a result obtained in earlier studies. Living in the suburbs or in rural areas also reduces the risk of dropping out, but the estimated coefficients are not statistically significant.

*Exploring gender differences*

Separate analyses have been undertaken for females and males in order to investigate gender differences in the impact of the explanatory variables (see Table 3). In general, there is a close correspondence between the two sets of results. For several variables, however, the magnitude of

[15] Variables representing mother's occupation were found to be statistically insignificant.

the impact varies substantially between females and males. The most striking differences are as follows. First, although the risk of dropping out is significantly lower for blacks than for whites, the risk is lower for black females than for black males (42% for black females compared to 56% for black males). The corresponding risks for Hispanics are 53% for females and 75% for males. Second, the risk of dropping out is lower for students whose parents are more highly educated, but again the risk is lower for females than for males. For those whose parents have a college degree, for example, the risk of dropping out (compared to those whose parents did not graduate from high school) is 36% for females and 45% for males. Third, students who attend a non-public school are less likely to drop out than those attending a public school, but the impact is greater for females than for males (33% for females compared to 42% for males). These results indicate that the variables influencing the decision to drop out have a more powerful effect on females than on males. This explains why the baseline hazard is higher for females since the baseline hazard is calculated for the base group.

## VI. CONCLUSION

The aims of this paper were twofold: first, to identify the factors influencing the decision to drop out of high school; and second, to investigate how the risk of dropping out varies during a student's high school years. It has been possible to investigate these two tasks simultaneously by using data obtained from the *National Educational Longitudinal Study 1988-94*.

[16] Other school level variables, such as the proportion of pupils on free lunch, were found not to be statistically significant and were excluded from the estimated equations. A parsimonious approach is justified on the grounds that computer time increases rapidly as the number of explanatory variables increases.

The present paper exploits the longitudinal nature of the data provided by the NELS by using event history techniques to investigate the dropout decision. Using duration modelling techniques, we have estimated the effects of a wide range of personal characteristics, family background and school-related factors on the dropout decision. Unlike previous cross-sectional research, we have also allowed for unobserved heterogeneity in our modelling and this has enabled us to identify the duration dependence in the hazard of dropping out of high school.

The main findings of this study are as follows. First, the hazard of dropping out of high school rises steadily during the first three years of high school and then jumps onto a higher plane in the fourth and final year. Second, by far the most important determinant of dropping out is a person's educational attainment level. The risk of dropping out is far higher for those with low attainment levels than for those with high attainment levels. The estimates suggest that a 10 percentage point reduction in the test score (i.e. one standard deviation) increases the risk of dropping out by 40%. This result is reinforced by the enormous increase in the risk of dropping out for those held back a year (or more) in school. Third, the risk of dropping out is substantially lower for blacks and Asians than for whites. The risk is also lower for Hispanics and American Indians (than for whites) but the differences are less marked in these two cases. The substantially lower estimated risk of dropping out for blacks is apparently at odds with the far higher *actual* dropout rate for blacks than for whites. But this discrepancy is explained by the fact that blacks have a much higher probability of dropping out because their educational attainment on entry to high school is much lower than whites. In addition, blacks are disadvantaged by factors such as the educational attainment levels achieved by their parents and by the family's economic circumstances. Fourth, it is clear from the estimated models that family background has a powerful impact on the risk of dropping out.

The results obtained in this paper have several implications for policy. First and foremost, there is an obvious need to improve the educational attainment of youths before they enter high school since this is very highly correlated with educational outcomes later on. Improving educational outcomes at the pre-high school stage will have the dual effect of reducing the overall dropout rate and reducing the dropout gap between blacks and whites. The second main implication is that the high dropout rates are part of a much wider problem of economic and social disadvantage. It is unlikely that major inroads into reducing the dropout rate will be made unless these disadvantages can be substantially reduced. Third, since those at greatest risk of dropping out are more likely to come from families in which the parents have poor educational attainment levels and low incomes, there is a need to focus school-based methods of reducing the dropout rate on students with these specific family background characteristics. Fourth, since the risk of dropping out rises dramatically in the final year of high school, this suggests that greater efforts are needed to discourage students from dropping out so late in the day.

REFERENCES

Andrews, M., Bradley, S., and Stott, D., (2001), 'The school-to-work transition, skill preferences and matching', mimeo, Department of Economics, Lancaster University.

Andrews, M., Bradley, S., and Upward, R., (2001), 'Employer search, vacancy duration and skill shortages', mimeo, Department of Economics, Lancaster University.

Astone, Nan, Marie and McLanahan, Sara, S., (1991), 'Family structure, parental practices and high school completion', *American Sociological Review*, Vol. 56, pp. 309-320.

Becker, G. S. (1964), *Human Capital: A Theoretical And Empirical Analysis, With Special Reference To Education*, Columbia University Press, New York.

Becker, G. S. (1967), 'Human capital and the personal distribution of income: an analytical approach', Woytinsky Lecture No.1, Institute of Public Administration, University of Michigan.

Becker, G.S., (1993), *Human Capital: A Theoretical And Empirical Analysis, With Special Reference to Education*, University of Chicago Press, Chicago.

Becker, G. S. and Tomes, N., (1986), 'Human capital and the rise and fall of family', *Journal of Labour Economics*, Vol. 4 (3), pp. S1-39.

Bishop, J. H. and Mane, F., (2001), 'The impacts of minimum competency exam graduation requirements on high school graduation, college attendance and early labor market success', *Labour Economics*, Vol. 8, 203-222.

Booth, L., A. and Datchell E. Stephen (1995), 'The hazards of doing a PhD: an analysis of completion and withdrawal rates of British PhD students in the 1980s', *Journal of the Royal Statistical Society A*, Vol. 158, pp. 297-318.

Cameron, and Heckman, J. (1993), 'The non-equivalence of high school equivalents', *Journal of Labor Economics*, Vol.11 (1), pp. 1-47.

Chuang, Hwei-Lin, (1997), 'High school youths' dropout and re-enrolment behaviour', *Economics of Education Review,* Vol. 16 (2), pp. 171-186.

Currie, J, and Thomas , D., (1999), 'Early test scores, socio-economic status and future outcomes', NBER Working Paper No. W6943.

DesJardins S.L., Ahlburg D.A. and McCall B.P. (1999), 'An event history model of student departure', *Economics of Education Review*, Vol. 18, pp. 375-390.

Ehrenberg G. Ronald and Mavros G. Panagiotos (1995), 'Do doctoral students' financial support patterns affect their time-to-degree and completion probability', *Journal of Human Resources*, Vol. 30, pp. 581-609.

Evans, William, N., Oates, Wallace, E., and Schwab, Robert, M., (1992), 'Measuring peer group effects: A study of teenage behavior', *Journal of Political Economy*, Vol. 100 (5), pp. 966-991.

Evans W., and Schwab R., (1995), 'Finishing high school and starting college: Do catholic schools make a difference?', *Quarterly Journal of Economics*, Vol. 110, pp. 941-974.

Hanushek, E., (1986), 'The economics of schooling', *Journal of Economic Literature*, Vol. 24, pp. 1141-1177.

Hanushek, E., (1992), 'The trade-off between child quantity and quality' Journal of Political Economy, Vol. 100(1), pp. 84117.
Hanushek, E., (1996), 'School resources and student performance' in *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by Burtless, G., Brookings Institution Press, Washington.

Haveman, R., Wolfe, B., and Spaulding, J., (1991), 'Childhood events and circumstances influencing high school completion', *Demography*, Vol. 28 (1), pp. 133-57.

Haveman, R., and Wolfe, B., (1995), 'The determinants of children's attainments: A review of methods and findings', *Journal of Economic Literature*, Vol. 33, pp. 1829-78.

Heckman, J., and Singer, B., (1984a), 'Econometric duration analysis', *Journal of Econometrics*, Vol. 24, pp. 63-132.

Heckman, J., and Singer, B., (1984b), 'A method for minimising the impact of distributional assumptions in econometric models of duration', *Econometrica*, Vol. 52 (2), pp. 271-320.

Hill, Russell C., (1979), 'Capacities, opportunities and educational investments: The case of the high school dropout', *Review of Economics and Statistics*, Vol. 61 (1), pp. 9-20.

Kane, Thomas, J. (1994), 'College entry by blacks since 1970: The role of college costs, family background, and the returns to education', *Journal of Political Economy*, Vol. 102 (5), pp. 878-991.

Lancaster, T., (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Leibowitz, A., (1974), 'Home Investments in children', *Journal of Political Economy*, Vol. 82, No. 2, part 2, pp. S111-S131.

Manski, C., Sandefur, G., McLanahan, S., and Powers D., (1992), 'Alternative estimates of the effects of family structure during adolescence on high school graduation', *Journal of the American Statistical Association*, Vol. 87 (417), pp. 25-37.

Mare, Robert, D., (1980), 'Social background and school continuation decisions', *Journal of the American Statistical Association*, Vol. 75 (370), pp. 295-305.

McElroy, Susan W., (1996), 'Early childbearing, high school completion, and college enrollment: Evidence from 1980 high school sophomores' *Economics of Education Review*, Vol. 15 (3), pp. 303-324.

McLanahan, Sara, (1985), 'Family structure and the reproduction of poverty', *American Journal of Sociology*, Vol. 90, pp. 873-901

Meyer, Bruce D., (1990), 'Unemployment insurance and unemployment spells', *Econometrica*, Vol. 58 (4), pp. 757-782.

Meyer, Bruce D., (1995), 'Semiparametric estimation of hazard models', unpublished paper, Department of Economics, Northwestern University.

Narendrathan, W., and Stewart M. B., (1993), 'Modelling the probability of leaving unemployment: Competing risks models with flexible baseline hazards', *Applied Statistics*, Vol. 42(1), pp.63-83.

National Center for Education Statistics, (2000), *Dropout Rates in the United States: 1999,* NCES 2001-022, by P Kaufman, J. Y. Kwon, S. Klein and C.D. Chapman. Washington, D.C.

Neal, Derek, (1997), 'The effects of catholic secondary schooling on educational achievement', *Journal of Labour Economics*, Vol. 15 (1), pp. 98-123.

Ours, J. van and Ridder, G. (2000), 'Fast track or failure: A study of the completion rates of graduate students in economics', Discussion Paper No. 2363, Centre for Economic Policy Research.

Prentice, R.L, and Gloeckler, L.A. (1978), 'Regression analysis of grouped survival data with application to breast cancer data' *Biometrics*, Vol. 34, pp. 57-67.

Rosenthal B. S., (1998), 'Non-school correlates of dropout: An integrative review of the literature'**,** *Children and Youth Services Review,* Vol. 20 (5), pp. 413-433

Rumberger, Russell W., 1983, 'Dropping out of high school: The influence of race, sex, and family background', *American Educational Research Journal*, Vol. 20 (2), pp. 199-220.

Sander, W., and Krautmann, A. C., (1995), 'Catholic schools, dropout rates and educational attainment', *Economic Inquiry*, Vol. 33, pp. 217-233.

Stewart, M., (1996), 'Heterogeneity specification in unemployment duration models', unpublished paper, Department of Economics, University of Warwick, Coventry, UK.

Wössman, L. (2000), 'Schooling resources, educational institutions and student performance: the international evidence,' Kiel Working Paper No. 983, Kiel Institute of World Economics.

Zemensky, R., Shapiro, D., Iannozi, M., Capelli, P. and Bailey, T. (1998) *The Transition from Initial Education to Working Life in the United States of America*, O.E.C.D.

TABLE 1. PROPORTION OF STUDENTS DROPPING OUT OF HIGH SCHOOL FOR ALL EXPLANATORY VARIABLES

| Characteristic | Sample proportions | Proportion dropping out | |
|---|---|---|---|
| | | Mean | Standard deviation |
| Female | 0.515 | 0.112 | 0.315 |
| Male | 0.485 | 0.117 | 0.321 |
| *Age* | | | |
| Born before September 1972 | 0.019 | 0.583 | 0.494 |
| Born September 1972 to August1973 | 0.141 | 0.291 | 0.454 |
| Born September 1973 to August 1974 | 0.704 | 0.079 | 0.269 |
| Born after August 1974 | 0.136 | 0.050 | 0.219 |
| *Ethnic group* | | | |
| Asian | 0.070 | 0.044 | 0.206 |
| White | 0.684 | 0.101 | 0.301 |
| Black | 0.102 | 0.156 | 0.363 |
| Hispanic | 0.130 | 0.179 | 0.383 |
| American Indian | 0.013 | 0.231 | 0.423 |
| *Family structure* | | | |
| Single parent | 0.155 | 0.198 | 0.399 |
| Parent plus partner | 0.160 | 0.181 | 0.385 |
| Both parents present | 0.685 | 0.080 | 0.271 |
| *Siblings* | | | |
| No siblings | 0.060 | 0.102 | 0.302 |
| 1 sibling | 0.319 | 0.075 | 0.264 |
| 2 siblings | 0.272 | 0.101 | 0.302 |
| 3 siblings | 0.157 | 0.133 | 0.340 |
| 4 or more siblings | 0.191 | 0.186 | 0.389 |
| *Parental education (highest level)* | | | |
| Less than high school / not known | 0.119 | 0.275 | 0.447 |
| High school | 0.212 | 0.144 | 0.351 |
| Some college | 0.392 | 0.112 | 0.315 |
| College graduate | 0.143 | 0.038 | 0.191 |
| Master's degree or PhD | 0.134 | 0.015 | 0.123 |
| *Father's occupation* | | | |
| Unskilled / semi-skilled | 0.289 | 0.170 | 0.376 |
| Skilled manual | 0.169 | 0.114 | 0.318 |
| Skilled non-manual | 0.183 | 0.099 | 0.299 |
| Professional or managerial | 0.248 | 0.036 | 0.166 |
| *Parents' labour market status* | | | |
| Both parents unemployed | 0.034 | 0.304 | 0.460 |
| *Type of school attended* | | | |
| Public | 0.876 | 0.128 | 0.335 |
| Non-public | 0.065 | 0.015 | 0.121 |
| *Test result (8$^{th}$ grade)* | | | |
| Lowest quintile | 0.200 | 0.278 | 0.448 |
| 2$^{nd}$ quintile | 0.200 | 0.142 | 0.349 |
| 3$^{rd}$ quintile | 0.200 | 0.094 | 0.291 |
| 4$^{th}$ quintile | 0.200 | 0.039 | 0.193 |
| Highest quintile | 0.200 | 0.019 | 0.136 |
| *School size* | | | |
| Under 800 students | 0.342 | 0.088 | 0.283 |
| 800-1200 students | 0.249 | 0.121 | 0.326 |
| 1201-1600 students | 0.152 | 0.123 | 0.328 |

| | | | |
|---|---|---|---|
| 1601-2000 students | 0.117 | 0.140 | 0.347 |
| Over 2000 students | 0.140 | 0.137 | 0.344 |
| *Type of area* | | | |
| Urban | 0.285 | 0.120 | 0.325 |
| Sub-urban | 0.397 | 0.101 | 0.301 |
| Rural | 0.319 | 0.126 | 0.332 |
| *Region* | | | |
| South | 0.351 | 0.101 | 0.301 |
| Other regions | 0.649 | 0.140 | 0.347 |
| n | 12551 | | |

*Note*: Dropout rate = proportion of students in the NELS sample who had ever dropped out of high school by the end of the 12th grade.
*Source*: National Educational Longitudinal Study 1988-94.

TABLE 2. HAZARD MODEL ESTIMATES

| | Nonparametric baseline and homogeneous model: model 1 | | | Nonparametric baseline and Gamma Unobserved heterogeneity: model 2 | | | Nonparametric baseline and Gaussian unobserved heterogeneity: model 3 | | | Nonparametric baseline and Heckman-Singer unobserved heterogeneity: model 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coefficient | Relative risk | P value | Coefficient | Relative risk | P value | Coefficient | Relative risk | P value | Coefficient | Relative risk | P value |
| *Personal characteristics* | | | | | | | | | | | | |
| Female | 0.242 | 1.274 | 0.000 | 0.301 | 1.352 | 0.000 | 0.302 | 1.352 | 0.000 | 0.283 | 1.327 | 0.000 |
| Born Sept 1972 to August1973 | 1.318 | 3.735 | 0.000 | 1.959 | 7.090 | 0.000 | 1.943 | 6.979 | 0.000 | 1.899 | 6.681 | 0.000 |
| Born Sept 1973 to August 1974 | 0.696 | 2.006 | 0.000 | 0.902 | 2.465 | 0.000 | 0.922 | 2.515 | 0.000 | 0.879 | 2.408 | 0.000 |
| Born after August 1974 | -0.337 | 0.714 | 0.003 | -0.391 | 0.677 | 0.003 | -0.400 | 0.670 | 0.003 | -0.377 | 0.686 | 0.002 |
| Black | -0.513 | 0.598 | 0.000 | -0.660 | 0.517 | 0.000 | -0.677 | 0.508 | 0.000 | -0.664 | 0.515 | 0.000 |
| Asian | -0.358 | 0.699 | 0.036 | -0.445 | 0.641 | 0.022 | -0.461 | 0.631 | 0.021 | -0.435 | 0.648 | 0.021 |
| Hispanic | -0.139 | 0.870 | 0.088 | -0.175 | 0.839 | 0.078 | -0.177 | 0.838 | 0.083 | -0.194 | 0.824 | 0.046 |
| American Indian | -0.269 | 0.764 | 0.147 | -0.158 | 0.854 | 0.504 | -0.197 | 0.821 | 0.410 | -0.120 | 0.887 | 0.600 |
| Test score in $8^{th}$ grade | -0.038 | 0.963 | 0.000 | -0.043 | 0.958 | 0.000 | -0.045 | 0.956 | 0.000 | -0.041 | 0.960 | 0.000 |
| *Family background* | | | | | | | | | | | | |
| Single parent | 0.322 | 1.380 | 0.000 | 0.401 | 1.493 | 0.000 | 0.419 | 1.520 | 0.000 | 0.382 | 1.466 | 0.000 |
| Parent with partner | 0.221 | 1.247 | 0.001 | 0.305 | 1.356 | 0.000 | 0.307 | 1.359 | 0.000 | 0.288 | 1.334 | 0.000 |
| 1 sibling | -0.103 | 0.902 | 0.436 | -0.170 | 0.844 | 0.282 | -0.163 | 0.850 | 0.312 | -0.160 | 0.852 | 0.288 |
| 2 siblings | -0.076 | 0.926 | 0.558 | -0.147 | 0.863 | 0.347 | -0.134 | 0.875 | 0.402 | -0.149 | 0.862 | 0.320 |
| 3 siblings | -0.049 | 0.953 | 0.719 | -0.105 | 0.901 | 0.519 | -0.087 | 0.917 | 0.599 | -0.086 | 0.917 | 0.577 |
| 4 or more siblings | 0.029 | 1.030 | 0.821 | 0.050 | 1.051 | 0.749 | 0.061 | 1.063 | 0.699 | 0.046 | 1.047 | 0.757 |
| Parents: high school graduate | -0.174 | 0.840 | 0.025 | -0.269 | 0.764 | 0.006 | -0.267 | 0.766 | 0.007 | -0.286 | 0.751 | 0.002 |
| Parents: some college | -0.187 | 0.829 | 0.011 | -0.281 | 0.755 | 0.002 | -0.280 | 0.756 | 0.003 | -0.298 | 0.742 | 0.001 |
| Parents: college graduate | -0.483 | 0.617 | 0.001 | -0.597 | 0.551 | 0.000 | -0.606 | 0.545 | 0.000 | -0.605 | 0.546 | 0.000 |
| Parents: higher degree | -0.763 | 0.466 | 0.000 | -0.866 | 0.421 | 0.000 | -0.873 | 0.418 | 0.000 | -0.883 | 0.414 | 0.000 |
| Father: professional / managerial | -0.378 | 0.685 | 0.001 | -0.386 | 0.680 | 0.002 | -0.397 | 0.672 | 0.002 | -0.379 | 0.685 | 0.002 |
| Father: skilled manual | -0.060 | 0.942 | 0.436 | -0.118 | 0.889 | 0.205 | -0.118 | 0.889 | 0.218 | -0.112 | 0.894 | 0.215 |
| Father: non-manual manual | -0.041 | 0.960 | 0.609 | -0.082 | 0.921 | 0.385 | -0.080 | 0.923 | 0.413 | -0.087 | 0.917 | 0.339 |
| Both parents unemployed | 0.031 | 1.031 | 0.764 | 0.138 | 1.148 | 0.294 | 0.113 | 1.120 | 0.391 | 0.172 | 1.188 | 0.187 |
| *Neighbourhood* | | | | | | | | | | | | |
| Live in suburbs | -0.083 | 0.920 | 0.263 | -0.142 | 0.867 | 0.111 | -0.139 | 0.870 | 0.131 | -0.138 | 0.871 | 0.108 |
| Live in rural area | -0.076 | 0.927 | 0.373 | -0.130 | 0.878 | 0.200 | -0.129 | 0.879 | 0.217 | -0.140 | 0.869 | 0.146 |
| Live in South | 0.199 | 1.220 | 0.001 | 0.238 | 1.269 | 0.001 | 0.247 | 1.280 | 0.001 | 0.227 | 1.255 | 0.001 |
| *School factors* | | | | | | | | | | | | |
| Non-public school in $9^{th}$ grade | -0.891 | 0.410 | 0.000 | -0.966 | 0.381 | 0.000 | -0.997 | 0.369 | 0.000 | -0.944 | 0.389 | 0.000 |
| Pupils in school: 800-1200 | 0.190 | 1.209 | 0.016 | 0.177 | 1.194 | 0.061 | 0.183 | 1.201 | 0.058 | 0.151 | 1.164 | 0.103 |
| Pupils in school: 1200-1600 | 0.056 | 1.058 | 0.545 | 0.046 | 1.047 | 0.677 | 0.052 | 1.053 | 0.650 | 0.022 | 1.022 | 0.838 |
| Pupils in school: 1600-2000 | 0.054 | 1.055 | 0.599 | 0.050 | 1.051 | 0.682 | 0.051 | 1.053 | 0.682 | 0.029 | 1.030 | 0.806 |
| Pupils in school: 2000+ | -0.089 | 0.915 | 0.408 | -0.070 | 0.933 | 0.584 | -0.084 | 0.920 | 0.522 | -0.081 | 0.922 | 0.513 |
| Time period dummy: Sept-Dec | -2.851 | 0.058 | 0.000 | -3.139 | 0.043 | 0.000 | -3.223 | 0.040 | 0.000 | -3.064 | 0.047 | 0.000 |
| Time period dummy: May-Aug | 0.633 | 1.884 | 0.000 | 0.710 | 2.033 | 0.000 | 0.746 | 2.108 | 0.000 | 0.663 | 1.941 | 0.000 |

TABLE 2 (continued)

| | | | | | | |
|---|---|---|---|---|---|---|
| *Heterogeneity parameters* | | | | | | |
| Variance | | 0.573 | 0.000 | 0.812 | 0.000 | 0.601 |
| Location of first mass point | | | | | | -1.248 |
| Probability of first mass point | | | | | | 0.278 |
| Location of second mass point | | | | | | 0.481 |
| Probability of second mass point | | | | | | 0.722 |
| Log likelihood | -6164 | -6147 | | -6148 | | -6146 |
| Number of students in sample | 12551 | 12551 | | 12551 | | 12551 |
| Number of person-periods | 275118 | 275118 | | 275118 | | 275118 |

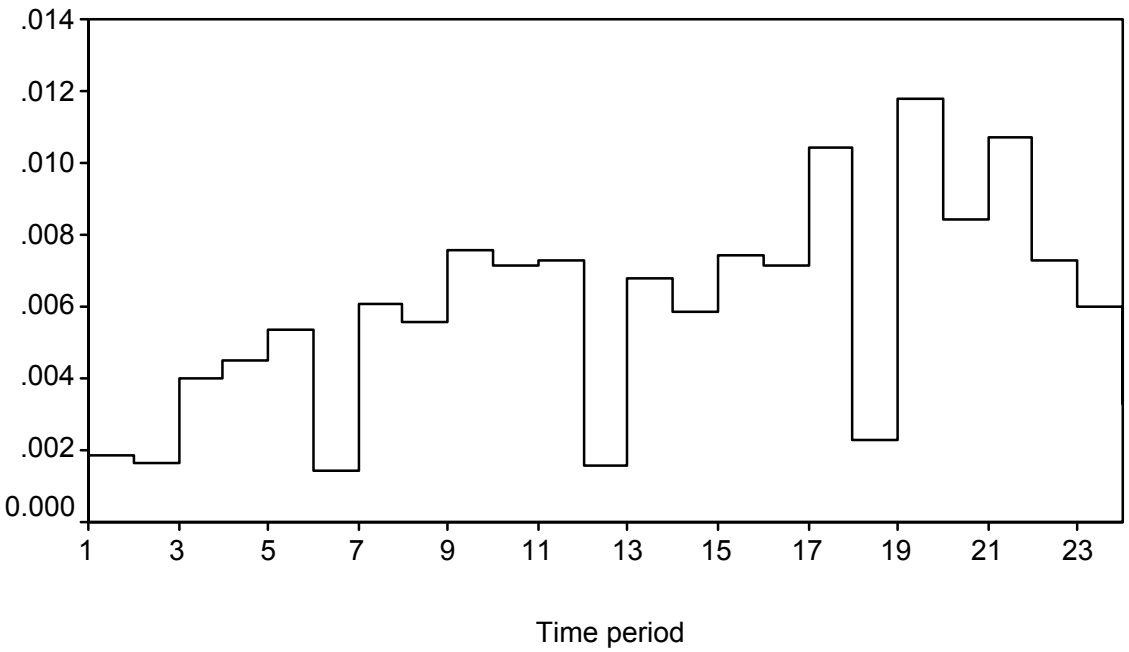
*Note*: 1. All explanatory variables were obtained from the survey undertaken when the respondents were in the 8$^{th}$ grade except for the number of pupils in the high school attended and type of school (public or private). 2. The base group of respondents has the following characteristics: did not drop out of high school, male, born during September 1974 to August 1975, white, both parents present, parents did not complete high school, father unskilled/semi-skilled or economically inactive, at least one parent employed, live in urban area, does not live in the South, attended public school in 10$^{th}$ grade, under 800 pupils at high school attended, time period = January to April.

TABLE 3. HAZARD MODEL ESTIMATES: FEMALES AND MALES

| | Nonparametric baseline and Heckman-Singer unobserved heterogeneity: model 4 Females | | | Nonparametric baseline and Heckman-Singer unobserved heterogeneity: model 4 Males | | |
|---|---|---|---|---|---|---|
| | Coefficient | Relative risk | P value | Coefficient | Relative risk | P value |
| *Personal characteristics* | | | | | | |
| Born Sept 1972 to August1973 | 1.880 | 6.552 | 0.000 | 1.964 | 7.128 | 0.000 |
| Born Sept 1973 to August 1974 | 0.902 | 2.464 | 0.000 | 0.879 | 2.407 | 0.000 |
| Born after August 1974 | -0.320 | 0.726 | 0.040 | -0.426 | 0.653 | 0.042 |
| Black | -0.825 | 0.438 | 0.000 | -0.553 | 0.575 | 0.000 |
| Asian | -0.644 | 0.525 | 0.040 | -0.290 | 0.748 | 0.219 |
| Hispanic | -0.285 | 0.752 | 0.035 | -0.111 | 0.895 | 0.436 |
| American Indian | -0.467 | 0.627 | 0.145 | 0.115 | 1.122 | 0.724 |
| Test score in 8th grade | -0.042 | 0.959 | 0.000 | -0.040 | 0.961 | 0.000 |
| *Family background* | | | | | | |
| Single parent | 0.361 | 1.434 | 0.002 | 0.409 | 1.506 | 0.003 |
| Parent with partner | 0.354 | 1.425 | 0.002 | 0.257 | 1.293 | 0.025 |
| 1 sibling | -0.018 | 0.982 | 0.931 | -0.214 | 0.807 | 0.333 |
| 2 siblings | -0.080 | 0.923 | 0.699 | -0.115 | 0.891 | 0.611 |
| 3 siblings | 0.172 | 1.187 | 0.420 | -0.239 | 0.787 | 0.294 |
| 4 or more siblings | 0.166 | 1.181 | 0.417 | -0.025 | 0.976 | 0.909 |
| Parents: high school graduate | -0.315 | 0.730 | 0.016 | -0.209 | 0.812 | 0.126 |
| Parents: some college | -0.416 | 0.660 | 0.001 | -0.200 | 0.819 | 0.125 |
| Parents: college graduate | -0.697 | 0.498 | 0.002 | -0.487 | 0.614 | 0.028 |
| Parents: higher degree | -1.023 | 0.359 | 0.003 | -0.807 | 0.446 | 0.009 |
| Father: professional / managerial | -0.521 | 0.594 | 0.002 | -0.241 | 0.786 | 0.167 |
| Father: skilled manual | -0.142 | 0.867 | 0.258 | -0.077 | 0.926 | 0.556 |
| Father: non-manual manual | -0.062 | 0.940 | 0.623 | -0.138 | 0.871 | 0.304 |
| Both parents unemployed | 0.192 | 1.212 | 0.309 | 0.163 | 1.177 | 0.383 |
| *Neighbourhood* | | | | | | |
| Live in suburbs | -0.167 | 0.846 | 0.168 | -0.122 | 0.886 | 0.320 |
| Live in rural area | -0.215 | 0.807 | 0.105 | -0.099 | 0.906 | 0.479 |
| Live in South | 0.336 | 1.399 | 0.001 | 0.171 | 1.187 | 0.079 |
| *School factors* | | | | | | |
| Non-public school in 9th grade | -1.105 | 0.331 | 0.003 | -0.862 | 0.423 | 0.006 |
| Pupils in school: 800-1200 | 0.173 | 1.189 | 0.181 | 0.112 | 1.118 | 0.406 |
| Pupils in school: 1200-1600 | 0.039 | 1.040 | 0.796 | 0.002 | 1.002 | 0.988 |
| Pupils in school: 1600-2000 | 0.059 | 1.060 | 0.724 | -0.025 | 0.975 | 0.889 |
| Pupils in school: 2000+ | -0.136 | 0.872 | 0.440 | -0.047 | 0.954 | 0.789 |
| Time period dummy: Sept-Dec | -3.234 | 0.039 | 0.000 | -2.935 | 0.053 | 0.000 |
| Time period dummy: May-Aug | 0.613 | 1.847 | 0.000 | 0.662 | 1.939 | 0.000 |
| | | | | | | |
| *Heterogeneity parameters* | | | | | | |
| Variance | 0.669 | | | 0.579 | | |
| Location of first mass point | -1.649 | | | -1.011 | | |
| Probability of first mass point | 0.198 | | | 0.361 | | |
| Location of second mass point | 0.459 | | | 0.573 | | |
| Probability of second mass point | 0.802 | | | 0.639 | | |
| Log likelihood | -3068 | | | -3056 | | |
| Number of students in sample | 12551 | | | 12551 | | |
| Number of person-periods | 141294 | | | 133824 | | |

*Note*: 1. All explanatory variables were obtained from the survey undertaken when the respondents were in the 8th grade except for the number of pupils in the high school attended and type of school (public or private). 2. The base group of respondents has the following characteristics: did not drop out of high school, born during September 1974 to August 1975, white, both parents present, parents did not complete high school, father unskilled/semi-skilled or economically inactive, at least one parent employed, live in urban area, does not live in the South, attended public school in 10th grade, under 800 pupils at high school attended, time period = January to April.

Figure 1. The empirical hazard of dropouts, 1988-94.




Note: Data are aggregated to two monthly intervals.

Figure 2. A comparison of the baseline hazards: models 1 and 4

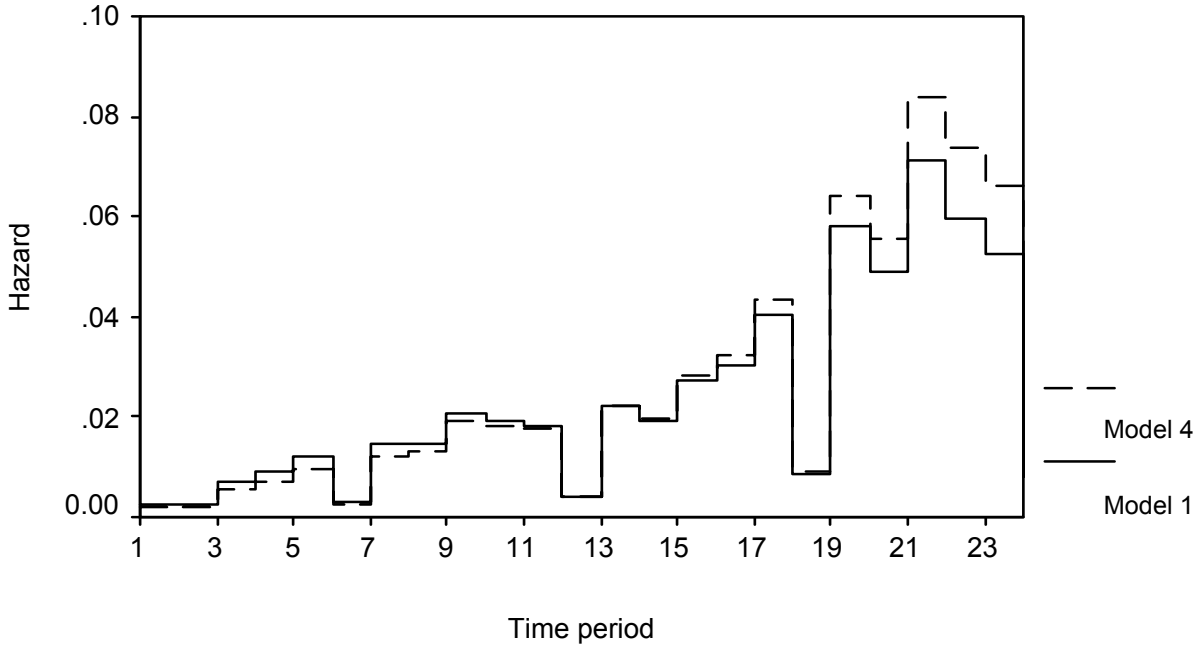


Figure 3. Baseline hazards for males and females: model 4