

11:2008 WORKING PAPER

Anders Holm  
Mads Meier Jæger

## SELECTION BIAS IN EDUCATIONAL TRANSITION MODELS: THEORY AND EMPIRICAL EVIDENCE

RESEARCH DEPARTMENT OF SOCIAL POLICY AND WELFARE SERVICES

# Selection Bias in Educational Transition Models: Theory and Empirical Evidence

Anders Holm & Mads Meier Jæger

Social Policy and Welfare Services  
Working Paper 11:2008

The Working Paper Series of The Danish National Centre for Social Research contain interim results of research and preparatory studies. The Working Paper Series provide a basis for professional discussion as part of the research process. Readers should note that results and interpretations in the final report or article may differ from the present Working Paper. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including ©-notice, is given to the source.

## **Selection Bias in Educational Transition Models: Theory and Empirical Evidence**

Anders Holm\* and Mads Meier Jæger\*\*

\* Department of Sociology, University of Copenhagen ([anders.holm@sociology.ku.dk](mailto:anders.holm@sociology.ku.dk))

\*\* The Danish National Centre for Social Research ([mads@sfi.dk](mailto:mads@sfi.dk))

**15 July 2008**

Most studies using Mare's (1980, 1981) seminal model of educational transitions find that the effect of family background decreases across transitions. Recently, Cameron and Heckman (1998, 2001) have argued that the "waning coefficients" in the Mare model are driven by selection on unobserved variables. This paper, first, explains theoretically how selection on unobserved variables leads to waning coefficients and, second, illustrates empirically how selection leads to biased estimates of the effect of family background on educational transitions. Our empirical analysis using data from the United States, United Kingdom, Denmark, and the Netherlands shows that when we take selection into account the effect of family background variables on educational transitions is largely constant across transitions. We also discuss several difficulties in estimating educational transition models which deal effectively with selection.

Word count: 7,674.

## 1. Introduction

Robert Mare's (1979; 1980; 1981) model of educational transitions represents one of the major methodological contributions to the literature on family background and educational outcomes. Instead of years of completed schooling and linear regression models, Mare suggested to treat educational attainment as a sequence of transitions from lower to higher educational levels. The principal advantages of Mare's educational transition model are, first, that it conforms better to the way most sociologists think about educational attainment (as a sequence of transitions) and, second, that it allows researchers to model the effect of family background variables on the probability of making successive educational transitions.

The Mare model is, and for long time has been, highly influential in applied research (e.g., Garnier and Raffalovich 1984; Cobalti 1990; Heath and Clifford 1990; Shavit and Blossfeld 1993; Hansen 1997; Shavit and Westerbeek 1998; Vaid 2004). One of the recurring findings from empirical studies using the Mare model is that the effect of family background variables decreases or "waned" across educational transitions. Several theories such as the theories of Maximally Maintained Inequality (Raftery and Hout 1993) and Effectively Maintained Inequality (Lucas 2001) have been proposed as substantive explanations of this "waning coefficients" phenomenon.

In two influential papers Cameron and Heckman (1998, 2001) argue that the waning coefficients in the Mare model are an artifact of an arbitrary choice of parameterization and selection on unobserved variables. Selection implies that the group at risk of making educational transitions becomes increasingly selective at higher transitions. For example, it is reasonable to assume that the group of youth that makes the transition from elementary school to high school on average has lower academic ability than the group that makes the transition from high school to college. If academic ability or some other important factor that affects the transition probability is not observed, the effect of this unobserved variable causes systematic bias in the estimated effects of family background variables on the probability of making successive educational transitions. Mare (1993) has observed that selection on unobserved variables leads to bias.

In this paper we address the crucial question of how selection on unobserved variables affects estimates of the effect of family background variables in educational transition models. Given the popularity of the Mare model in applied sociological research but, at the same time, the lack of

attention to the role of selection in this type of model, it seems essential to explicate how selection affects empirical estimates of the effect of family background on educational transitions.<sup>1</sup> The paper seeks to contribute to the existing literature in two areas: *theoretical clarification* of the role of selection on unobserved variables in the Mare model and *empirical illustration* of the impact of selection bias in empirical analysis.

First, we show theoretically how selection on unobserved variables leads to bias in estimates of the effect of explanatory variables on the probability of making successive educational transitions. For simplicity, we propose a Mare model with only two transitions: (1) the transition from elementary school to high school (or a similar type of upper secondary education) and, conditional on having made the first transition, (2) the transition from high school to college (or a similar type of higher education). The theory behind our approach easily generalizes to situations with more than two educational transitions. We show that parameter bias in the Mare model originates in two phenomena: *selection on unobserved variables* which leads to downward bias in the effect of family background variables at higher educational transitions (i.e., waning coefficients) and *scaling effects* (different variances in the distributions of unobservables in the populations at risk of making successive educational transitions) which lead to upward bias. In empirical applications it is usually not possible to distinguish between bias arising from selection on unobserved variables and scaling effects (Mare 2006).

Second, using data from four countries (the United States, the United Kingdom, Denmark, and the Netherlands) we provide empirical illustrations of how selection on unobserved variables affects estimates of the effect of family background on educational transitions. Our analysis is built around trying to distinguish between two alternative hypotheses: a “waning coefficients” hypothesis claiming that the effect of family background declines across educational transitions (Raftery and Hout 1993; Lucas 2001) and a “constant inequality” hypothesis arguing that, due to selection on

---

<sup>1</sup> Several studies deal with other aspects of the Mare model such as improvements in identification from repeated measurements of family background variables (Lucas 2001) and a more parsimonious formulation of the Mare model along the lines of Anderson’s (1984) Stereotype Ordered regression model (Hauser and Andrew 2006). With the exception of Breen and Jonsson (2000) we are not familiar with any study that addresses selection on unobserved variables in educational transition models. In the economics literature selection bias in educational transition models is more often dealt with (e.g., Chevalier and Lanot 2002; Lauer 2003; Arends-Kuenning and Duryea 2006; Colding 2006)

unobserved variables, the effect of family background is constant across transitions. We use non-parametric Manski bounds (Manski 1995; Horowitz and Manski 1998) to show that, when not imposing any parametric assumptions about the transition probabilities in the four countries under study, our data alone cannot distinguish between the waning coefficients and the constant inequality hypotheses. In other words, the data is equally consistent with both hypotheses. To distinguish between the two hypotheses we need to apply parametric assumptions similar to those in the Mare model. We assume bivariate normal distributions for the two educational transitions and estimate simple and bivariate probit transition models. When estimating simple probit models for the individual countries we find the typical waning coefficients pattern reported in previous studies. When estimating bivariate probit models for each country which allow for a correlation between the unobserved variables in each transition we also find waning coefficients. However, due to small sample sizes the correlation between the unobservables in the two educational transitions is very poorly identified. Finally, when pooling the data for three of the four countries (the Netherlands is excluded in this part of the analysis) and using country-specific transition rates at each of the two transitions as instruments our bivariate probit model shows that the effect of family background is largely constant across educational transitions. The models which use the pooled data also provide credible estimates of the correlation between the unobservables in the two transitions.

We conclude, first, that selection (and scaling) effects appear to have a significant impact on substantive conclusions regarding the impact of family background on educational transitions. Second, we find that it is necessary to impose parametric assumptions on transition probabilities to deal with selection and to distinguish between competing hypotheses. Third, we conclude that it is difficult to estimate educational transition models which deal effectively with selection on unobserved variables.

## **2. Selection in Educational Transition Models**

### *2.1 The Basic Model*

Mare's (1979, 1980, 1981) original transition model consists of a sequence of binary logit models in which the dependent variables are dummy variables for making the  $j$ 'th educational transition conditional on previously having made the  $j-1$ 'th transition. Our model consists of only two transitions and, for reasons that will become clear shortly, we prefer the probit specification to the logit specification used by Mare.

Define the two latent stochastic variables  $y_1^*$  and  $y_2^*$  which capture the propensity to make transition 1 and 2 in an educational system. As described previously, transition 1 represents the transition from elementary school to high school (or equivalent upper secondary schooling such as A levels in the UK or the *Gymnasium* in Denmark) and transition 2 represents the transition from high school to higher education (for example, college, university, or university-college education). These types of transitions exist in most Anglo-American and Western European educational systems. We assume that in order for individuals to complete transition 2 they must first successfully complete transition 1. We never observe the latent variables  $y_1^*$  and  $y_2^*$  but instead two binary variables  $y$  indicating if individuals make each of the educational transitions. These binary variables are defined as  $y_j = 1$  if  $y_j^* > 0$  and 0 otherwise,  $j = 1, 2$ , with  $j$  indexing transitions.

The likelihood that an individual makes each educational transition depends on a set of observed variables (for example, parents' education and income, academic ability and motivation) and some unobserved variables. The process of educational transitions can be represented by the following system of regression equations

$$\begin{aligned} y_1^* &= \beta_1 x_1 + e_1 \\ y_2^* &= \beta_2 x_2 + e_2 \end{aligned}$$

where  $x_j, j = 1, 2$  represents observed variables for each transition and  $e_j, j = 1, 2$  represent error terms that capture the effect of unobserved variables. We assume that the error terms are distributed as

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \square N(\mathbf{0}, \mathbf{\Sigma}),$$

where

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The coefficient  $\rho$  captures the correlation between the unobserved variables in each educational transition. Our reason for choosing the probit specification over the logit used by Mare (1980, 1981) is that the probit specification allows us to estimate  $\rho$  and thereby to take into account the fact that the unobserved variables that affect the propensity to make educational transitions are correlated across transitions. We name this model the bivariate probit selection model. In empirical analysis we cannot identify the variance of either of the error terms  $e$  and, as is always the case in the probit model, they are normalized to 1.

## 2.2 Selection Effects

The fundamental problem in analyzing educational transitions in the model presented above is that the probability of making the second transition depends on whether or not individuals have previously made the first transition; i.e., the group at risk of making the second transition is a *selected sample*. In this section we explain how sample selection might lead to bias in educational transition models. We first give an intuitive explanation of how selection works and then present a formal statistical account of selection bias.

-- FIGURE 1 ABOUT HERE --

Figure 1 shows the relationship between the latent propensity to make the first transition,  $y_1^*$  and a hypothetical explanatory variable,  $x$ . The shaded area represents the distribution of the data. In Figure 1 the threshold for making the first transition is represented by the horizontal axis. Empirically, educational systems differ with regard to the strength of the selection at the first transition. This difference can be conceptualized by shifting the horizontal axis up or down. Students with high values of  $x$ , for example parental education, have a higher propensity to make the first transition than students with low values of  $x$  and thus a higher probability of actually making this transition. However, students with identical values of  $x$  also have different probabilities of making the first transition. This difference is caused by unobserved factors, for example academic ability or motivation which, in addition to  $x$ , also affect whether students make the first transition. Consequently, student with low values of  $x$  but high values of unobserved factors tend also to make the first transition. This is not the case for students with high values of  $x$  who tend to make the first transition irrespective of their values on the unobserved factors. Thus, among students with low values of  $x$  we typically expect a large amount of selection on unobserved



characteristics whereas is not to the case for students with high values of  $x$ . This difference in the impact of the unobservables is important when we later inspect non-parametric bounds for the relationship between  $x$  and the probability of making the second transition.

-- FIGURE 2 ABOUT HERE --

From Figure 1 we are able to estimate the true relationship between  $x$  and  $y_1$  because we have available the whole sample. Now imagine a high correlation between  $y_1^*$  and  $y_2^*$  meaning that the  $x$ - $y_1$  and the  $x$ - $y_2$  plots look very similar. We would then illustrate the relationship between  $x$  and  $y_2$  as shown in Figure 2. The dotted horizontal line shows the threshold for making the first transition.<sup>2</sup> Hence, the distribution above the horizontal dotted line is the empirical relationship between  $x$  and  $y_2^*$  that is observed in the data (since we only observe whether  $y_2^*$  is above the threshold for making the second transition, here indicated by the  $x$ -axis). From the figure it is evident that because of the selection at the first transition which leaves only a subpopulation of the whole sample at risk of making the second transition we estimate a biased regression slope for the relationship between  $x$  and  $y_2^*$  (the dashed slope) if we only use data for those individuals who made the first transition. The reason why we observe this pattern is that students with low values of  $x$  in the sample of individuals who made the first transition have higher values on the unobservables than students in the whole sample (and vice versa for students with high values of  $x$ ). Sample selection thus means that the empirical relationship between  $x$  and  $y_2$  is biased downwards. Note that this empirical relationship is different from the true relationship between  $x$  and the probability of making the second transition which is shown by the solid slope.

### 2.3 *Scaling Effects*

Selection potentially induces two different types of bias in the estimated effects of family background variables on educational transitions. The first type of bias originates in the problem of using a selected sample and leads to downward bias in parameter estimates. This bias is illustrated above. The second bias originates in a scaling problem which is caused by the fact that the distribution of the unobservables is different in the selected sample and in the whole sample. It

---

<sup>2</sup> Due to the probabilistic relationship between  $y_1^*$  and  $y_2^*$  this thresholds will not be a horizontal line but rather a jittered curve along the horizontal axis. However, for simplicity it is represented by a straight line.

turns out that the variance in the selected sample is smaller than the variance in the whole sample (intuitively, this happens because students become more similar on unobserved characteristics at higher transitions). In binary probability models such as the Mare model the variance of the error term is not identified and must be normalized (in the probit model the variance is normalized to 1 and in the logit model the variance is normalized to  $\pi^2/3$ ). Furthermore, binary probability models do not identify the actual regression coefficients associated with explanatory variables but only the regression coefficients divided by the error variance in the probability (probit/logit) model. However, since the variance in the selected sample is lower than the variance in the whole sample the regression coefficients for the effect of family background variables in the second transition are upwardly biased when analyzing the selected sample because the denominator (in the case of the probit) is smaller than 1 (assuming that the variance in the selected sample is 0.8 and the true regression coefficient is 1 it is easy to illustrate the upward bias from scaling. For transition 1 we get:  $\tilde{\beta} = 1/1 = 1$  and for transition 2 we get:  $\tilde{\beta} = 1/0.8 = 1.25$ ). We formally show the bias from scaling in Appendix 1.

#### 2.4 Selection and Scaling Effects in a Formal Model

We now present the selection problem in a formal model. This formal model leads to a natural statistical way of dealing with the selection problem. Our aim is to get insights into how selection and scaling affects the estimated effects of family background variables on the second transition (and, possibly, later transitions).

We study the conditional probability of making the second transition and decompose this probability into three different components: (1) the true effect of the observed family background variables (the  $x$ 's), (2) the selection effect, and (3) the scaling effect. Our decomposition is based on the approximation suggested by Nicoletti and Peracchi (2001) who shows that it works well for  $\rho$  correlation coefficients up to 0.8 (see Appendix 2 for a derivation of the approximation). The approximation is a convenient way of representing *attenuation bias* (the combined effect of selection and scaling) and has the following form

$$P(Y_2 = 1 | Y_1 = 1) \approx \Phi \left( \frac{\beta_2 x_2 + \rho \lambda(\beta_1 x_1)}{\sqrt{1 - \rho^2 \{ \beta_1 x_1 \lambda(\beta_1 x_1) + \lambda(\beta_1 x_1)^2 \}}} \right). \quad (1)$$

Equation 1 shows the approximation of the probability of making the second transition conditional on having made the first transition. The true effects of the family background variables on the probability of making the second transition is represented by the term  $\beta_2 x_2$ . Unfortunately, because of selection and scaling we do not estimate these true effects but instead biased effects. The selection term is  $\rho\lambda(\beta_1 x_1)$  and arises from selection in the first transition. The scaling term is  $\sqrt{1 - \rho^2 \{ \beta_1 x_1 \lambda(\beta_1 x_1) + \lambda(\beta_1 x_1)^2 \}}$  and captures the variance in the selected sample. What we then estimate in empirical applications is the combination of the true and the attenuation effects. That is, we estimate  $P(Y_2 = 1 | Y_1 = 1) = \Phi(\alpha x_2)$  where

$$\alpha x_2 \approx \frac{\beta_2 x_2 + \rho\lambda(\beta_1 x_1)}{\sqrt{1 - \rho^2 \{ \beta_1 x_1 \lambda(\beta_1 x_1) + \lambda(\beta_1 x_1)^2 \}}} . \quad (2)$$

Since most often we do not have any information on the actual magnitude of the selection and scaling effect we cannot determine the severity of the parameter bias. However, to investigate the likely direction of the selection bias we note (following Wooldridge 2002) that

$$\frac{\partial \lambda(\beta_1 x_1)}{\partial x_1} = -\beta_1 \lambda(\beta_1 x_1) (\beta_1 x_1 + \lambda(\beta_1 x_1)) .$$

This means that if we assume that that  $\beta_1 > 0$  (i.e., some of the family background variables have a positive effect on the probability of making the first transition) and that the correlation between the unobservables in the two transitions is positive,  $\rho > 0$  (which makes intuitive sense), we find

that  $\frac{\partial \lambda(\beta_1 x_1)}{\partial x_1} < 0$ . This finding implies that for high values of  $x$  we get a relatively small

selection effect and for small values of  $x$  we get a relatively large selection effect (as is illustrated in Figure 2). Consequently, the total effect of  $x$  in the denominator is likely to be smaller than the “true” effect  $\beta_2$ . Hence, from the selection effects we expect that  $\alpha < \beta_2$ .

With regard to the scaling effect we find that  $Var(Y_2^* | Y_1 = 0) = 1 - \rho^2 \{ \beta_1 x_1 \lambda(\beta_1 x_1) + \lambda(\beta_1 x_1)^2 \}$  (cf. Appendix 1). This means that the conditional variance is always smaller than the unconditional variance. Accordingly, the scaling effect, the denominator in Equation (2), tends to inflate the estimate of the combined effect  $\alpha$  compared to the true effect  $\beta_2$ .

In summary, we have two interrelated processes that might generate attenuation bias in the estimated effects of family background variables in the second transition: selection effects which lead to downward bias and scaling effects which lead to upward bias. Furthermore, in addition to the observed data we also need to impose parametric assumptions on the model governing the selection and true effects to distinguish between true and attenuation effects.

### 2.5 Manski Bounds

Cameron and Heckman (1998) argue that the Mare model imposes a set of arbitrary parametric assumptions. In order to investigate the extent to which parametric assumptions in the Mare model affect substantive conclusions we implement a simple non-parametric approach which does not involve any parametric assumptions about the relationship between the  $x$  variables and the probability of making the second educational transition. Manski (1995) and Horowitz and Manski (1998) show how one can use so-called “Manski bounds” to bound true probabilities in data with selection without making any distributional assumptions. In our application we use Manski bounds to bound the true probability of making the second transition  $P(Y_2 = 1 | x)$  using only the observed data. The Manski bounds are useful because they represent a baseline model which can be used to determine if the data alone allows us to distinguish between the waning coefficients and the constant inequality hypotheses.

The Manski bounds are defined as

$$\underbrace{P(Y_2 = 1 | x, Y_1 = 1) P(Y_1 = 1 | x)}_{\text{data}} \leq P(Y_2 = 1 | x) \leq \underbrace{P(Y_2 = 1 | x, Y_1 = 1) P(Y_1 = 1 | x) + P(Y_1 = 0 | x)}_{\text{data}}$$

These bound exist because

$$\begin{aligned} P(Y_2 = 1 | x) &= P(Y_2 = 1 | x, Y_1 = 1)P(Y_1 = 1 | x) + P(Y_2 = 1 | x, Y_1 = 0)P(Y_1 = 0 | x) \\ &\leq P(Y_2 = 1 | x, Y_1 = 1)P(Y_1 = 1 | x) + P(Y_1 = 0 | x) \end{aligned}$$

and

$$\begin{aligned} P(Y_2 = 1 | x) &= P(Y_2 = 1 | x, Y_1 = 1)P(Y_1 = 1 | x) + P(Y_2 = 1 | x, Y_1 = 0)P(Y_1 = 0 | x) \\ &\geq P(Y_2 = 1 | x, Y_1 = 1)P(Y_1 = 1 | x) \end{aligned}$$

Furthermore, the bounds are informative because

$$\underbrace{P(Y_2 = 1 | x, Y_1 = 1)P(Y_1 = 1 | x) + P(Y_1 = 0 | x)}_{\text{upper bound}} \geq \underbrace{P(Y_2 = 1 | x, Y_1 = 1)P(Y_1 = 1 | x)}_{\text{lower bound}}$$

The Manski bounds provide a basis for assessing the effects of family background variables on the probability of making the second educational transition. The bounds may be so wide that they are equally consistent with both the waning coefficients and the constant inequality hypotheses. As shown above the width of the bound depends on the fraction of students that does not make the first transition  $P(Y_1 = 0 | x)$ . This makes sense since this fraction determines the magnitude of the selection at the first transition. Consequently, the lower the selection at this transition the narrower the Manski bounds will be. In order to further narrow the bounds one has to make assumptions regarding the transition probabilities. The Mare model assumes, first, that the effects of observed family background variables are homogenous, second, that there is no selection on unobserved variables and, third, that the transition probabilities are modeled either as probits or logits. The bivariate probit selection model we presented in section 2.1 relaxes the assumption of no selection on unobservables but maintains the other two assumptions (see also Poirier (1980) and Wynand and van Praag (1981)).

### 3. Empirical Application

In the first part of the paper we hope to have shown theoretically how selection on unobserved variables and scaling effects might lead to bias in the estimated effects of family background variables on the probability of making successive educational transitions. Furthermore, we have presented the bivariate probit selection model as an alternative approach which corrects for selection on unobserved variables. Finally, we have proposed Manski bounds as a way of bounding

the effect of family background on educational transitions without making any distributional assumptions.

In the second part of the paper we provide empirical illustrations of how selection on unobserved variables affects estimates of the effect of family background on the probability of making educational transitions and how different empirical strategies yield different conclusions. We analyze data from four countries: the United States (US), the United Kingdom (UK), Denmark, and the Netherlands. These countries were chosen because they have comparable educational systems but differ substantially with regard to how large a proportion of a cohort of youth that makes the first and second transitions. In the US a large proportion of students make the first transition (high school) but a smaller proportion of the students who make the first transition also make the second transition (higher education) (see Table 1). By contrast, in the UK, Denmark, and the Netherlands much fewer students make the first transition but, if they make the first transition, they have a relatively high probability of also making the second transition. The four countries thus differ substantially with regards to the degree of selection at the first and second transitions. The four countries were also chosen because the available data for each country (except the US) is a cohort study (which reduces cohort heterogeneity) and because the data includes information on respondents' academic ability (which is typically seen as one of the major unobserved variables). The datasets are presented below.

Our empirical analysis is built around attempting to distinguish between the waning coefficients and the constant inequality hypotheses. The former hypothesis states that the effect of family background decreases across educational transitions and the latter hypothesis states that the effect is constant. Distinguishing between these two different hypotheses is important for theoretical and substantive reasons and has been a recurring theme in the literature using the Mare model. We begin the empirical analysis in section 4 below by presenting the results from the Manski bounds approach which does not impose any assumptions. We then proceed by estimating more complex parametric models such as the Mare model and the bivariate probit selection model.

### *3.1 Data*

#### *3.1.1 The US*

For the US we use data from the National Longitudinal Survey of Youth (NLSY79). The NLSY79 is a longitudinal study of a nationally representative sample of approximately 12,700 young men and women who were between 14 and 22 years old when they were first interviewed in 1979 (see Center for Human Resource Research 2006). In the empirical analysis we only use the 6,111 respondents from the cross-sectional sample and exclude respondents from the supplementary and military samples. The variables used are described below

### *3.1.2 The UK*

For the UK we use data from the National Child Development Study (NCDS). The NCDS is a longitudinal study of all children (approximately 17,500) born during the first week of March 1958 in the UK (see Plewis et al. 2004). The NCDS respondents have been followed since birth and surveys have been carried out in 1965, 1969, 1974, 1981, 1991, and 1999/2000.

### *3.1.3 Denmark*

For Denmark we use data from the Danish Youth Longitudinal Study (DYLS). The DYLS is a longitudinal study of a nationally representative sample of 3,151 men and women who were born in or around 1954 (see Jæger and Holm 2007). The DYLS respondents were first interviewed in 1968 at age 14 and have since been interviewed in 1970, 1973, 1976, 1992, 2001, and 2004.

### *3.1.4 The Netherlands*

For the Netherlands we use data from the Noord-Brabant cohort study (NB). The NB is a longitudinal study of a nationally representative sample of 5,771 men born in or around 1940 in the province of Noord-Brabant in the Netherlands. The NB respondents were first interviewed in 1952 and have since been interviewed in 1957-59, 1983, and 1993 (see van Praag 1992). Unlike the other data sets attrition has been quite considerable in the NB and in the empirical analysis we use a sample of 1,440 respondents.

-- TABLE 1 ABOUT HERE --

## *3.2 Variables*

### *3.2.1 Dependent Variables: Educational Transitions*

We construct dummy variables for all four countries which take the value 1 if respondents have completed educational transition 1 (into upper secondary education) and 2 (into higher education), and 0 otherwise.

In the analysis we consider respondents to have completed upper secondary education (transition 1) if they report the following degrees:

- NLSY79: High school or GED degree.
- NCDS: A-level qualifications (see Jackson et al. 2007).
- DYLS: Upper secondary education or equivalent degree (see Jæger and Holm 2007).
- NB: 5-6 years of high school (HAVO, VWO) or more (see de Haan 2005).

We consider respondents to have completed higher education (transition 2) if, in addition to having completed upper secondary education (transition 1), they report the following degrees:

- NLSY79: College degree.
- NCDS: Higher qualifications or University degree.
- DYLS: Intermediate or higher (University) tertiary education degree.
- NB: Higher vocational education (HBO) or University degree.

Table 1 show that in the US almost 85 percent of the respondents make the first transition. By contrast, only around 26 percent of the respondents make this transition in Denmark. Transition probabilities for the first transition in the UK and the Netherlands are around 37 percent. The table also shows that the conditional transition probabilities into higher education are much higher in the Netherlands, UK, and Denmark compared to in the US. Consequently, it appears that selection is low on the first transition and high on the second transition in the US but that the opposite is the case in the other countries.

-- TABLE 2 ABOUT HERE --

### *3.2.2 Explanatory Variables*



We include five family background and two individual-level variables in the analyses. These variables have often been used in previous studies on educational transitions and are, with noted exceptions, comparable across the four datasets. Descriptive statistics are shown in Table 2.

Family background variables one and two variables are *father* and *mother's education*. In the US, UK, and Danish data parents' education is measured by years of completed schooling. In the Dutch NB data parents' education is measured using a five-point ordinal scale with the values one through five indicating increasing levels of education (1 = First level school, 2 = Second level, first stage (LAVO, VGLO, MAVO), 3 = Second level, second stage (HBS, MMS, HAVO, VWO), 4 = Third level, first stage (HBO), and 5 = Third level, second stage (WO)).

The third variable is *gross monthly family income*. In the NLSY79 family income is measured in US dollars in 1980 (i.e., when respondents were 15-23 years old) and, if data on family income was missing in 1980, data for 1978 or 1979 (indexed to 1980 level) was used. In the NCDS family income is measured in Pounds Sterling when respondents were 16 years old. In the DYLS family income is measured in Danish Kroner when respondents were 14 years old. In the NB the only available measure of family income is a dummy variable indicating if family income exceeds 6,000 Dutch Guilders. However, as many cases do not have any information on family income we chose not to include this variable in the analysis for the Dutch data. In the empirical analysis we standardize the family income variables to have mean 0 and standard deviation 1.

The fourth and fifth family background variables are, respectively, *family type* (with a dummy variable for having been raised in a single-parent household) and *number of siblings*.

It was not possible to construct a measure of Socioeconomic Status (SES) that was comparable across the four surveys.

The two individual-level variables are the respondent's *sex* (with a dummy variable for female) and *cognitive ability*. In the NLSY79 we measure the respondent's cognitive ability by his or her 1980 score on the Armed Forces Qualification Test (AFQT). In the NCDS we use the respondent's total score on the General Ability Test (carried out at age 11). In the DYLS we use as our measure of cognitive ability the individual scores from a Principal Component Analysis of three tests of math,

reading, and spatial ability carried out when respondents were 14 years old. Finally, in the NB we use a similar measure of cognitive ability extracted from a factor analysis of 10 test items in different subjects carried out when respondents were 12 years old. In the empirical analysis we standardize all cognitive ability variables to have mean 0 and standard deviation 1.

#### 4. Results

The results section is divided into three subsections. In the first subsection we illustrate the magnitude of the selection problem. In the second subsection we present results from the initial analysis using the Manski bounds. In the third subsection we present results from simple probit and bivariate probit selection models.

-- FIGURE 3 ABOUT HERE--

In this first section we provide a simple illustration of the magnitude of selection in the four countries. Figure 3 displays significance levels (measured by  $t$ -statistics) from simple probit regressions for the effect of father and mother's education and family income on the probability of making the second transition plotted against the fraction of youth that makes the first transition. The figure shows that the larger is the proportion of youth that makes the first transition the stronger is the significance of the family background variables in the second transition. This finding clearly illustrates that a selection process exists since an "easy pass" in the first transition (with respect to the impact of family background) leads to a "tough" pass in the second transition and vice versa.

In order to further develop the analysis we move on to the non-parametric Manski bounds for the relationship between father's education (averaged by mother's education) and the probability of making the second transition. We use father's education as the explanatory variable in this part of the analysis because most previous studies find that father's education is one of the major family-background determinants of educational outcomes. The Manski bounds for the effect of father's education in our four countries are shown in Figures 4-7.

-- FIGURES 4-7 ABOUT HERE --

In addition to the Manski bounds the figures also plot the estimated effects of father's education on the probability of making the second transition (conditional on having made the first transition) from a simple probit model which does not account for selection at the first transition. In the probit model we control for mother's education. This is appropriate since the Manski bounds are averaged by mother's education.

In the US which has the lowest level of selection at the first transition we find that the Manski bounds are quite narrow and suggest a positive relationship between father's education and the probability of making the second transition. By contrast, the largely flat line for this effect in the simple probit model suggests that father's education does not have any effect on the probability of making the second transition. In the other three countries which have relatively strong selection at the first transition both the Manski bounds (which are much wider than in the US) and the simple probit analyses suggest that, if students make the first transition, father's education does not have any significant effect on the probability of making the second transition. Substantively, these results support the waning coefficients hypothesis since, as can be seen in Table 3 below, simple probit models indicate that father's education has a strong effect on the probability of making the first transition.

Our simple non-parametric approach suggests that the waning coefficients hypothesis is supported by the data. However, this approach does not address the problem that the population at risk of making the second transition in each country is selective. In order to address the selection problem we need to make distributional assumptions about the processes that govern educational transitions. As presented in section 2.1, we make the assumption that the latent propensity to make both educational transitions is jointly normally distributed. By doing so we propose a parametric structure which allows us to model the selection on unobserved variables as a bivariate normal process and to estimate the correlation between the unobserved variables in the two transitions,  $\rho$ . It should be kept in mind that although our bivariate probit selection model in theory deals with selection on unobserved variables there is no guarantee that our empirical data will accurately identify the selection process and the correlation between the unobservables. Selection models are often difficult to estimate and it is well-known that the correlation coefficient  $\rho$  has low statistical power (Copas and Li 1997; Angrist 2001). This turns out also to be the case in our analysis.

However, when pooling the datasets for the US, UK, and Denmark and including instrumental variables we get stable results.

-- TABLE 3 ABOUT HERE --

Table 3 shows results of the simple and bivariate probit models estimated separately for each country. In addition, we estimate models both with and without cognitive ability to account for the indirect impact of family background on educational transitions that runs through cognitive ability (“primary effects”; see Jackson et al. 2007).<sup>3</sup> From Table 3 we find that the difference in model fit (evaluated by the log-likelihood) between the simple and bivariate probit models is very small. For the UK, Denmark and the Netherlands we obtain very small chi-square test statistics and thus insignificant likelihood ratio tests when comparing the fit of the simple and bivariate probit selection models. For the US we find that the bivariate model has a significantly better fit than the simple model. However, apart from the correlation coefficients  $\rho$  (which only appears in the bivariate model) the estimated effects of the family background variables on the probability of making the two educational transitions is very similar in the simple and bivariate models. This result suggests that they explain the observed patterns in the data about equally well. Substantively, with the exception of the US, we find that the effects of father and mother’s education and family income decline from the first to the second transition. This pattern is consistent with the waning coefficients hypothesis, especially in the models that also control for cognitive ability. Finally, we find that none of the correlation coefficients  $\rho$  in the bivariate probit models are significant. This does not imply that there is no selection but rather that the sample sizes in the individual datasets are too small to obtain reliable estimates of the correlation between the unobservables. In other words, we might have insufficient data to properly identify the selection on unobserved variables. This result is substantively important because the datasets used to analyze educational transitions often are of comparable size to the ones we use in this analysis.

In order to improve our ability to identify the selection on unobserved variables in the bivariate probit selection model we pool the data from the NLSY79, NCDS, and DYLS into a single dataset.

---

<sup>3</sup> The bivariate probit selection models were estimated using STATA’s “heckprob” routine. The “biprobit” routine can also be used. The QLIM procedure in SAS also allows users to estimate bivariate probit models.

The Dutch NB data was not included in this pooled dataset because some of the explanatory variables were missing or were not comparable to the other datasets. The pooled dataset is much larger ( $N = 9,552$ ) and provides some further opportunities to study the selection process.

In the bivariate probit selection models reported in Table 3 the selection model is only identified on the basis of the parametric assumption of joint normality for the error terms. This is a strong assumption and the bivariate models might also be difficult to estimate empirically. To improve identification we exploit the availability of data from different countries in the pooled dataset to construct instrumental variables. The instrumental variables are the country-specific transition rates at the first and second transitions calculated as the fractions of respondents in our samples that pass the two educational transitions.<sup>4</sup> For the second transition, we furthermore construct variables that calculate the fraction of the *whole* sample that makes the second transition and the fraction that makes the second transition conditional on having made the first transition (i.e., the *conditional* fraction). The idea behind using the fractions that passes each transition as instruments is that we capture cross-national variation in the structural difficulty in passing the two transitions.<sup>5</sup> We then use the fraction passing the two transitions as instrumental variables in the bivariate probit selection model to obtain non-parametric identification of the correlation coefficient,  $\rho$ . This approach works because, as shown in Equation (2), the effect of the independent variables in the first transition are now linearly independent of the effect of the independent variables in the second transition. We can tentatively investigate instrument validity by examining changes in parameter estimates in the models using respectively fractions and country dummy variables as instruments and by comparing model fit in the two model specifications.

-- TABLE 4 ABOUT HERE --

---

<sup>4</sup> Instrumental variables are variables that enter only one equation in a simultaneous-equation model in order to identify other equations in the model (see Pearl 2000). In our application we cannot use country dummy variables as instrumental variables because they would appear (in identical form) in both equations in our model.

<sup>5</sup> In the present analysis we were unable to find any instruments measured at the individual level and which were available in all three datasets. However, future research on selection problems in educational transition models should develop instruments that differ across transitions.

Table 4 shows results from the simple and bivariate probit selection models estimated on the pooled dataset. Again, we estimate models with and without cognitive ability to gauge the indirect effects of the family background variables running through cognitive ability.

From the table we find, first, that there is a significant difference in model fit when we compare the models which include country dummy variables and the models which include fractions passing upper secondary and higher education. However, judging from the small differences between the estimated effects of the family background variables in the two different model types we conclude that, although significantly different in terms of model fit, the two models yield similar substantive results.

There is an interesting difference in the effect of the conditional fractions instruments in the simple and bivariate probit selection models. The effect of the conditional fraction that makes the second transition on the probability of making the second transition is positive in both models but substantially larger in the simple probit than in the bivariate models. This is a consequence of selection bias. In the simple probit model the effect of the conditional fraction instrument captures both the structural difficulty of making the second transition but also how many people failed the first transition. The higher is the fraction that makes the first transition (as in the US) the lower is the fraction that is able to make the second transition (and vice versa). The bivariate probit model takes unobserved factors into account and here the conditional fractions instrument captures how difficult the second transition is for students *with identical values on observed and unobserved characteristics*.

The instrument measuring the fraction of the *total* sample that makes the second transition has a negative effect on the probability of making the second transition in the simple probit model which controls for cognitive ability but is insignificant in the model without cognitive ability. This result can be explained by the fact that the conditional fractions instrument is already included as an explanatory variable. Consequently, the total fractions instrument captures the effect of the likelihood of making the second transition over and above the effect captured by the conditional fractions instrument; i.e., it measures the effect of having a large group of students that has made

the first transition and that has the opportunity to make the second transition.<sup>6</sup> In the simple probit model the total fraction is either insignificant (when we do not control for cognitive ability) or negative (when we control for cognitive ability). In the bivariate probit selection model the total fraction instrument is highly significant and positive indicating that, once we account for selection, the larger is the fraction of students that has the opportunity to make the second transition the larger is the fraction that actually makes this transition.

Moving on the effects of the family background variables we find clear differences between the simple and bivariate probit selection models. In the simple probit models we find evidence of waning coefficients and lower significance levels for the effects of father and mother's education and, especially, for the effect of family income. In the simple probit model that includes cognitive ability family income has a highly significant positive effect on the probability of making the first transition ( $\tilde{\beta} = 0.087$ ,  $t = 4.73$ ). In the second transition family income is no longer significant ( $\tilde{\beta} = 0.027$ ,  $t = 1.33$ ). In the model that does not include cognitive ability family income is significant at the second transition, but here the effect is upwardly biased due to the omission of cognitive ability. The effects of parents' education and their significance also decrease at transition 2 compared to at transition 1 in the simple probit models. Together, the results from the simple probit models support the waning coefficients hypothesis.

In the bivariate probit models in which we allow for a correlation among the unobservables across transitions we find little or no evidence of waning coefficients. The effects of parents' (and especially mother's) education decrease slightly at transition 2 but remain highly significant. We observe the same pattern across the different specifications of the bivariate probit model. Most striking, however, is the fact that the effect of family income which "waned" considerably in the second transition in the simple probit model is very strong and highly significant in the bivariate selection model. Consequently, our analysis suggests that in our pooled dataset selection on

---

<sup>6</sup> There are two educational strategies a country can pursue in order to have a large fraction passing the second transition: either by making the second transition easy (as measured by the conditional fraction making the second transition) or by having a large pool of students who are eligible for making the second transition (as measured by the total fraction).

unobserved variables appears to be particularly strongly related to families' economic resources.<sup>7</sup> In sum, our analysis using bivariate probit selection models suggests that the constant inequality hypothesis is more plausible than the waning coefficients hypothesis.

Finally, we note an interesting feature of the estimated correlation coefficients  $\rho$  in the bivariate probit selection models. The correlation coefficient  $\rho$  measures the degree to which students' unobserved characteristics are correlated across educational transitions. In the bivariate model that does not control for cognitive ability we find a highly significant correlation coefficient of about 0.8 indicating that approximately two thirds ( $\rho^2 = 0.811^2 = 0.658$ ) of the unobserved characteristics of students are common across transitions. In the model that controls for cognitive ability this share is only about one fourth ( $\rho^2 = 0.512^2 = 0.262$ ). Consequently, cognitive ability accounts for about 40 percent of the unobserved characteristics in the model that does not control for cognitive ability. This number is very high and suggests that cognitive ability is a very important determinant of educational attainment.

## 5. Conclusion

The Mare model of educational transitions has been highly influential in applied research on the impact of family background on educational success. Although a major conceptual and empirical improvement over previous models the Mare model (and educational transition models in general) is inherently susceptible to bias arising from selection on unobserved variables. Selection bias might lead to erroneous results concerning the effect of family background on educational success.

In this paper we study the impact of selection on unobserved variables in the educational transition model with two transitions. In the first part of the paper we show theoretically that biased parameter estimates of the effect of family background variables may arise from two phenomena: selection on unobserved variables (which, due to increasing selectivity, leads to downward bias) and scaling effects (which, due to different sample variances in the distributions of unobservables, leads to upward bias).

---

<sup>7</sup> We also included interactions between family income, the instruments, and the country dummies in the models to test for country-specific effects of family income. We only found a statistically significant interaction term for the UK. However, this effect is of little substantive importance.



In the second part of the paper we present results from an empirical analysis which illustrates the impact of selection bias on substantive results in the model with two educational transitions. We seek to distinguish between a “waning coefficients” and a “constant inequality” hypothesis. We use data from four countries and different empirical strategies to determine how many assumptions regarding the data generating process we need to invoke to distinguish between the two competing hypotheses. Using non-parametric Manski bounds we find that in three out of four countries the data is equally consistent with both hypotheses. We then impose a parametric structure on the data generating process and estimate simple and bivariate probit models. The bivariate probit model accounts for selection on unobserved variables by assuming a bivariate normal structure for the unobservables in the two educational transitions. Our empirical results using the simple probit model (similar to the traditional Mare model) support the waning coefficients hypothesis. However, when we pool the data our bivariate probit model suggests that the effect of family background is largely constant across educational transitions. Consequently, when we take selection on unobserved variables into account our results support the constant inequality hypothesis. Our analysis, although only illustrative, then suggests that selection on unobserved variables has a substantial impact on the estimated effects of family background variables on the probability of making successive educational transitions.

The main contribution of this paper is to show that selection on unobserved variables matters. Our empirical results indicate that the waning coefficients pattern found in previous studies using the Mare model, at least to some extent, might be driven by selection on unobserved variables. Consequently, in line with Cameron and Heckman (1998, 2001) we urge analysts which use the Mare model to pay explicit attention to selection bias.

The second contribution of the paper is to show that our ability to distinguish between competing hypotheses regarding the effect of family background on educational success depends critically on the number of assumptions we make regarding the data generating process. If we do not wish to make any assumptions we are typically not able to distinguish very accurately between competing hypotheses (unless we have very large samples). The Mare model imposes a parametric structure on the data generating process and assumes no selection on unobserved variables. We extended the Mare model by means of a bivariate probit model to allow for selection on unobserved variables.

Our bivariate model is easily estimated using standard software. However, the parametric assumption regarding the distribution of the unobservables in the bivariate probit model is not testable. This means that we have no way of testing whether our assumption regarding the unobserved part of the selection process is plausible. Furthermore, in the empirical analysis we find that we need auxiliary information in the form of instrumental variables to properly estimate the correlation between the unobservables in the two educational transitions.

In summary, our bivariate approach is no “magic bullet” which solves the problem of selection on unobserved variables in educational transition models. Essentially, in the future we need more and better data to estimate educational transition models. Bigger datasets would enable a more accurate identification of the underlying distributions of the unobservables. Better datasets would include variables explicitly designed to capture the selection process at the different educational transitions.

## Appendix 1. The Effect of Scaling on the Estimated Parameters in the Second Educational Transition

Define the inverse mills ratio as

$$\lambda(\beta_1 x_1) = \frac{\varphi(\beta_1 x_1)}{\Phi(\beta_1 x_1)},$$

where  $\varphi(\cdot)$  and  $\Phi(\cdot)$  are the standard normal density and distribution functions (Heckman 1979).

Note that  $Var(Y_2^*) = \sigma^2 \neq Var(Y_2^* | Y_1^* > 0)$  because

$$Var(Y_2^* | Y_1^* > 0) = Var(Y_2^* | Y_1 = 0) = \sigma^2 \left( 1 - \rho^2 \left\{ \beta_1 x_1 \lambda(\beta_1 x_1) + \lambda(\beta_1 x_1)^2 \right\} \right)$$

(Maddala 1983: 269; Heckman 1979). Indeed,  $\sigma^2 \geq Var(Y_2^* | Y_1 = 0)$  as the term

$0 \leq \beta_1 x_1 \lambda(\beta_1 x_1) + \lambda(\beta_1 x_1)^2 \leq 1$ . Note that this relationship only exists because of the assumption of joint normality.

## Appendix 2. The Nicoletti/Perracchi Approximation

In this appendix we derive the expression in Equation (1) which we use to show analytically the effect of selection and scaling on the conditional probability of making the second transition. By the definition of the conditional probability of making the second transition we get

$$\begin{aligned}
P(Y_2 = 1 | Y_1 = 1) &= P(Y_2^* > 0 | Y_1^* > 0) = \frac{P(Y_2^* > 0, Y_1^* > 0)}{P(Y_1^* > 0)} \\
&= \frac{1}{P(Y_1^* > 0)} \int_0^\infty \int_0^\infty f(r - \beta_1 x_1, s - \beta_2 x_2) dr ds \\
&= \frac{1}{F_1(\beta_1 x_1)} \int_{-\beta_1 x_1}^\infty \left\{ \int_{-\beta_2 x_2}^\infty f_{2|1}(u | v) du \right\} f_1(v) dv \\
&= \frac{1}{F_1(\beta_1 x_1)} \int_{-\beta_1 x_1}^\infty \left\{ F_{2|1}(-\beta_2 x_2 | v) \right\} f_1(v) dv \\
&= \frac{1}{\Phi(\beta_1 x_1)} \int_{-\beta_1 x_1}^\infty \left\{ \Phi \left( \frac{\beta_2 x_2 - \rho v}{\sqrt{1 - \rho^2}} \right) \right\} \varphi(v) dv \\
&\approx \Phi \left( \frac{\beta_2 x_2 + \rho \lambda(\beta_1 x_1)}{\sqrt{1 - \rho^2} \left\{ \beta_1 x_1 \lambda(\beta_1 x_1) + \lambda(\beta_1 x_1)^2 \right\}} \right),
\end{aligned}$$

where  $f(.,.)$  is the joint distribution of the unobservables,  $f_1$  and  $F_1(.)$  are the marginal density and distribution of the unobservables in the first transition,  $f_{2|1}$  and  $F_{2|1}$  are the conditional density and distribution of the unobservables in the second transition. The unobservables are assumed bivariate normal.

## References

- Anderson, John A. (1984): "Regression and ordered categorical variables". *Journal of the Royal Statistical Society Series B*, 46: 1-30.
- Angrist, Joshua D. (2001): "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice". *Journal of Business and Economic Statistics*, 19: 1-16.
- Arends-Kuening, Mary and Suzanne Duryea (2006): "The Effect of Parental Presence, Parents' Education, and Household Headship on Adolescents' Schooling and Work in Latin America". *Journal of Family and Economic Issues*, 27(2): 263-286.
- Breen, Richard and Jan O. Jonsson (2000): "Analyzing Educational Careers: A Multinomial Transition Model". *American Sociological Review*, 65(5): 754-772.
- Cameron, Stephen V. and James J. Heckman (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males". *Journal of Political Economy*, 106(2): 262-333.
- Cameron, Stephen V. and James J. Heckman (2001): "The Dynamics of Education Attainment for Black, Hispanic, and White Males". *Journal of Political Economy*, 109(31): 455-499.
- Center for Human Resource Research (2006): *NLSY79 User's Guide*. Columbus, OH: Center for Human Resource Research, Ohio State University.
- Chevalier, Arnaud and Gauthier Lanot (2002): "The Relative Effect of Family Characteristics and Financial Situation on Educational Achievement". *Education Economics*, 10(2): 165-181.
- Cobalti, Antonio (1990): "Schooling inequalities in Italy: trends over time". *European Sociological Review*, 6(3): 199-214.
- Colding, Bjørg (2006): "A dynamic analysis of educational progression of children of immigrants". *Labour Economics*, 13(4): 479-492.
- Copas, John and Li, H. G. (1997): "Inference for non-random samples (with discussion)". *Journal of the Royal Statistical Society B*, 59: 55-96.

- Garnier, Maurice A. and Lawrence E. Raffalovich (1984): "The Evolution of Educational Opportunities in France". *Sociology of Education*, 57: 1-10.
- Hansen, Marianne Nordli (1997): "Social and Economic Inequality in Educational Career: Do the Effects of Social Background Characteristics Decline?". *European Sociological Review*, 13(3): 305-320.
- Hauser, Robert M. and Megan Andrew (2006): "Another look at the stratification of educational transitions: The logistic response model with partial proportionality constraints". *Sociological Methodology*, 36: 1-26
- Heath, Anthony F. and Peter Clifford (1990): "Class inequalities in education in the twentieth century". *Journal of the Royal Statistical Society A*, 153(1): 1-16.
- Horowitz, Joel L and Charles F. Manski (1998): "Censoring of Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations". *Journal of Econometrics*, 84(1): 37-58.
- Jackson, Michelle, Robert E. Erikson, John H. Goldthorpe, and Meir Yaish (2007): "Primary and Secondary Effects in Class Differentials in Educational Attainment". *Acta Sociologica*, 50(3): 211-229.
- Jæger, Mads Meier and Anders Holm (2007): "Does parents' economic, cultural, and social capital explain the social class effect on educational attainment in the Scandinavian mobility regime?". *Social Science Research*, 36: 719-744.
- Lauer, Charlotte (2003): "Family background, cohort and education: A French-German comparison based on a multivariate ordered probit model of educational attainment". *Labour Economics*, 10(2): 231-251.
- Lucas, Samuel R. (2001): "Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects". *American Journal of Sociology*, 106(6): 1642-1690.
- Manski, Charles F. (1995): *Identification problems in the social sciences*. Cambridge, MA: Harvard University Press.

- Mare, Robert D. (1979): "Social Background Composition and Educational Growth". *Demography*, 16(1): 55-71.
- Mare, Robert D. (1980): "Social Background and School Continuation Decisions". *Journal of the American Statistical Association*, 75(370): 295-305.
- Mare, Robert D. (1981): "Change and Stability in Educational Stratification". *American Sociological Review*, 46(1): 72-87.
- Mare, Robert D. (1993): "Educational Stratification on Observed and Unobserved Components of Family Background", in Shavit, Yossi and Blossfeld, Hans-Peter (eds.): *Persistent Inequality. Changing Educational Attainment in Thirteen Countries*. Boulder, CO: Westview Press, pp. 351-376.
- Mare, Robert D. (2006): "Statistical Models of Educational Stratification: Hauser and Andrew's Models of School Transitions". *Sociological Methodology*, 36: 27-37.
- Nicoletti, C. and F. Perracchi (2001). *Two-step estimation of binary response models with sample selection*, unpublished working paper.
- Pearl, J. (2000): *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
- Plewis, Ian, Lisa Caldergood, Denise Hawkes, and Gad Nathan (2004): *National Child Development Study and 1970 British Cohort Study Technical Report*. London: Centre for Longitudinal Studies, Institute of Education, University of London.
- Raftery, Adrian E. and Michael Hout (1993): "Maximally Maintained Inequality: Expansion, Reform, and Opportunity in Irish Education, 1921-75". *Sociology of Education*, 66: 41-62.
- Shavit, Yossi and Hans-Peter Blossfeld (1993) (eds.): *Persistent Inequality. Changing Educational Attainment in Thirteen Countries*. Boulder, CO.: Westview Press.
- Shavit, Yossi and Karin Westerbeek (1998): "Educational Stratification in Italy. Reforms, Expansions, and Equality of Opportunity". *European Sociological Review*, 14(1): 33-47.

Vaid, Divya (2004): "Gendered Inequality in Educational Transitions". *Economic and Political Weekly*, August: 3927-3938.

van Praag, Bernard M. S. (1992): *Zomaar een dataset. "De Noordbrabantse Zesde Klassers". Een Presentatie van 15 jaar Onderzoek*. Amsterdam: Vakgroep Micro-economie, Universiteit van Amsterdam.

Van de Ven, Wynand P. M. M. and Bernard M. S. van Praag (1981): "The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection". *Journal of Econometrics* 17(2): 229-252.



Figure 1: The relationship between passing the first transition and background characteristics

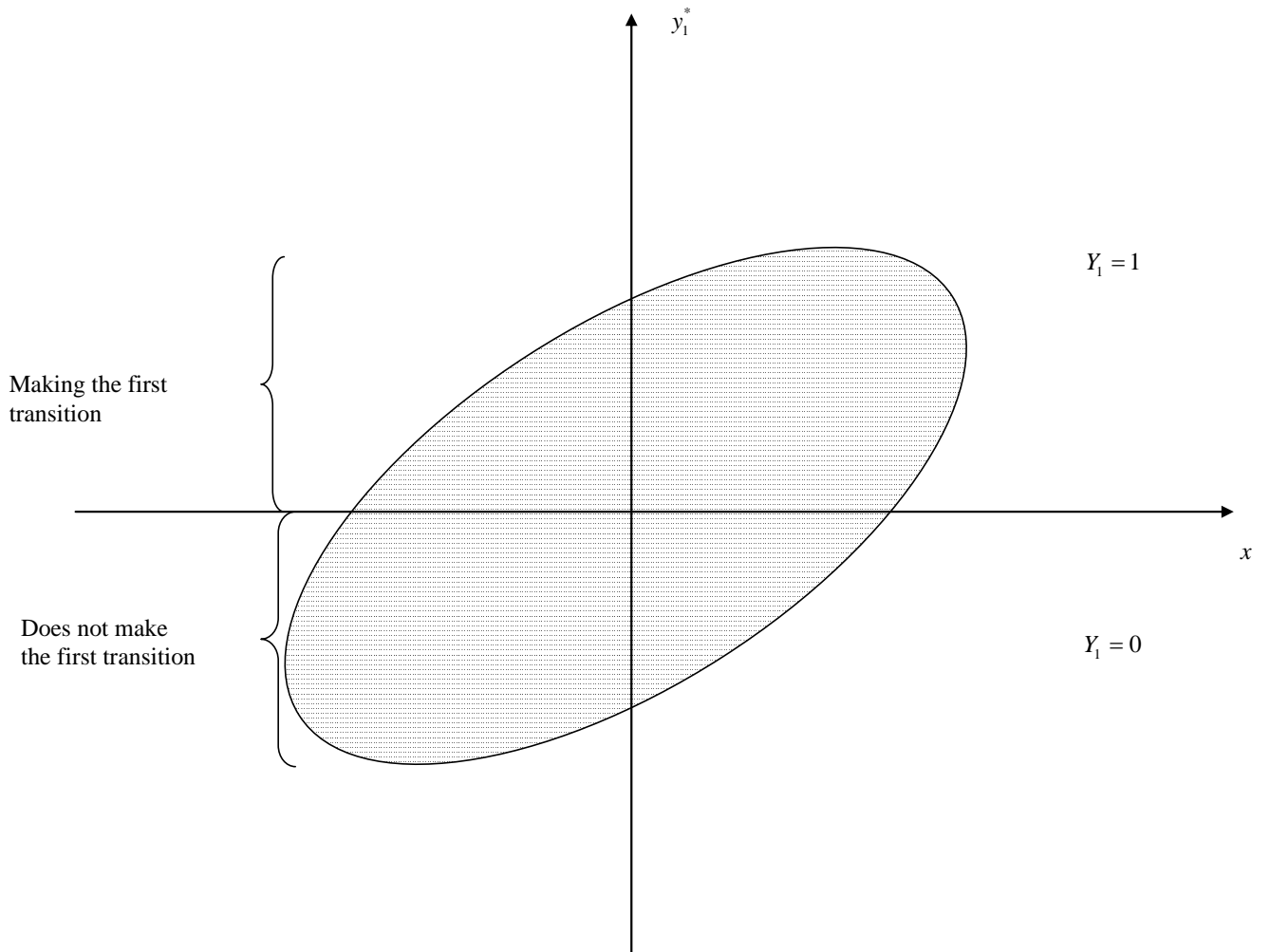


Figure 2. Relationship between passing the second transition and background characteristics

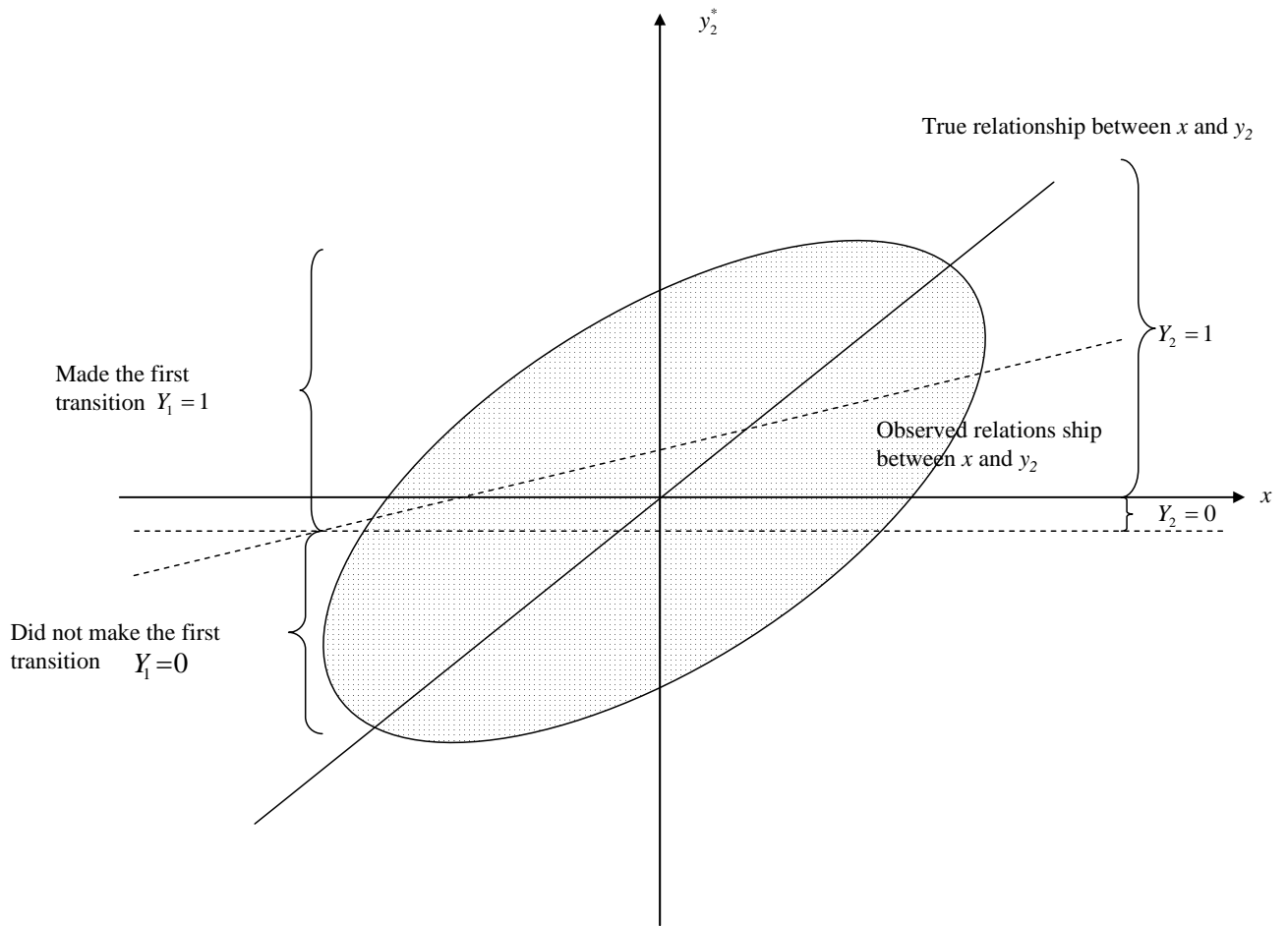


Figure 3. Significance of parental background on completing higher education as fraction of completing US

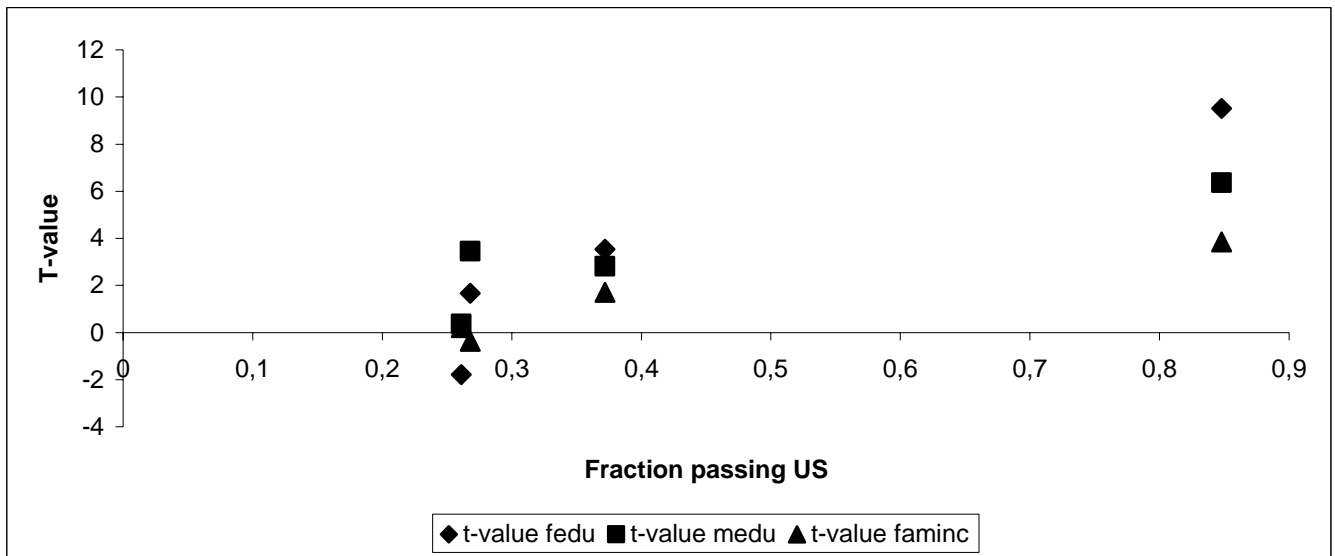


Figure 4. Manski bounds for the probability of making the second transition, NLSY79. Fathers' education averaged by mother's education

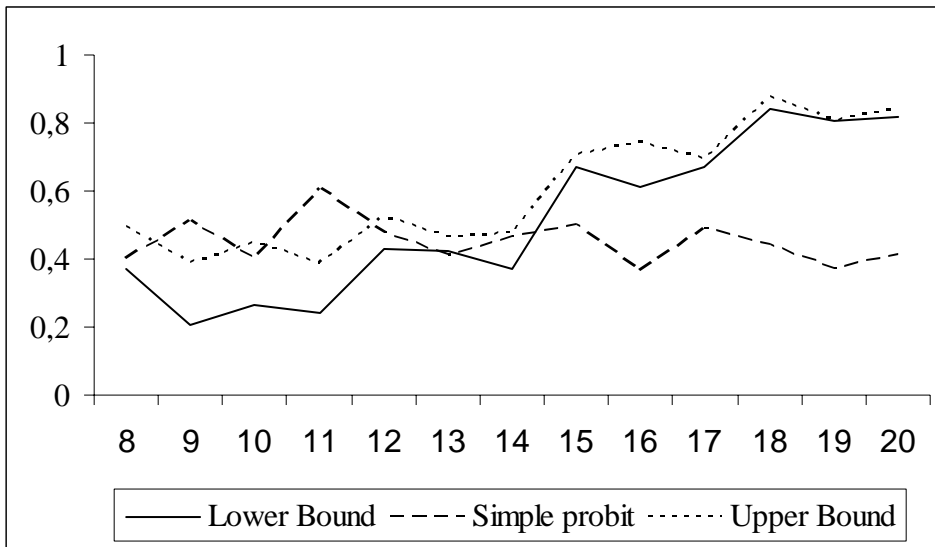


Figure 5. Manski bound for the probability of making the second transition, NCDS. Fathers' education averaged by mother's education

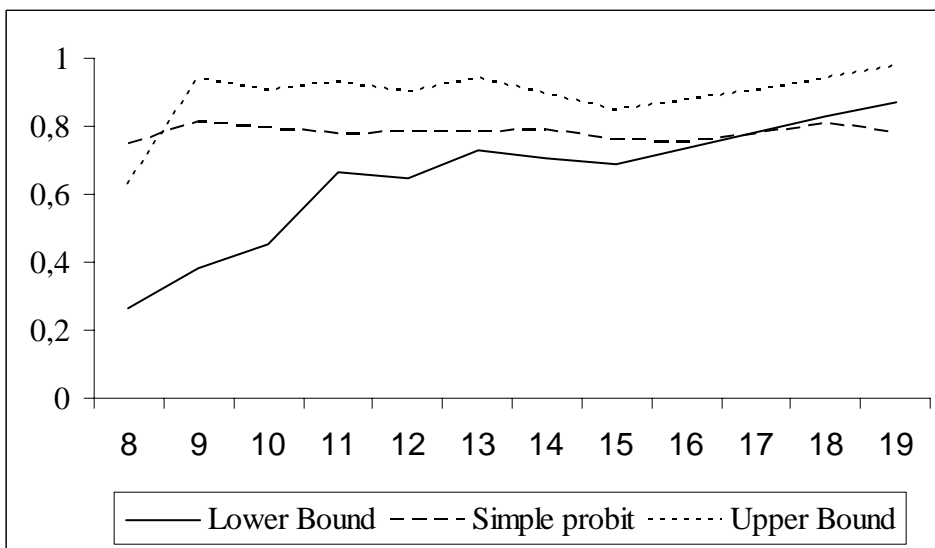


Figure 6. Manski bounds for the probability of making the second transition, DYLS. Fathers' education averaged by mother's education

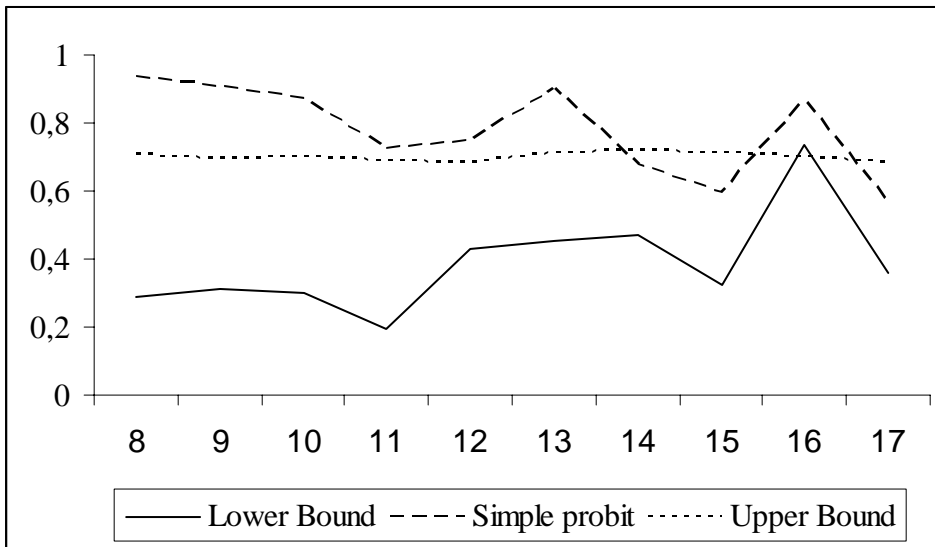


Figure 7. Manski bounds for the probability of making the second transition, Noord-Brabant Cohort. Fathers' education averaged by mother's education

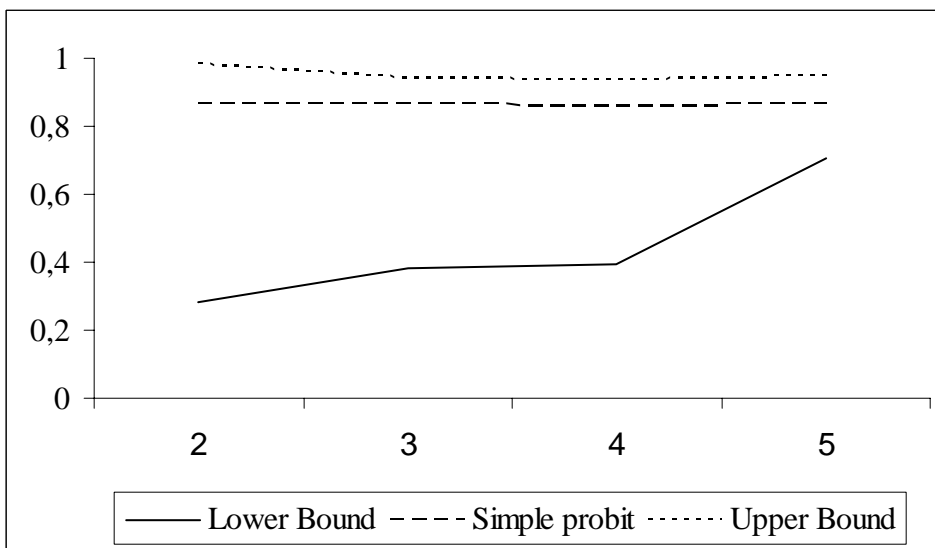


Table 1. Transition probabilities into Upper Secondary (US) and Higher Education

	Sample			
	NLSY79	NCDS	DYLS	NB
Proportion completing upper secondary	0.848	0.372	0.261	0.375
Proportion completing higher education (of those who have completed US)	0.403	0.796	0.663	0.863
Sample size	6,029	7,903	2,660	1,440

Table 2. Descriptive statistics. Means and Standard Deviations

	Sample							
	NLSY79		NCDS		DYLS		NB	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>Total sample</i>								
Father's education (Years)	11.90	3.58	10.03	2.02	9.62	2.70	2.43*	0.85
Mother's education (Years)	11.69	2.71	9.99	1.60	8.88	2.39	2.15*	0.64
Family income	0.06	0.99	0.79	0.68	0.08	0.99	-	-
Gender (1 = female)	0.47	-	0.55	-	0.46	-	-	-
Lone parent (1 = not living with both biological parents)	0.20	-	0.09	-	0.03	-	-	-
No. siblings	3.16	2.12	2.37	1.78	2.19	1.41	5.83	3.34
Cognitive ability	0.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00
<i>Sample completing upper secondary education</i>								
Father's education	12.19	3.45	10.91	2.73	11.18	3.01	2.87*	1.16
Mother's education	11.94	2.59	10.67	2.16	10.21	2.71	2.51*	0.91
Family income	0.14	1.00	0.98	0.73	0.42	1.25	-	-
Gender (1 = female)	0.50	-	0.51	-	0.43	-	-	-
Lone parent (1 = not living with both biological parents)	0.18	-	0.06	-	0.03	-	-	-
No. siblings	3.04	2.02	1.89	1.36	1.81	1.06	5.55	2.61
Cognitive ability	0.17	0.94	0.45	0.77	0.40	0.57	0.42	0.63

Note: \* Father and mother's education are measured by five ordered levels coded as 0-4.

Table 3. Results from simple and bivariate probit selection models. Parameter estimates with *t*-statistics in parenthesis

	NLSY79				NCDS				DYLS				NB			
	Simple probit	Simple probit	Bivariate probit	Bivariate probit	Simple probit	Simple probit	Bivariate probit	Bivariate probit	Simple probit	Simple probit	Bivariate probit	Bivariate probit	Simple probit	Simple probit	Bivariate probit	Bivariate probit
	Transition to upper secondary education															
Father's education	0.039 (3.95)	0.009 (0.58)	0.0415 (4.13)	0.009 (0.79)	0.159 (10.68)	0.120 (7.69)	0.160 (10.71)	0.120 (7.68)	0.111 (6.50)	0.072 (3.82)	0.110 (6.44)	0.073 (3.86)	0.452 (7.47)	0.418 (6.83)	0.456 (7.54)	0.422 (6.85)
Mother's education	0.077 (6.10)	0.045 (3.32)	0.081 (6.32)	0.048 (3.54)	0.202 (10.40)	0.157 (7.75)	0.201 (10.33)	0.127 (7.74)	0.104 (5.74)	0.105 (5.27)	0.106 (5.86)	0.103 (5.17)	0.327 (3.98)	0.327 (3.91)	0.317 (3.96)	0.322 (3.69)
Family income	0.178 (4.99)	0.053 (1.39)	0.173 (4.90)	0.055 (1.45)	0.163 (4.60)	0.096 (2.53)	0.163 (4.62)	0.096 (2.53)	0.065 (1.63)	0.068 (1.58)	0.065 (1.66)	0.071 (1.63)	-	-	-	-
Gender	0.302 (5.48)	0.304 (5.19)	0.306 (5.56)	0.303 (5.19)	-0.186 (4.24)	-0.299 (6.31)	-0.186 (4.25)	-0.299 (6.31)	-0.021 (0.30)	0.029 (0.37)	-0.037 (0.52)	0.043 (0.55)	-	-	-	-
Lone parent	-0.401 (6.38)	-0.343 (5.11)	-0.413 (6.56)	-0.359 (5.38)	-0.331 (4.00)	-0.254 (2.85)	-0.333 (4.03)	-0.254 (2.85)	-0.126 (0.56)	-0.146 (0.60)	-0.140 (0.62)	-0.148 (0.61)	-	-	-	-
No. siblings	-0.040 (3.21)	-0.016 (1.25)	-0.041 (3.32)	-0.019 (1.45)	-0.152 (10.80)	-0.099 (6.47)	-0.152 (10.81)	-0.099 (6.47)	-0.087 (3.08)	-0.091 (2.91)	-0.085 (3.03)	-0.091 (2.93)	-0.021 (1.34)	-0.065 (4.47)	-0.019 (1.26)	-0.064 (4.34)
Cognitive ability	-	0.636 (15.70)	-	0.637 (15.84)	-	0.679 (23.77)	-	0.679 (23.76)	-	1.023 (15.41)	-	1.027 (15.48)	-	0.256 (6.68)	-	0.256 (6.65)
	Transition to higher education															
Father's education	0.083 (9.52)	0.052 (5.56)	0.085 (9.79)	0.051 (5.61)	0.072 (3.54)	0.059 (2.77)	0.043 (1.54)	0.061 (1.71)	-0.048 (1.79)	-0.057 (2.07)	-0.099 (4.22)	-0.024 (0.55)	-0.064 (0.70)	-0.080 (0.87)	0.359 (5.24)	0.251 (0.40)
Mother's education	0.074 (6.36)	0.039 (3.22)	0.081 (6.32)	0.044 (3.70)	0.073 (2.82)	0.064 (2.36)	0.037 (1.04)	0.066 (1.41)	0.008 (0.29)	0.009 (0.32)	-0.049 (1.91)	0.043 (1.03)	0.119 (0.97)	0.121 (0.98)	0.300 (3.83)	0.281 (1.27)
Family income	0.098 (3.85)	0.024 (0.91)	0.173 (4.90)	0.030 (1.14)	0.102 (1.70)	0.072 (1.13)	0.060 (0.91)	0.073 (1.06)	0.003 (0.06)	0.011 (0.20)	-0.028 (0.60)	0.031 (0.56)	-	-	-	-
Gender	0.137 (3.49)	0.222 (4.66)	0.306 (5.56)	0.251 (5.46)	-0.150 (1.97)	-0.229 (2.85)	-0.104 (1.27)	-0.233 (2.00)	0.124 (1.02)	0.151 (1.23)	0.089 (0.90)	0.132 (1.10)	-	-	-	-
Lone parent	-0.074 (1.21)	0.009 (0.14)	-0.413 (6.56)	0.047 (0.78)	-0.129 (0.81)	-0.124 (0.75)	-0.042 (0.25)	-0.128 (0.70)	-0.212 (0.60)	-0.192 (0.54)	-0.091 (0.32)	-0.232 (0.71)	-	-	-	-
No. siblings	-0.036 (2.98)	-0.016 (1.33)	-0.41 (3.32)	-0.019 (1.51)	-0.041 (1.53)	-0.009 (0.31)	0.000 (0.01)	-0.010 (0.23)	-0.061 (1.06)	-0.064 (1.10)	0.020 (0.36)	-0.094 (1.66)	0.046 (1.32)	0.055 (1.55)	-0.007 (0.40)	-0.023 (0.12)
Cognitive ability	-	0.627 (21.24)	-	0.681 (24.00)	-	0.577 (10.93)	-	0.588 (2.72)	-	0.229 (2.11)	-	0.560 (1.83)	-	0.179 (1.22)	-	0.280 (3.16)
$\rho$	-	-	0.842 (0.34)	0.934 (0.10)	-	-	-0.373 (1.24)	0.024 (0.05)	-	-	-0.904 (0.89)	0.557 (0.89)	-	-	0.995 (1.28)	0.924 (0.27)
Log-Likelihood	-3445.0	-3049.7	-3441.6	-3046.6	-2919.4	-2527.0	-2918.7	-2527.0	-1141.1	-987.1	-1140.4	-986.7	-690.0	-667.7	-689.6	-667.3
Sample size			3,949				3,990				1,613				1,440	



Table 4. Results from simple and bivariate probit selection models with instruments. Parameter estimates with *t*-statistics in parenthesis

	Models without cognitive ability				Models with cognitive ability			
	Simple probit	Bivariate probit	Simple probit, country dummies	Bivariate probit, country dummies	Simple probit	Bivariate probit	Simple probit, country dummies	Bivariate probit, country dummies
Transition to upper secondary education								
Father's education	0.079 (11.02)	0.084 (11.91)	0.077 (10.60)	0.082 (11.45)	0.049 (6.24)	0.056 (7.65)	0.045 (574)	0.0472 (6.06)
Mother's education	0.097 (11.02)	0.100 (11.57)	0.102 (11.52)	0.105 (12.02)	0.069 (7.38)	0.067 (7.61)	0.076 (8.03)	0.0775 (8.26)
Family income	0.142 (8.27)	0.139 (8.29)	0.142 (8.23)	0.138 (8.23)	0.087 (4.73)	0.037 (1.97)	0.087 (4.69)	0.089 (4.04)
Gender	-0.009 (0.30)	0.015 (0.50)	0.002 (0.08)	0.025 (0.83)	-0.067 (2.07)	-0.011 (0.36)	-0.052 (1.60)	-0.037 (1.14)
Lone parent	-0.360 (7.43)	-0.352 (7.36)	-0.358 (7.37)	-0.352 (7.32)	-0.283 (5.46)	0.276 (5.51)	-0.282 (5.40)	-0.292 (5.61)
No. siblings	-0.076 (9.12)	-0.076 (9.14)	-0.077 (9.13)	-0.076 (9.17)	-0.038 (4.24)	-0.040 (4.69)	-0.039 (4.31)	-0.040 (4.43)
Cognitive ability	-	-	-	-	0.692 (32.38)	0.674 (34.28)	0.699 (32.42)	0.696 (32.44)
US	-	-	1.608 (32.04)	1.606 (32.20)	-	-	2.088 (36.91)	2.09 (36.91)
UK	-	-	0.136 (3.29)	0.143 (3.45)	-	-	0.183 (4.17)	0.126 (2.74)
Fraction passing upper secondary	2.936 (39.51)	3.015 (35.28)	-	-	3.785 (43.50)	3.749 (43.57)	-	-
Transition to higher education								
Father's education	0.071 (9.37)	0.083 (12.04)	0.071 (9.37)	0.082 (11.85)	0.042 (5.27)	0.046 (5.95)	0.042 (5.27)	0.045 (5.83)
Mother's education	0.060 (6.19)	0.083 (9.33)	0.060 (6.19)	0.084 (9.38)	0.036 (3.49)	0.045 (4.52)	0.036 (3.49)	0.046 (4.53)
Family income	0.069 (3.52)	0.113 (6.25)	0.069 (3.52)	0.112 (6.16)	0.027 (1.33)	0.044 (2.16)	0.027 (1.33)	0.043 (2.10)
Gender	0.081 (2.20)	0.062 (1.89)	0.081 (2.20)	0.067 (2.03)	0.112 (2.91)	0.098 (2.61)	0.112 (2.91)	0.102 (2.72)
Lone parent	-0.105 (1.89)	-0.186 (3.62)	-0.105 (1.89)	-0.184 (3.57)	-0.023 (0.39)	-0.055 (0.96)	-0.023 (0.39)	-0.052 (0.92)
No. siblings	-0.039 (3.70)	-0.057 (5.97)	-0.039 (3.70)	-0.057 (5.92)	-0.018 (1.58)	-0.023 (2.16)	-0.018 (1.58)	-0.023 (2.13)
Cognitive ability	-	-	-	-	0.594 (24.22)	0.678 (25.28)	0.594 (24.22)	0.676 (24.84)
US	-	-	-0.769 (11.32)	-0.009 (0.08)	-	-	-0.493 (7.05)	-0.059 (0.48)
UK	-	-	0.409 (5.77)	0.358 (5.90)	-	-	0.421 (5.82)	0.425 (6.29)
Total fraction passing higher education	0.078 (0.16)	2.528 (6.31)	-	-	-0.827 (1.66)	2.521 (4.46)	-	-
Conditional fraction passing higher education	3.009 (20.13)	1.255 (4.34)	-	-	2.431 (15.58)	1.555 (5.70)	-	-
$\rho$	-	0.811 (5.01)	-	0.799 (5.05)	-	0.512 (3.91)	-	0.488 (3.82)
<i>Log-Likelihood</i>	-7697.9	-7674.6	-7687.2	-7664.7	-6761.9	-6751.4	-6746.7	-6737.0
Sample size	9,552							