Multiple paths in educational transitions: A multinomial transition model with unobserved heterogeneity

Karlson, Kristian Bernt

Published in:
Research in Social Stratification and Mobility

DOI:
10.1016/j.rssm.2011.02.001

Publication date:
2011

Multiple paths in educational transitions: A multinomial transition model with unobserved heterogeneity

THIS PAPER IS PUBLISHED IN
RESEARCH IN SOCIAL STRATIFICATION AND MOBILITY, 2011
VOL. 29, 323-341
Link: http://www.sciencedirect.com/science/article/pii/S0276562411000072
This is a post-print (i.e. final draft post-refereeing) version according to SHERPA/ROMEO

Kristian Bernt Karlson*

Revised, January 7, 2011

(Original, August 20, 2010)

8520 words (including all text, notes, and references)

5 tables, 1 figure, 1 appendix

Running head: Multiple paths in educational transitions

* The Danish School of Education, Aarhus University, Tuborgvej 164, DK-2400 Copenhagen NV, Denmark. Telephone: 0045 23369285. Fax: 0045 88889001. SFI – The Danish National Centre for Social Research, Herlufs Trolle Gade 11, DK-1052 Copenhagen K. Email: kbk@dpu.dk.

Acknowledgements: I thank Anders Holm and Mads Meier Jæger for helpful comments and suggestions.

Multiple paths in educational transitions:

A multinomial transition model with unobserved heterogeneity

Abstract

In many countries educational branching points consist of more than two qualitatively different alternatives, and only some alternatives provide the opportunity of continuing into higher education. I develop a multinomial transition model for modeling the effects of family background characteristics and individual characteristics on these complex educational careers. The model controls for unobserved heterogeneity that may, if ignored, result in biased estimates. Compared to previous research, I explicitly include instrumental variables that ensure identification of the unobserved component. I apply the model to the Danish case and analyze data which covers the educational careers of a cohort of Danes born around 1954. I find that the model brings forward non-trivial heterogeneity in the influence of family background and ability on qualitatively different choice alternatives both within and across transitions. I also find that not controlling for unobserved heterogeneity leads to marked underestimation of the family background effect on both early and late transitions in the educational career.

Keywords: multinomial transition model, unobserved heterogeneity, educational transitions, educational decisions, inequality of educational opportunity

Multiple paths in educational transitions:

A multinomial transition model with unobserved heterogeneity

1. Introduction

Individuals obtain educational qualifications through various routes in the educational system. These routes are defined by the structure of the educational system and can be thought of as comprising a set of sequential branching points (Boudon, 1974; Gambetta, 1987). In many countries particularly in Western Europe these branching points involve more than two choice alternatives. In these situations, the standard sequential logit model (SLM) for educational transitions developed by Mare (1980, 1981) is inappropriate because it does not capture the multiple and unordered nature of choice alternatives. Consequently, using the SLM for diversified educational systems may ignore important heterogeneity in the ways family background influences educational decisions. In an influential paper Breen and Jonsson (2000) proposed a multinomial transition model (MTM) to accommodate the multiple and unordered choice alternatives of diversified educational systems.[1] In this paper I follow Breen and Jonsson, but compared to their strategy, I explicitly account for unobserved heterogeneity, and I include two instrumental variables that ensure identification of the unobserved component. In contrast to the finding of Breen and Jonsson that results are robust to unobserved heterogeneity, I find that not accounting for unobserved heterogeneity may downwardly bias estimates at both early and late transitions.

While the MTM suggested by Breen and Jonsson more than a decade ago is a viable alternative to the SLM, the model has so far not been adopted in mainstream stratification research. Rather, stratification researchers using multinomial logit models typically focus on

[1] Other alternatives to the SLM have been suggested, in particular the ordered logit or probit model (see Cameron & Heckman, 1998; Lucas, 2001; Breen *et al.*, 2009; Ballarino and Shadee, 2010).

either the transition from primary to secondary education or the transition from secondary to tertiary education. For the transition from primary to secondary education, Need and de Jong (2001) provide results for the Netherlands, Becker (2003) for Germany, Hansen (2007) for Norway, Kreidl (2004) for the Czech Republic, Jao and McKeever (2006) for Taiwan, Ayalon and Shavit (2004) for Israel, and Jæger (2009) for Denmark. For the transition from secondary to tertiary or higher education, Tieben and Wolbers (2010) and Tolsma *et al*. (2010) provide results for the Netherlands, Becker and Hecken (2009) for Germany, and Mastakaasa (2006) for Norway. Modeling two or more transitions with a MTM has thus not found its way into mainstream practice of stratification research. This tendency is problematic for two reasons. First, the standard multinomial logit model is based on the restrictive assumption of Independence from Irrelevant Alternatives (IIA).[2] Second, applying multinomial logit models to later transitions (e.g., from secondary to tertiary education) ignores the fact that individuals who face these decisions represent a selective sample. Regression coefficients estimated on selective samples may be influenced by unobserved heterogeneity and may therefore suffer from sample selection bias (cf. Heckman, 1979).

Using measures of peer group influence as instrumental variables, I estimate a MTM that explicitly accounts for unobserved heterogeneity. I name this model the multinomial transition model with unobserved heterogeneity (MTMU). The model is a flexible finite mixture model that accommodates both selection bias and violations of the IIA assumption. With the MTMU I model the effects of family background and individual characteristics on the probability of making two transitions using data from the Danish Longitudinal Survey of Youth (DLSY). First, in the transition from primary to secondary education, individuals complete the academic track, complete the vocational track, or leave the educational system.

[2] Similar to the model proposed in this paper, Jæger (2009) uses a finite mixture model to overcome this limitation. However, Jæger (2009) is an exception to the rule.

Second, in the transition from secondary to tertiary education, individuals complete the university track, complete the short-cycle track, or leave the educational system. Thus, in contrast to the model by Breen and Jonsson (2000), my MTMU is simpler in terms of the number of branching points to be estimated and in terms the possible pathways to pursue. Moreover, in my analysis the academic track in secondary education is an absorbing state, which means that only individuals completing academic secondary education "survive" to make the transition into tertiary education. This property is not an inherent part of the model, but rather reflects the institutional structure of the educational system in Denmark in the 1960s and 1970s. The model consequently extends to other educational systems with different structures than the one analyzed in this paper (e.g., such as the structure in Breen and Jonsson (2000)). I use Stata command *gllamm* to estimate my model, and sample data and code is available from the journal website. I proceed as follows. First, I present the multinomial transition model with unobserved heterogeneity. Second, I introduce the data from the DLSY. Third, I present the results. Fourth, I conclude with a discussion of the advantages of the MTMU.

2. A multinomial transition model with unobserved heterogeneity

In this section I present the MTMU. The MTMU is an extension of the SLM popularized by Mare (1980, 1981) which, first, allows for more than two choice alternatives at two or more branching points and, second, controls for the possible selection bias caused by unobserved heterogeneity. I capture these unobserved variables with a finite number of latent classes. This specification is highly flexible and makes the model a finite mixture model. Conceptually, the model may be thought of as two or more multinomial logit models with a common, unobserved variable affecting each choice alternative relative to a baseline alternative for each transition. I identify the model by including two alternative-specific instrumental variables at the first transition. However, because the validity of instrumental variables cannot be proved but rather rests on arguments, I also perform a robustness analysis using another identification strategy inspired by Cameron and Heckman (1998). In the next subsections I present the model. Because multinomial transition models are nothing more than two or more multinomial logit models, I first present the multinomial logit model. I then give the intuition behind the MTMU, and I end the section with a formal presentation of the MTMU.

2.1 The multinomial logit model

I first consider a multinomial logit model presented in a latent variable framework (cf. McFadden, 1974; Powers & Xie, 2000:238-9). Let y^*_{ia} be a continuous latent propensity of individual i to choose the ath educational alternative, where $i = 1,..,N$ and $a = 1,..,A$. Let x_{ij} be the jth explanatory variable for individual i, where $j = 1,\ldots,J$. Let y^*_{ia} be a linear function function of x_{ij} and an alternative-specific random error term ξ_{ia}:

$$y^*_{ia} = \sum_{j=1}^{J} b_{aj} x_{ij} + \xi_{ia} \tag{1}$$

b_{aj} is the effect of x_{ij} on the latent propensity for alternative a. In (1) each individual has an unobserved propensity to choose the ath alternative, but I only observe which of the A alternatives the individual *actually* chooses. To identify the model, I need to assume that

$$y_{ia} = a \ \ if \ \ y^*_{ia} > y^*_{ia'} \quad for\,all\, a \neq a'. \tag{2}$$

In other words, I assume that the individual chooses the alternative for which he or she has the largest propensity. Moreover, I assume that the random error term, ξ_{ia}, is uncorrelated across alternatives and that it follows a standard type-I extreme value distribution (Train, 2009). The assumption of uncorrelated error terms is also known as the assumption of Independence from

Irrelevant Alternatives (IIA) (McFadden, 1986). IIA implies that if I remove one alternative individuals who would have chosen this alternative are randomly distributed among the remaining alternatives (McFadden, 1974; Wooldridge, 2002:501-2).[3] Given the assumptions on the error terms, the probability of choosing *a* can be written as

$$\Pr(y_i = a \mid x_{ij}) = \frac{\exp(\sum_{j=1}^{J} \beta_{aj} x_{ij})}{\sum_{s=1}^{A} \exp(\sum_{j=1}^{J} \beta_{sj} x_{ij})}, \quad (3)$$

where $\sum_{a=1}^{A} \Pr(y_i = a) = 1$. The coefficients, β_{aj}, are well-known logit coefficients (i.e., log odds-ratios). These coefficients have a certain relation to the coefficients in the underlying linear model in (1), because β_{aj} are the underlying coefficients from that model divided by a scale parameter, $\beta_{aj} = \frac{b_{aj}}{\sigma}$. The scale parameter can be thought of as reflecting the unobserved portion of the model (see Train, 2009:40). This scale identification restriction is a property of the logit model and means that comparisons of coefficients across models or samples are hampered. If scales vary across models, then we cannot know whether differences in regression coefficients across alternatives are due to differences in residual variance or in the underlying regression coefficients (cf. Swait & Louviere, 1993; Allison, 1999). I return to this issue in the results section, where I report average partial effects, because these are less sensitive to this scale identification issue.

[3] To fix ideas, imagine an educational system with three choice options in the transition from primary to secondary education: exit, vocational track, and academic track. If IIA holds, then the consequence of closing down the academic track (i.e., removing that choice alternative) is that individuals that would have chosen the academic track (i.e., individuals with a high propensity for doing so) would be distributed randomly across the two remaining tracks. This assumption is not realistic in this example because we would expect the affected individuals to have a higher propensity to enroll in the vocational track than to exit the educational system. The multinomial transition model with unobserved heterogeneity I present below relaxes the IIA assumption.

I normalize the model in (3) such that $\beta_{1j} = 0$, which is equivalent to stating that alternative *a = 1* is a reference category that defines the contrast alternative against which the other alternatives are defined. I may thus rewrite equation (3) such that

$$\Pr(y_i = a \mid x_{ij}) = \frac{\exp(\sum_{j=1}^{J} \beta_{aj} x_{ij})}{1 + \sum_{s=2}^{A} \exp(\sum_{j=1}^{J} \beta_{sj} x_{ij})} \quad \text{for } a > 1. \tag{4}$$

Taking the log-odds of the probability in (4) returns the familiar multinomial logit model:

$$\text{logit}[\Pr(y_i = a \mid x_{ij})] = \sum_{j=1}^{J} \beta_{aj} x_{ij} \quad \text{for } a > 1. \tag{5}$$

The multinomial transition model is an extension of the model in (4) and (5).

2.2 Intuition behind the MTMU

Before I proceed to the MTMU, I give an account of the intuition behind the model, in particular the role played by the unobserved variable. Imagine two transition points, each with three choice alternatives: primary to secondary education (exit, vocational track, academic track) and secondary to tertiary education (exit, short-cycle track, university track). Assume that only students who complete the academic track at the first transition are allowed to make the second transition (i.e., they "survive" the first transition). Similar to Cameron and Heckman (1998:296), I assume that the population of students can be divided into two mutually exclusive types (or classes). To fix ideas, let the first type be characterized by low educational aspirations, while the other type is characterized by high educational aspirations.[4] The researcher does not observe whether an individual belongs to one type or the other, i.e., the aspiration variable is unobserved. In addition, imagine that the researcher observes the social class membership of the student (low/high), i.e., the social class variable is observed.

[4] I interpret the unobserved variable in terms of aspirations only to illustrate the logic of the model. Because the variable is inherently unobserved, I cannot know whether the omitted variable in fact measures aspirations.

I expect that, compared to students with low aspirations, students with high aspirations are more likely to complete the academic track in the first transition and, if they "survive," also more likely to complete the university track in the second transition. I also expect that, compared to lower class students, higher class students are more likely to complete these tracks. From these expectations and assumptions, it follows that students who survive to face the choices of the second transition have higher aspirations and tend to come more from the high social class than those who do not survive. This selection or sorting mechanism induces a *negative* correlation between aspirations and social class in the sample of those who survive (Cameron & Heckman, 1998:276; cf. Mare, 1980:298f, 1981:82). Because omitted variables (aspirations) which are correlated with both the observed variables (social class) and the outcome (completing the university track in the second transition) give rise to bias in the estimates of the observed variables, the selection mechanism obscures the estimates of the influence of social class on the second transition (cf. Heckman 1979). Cameron & Heckman (1998) refer to this kind of selection bias as dynamic selection bias.

The consequences of dynamic selection bias on the estimates at later transitions may be severe (depending on the magnitude of the induced correlation between the observed and unobserved variables and on the magnitude of the effect of the unobserved variable on the outcome). However, in the multinomial case matters are even more complicated. I identify a standard multinomial logit model as in (5) through the assumption of IIA. If this assumption does not hold, I expect bias to arise in the estimates of the multinomial logit model. A similar logic holds for a multinomial transition model.[5] I would expect that, compared to students

[5] Olsen (1982) shows that the consequences of violations of IIA assumption for the estimates are highly similar to the consequences of sample selection. McFadden (1986) and Train (2009) also explain the consequences of violating IIA assumption. The intuition is that if the alternative-specific errors in the latent model are correlated, i.e., if the IIA assumption does not hold, then a part of the effect of a predictor on one alternative is due to the effects on the other alternatives. The mixed multinomial logit model controls for this bias by reweighing the model such that the effect of a predictor on one alternative is purged from the confounding influence of another

with low aspirations, students with high aspirations are more likely to complete the academic track in the first transition and university track in the second transition. A consequence of this expectation is that students would not distribute themselves randomly across the remaining alternatives if one of the choice alternatives was removed. The IIA assumption is thus violated: If the academic track at the first transition was removed then I would expect those with high aspirations to opt for the vocational track more so than those with low aspirations. I would expect the same with respect to social class, thereby inducing a correlation between the unobserved (aspirations) and observed (social class) variables that may result in biased estimates of the observed variables. Thus, unobserved heterogeneity may, if not corrected for, bias estimates both through dynamic selection and through violations of the IIA assumption. The model I now turn to corrects for both sources of bias.

2.3 The multinomial transition model with unobserved heterogeneity

In this subsection I extend the model in (4) and (5) to include two or more transitions and to accommodate unobserved heterogeneity. Let y^*_{iak} be a continuous latent propensity of individual i associated with choice of the ath educational alternative at the kth transition, where $k = 1,\dots,K$. I define x_{ij} as before. I now decompose the error term similar to the one in (1) into a systematic and random component: $\xi_{iak} = u_{iak} + \varepsilon_{iak}$. u_{iak} is approximated with a discrete variable, u_{akw}, drawn from a discrete distribution with W latent classes, where $w = 1,...,W$ and π_w is the share in class w and where $\sum_{w=1}^{W} \pi_w = 1$. The W latent classes can be thought of as groups of individuals that have similar unobserved characteristics which lead them to make similar educational choices. u_{akw} can be thought of as the effects of the latent classes on

alternative. An alternative to the mixed multinomial logit model is the multinomial probit model, which, however, places quite different assumptions on the error terms (see Greene, 2003).

alternative *a* at transition *k* and are often called the location parameters (Heckman & Singer, 1982). In my MTMU the number of location parameters is four, because I use $W = 2$ latent classes and model two transitions each with two contrasts. This means that I use five parameters to estimate the unobserved part of the MTMU.[6] Following the specification of the multinomial logit model defined in (1)-(4), I write the conditional multinomial probability of choice *a* on transition *k* as:

$$\Pr(y_{ik} = a \mid x_{ij}, v_{akw}) = \frac{\exp\left(\sum_{j=1}^{J} \beta_{akj} x_{ij} + v_{akw}\right)}{1 + \sum_{s=2}^{A} \exp(\sum_{j=1}^{J} \beta_{skj} x_{ij} + v_{skw})}, \quad (6)$$

where β_{akj} is the logit coefficient of x_{ij} for alternative *a* at transition *k*, and v_{akw} captures the effect of the unobserved variable for the *w*'th latent class for alternative *a* at transition *k*, i.e., the location parameters.[7] In the analysis I model two transitions (i.e., $K = 2$), and I therefore define the joint probability of making two consecutive transitions as

$$\Pr(y_{i1} = a \mid x_{ij}, v_{a1w}) \times \Pr(y_{i2} = a' \mid x_{ij}, v_{a'2w}). \quad (7)$$

Finally, I write the multivariate probability unconditional on unobserved variables (i.e., they are averaged or integrated out), $\Pr(y_i = a, a')$, as a finite mixture model:

---

[6] I arrive a five parameters the following way: Two latent classes are equivalent to a single weight parameter, π_w, because $\sum_{w=1}^{W} \pi_w = 1$. I estimate four location parameters, because the latent variable is a dummy variable, and hence it is redundant to estimate location parameters for each class. A single weight parameter and four location parameters equal five parameters.

[7] Note that, because I use a latent variable formulation, it holds that: $v_{akw} = \frac{u_{akw}}{\sigma_k}$. In other words, the effect of the unobserved variable is identified only up to scale.

$$
\begin{aligned}
&\Pr(y_{ik} = a, a' \mid x_{ij}) = \\
&\sum_{w=1}^{W} \Pr(y_{i1} = a \mid x_{ij}, \nu_{a1}) \times \Pr(y_{i2} = a' \mid x_{ij}, \nu_{a'2})^{I} \pi_w = \\
&\sum_{w=1}^{W} \frac{\exp(\sum_{j=1}^{J} \beta_{a1j} x_{ij} + \nu_{a1w})}{1 + \sum_{s=2}^{A} \exp(\sum_{j=1}^{J} \beta_{s1j} x_{ij} + \nu_{s1w})} \times \left(\frac{\exp(\sum_{j=1}^{J} \beta_{a'2j} x_{ij} + \nu_{a'2w})}{1 + \sum_{s'=2}^{A'} \exp(\sum_{j=1}^{J} \beta_{s'2j} x_{ij} + \nu_{s'2w})} \right)^{I} \pi_w
\end{aligned} \tag{8}
$$

The unconditional joint probability in (8) is given by a weighted (over the W latent classes) product of the conditional probability of completing choice a at transition k (cf. Wedel & DeSarbo, 1995, 2002; McLachlan & Peel, 2000:145f). I is an indicator taking on the value 1 for those who survive to face the second transition (k = 2), and taking on the value 0 for those who do not survive. The MTMU in (8) jointly models two transitions and accommodates for unobserved heterogeneity. If I estimated a model that did not correct for unobserved heterogeneity, then the initial sorting of individuals on observed and unobserved characteristics would result in bias of the estimates at later transitions (Cameron & Heckman, 1998; Holm & Jæger, this issue).[8] Moreover, correcting for unobserved heterogeneity in the multinomial case also relaxes the assumption of IIA, which is a well-known result in the literature on random utility models (cf. Olsen 1982; Hensher & Greene, 2003; Train, 2009).[9]

I identify the parameters in model (8) that I estimate in my analysis by including alternative-specific instrumental variables at the first transition (see the data description). This identification strategy provides me with the necessary exclusion restrictions for identifying

[8] In the duration model literature, this dynamic selection problem is known as *frailty*. It refers to the identification problem that in duration models changes in survival probabilities can be a mixture of unobserved population heterogeneity and state dependence (cf. Vaupel & Yashin, 1985; Trussel & Richards, 1985:245; Yamaguchi, 1987:78; Lancaster, 1990:64). In my example, the problem can also be conceived of as a sample selection problem, because only select individuals experience later transitions (see Heckman, 1979; Berk, 1983; Winship & Mare, 1992).

[9] Notice that the model in (8) also corrects for rescaling bias induced by not including ν_{akw} (Cameron & Heckman, 1998:282; Nicoletti & Rondinelli, 2006). This problem is related to the fact that logit coefficients are identified up to scale and thus depends on the included variables in the model (Amemiya, 1975; Winship & Mare, 1984; Yatchew & Griliches, 1985).

the unobserved variables.[10] Because the credibility of my identification strategy depends on the validity of my two instrumental variables, described in the Data and variables section, I test whether my results are robust to different specifications. Following Cameron and Heckman (1998), I omit the instrumental variables and constrain the effect of the unobserved variable to be the same across both contrasts in both transitions. Under this assumption, transition-specific effects are identified. The Appendix explains this strategy in more detail and also provides the results of the analysis. I do the estimations with *gllamm* for Stata (Rabe-Hesketh *et al.*, 2004). A worked example with sample data and code is available from the journal website.

3. Data and variables

I analyze data from the Danish Longitudinal Survey of Youth (DLSY) (Hansen, 1995). I refer to Jæger and Holm (2007) and Jæger (2007) for a detailed data description. The DLSY follows the life course of 3,151 children born in or around 1954 who were all attending the 7th grade of comprehensive school when they were first interviewed in 1968. The DLSY is based on cluster sampling and respondents were sampled from 151 complete school classes. The survey contains information on family background and ability, and the longitudinal data structure enables me to reconstruct the educational careers of the individuals. My final sample consists of 1,900 individuals, i.e., 40 percent of the original sample is set to missing. This non-response is a consequence of drop-out of the survey and of non-response on both dependent and explanatory variables. I take the sample to be representative of the 1954 birth cohort.

---

[10] Cameron and Heckman (1998) discuss different identification strategies in the binary case. In the multinomial case, another identification strategy that establishes the necessary exclusion restrictions is the inclusion of time-varying covariates (see Holm & Jæger and Lucas, this issue). In the multinomial case, alternative specific variation is also a necessary condition for identification (see Train, 2009).

3.1 Dependent variables

Figure 1 shows the institutional structure of the Danish educational system and the flow of students born in 1954 as they progress through the educational system (see Table 1 for the marginal distributions).[11] Students first complete comprehensive school after 7-10 years of schooling. After completing comprehensive school, at ages 14-17, the individual can choose between three alternatives in secondary education: Leave school, enroll in a vocational track (apprenticeship based education, typically three-four years), or enroll in an academic track (*Gymnasium*, a three year program). Of the total sample, around one fifth leaves the educational system after ending primary education, around half completes the vocational track, and around one third completes the academic track. Those who complete the academic track face the tertiary education decision, around ages 19-20: Leave school, enroll in a short-cycle track (typically aiming at the professions such as teacher or nurse, two-four year programs), or enroll in a university track (five year programs). Of the 616 individuals completing the academic track, around one fifth leaves the educational system with the degree, around half completes a short cycle education, and around one third completes a university education.

-- FIGURE 1 HERE –

-- TABLE 1 HERE --

[11] The presentation in Figure 1 is simplified. According to the Danish Education Act of 1958 ("Skoleloven 1958") those students who did not leave comprehensive school after 7 years of schooling were divided into two tracks (of two to three years of length): a theoretically oriented track ("Realafdelingen") or a practically oriented track. Completion of either track gave the opportunity to choose the academic track in secondary education. To keep my transition model as simple as possible and to keep it comparable to the one in Breen and Jonsson (2000), I do not include this early tracking. Moreover, at that time students who completed the theoretically oriented track in comprehensive school were allowed to enroll in short-cycle tertiary education programs. Thus, a fraction of the birth cohort did enroll in tertiary education without completing the academic track in secondary education, thereby compromising the logic of the academic track in secondary education being an absorbing state. However, they account for a minor fraction of the total sample. I exclude these individuals in my analysis.

3.2 Explanatory variables

I include both family background characteristics and individual characteristics as explanatory variables in my analysis. *Parental highest social class* is measured with the EGP scheme divided into five classes (EGP-5) (Halpin, 1999; Jæger, 2007; cf. Erikson & Goldthorpe, 1992): I/II (professional and managerial employees and self-employed with 10 or more employees), III (routine non-manual professionals), IV (self-employed and small employers (1-9 employees), V/VI (skilled workers), and VII (unskilled and semi-skilled workers). *Parental highest education* is the number of years of completed schooling for the parent with the highest level of education. *Non-intact family* is a dummy variable indicating whether the child did not live with both biological parents at age 14. *Boy* indicates the gender of the child. *Ability* is a measure of the academic skills of the student at age 14 and is constructed as the principal component from a principal component analysis on three test scores in a verbal test, spatial test, and inductive test (each measured by the number of correct answers on the test). The principal component accounts for 66.4 percent of the total variation in the three items. Ability is standardized to have mean zero and variance of unity in the final sample, where higher scores indicate higher ability.

Table 2 shows descriptive statistics for the explanatory variables for the total sample and the sample that survives to face the tertiary education decision. The sample becomes more selective as respondents progress through the educational system. For example, 12.8 percent originates in social classes I and II in the total sample, while 24.7 percent does so in the selected sample. I also see that average number of years of parental schooling changes from 9.8 years to 11.5 years. These changes show that students from socioeconomically well-off families have a higher propensity to complete the academic track in secondary education. Moreover, the average ability score is 0.66 standard deviations larger for the selected sample than for the total sample (from 0 to 0.66 points), and the standard deviation of ability decreases from 1 to 0.92. This selection pattern suggests that students with higher ability have a higher propensity to complete the academic track, and that these students are more homogeneous than the total sample in terms of academic ability.

-- TABLE 2 HERE --

3.3 Instrumental variables

I include two instrumental variables for the first transition in the MTMU, one for each choice alternative (vocational and academic). These variables ensure identification of the MTMU model, and their distributions are described in the bottom columns of Table 2. Including these variables in the model makes my model deviate from that of Breen and Jonsson (2000). In their analyses, Breen and Jonsson control for unobserved heterogeneity as a robustness check and find that their results are robust to unobserved heterogeneity. However, one explanation of their result might be that the unobserved component is not well identified. Thus, including instrumental variables provides me with better identification. However, because the validity of instrumental variables relies on assumptions often difficult to defend (see footnote 12), I perform a robustness check of my results in the Appendix using another identification strategy based on the suggestions in Cameron and Heckman (1998). I return to this check in the Results section.

The instrumental variable for the vocational track is the share of the respondent's school class in comprehensive school that chooses the vocational track. Similarly, for the academic track I use the share of the respondent's school class that chooses the academic track. In the construction of both variables 151 school classes are used, and the respondent is omitted in the calculation of the class mean (thereby avoiding tautological inferences). I thus exploit the cluster design of the DLSY in which respondents are nested in school classes. I interpret my two instruments as indicators of the *influence of peers* on the educational decision.[12] This influence operates through the revealed preferences for secondary education choices of the school class peers in comprehensive school (i.e., in primary education). Such interpretation is in line with the sociological evidence on the influence of peers on educational attainment (e.g., Sewell *et al.*, 1969). Moreover, because I control for family background and ability, the peer influence operates net of these potentially confounding characteristics. Thus, I net out the potential sorting into school classes on parental characteristics and child characteristics.

I only include instrumental variables at the first transition in order to establish the necessary exclusion restrictions. Because school classes in comprehensive school dissolve after the completion of comprehensive school, and the influence of those peers therefore markedly decreases later in the educational career, this model strategy appears credible. Moreover, at later points in the educational career I expect new peer groups to have formed, and I expect these new peer groups, rather than the old ones, to influence the later educational decisions. In addition to this, I expect each instrument to affect only the chosen alternative on the first transition (i.e., the instruments are alternative-specific). Thus, the instrument for the vocational track only affects the respondent's propensity to choose the vocational track, not the academic track, and vice versa for the academic track. This assumption may be violated if school class spillover effects exist, but given the control for family background and ability, such effects should be negligible. A robustness check reported below supports this contention.

[12] Because these two variables work as instruments in my model, I assume that they (A) directly affect their respective choice alternatives on the first transition, but do not, directly or indirectly, (B) either affect the choices made in the second transition or the other choice alternative in the first transition. I present another identification strategy, which I use for a robustness check of my results, in the Appendix.

4. Results

In this section I present the results from my MTMU with two latent classes and compare the results with a standard MTM (i.e., a model without unobserved heterogeneity). I first report the results in logit coefficients (i.e., log odds-ratios) and thereafter report normalized average partial effects. Deriving these average partial effects is not trivial in the MTMU, and I therefore provide the derivations in the Appendix.[13] In the Appendix I also show how I normalize the average partial effects to have an interpretation relative to the baseline alternative for each transition. In the analysis I pay particular attention to the influence of parental social class, parental education, and the student's academic ability on the choices alternatives at each of the two branching points (see Figure 1 to recall the institutional structure of the educational system).

4.1 The MTMU

Table 3 shows the logit coefficients from a MTM and a MTMU with two latent classes.[14] Although the logit coefficients from these two models cannot be directly compared (because they are measured on different scales), I report some notable differences between the estimates of the two models.[15] In general the MTM underestimates the effects compared to the MTMU, although exceptions exist. For example, the effect of ability on the university track in tertiary education (Panel D in Table 3) is about 60 percent larger for the MTMU (0.285) than for the MTM (0.174). Such difference reflects considerable underestimation of

[13] I thank Anders Holm for assisting me in the derivations of these quantities.

[14] I tried to estimate a model with three latent classes, but the model did not converge. One explanation is the relatively small sample of mine.

[15] Given the nature of rescaling bias in logit models, we would expect—all other things being equal—the coefficient to increase between MTM and MTMU, because we divide the logit coefficients with a smaller number in the MTMU than in the MTM (because the MTMU explains "more variation" in the outcome and thus reduces the underlying residual standard deviation). Thus, whether the percent change from MTM to MTMU reflects a change in the underlying "causal" effects, or simply is a consequence of rescaling, is not possible to confirm here (for a thorough discussion of this identification problem, see Karlson, Holm, & Breen 2010).

the MTM estimates. Moreover, in some cases the significance of estimates changes, thereby returning qualitatively different conclusions. For example, the gender effect on the academic track in secondary education (Panel B in Table 3) is statistically insignificant in the MTM (0.139), but significant on a 10 percent level in the MTMU (0.441), and the effect is about 3 times larger. Another example is the effect of parental education on the vocational track in secondary education (Panel A in Table 3). The effect in the MTM is insignificant (0.019), but significant on a 10 percent level in the MTMU (0.133). The effect in the MTM is moreover severely underestimated, with the MTMU estimate being around 7 times larger than the MTM estimate. Thus, had I not controlled for unobserved heterogeneity, I would have drawn somewhat erroneous conclusions with respect to the effects of some of the included variables in my model. I return to this issue below, when I report average partial effects.

For now, however, I report the results from the MTMU (and not the MTM), because I consider this model to be my preferred model (i.e., the model on which I will base my inferences). For the vocational track at the first transition (Panel A in Table 3), all social classes, except classes V/VI, are more likely than class VII to complete the track (a joint Wald-test confirms that the joint social class effect is statistically significant), and the effect of parental education is significant on a 10 percent level. The effect of ability is highly significant and positive. Thus, the vocational track appears to be both socially and academically selective. For the academic track at the first transition (Panel B in Table 3), the effects of parental social class are positive and, except classes V/VI, significant. The effects of parental education and student ability are all positive and significant. The effects for this track are also larger than for the vocational track, indicating stronger selectivity. I return to the issue of comparing effect magnitudes, when I report average partial effects below.

-- TABLE 3 HERE --

For the short-cycle track in tertiary education (Panel C in Table 3), the effect of each social class is insignificant and their joint contribution is also insignificant (confirmed by a joint Wald-test). In addition, the effect of parental education is negative, although insignificant. This pattern suggests that the social selectivity in completing the short-cycle track is negligible,[16] if not even reversed in such a way that students of well-educated parents are less likely to complete the track (relative to leaving school) than students of less-educated parents. Moreover, the effect of ability on the short-cycle track is insignificant, indicating that ability does not matter for completing the track. One explanation of these negligible effects is that completing the short-cycle track in tertiary education is just as difficult (if not less than) as completing the academic track in secondary education for the cohort under study. Thus, the selective nature of the academic track renders the influence of family background and ability less important for completing the short-cycle track. This finding supports the conclusions drawn for the somewhat more selective university track to which I now turn.

For the university track in tertiary education (Panel D in Table 3), the effects of social class, except classes V/VI, are positive and statistically significant, while parental education is insignificant. Consequently, the socioeconomic family influence on the completion of this track appears to run through social class. Contrary to what might be expected, the effect of ability is not statistically significant, once again indicating the selective nature of the academic track in secondary education. Thus, while I cannot trace any social or academic selectivity in the short-cycle track, I do trace social selectivity in the university track.

[16] Note that the main reason for the effects being insignificant is the low number of individuals surviving to face the tertiary education decision (N = 616). More data would provide me with more efficient estimates.

Because the results of my MTMU depend on the identifying assumptions (i.e., the validity of my instrumental variables, cf. footnote 12), I perform a robustness check by using another identification strategy that does not rely on instrumental variables. I explain this strategy in the Appendix, where I also report the results from the model. The overall finding from this alternative analysis is that the results of my MTMU are robust to different specifications. In fact, the results of the two strategies are so similar that virtually no differences exist. The reported results in this paper consequently appear to be robust to the identifying assumptions maintained (cf. footnote 12).

-- TABLE 4 HERE --

Table 4 describes the distribution and effects of the binary unobserved variable in the MTMU. Each category in this variable can be equated with the unobserved types described in a previous section (e.g., unobserved aspirations). The first type or class comprises around 7 percent of the population, while the second type comprises the remaining 93 percent. Members in the first type have a persistently lower likelihood of completing the four educational tracks in the model (two on each transition) relative to the baseline group defined by the intercepts in the MTMU. By way of contrast, members in the second type have a higher likelihood of completing the tracks. Thus, if I follow my interpretation of the binary unobserved variable as capturing educational aspirations might be appropriate, then the population consists of low-aspiring individuals (type 1) and high-aspiring individuals (type 2), who differ in success rates at the different tracks.[17] Omitting the variable capturing these

[17] Notice that the unobserved variable involves counterfactual statements. For example, a type-1 student that in fact chose vocational track would have performed poorly on the academic track, had he chosen the academic track. Similar counterfactuals can be constructed. However, the general conclusion to be drawn here is that no

types may have consequences for the estimates of a MTM and should therefore be included as in the MTMU.

4.2 Normalized average partial effects

Table 5 presents normalized average partial effects of the MTM and MTMU. In the Appendix I derive these average partial effects and I explain the normalization of these effects. As an effect measure, the average partial effect has two notable properties. First, it states the effects on the probability scale (from zero to one), i.e., how a one unit change in x changes the probability of the outcome, Pr(y=1). Second, it is less sensitive to the scale identification than logit coefficients (see Cramer, 2007).[18] Because Table 5 contains as many estimates as Table 3, I only report the average partial effects of social class (indicating the social selectivity of the tracks). Using estimates from the MTMU, I first report differences across tracks both within and across transitions. Thereafter I emphasize how the MTM considerably underestimates the social class effects for all tracks compared to the MTMU.

Looking at the first transition, I find that the social class effects are substantially larger for the academic track than for the vocational counterpart (compare Panels A and B in Table 5). For example, the effect of social classes I/II on completing the vocational track is about 20 percentage points (with social class VII being the reference). The effect on completing the academic track is 25 percentage points, reflecting a difference in the social selectivity of the two tracks. I reach a similar conclusion for the second transition. Here the

---

matter which track factually completed, the type-1 (type-2) students would have had lower (higher) completion rates on the other tracks, had they pursued them, than the baseline group defined by the intercepts in the MTMU.

[18] Average partial effects are, however, not always suited for cross-model comparisons (for a formal discussion, see Karlson, Holm, & Breen, 2010). However, in this application, average partial effects appear to be the best alternative available.

social class effects are much more pronounced for the university track than for the short-cycle track (compare Panels C and D in Table 5). For example, the effect of social class IV on completing the short-cycle track is about 16 percentage points, while this effect is about 29 percentage points for the university track. Thus, the overall conclusion to be drawn from the MTMU estimates stated as average partial effects is that the academic track at the first transition and the university track are more socially selective than the other tracks at the respective transitions. Among all tracks, the university track appears to be the most socially selective.

-- TABLE 5 HERE --

The average partial effects reported in Table 5 clearly show that controlling for unobserved heterogeneity increases the influence of family background at all transitions (see the final column in Table 5 which contains the difference between the MTM and MTMU average partial effects). The MTM tends to underestimate the parameters given by the MTMU. This finding suggests that the MTMU controls not only for dynamic selection bias, but also for violations of the IIA assumption. For example, the social class effects for the vocational track (Panel A in Table 5) are underestimated with up to 11 percentage points. The social class effects for the university track (Panel D in Table 5) are underestimated with up to eight percentage points. Such differences are considerable. Thus, had I not controlled for unobserved heterogeneity, I would report an effect of social classes I/II on university track of 22 percentage points. However, correcting for unobserved heterogeneity reveals an effect of 30 percentage points, indicating more pronounced social selectivity. Such a difference in effects between the MTM and the MTMU must be considered to “make a difference” in terms of the social selectivity of the university track.

5. Discussion

The multinomial transition model with unobserved heterogeneity is a flexible tool for modeling the influence of family background characteristics and individual characteristics on complex educational pathways in diversified educational systems. Compared with a standard SLM, the model allows for branching points with more than two choice alternatives (e.g., exit or continue), it controls for the selection bias at later transitions induced by the selective nature of educational systems, and it relaxes the often unrealistic assumption of IIA on which the standard multinomial logit model is based. Thus, the MTMU provides researchers with insight into the social and individual heterogeneity in educational decision making in diversified systems and with better (i.e., unbiased *ceteris paribus*) estimates.

In the paper I estimate the MTMU on longitudinal survey data from a cohort born in 1954 in Denmark. I find marked social selectivity for the academic track in secondary education and for the university track in tertiary education, while the social selectivity is less pronounced for the vocational track in secondary education and more or less non-existent for the short-cycle track in tertiary education. Moreover, academic skills appear to matter for completion of the secondary education tracks, in particular for the academic track, but not for the completing the tracks in tertiary education. Thus, compared to the standard SLM, multinomial transition models have the potential of revealing important heterogeneity in the social and academic selectivity across tracks and transitions in diversified educational systems. Moreover, using a MTMU compared to a MTM provides estimates that are controlled for the potential bias caused by selection bias and violations of the IIA assumption.

The MTM generally underestimates the estimates given by the MTMU, a finding much different than that of Breen and Jonsson (2000), who find their results robust to unobserved heterogeneity. In terms of normalized average partial effects, as defined in the Appendix, some of these social class effects are underestimated with up to 11 percentage points. If stratification researcher are to inform policy-makers such considerable differences may “make a difference” in terms of policy interventions to be constructed and implemented. Thus, researchers may have good reasons for adopting the MTMU rather than the conventional MTM.

Despite the apparent advantages of the MTMU and the fact that Breen and Jonsson (2000) presented the model more than a decade ago, the model has not diffused into mainstream stratification research. In this paper I have tried to address this problem by applying a MTMU on the educational careers of a Danish cohort born 1954. However, although the Danish educational system at that time had a specific institutional structure (cf. Figure 1), the MTMU can be accommodated to almost any diversified educational system. Future research on educational transitions should therefore exploit the opportunities and flexibility of the model to study the selectivity of educational decisions in diversified educational systems.

References

Allison, Paul D. (1999). Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research*, 28, 186-208.

Amemiya, T. (1975). Qualitative Response Models. *Annals of Economic and Social Measurement*, 4, 363-388.

Ayalon, H. & Shavit, Y. (2004). Educational Reforms and Inequalities in Israel: The MMI Hypothesis Revisited. *Sociology of Education*, 77, 103-120.

Ballarino, G. & Schadee, H. (2010). Allocation and distribution: A discussion of the educational transition model, with reference to the Italian case. *Research in Stratification and Social Mobility*, 28, 45-58.

Becker, R. (2003). Educational Expansion and Persistent Inequalities of Education: Utilizing Subjective Expected Utility Theory to Explain Increasing Participation Rates in Upper Secondary School in the Federal Republic of Germany. *European Sociological Review,* 19, 1-24.

Becker, R. & Hecken, A.E. (2008). Why are Working-class Children Diverted from Universities?—An Empirical Assessment of the Diversion Thesis. *European Sociological Review*, 25, 233-250.

Berk, R.A. (1983). An Introduction to Sample Selection Bias in Sociological Data. *American Sociological Review*, 48, 386-398.

Boudon, R. (1974). *Education, Opportunity and Social Inequality*. New York: Wiley.

Breen, R. & Jonsson, J.O. (2000). Analyzing Educational Careers: A Multinomial Transition Model. *American Sociological Review*, 65, 754-772.

Cameron, S.V. & Heckman, J.J. (1998). Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy*, 106, 262-333.

Cramer, J.S. (2007). Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specficied Disturbances. *Oxford Bulletin of Economics and Statistics*, 69, 545-555.

Erikson, R. & Goldthorpe, J.H. (1992). *The Constant Flux*. Oxford: Oxford University Press.

Gambetta, D. (1987). *Where they pushed or did they Jump? Individual decision mechanisms in education*. Cambridge: Cambridge University Press.

Greene, W.H. (2003). *Econometric Analysis*. 5th Edition. Upper Saddle River, NJ: Prentice Hall.

Halpin, B. (1999). Is Class Changing? AWork-life History Perspective on the Salariat. *Sociological Research Online*, 4.

Hansen, E. J. (1995). *En generation blev voksen*. [A Generation Grew Up]. Copenhagen: SFI – Danish Institute of Social Research.

Hansen, M.N. (2007). Rational Action Theory and Educational Attainment: Changes in the Impact of Economic Resources. *European Sociological Review*, 24, 1-17.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153-161.

Heckman, J.J. & Singer, B. (1982). Population heterogeneity in demographic models. In K. Land & A. Rogers (Eds.), *Multidimensional Mathematical Demography* (pp. 567-599). New York: Academic Press.

Hensher, D.A. & Greene, W.H. (2003). The Mixed Logit model: The state of practice. *Transportation*, 30, 133-176.

Jao, J.C. & McKeever, M. (2006). Ethnic Inequalities and Educational Attainment in Taiwan. *Sociology of Education*, 79, 131-152.

Jæger, M.M. (2007). Educational Mobility Across Three Generations: The Changing Impact of Parental Social Class, Economic, Cultural and Social Capital. *European Societies*, 9, 527-550.

Jæger, M.M. (2009). Equal Access but Unequal Outcomes: Cultural Capital and Educational Choice in a Meritocratic Society. *Social Forces*, 87, 1943-1971.

Jæger, M.M. & Holm, A. (2007). Does parent's economic, cultural, and social capital explain the social class effect on educational attainment in the Scandinavian mobility regime? *Social Science Research*, 36, 719-744.

Karlson, K. B., Holm, A., & Breen, R. (2010). Comparing Regression Coefficients Between Models using Logit and Probit: A New Method. *Unpublished paper*. [Draft available from authors].

Kreidl, M. (2004). Politics and Secondary School Tracking in Socialist Czechoslovakia, 1948-1989. *European Sociological Review*, 20, 123-139.

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

Lucas, S.R. (2001). Effectively Maintained Inequality: Education Transitions, Track Mobility, and Family background Effects. *The American Journal of Sociology*, 106, 1642-1690.

Maastekaasa, A. (2006). Educational Transitions at Graduate Level: Social Origins and Enrolment in PhD Programmes in Norway. *Acta Sociologica*, 49, 437-453.

Mare, R. D. (1980). Family background and School Continuation Decisions. *Journal of the American Statistical Association*, 75, 295-305.

Mare, R.D. (1981). Change and Stability in Educational Stratification. *American Sociological Review*, 46, 72-87.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press.

McFadden, D. (1986). The Choice Theory Approach to Market Research. *Marketing Science*, 5, 275-297.

McLachlan, G. & Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.

Need, A. & de Jong, U. (2001). Educational Differentials in the Netherlands: Testing Rational Action Theory. *Rationality and Society*, 13, 71-98.

Nicoletti, C. & Rondinelli, C. (2006). The (mis)specification of discrete time duration models with unobserved heterogeneity: a Monte Carlo Study. *ISER Working Paper* (2006-53). Colchester: Institute for Social & Economic Research, University of Essex.

Olsen R.J. (1982). Independence from Irrelevant Alternatives and Attrition Bias: Their Relation to One Another in the Evaluation of Experimental Programs. *Southern Economic Journal*, 49, 521-535.

Powers, D. & Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis*. Boston: Academic Press.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). GLLAMM Manual. *U.C. Berkeley Division of Biostatistics Working Papers Series, WP 160*. California: University of California, Berkeley.

Sewell, W.H., Haller, A.O., & Portes, A. (1969). The Educational and Early Occupational Attainment Process. *American Sociological Review*, 34, 82-92.

Swait, J. & Louviere, J. (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research* 30, 305-314.

Tieben, N. & Wolbers, M.H.J. (2010). Transitions to post-secondary and tertiary education in the Netherlands: a trend analysis of unconditional and conditional socio-economic background effects. *Higher Education*, 60, 85-100.

Tolsma, J., Need, A., & de Jong, U. (2010). Explaining Participation Differentials in Dutch Higher Education: The Impact of Subjective Success Probabilities on Level Choice and Field Choice. *European Sociological Review*, 26, 235-252.

Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Trussel, J. & Richards, T. (1985). Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman – Singer Procedure. *Sociological Methodology*, 15, 242-276.

Vaupel, J. W. & Yashin, A.I. (1985). Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics. *American Statistician*, 39, 176-185.

Wedel, M. & DeSarbo, W.S. (1995). A Mixture Likelihood Approach for Generalized Linear Models. *Journal of Classification*, 12, 21-55.

Wedel, M. & DeSarbo, W.S. (2002). Mixture Regression Models. In J.A. Haganaars, & A.L. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 366-382). Cambridge: Cambridge University Press.

Winship, C. & Mare R.D. (1984). Regression Models for Ordinal Variables. *American Sociological Review*, 49, 512-525.

Winship, C. & Mare R.D. (1992). Models for Sample Selection Bias. *Annual Review of Sociology,* 18, 327-350.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yamaguchi, K. (1987). Event-History Analysis: Its Contributions to Modeling and Causal Inference. *Sociological Theory and Methods*, 2, 61-82.

Yatchew, A. & Griliches, Z. (1985). Specification Error in Probit Models. *The Review of Economics and Statistics*, 67, 134-139.

Appendix

Deriving average partial effects

I derive the marginal effect of the *x*, that is, the derivative of the expectation with respect to *x*, where *x* is a predictor of interest. If the sample is a random sample of the population, the average of these marginal effects is the average partial effect (Wooldridge, 2002). I derive the marginal effects for the multinomial logit model and for a counterpart with unobserved heterogeneity captured by *W* latent classes. The derivations thus hold for a single transition, but can be equally applied to other transitions in a multinomial transition model. I express my deepest thanks to Anders Holm for assisting me in these derivations. Greene (2003:722) also gives the derivatives for the multinomial logit model.

Omitting individual subscripts, in the multinomial logit model with three alternatives, the predicted probability of alternative *a* is:

$$P(y=a) = \frac{\exp(\sum_{j=1}^{J}\beta_{aj}x_j)}{1+\sum_{s=2}^{A}\exp(\sum_{j=1}^{J}\beta_{sj}x_j)}, \text{ where } \beta_{1j}=0 . \quad \text{(A1)}$$

The marginal effect on the probability of alternative *a* of predictor *j* is

$$\partial P(y=a)\big/\partial x_j = \left(\beta_{aj}-t_j\right)P(y=a), \quad \text{(A2)}$$

where

$$t_j = \frac{\sum_{s=2}^{A}\beta_{sj}\exp(\sum_{j=1}^{J}\beta_{sj}x_j)}{1+\sum_{s=2}^{A}\exp(\sum_{j=1}^{J}\beta_{sj}x_j)}. \quad \text{(A3)}$$

Taking the mean of (A2) over the sample returns the average partial effect.

In the finite mixture multinomial logit model with three alternatives and *W* latent classes, the predicted probability of alternative *a* is

$$P(y=a)=\sum_{w=1}^{W}\frac{\exp(\sum_{j=1}^{J}\beta_{aj}x_j+\nu_{aw})}{1+\sum_{s=2}^{A}\exp(\sum_{j=1}^{J}\beta_{sj}x_j+\nu_{sw})}\pi_w\text{, where }\beta_{1j}=0\,. \qquad \text{(A4)}$$

The marginal effect on the probability of alternative *a* of predictor *j* is

$$\partial P(y=a)\big/\partial x_j=\left(\beta_{aj}-t_j\right)\sum_{w=1}^{W}P(y=a\,|\,\nu_{aw})\pi_w\,, \qquad \text{(A5)}$$

where

$$t_j=\frac{\sum_{s=2}^{A}\beta_{sj}\exp(\sum_{j=1}^{J}\beta_{sj}x_j+\nu_{sw})}{1+\sum_{s=2}^{A}\exp(\sum_{j=1}^{J}\beta_{sj}x_j+\nu_{sw})}\,. \qquad \text{(A6)}$$

Compared to the marginal effect in (A2), the marginal effect in (A5) weighs the effect over the *W* latent classes. Taking the mean of (A5) over the sample returns the average partial effect. I notice that post-estimation command *gllapred* for Stata command *gllamm* can compute the weighted predicted probability, $\sum_{w=1}^{W}P(y=a\,|\,\nu_{aw})\pi_w$, making computations of the marginal effect in these complex models straightforward. The worked example available through the journal website illustrates the use of these commands.

Normalizing average partial effects

In Table 5 I report normalized average partial effects, with the exit alternative being the baseline alternative. However, when I calculate the raw average partial effects, I obtain effects for each alternative, not each contrast as in the logit case. For example, with three alternatives (*A*), I obtain for predictor *j* three (*A*) average partial effects, while I obtain two logit coefficients (*A-1*), because of the baseline normalization described in the main text. Therefore, average partial effects (or marginal effects) cannot as such be translated into the logit counterparts, and may moreover be of different direction. To see this, simply note that, in a three alternative case, I have that

$$P(Y=1)=1-\left[P(Y=2)+P(Y=3)\right]$$
$$P(Y=2)=1-\left[P(Y=1)+P(Y=3)\right]$$
$$P(Y=3)=1-\left[P(Y=1)+P(Y=2)\right].$$

In other words, the "contrasts" for each alternative is the sum of the two other alternatives, thereby leaving out a baseline contrast, which is present in the logit specification. To normalize the average partial effects relative to a baseline alternative, I therefore subtract the average partial effect for the baseline alternative (here $a = 1$) from the average partial effects of each of the other two alternatives. Let $APE_a(x_j)$ be the average partial effect on alternative a of predictor j. With three alternatives, I obtain the normalized effects for alternative $a = 2$ and $a = 3$ as:

$$a = 2: \; APE_{2-1}(x_j) = APE_2(x_j) - APE_1(x_j)$$

$$a = 3: \; APE_{3-1}(x_j) = APE_3(x_j) - APE_1(x_j)$$

These are the effects reported in Table 5. They state the effect of predictor j on probability of choosing alternative a relative to the baseline alternative. In Table A1 below I report for inspection the raw average partial effects on which the normalized counterparts in Table 5 are based.

-- TABLE A1 HERE --

Robustness check of results from MTMU

Here I report the results from a MTMU *without* instrumental variables. I use a different identification strategy. Following Cameron and Heckman (1998) and Tam (this issue), I restrict the effects of the unobserved variables to be the same across all contrasts and transitions. Cameron and Heckman (1998) argue that, in absence of any plausible exclusion restrictions in a Mare model with unobserved heterogeneity, transition-specific effects of time-invariant predictors are identified under the assumption of constant effects across transitions. This assumption restricts the number of parameters in the latent part to two (compared to five in the unrestricted model with instruments). Table A2 reports the results of three models: first, a model without instruments and without latent classes (MTM-R); second, a model without instruments, with two latent classes, and with restriction of constant effects across transitions and alternatives (MTMU-R), and the MTMU reported in Table 3. Table A3 describes the latent classes of MTMU-R.

Inspecting Table A2, I find that the conclusions based on my model with instrumental variables (MTMU) are robust. The MTMU-R returns results more or less identical to the results of the MTMU, and the differences between MTM-R and MTMU-R are highly similar to the differences between MTM and MTMU reported in the main text. From Table A3 I also find that the distribution of the latent classes is much similar to that in Table 4. The first type has a lower propensity to complete any educational alternative at any transition, while the second has a higher propensity. Thus, had I not had any believable instruments at my disposal, the strategy used in this robustness check would have provided me with the same conclusions reported in the main text.

-- TABLE A2 HERE --

-- TABLE A3 HERE --