

More Successful because of Math: Combining a Natural Experiment and a Structural Dynamic Model to Explore the Underlying Channels*

Juanna Schrøter Joensen	Helena Skyt Nielsen
Stockholm School of Economics	University of Aarhus and IZA
Juanna.Joensen@hhs.se	hnielsen@econ.au.dk

January 28, 2011
- very preliminary -

Abstract

In this paper, we exploit a high school pilot scheme to identify the channels through which advanced high school math causes more favorable academic and labor market outcomes. The pilot scheme reduced the costs of choosing advanced math because it allowed for a more flexible combination of math with other courses. We find clear evidence of a causal relationship between math and earnings for students who are induced to choose math after being exposed to the pilot scheme. The effect partly stems from the fact that these students end up with a higher education. To further explore these channels, we estimate a structural dynamic model of higher education choices and outcomes. This allows us to quantify the direct and indirect effects of advanced math and to simulate the impacts of education policies changing math and college entry requirements.

JEL Classification: I21, J24

Keywords: Math, High School Curriculum, Instrumental Variable, Structural Estimation.

*Financial support from the Danish Social Science Research Council and the Jan Wallander och Tom Hedelius Stiftelse is gratefully acknowledged. Thanks to seminar participants at the University of Aarhus for valuable comments and discussions. Thanks to Kristine Eibye Brandt and Tórbera Schrøter Joensen for competent research assistance. The usual disclaimer apply.

1 Introduction

The beneficial effect of high level Math courses on earnings has been documented by several authors, some of which maintain that the estimated effect of Math is a causal effect. A causal effect of Math on earnings may work through several channels: enhanced cognitive or noncognitive skills, changed preferences (or expectations) for enrolling into a higher education or enhanced productivity in completing a higher education. In this paper, we explore the channels through which high level Math influences earnings.

Based on cross-country data, Hanushek and Kimko (2000) and Jamison, Jamison, and Hanushek (2007) study the influence of Math and Science test scores on growth. They find that one standard deviation increase in the Math test score leads to an increase in annual growth rates of 0.5–1.0 percentage points, which only makes sense if there are strong externalities related to accumulating high-quality human capital. Hence, cross-country studies indicate a positive causal effect of high-quality Math skills on economic growth.

Based on individual data, a handful of papers investigate the influence of Math on individual earnings. Altonji (1995) pioneered this area of research by studying the effect of high school curricula on earnings. In a study based on the National Longitudinal Study of the High School Class of 1972 (NLS72), he uses the variation in curricula across U.S. high schools to identify the effects of coursework on wages and educational outcomes. He finds a negligible effect from specific coursework, including Math. Altonji (1995) uses the curriculum of the average student from the high school in question as an instrument for acquired Math qualifications. However, as he points out, this experiment is not a clean natural experiment because the curriculum of the average individual at a given high school may be correlated with the average family background, primary school preparation, ability of the student body, and the quality of the courses. Rose and Betts (2004) use data for the 1982 cohort from High School and Beyond (HSB). This is an improvement upon earlier studies, mainly because transcript data for the sampled individuals are more detailed, and because the individuals are observed when in their thirties rather than in their twenties. Rose and Betts (2004) find that Math matters. They disaggregate Math into six courses, and estimate the return to each of the six Math courses compared to the average curriculum. They perform a whole battery of robustness checks using OLS, Altonji's IV, and school fixed effects. Their most robust conclusion is that course credits in algebra/geometry significantly increase earnings. As

the authors themselves are aware, it is unclear whether the results of Altonji (1995) and Rose and Betts (2004) reflect selection or causality. However, Joensen and Nielsen (2009) improve upon these studies by getting closer to the causal impact of Math by applying a better instrument for acquired Math qualifications. They use an instrument based on a pilot scheme implemented for the cohorts starting in the Danish high school before the structural reform of 1988. They find clear evidence of a sizeable causal impact of Math on earnings for students who are induced to choose Math after being exposed to the pilot scheme. They find that the dominant part of the effect stems from the fact that these students end up with a higher education. In this paper, we show that the total causal effect of Math on earnings identified by IV comprises: (i) the causal effect of Math on earnings for those who not pursue a college education plus (ii) the difference in the causal effect of college enrollment (without graduating) on earnings between those who have advanced Math and those who do not, and (iii) the difference in the causal effect of college graduation (relative to college enrollment without graduating) on earnings between those who have advanced Math and those who do not. Consequently, a positive causal effect of Math on earnings may arise if students acquire skills (like clarity in expression and logical inference) that are useful in any future career *or* if the acquired skills are only particularly productive in augmenting the returns to college.

The main empirical challenge is how to separately identify and quantify these effects. The distinction may be crucial for education policy. If the causal effect of Math on earnings is mainly a direct causal effect of Math for those with no college degree then there is no need to encourage students with advanced Math to pursue college. On the other hand, if the indirect effect is more important and advanced Math students have much higher benefits of college enrollment and graduation, then it becomes crucial to encourage further academic success in order to reap the benefits of advanced Math courses in high school.

In this paper, we identify and quantify various channels through which Math causes higher earnings by exploiting the exogenous variation stemming from the instrument used by Joensen and Nielsen (2009) in combination with a structural dynamic discrete choice model. The idea of combining a natural experiment with a structural dynamic model has previously been strongly advocated by for example Card and Hyslop (2005), Meghir (2006) and Keane (2009). The combination with the natural experiment both allows us to: (i) estimate a richer structural model as also advocated by Attanasio,

Meghir and Santiago (2004), (ii) make more elaborate model validation as also advocated by Todd and Wolpin (2006) and by Keane and Wolpin (2007) using a “nonrandom holdout sample” (a sample that differs significantly from the estimation sample along the policy dimension the model is meant to forecast), (iii) simulate effects of potential policy interventions like Math and college entry requirements, and (iv) assess some experimental curriculum design issues.

The rest of this paper is structured as follows: In section 2, we describe the natural experiment which complements the structural model in identifying the effects of Math on educational choices and outcomes. In section 3, we set up and solve the structural model, and derive the likelihood function to be maximized. In section 4, we describe the data and the sample selection. In section 5, we present the empirical results, while section 6 concludes the paper.

2 The Natural Experiment in Danish High Schools

In this section, we briefly describe the environment of the natural experiment and the applied identification strategy. In the first subsection, we present the relevant Danish high school regime. Thereafter we describe the program constituting the natural experiment. In the final subsection, we discuss how this instrument may be used for identification in the present context of an exploration of channels through which advanced Math influences educational and labor market outcomes.

2.1 The pre-1988 high school¹

In the period 1961-1988, the Danish high school system was a “branch-based” high school system, where courses were grouped in restrictive course packages.² We focus on this period for two reasons. First, the supply of course packages gives us a useful exogenous variation in the cost of acquiring advanced Math. Second, focusing on students entering high school prior to 1988, means that our data set includes completed

¹Consult Joensen and Nielsen (2009) for additional details on the Danish high school regime during the relevant period.

²Available course packages were labelled: Social Science and Languages, Music and Languages, Modern Languages, Classical Languages, Math-Social Science, Math-Natural Science, Math-Music, Math-Physics and Math-Chemistry.

education spells as well as labor market outcomes when the individuals are in their thirties.

This system implied that students upon high school graduation would have achieved one of three Math levels available: high-, intermediate-, or low-level. The difference between the three levels is reflected in the number of lessons per week as well as in the content of the courses. For instance, the extent of geometry and algebra increases as the level becomes more advanced. In the empirical analysis, we focus on whether students choose the high-level Math course or not, meaning that the intermediate- and low-level Math courses are lumped together. The decision about which Math level to opt for is taken at the end of the first year at high school. The only way to obtain the high-level Math course was in combination with high-level physics, unless the student was enrolled at a pilot school, where the high-level Math course could also be obtained in combination with high-level chemistry. It is exactly this increased course flexibility which some students were unexpectedly exposed to at pilot schools that constitute the natural experiment that we exploit in this paper.

2.2 The Natural Experiment

The pilot scheme was implemented as an experimental curriculum at about half of the high schools prior to the 1988-reform. Table 1 gives an overview of the gradual implementation of the pilot scheme from 1984–87. The table is divided by types of high schools: schools with no pilot scheme ($PilotSchool = 0$), schools where the pilot scheme was introduced after enrollment of the relevant cohort ($PilotSchool = 1 \ \& \ PilotIntro = 1$), and schools where the pilot scheme was implemented prior to enrollment of the relevant cohort ($PilotSchool = 1 \ \& \ PilotIntro = 0$).

Schools were not randomly assigned to become pilot schools. Instead, from 1984–86, they could apply to the Ministry of Education for permission to adopt the experimental curriculum, whereas in 1987, the high school principals could make this decision without approval from the ministry. It is not possible to check whether the pilot schools represent a sample of schools which is essentially random with respect to Math ability.³ It is clear however, that students with a preference for advanced Math and chemistry may self-select into schools that are known to offer the pilot program before entrance. This

³Joensen and Nielsen (2009) elaborate more on the entrance procedures and whether the pilot status is essentially random.

is why we distinguish between students at pilot schools where the pilot scheme was unexpectedly introduced after they had enrolled in the high school ($PilotSchool = 1$ & $PilotIntro = 1$), and those who knew that the school was a pilot school before they applied to enter the school ($PilotSchool = 1$ & $PilotIntro = 0$).

The instrumental variable is based on the fact that the pilot program reduces the opportunity cost of choosing high-level Math since the students exposed to the scheme are not required to take the physics course together with advanced Math. Hence, first-year high school students enrolled at a school when it decided to introduce the pilot scheme were exposed to an exogenous cost shock, which induces more students to choose high-level Math compared to students at nonpilot schools. If the selection of newly participating schools is exogenous with respect to student ability, which Joensen and Nielsen (2009) argue it is, the pilot scheme may be seen as a natural experiment which provides exogenous variation in students' Math qualifications without influencing the outcomes of interest other than through the effect on Math qualifications.

The instrumental variable, $PilotIntro_i$, is equal to one if the individual enrolled in a high school which afterwards decided to introduce the experimental curriculum for the first time and takes the value zero otherwise. This instrument is valid if the pilot program is as good as randomly assigned to schools and if individuals are as good as randomly distributed across schools that have not yet decided to introduce the experimental curriculum. This assumption is violated only if the school decides to participate in the program based on the Math ability of local students. Joensen and Nielsen (2009) find that the introduction of the program in 1987 was not a valid instrument (given that the introduction in 1984-86 was a valid instruments), while no similar concerns are raised regarding the other cohorts. Therefore, we disregard the cohort of 1987 in the present study. The instrument is strong if the unexpected introduction of the scheme induces students to choose advanced Math, which is automatically tested in the empirical section. The instrument satisfies the monotonicity (or uniformity) condition if individuals who chose advanced Math when it could only be combined with physics also would have chosen high-level Math if they had unexpectedly had the option of also combining it with advanced chemistry. We are confident that the monotonicity assumption is reasonable in our application since all the options available at nonpilot schools also were available at schools that introduced the pilot program.

Our instrument exploits the exogenous variation in the exposure of students to the possibility of combining advanced Math courses with advanced chemistry. Hence, the

“treatment” that we investigate is the combined treatment of advanced Math with advanced chemistry. Because advanced Math and advanced chemistry are combined in a course package, we cannot separate the effect of advanced Math from that of advanced chemistry or from the potential synergy effect of the combination of Math with chemistry. However, the earlier literature suggests that if any specific course work matters, it is Math rather than Science courses; see e.g. Rose and Betts (2004) and Altonji (1995).

2.3 Using the Natural Experiment for Identification

We assume that individuals choose advanced Math if their expected gains exceed their expected cost of the investment; i.e. if $Y_{i1} - Y_{i0} - c_i \geq 0$, where Y_{i1} denotes the potential outcome with advanced Math, and Y_{i0} denotes the potential outcome without. Students at pilot schools have the option of choosing advanced Math at a lower cost, $(c_i | PS_i = 1) \leq (c_i | PS_i = 0)$. This cost may also include effort and psychological costs. The observed advanced Math choice can be decomposed as:

$$Math_i = Math_{i0} + PS_i (Math_{i1} - Math_{i0}), \quad (1)$$

where $Math_{i1}$ denotes the Math choice observed if $PS_i = 1$, meaning student i was exposed to the experimental curriculum, while $Math_{i0}$ is the observed Math choice if $PS_i = 0$. Monotonicity implies that: $Math_{i1} \geq Math_{i0}$, i.e. all those who would have chosen Math without being exposed to the experiment would also choose Math given the extra options available after introducing the pilot program. Let $P_{PS} \equiv P(PS_i = 1)$ denote the probability of attending a pilot school. The observed probability of choosing advanced Math, $P_M \equiv P(Math_i = 1)$, is given by:

$$P_M = P_{M0} + P_{PS} (P_{M1} - P_{M0}) \quad (2)$$

where $P_{M0} \equiv P(Math_i = 1 | PS_i = 0)$, $P_{M1} \equiv P(Math_i = 1 | PS_i = 1)$, and $P_{M1} \geq P_{M0}$ since students at pilot schools can choose Math at a lower cost.

Denote the counterfactual outcomes of college enrollment choices as $College_{i1}$ if $Math_i = 1$ and $College_{i0}$ if $Math_i = 0$. The causal effect of advanced Math on college enrollment is given by: $College_{i1} - College_{i0}$. The probability of college enrollment,

$P_C \equiv P(\text{College}_i = 1)$, is given by:

$$P_C = P_{C_0} + P_M(P_{C_1} - P_{C_0}) \quad (3)$$

where $P_{C_0} \equiv P(\text{College}_i = 1 \mid \text{Math}_i = 0)$ and $P_{C_1} \equiv P(\text{College}_i = 1 \mid \text{Math}_i = 1)$. Equivalently, observed college enrollment is given by:

$$\text{College}_i = \text{College}_{i0} + \text{Math}_i(\text{College}_{i1} - \text{College}_{i0}). \quad (4)$$

The counterfactual outcomes of college completion are denoted Grad_{i1} if $\text{Math}_i = 1$ and Grad_{i0} if $\text{Math}_i = 0$. The causal effect of advanced Math on college graduation is given by: $\text{Grad}_{i1} - \text{Grad}_{i0}$. The probability of college graduation, $P_G \equiv P(\text{Grad}_i = 1)$, is given by:

$$\begin{aligned} P_G &= P(\text{Grad}_i = 1 \mid \text{College}_i = 1) P_C \\ &= (P_{G_0} + P_M(P_{G_1} - P_{G_0})) P_C \end{aligned} \quad (5)$$

where $P_{G_0} \equiv P(\text{Grad}_i = 1 \mid \text{Math}_i = 0)$ and $P_{G_1} \equiv P(\text{Grad}_i = 1 \mid \text{Math}_i = 1)$. Observed college graduation is given by:

$$\begin{aligned} \text{Grad}_i &= \text{College}_i[\text{Grad}_{i0} + \text{Math}_i(\text{Grad}_{i1} - \text{Grad}_{i0})] \\ &= \text{Grad}_{i0} + \text{Math}_i(\text{Grad}_{i1} - \text{Grad}_{i0}) \end{aligned} \quad (6)$$

where the last equality follows from the fact that $\text{College}_i \text{Grad}_i = \text{Grad}_i$.

We want to identify the causal effect of advanced Math on two academic outcomes: college enrollment and college completion, as well as labor market success measured by earnings. First, we use an IV strategy to identify the causal effect of those who are induced to choose advanced Math because they could do so at a lower cost. This is labeled the local average treatment effect (LATE) by Imbens and Angrist (1994), and is given by:

$$\frac{E[Y_i \mid PS_i = 1] - E[Y_i \mid PS_i = 0]}{E[\text{Math}_i \mid PS_i = 1] - E[\text{Math}_i \mid PS_i = 0]} = E[Y_{i1} - Y_{i0} \mid \text{Math}_{i1} > \text{Math}_{i0}] \quad (7)$$

For the academic outcomes, college enrollment and college graduation, IV identifies the direct causal effect of advanced Math. However, for earnings the total causal effect

identified by IV is comprised of a direct effect on earnings, as well as indirect effects working through college enrollment and graduation. To see this note that for college enrollment, $College_i$, it follows naturally from (4) that under the assumptions that PS_i is a strong and valid instrument that affects $Math_i$ monotonously:

$$\begin{aligned}
& E[College_i | PS_i = 1] - E[College_i | PS_i = 0] \\
&= E[(Math_{i1} - Math_{i0})(College_{i1} - College_{i0})] \\
&= E[College_{i1} - College_{i0} | Math_{i1} > Math_{i0}] P(Math_{i1} > Math_{i0})
\end{aligned}$$

where the first equality follows from substitution of (4) and invoking the independence assumption, $PS_i \perp\!\!\!\perp College_i | Math_i$, the second equality follows from the monotonicity assumption (i.e. no defiers, $P(Math_{i1} < Math_{i0}) = 0$) and the law of iterated expectations (LIE). Lastly, under the assumption that PS_i is a strong instrument, we can rearrange to arrive at the expression in (7), since $P(Math_{i1} > Math_{i0}) = P(Math_{i1} = 1) - P(Math_{i0} = 1) = E[Math_i | PS_i = 1] - E[Math_i | PS_i = 0]$. Analogously, we can show that IV identifies the LATE for the college graduation outcome by substitution of (6).

Now let $Y_{i,Math,College,Grad}$ denote the counterfactual earnings outcomes for the various combinations of $Math_i, College_i, Grad_i \in \{0, 1\}$ in order to show that the causal effect of Math on earnings can be decomposed as:

$$\begin{aligned}
Y_{i1} - Y_{i0} &= \underbrace{(Y_{i100} - Y_{i000})}_{\text{causal effect of Math on Y for i with College=0}} \tag{8} \\
&+ \underbrace{College_{i1}(Y_{i110} - Y_{i100})}_{\text{causal effect of College enrollment on Y for i with Math = 1}} - \underbrace{College_{i0}(Y_{i010} - Y_{i000})}_{\text{causal effect of College enrollment on Y for i with Math = 0}} \\
&\quad \underbrace{\hspace{10em}}_{\text{difference in causal effect of College enrollment on Y for i over advanced Math level}} \\
&+ \underbrace{Grad_{i1}(Y_{i111} - Y_{i110})}_{\text{causal effect of College graduation (i.e. Grad) on Y for i with Math = 1}} - \underbrace{Grad_{i0}(Y_{i011} - Y_{i010})}_{\text{causal effect of College graduation (i.e. Grad) on Y for i with Math = 0}} \\
&\quad \underbrace{\hspace{10em}}_{\text{difference in causal effect of College graduation (i.e. Grad) on Y for i over advanced Math level}}
\end{aligned}$$

Substituting (8) into the right hand side of (7) we see that the total causal effect of advanced Math on earnings identified by IV comprises: (i) the causal effect of Math

on earnings for those who do not pursue a college education plus (ii) the difference in the causal effect of college enrollment (without graduating) on earnings between those who have advanced Math and those who do not, and (iii) the difference in the causal effect of college graduation (relative to college enrollment without graduating) on earnings between those who have advanced Math and those who do not. Consequently, a positive causal effect of Math on earnings may arise if students acquire skills (like clarity in expression and logical inference) that are useful in any future career *or* if the acquired skills are only particularly productive in augmenting the returns to college.

The main empirical challenge is how to separately identify and quantify these effects. The distinction may be crucial for policy recommendations. If the causal effect of Math on earnings is mainly a direct causal effect of Math for those with no college degree (i.e. $Y_{i100} - Y_{i000}$ large) then there is no need to encourage students with advanced Math to pursue college. On the other hand, if the indirect effect is more important and advanced Math students have much higher benefits of college enrollment and graduation, then it becomes crucial to encourage further academic success in order to reap the benefits of advanced Math courses in high school.

In the empirical analysis we first shed more light on the channels using an IV strategy based on the natural experiment to identify the total causal effect. We explore the effects on three outcomes: college enrollment, college graduation, and earnings. Second, we combine the natural experiment with a structural model to further disentangle the channels through which advanced Math causes more favorable labor market outcomes. We also illustrate how this combination of a natural experiment and a structural model allows us to (i) estimate a richer model specification, (ii) make more elaborate model validation, (iii) simulate effects of potential policy interventions, and (iv) assess some experimental curriculum design issues.

3 Model Setup

This section presents the structural model that specifies the educational environment in terms of choice sets and outcomes, and the labor market environment in terms of wage opportunities. Although all decisions are taken by individuals, the subscripts i are suppressed for notational ease.

The model has two decision stages. In the third stage, all individuals work on the

labor market receiving wages depending on their choices and outcomes in the first two stages. Figure 1 illustrates individual choices in each of the stages.

The first stage ($s = 1$) is the high school enrollment period, $t \in \{0, 1\}$. First, high school students are randomly assigned to high schools at $t = 0$: Pilot Schools ($PS = 1$) or Non-Pilot Schools ($PS = 0$). Second, at $t = 1$ high school students choose their Math level, $Math \in \{0, 1\}$, where the advanced level corresponds to $Math = 1$. The advanced Math choice depends on Pilot School status, PS , and unobserved Math abilities/motivation, ε_1^m :⁴

$$Math = \mathbf{1} [m(PS, \varepsilon_1^m) > 0] \quad (9)$$

The second stage ($s = 2$) is the potential college enrollment period, $t \in \{2, \dots, \tilde{t}\}$. Individuals choose whether to work or attend college, $College_t \in \{0, 1\}$, where attending college corresponds to $College_t = 1$. Let $\tilde{t}_{C_1}$ denote the time of first college enrollment. Hence, duration of schooling interruption after high school is given by $\tilde{t}_C = \tilde{t}_{C_1} - 2$. The college enrollment decision depends on Math level, $Math$, labor market experience, Exp_t , accumulated college degrees, G_t , prior period college attainment, $College_{t-1}$, and unobserved college motivation, ε_t^c :⁵

$$College_t = \mathbf{1} [c(X_t, \varepsilon_t^c) > 0] \quad (10)$$

where $X_t = (Math, Exp_t, G_t, t, \tilde{t}_C, College_{t-1})$ denotes the observed state variable and $S_t = (X_t, \varepsilon_t^c)$ denotes the state variable summarizing all relevant information. $College_{t-1}$ captures switching costs, whereas the other state variables capture various productivity differences. Initial conditions are given by $Exp_2 = G_2 = College_1 = 0$.⁶ Note that $Math$ is predetermined and hence fixed in the second stage. Experience evolves deterministically, $Exp_{t+1} = Exp_t + (1 - College_t)$, since work and college are mutually exclusive. Higher acquired college degrees evolve stochastically, $G_{t+1} = \max \{G_t, Grad_t\}$, since college graduation is a probabilistic outcome, $Grad_t \in$

⁴Recall from Section 2 that students at Pilot Schools have the option of choosing advanced Math at a lower cost. Hence, $P(Math = 1 \mid PS = 1) > P(Math = 1 \mid PS = 0)$ assuring that (2) is positive.

⁵Since we have three cohorts of high school graduates, we also plan to exploit regional and time variation in labor market opportunities (proxied by average youth unemployment rate in municipality of residence), $\bar{Z}_{lt}$. Hence, $College_t = \mathbf{1} [c(X_t, \bar{Z}_{lt}, \varepsilon_t^c) > 0]$

⁶This is not a restrictive assumption, since experience is practically zero on average at high school graduation.

$\{0, 2, 3, 4, 5\}$, where graduating corresponds to $Grad_t > 0$.⁷ Graduation is reflected by the underlying latent variable, $Grad_t^*$, depending on college enrollment, $College_t$, Math level, $Math$, labor market experience, Exp_t , accumulated college degrees, G_t , time since high school, t , duration of schooling interruption after high school, $\tilde{t}_C - 2$, and unobserved college abilities, ε_t^g :⁸

$$Grad_t^* = g(X_t, \varepsilon_t^g) \quad (11)$$

Naturally, the probability of graduating is zero for students who do not enroll in college, $P(Grad_t = 0 \mid X_t, College_t = 0) = 1, \forall X_t$.⁹

In the third stage ($s = 3$), $t \in \{\tilde{t}, \dots, T\}$, all individuals work and receive wages conditional on their educational histories and labor market experience:

$$\ln W_t = w(X_t, \varepsilon_t^w). \quad (12)$$

In order to solve and estimate the structural dynamic model we need to adopt explicit functional forms of $m(\cdot)$, $c(\cdot)$, $g(\cdot)$, and $w(\cdot)$, as well as distributional assumptions regarding the stochastic components $\varepsilon = (\varepsilon^m, \varepsilon^c, \varepsilon^g, \varepsilon^w)$.

3.1 Solution

An individual makes a sequence of choices to maximize the expected present value of utility. Utility is measured in real 2000 *DKK*, linear in consumption, and additively separable: $U_t^k(S_t) = C_t + d_s \varepsilon_t^j$, where $j = m$ at $t = 1$ and $j = c$ for $t \in \{2, \dots, \tilde{t}\}$. Consumption is equal to the wage or the net gain from education attendance: $C_t = (1 - d_s) \mathbf{1}[t \geq 2] W_t + d_s b_t$. Utility individuals receive from education attendance is stochastic; where $d_1 = Math$, $d_2 = College_t$, and $d_3 = 0$. The consumption value of

⁷Consequently, $\tilde{t} = \max\{\tilde{t}_0, \tilde{t}_2, \tilde{t}_3, \tilde{t}_4, \tilde{t}_5\}$ denotes stochastic time to dropout or graduation with highest completed college degree. $g \in \{0, 2, 3, 4, 5\}$ denotes ordained years of college education. Hence, 0 denotes dropouts, 2 denotes short cycle higher education, 3 denotes Bachelor's degrees, 4 denotes medium cycle higher education, and 5 denotes Master's degrees.

⁸We also plan to let college graduation probability depend on job opportunities (proxied by unemployment rate) in field of study, $\bar{Z}_{ft}$. Hence, $Grad_t = g(X_t, \bar{Z}_{ft}, \varepsilon_t^g)$.

⁹If enrolled $College_t = 1$, the probability of dropping out is $P(Grad_t = 0 \mid X_t, College_t = 1) = P(Grad_t^* \leq 0 \mid X_t, College_t = 1)$ and like any ordered choice model; e.g. the probability of obtaining a Master's degree is $P(Grad_t = 5 \mid X_t, College_t = 1) = P(\mu_4 \leq Grad_t^* \leq \mu_5 \mid X_t, College_t = 1)$ where μ_j , $j \in \{1, \dots, 5\}$, are unknown cut-off parameters to be estimated jointly with the structural parameters.

education, b_t , can be thought of as the value attached to learning less the psychological effort cost. The consumption value of advanced Math course attendance in high school (relative to other course attendance) is denoted $b_1 = b^m(PS)$, and assumed to depend on whether the option of taking Math at a potentially lower psychological effort cost was available; hence $b^m(0) \leq b^m(1)$. The consumption value of college attendance (relative to working) is given by $b_2 = b^c(X_t)$, and assumed to depend on Math qualifications, acquired degrees, and labor market opportunities.¹⁰ More specifically:

$$m(PS, \varepsilon_1^m) = b^m(PS) + \varepsilon_1^m \quad (13)$$

$$= \mu_0 + \mu_1 PS + \varepsilon_1^m$$

$$c(X_t, \varepsilon_t^c) = b^c(X_t) + \varepsilon_t^c \quad (14)$$

$$= \beta_0 + \beta_1 Math + \sum_{g=2}^5 \beta_g \mathbf{1}[G_t = g] + \beta_6 t + \beta_7 College_{t-1} + \varepsilon_t^c$$

$$g(X_t, \varepsilon_t^g) = \gamma_0 + \gamma_1 Math + \sum_{g=2}^5 \gamma_g \mathbf{1}[G_t = g] + \gamma_6 t + \gamma_7 \tilde{t}_C + \varepsilon_t^g \quad (15)$$

where ε_t^m , ε_t^c , and ε_t^g are logistically distributed. The Mincerian wage equation is:

$$\begin{aligned} w(X_t, \varepsilon_t^w) &= \alpha_0 + \alpha_1 Math + \sum_{g=2}^5 \alpha_g \mathbf{1}[G_t = g] + \alpha_6 Exp_t + \alpha_7 Exp_t^2 \\ &+ \sum_{g=2}^5 \alpha_{1g} Math * \mathbf{1}[G_t = g] + \alpha_{16} Math * Exp_t + \alpha_{17} Math * Exp_t^2 \\ &+ \sum_{g=2}^5 \alpha_{g6} \mathbf{1}[G_t = g] * Exp + \sum_{g=2}^5 \alpha_{g7} \mathbf{1}[G_t = g] * Exp_t^2 + \varepsilon_t^w \end{aligned} \quad (16)$$

where the error terms are assumed to be normally distributed, $\varepsilon_t^w \sim N(0, \sigma_w^2)$.

The optimization problem can be written as a dynamic programming problem via Bellman's principle of optimality. The value function, $V_t(S_t)$, is defined to be the maximum expected present value at time t , given state, S_t , and discount factor, δ :

$$V_t(S_t) = U_t^k(S_t) + \delta E[V_{t+1}(S_{t+1})] \quad (17)$$

¹⁰We also plan to exploit variation in local youth unemployment so that $b_2 = b^c(X_t, \bar{Z}_{lt})$. Presumably, the worse youth labor market opportunities, the lower the expected opportunity cost of college attendance; hence the higher the value of college attendance, $\frac{\partial b^c(\cdot)}{\partial \bar{Z}_{lt}} > 0$.

It completely summarizes optimal behavior from period t onward, and is a function of a current utility component and a future expected utility component. Consequently, it can be written as $V_t(S_t) \equiv \max_k V_t^k(S_t)$, where $V_t^k(S_t)$ denotes the alternative specific value functions:

$$V_t^k(S_t) = U_t^k(S_t) + \delta E[V_{t+1}(S_{t+1}) | S_t, k] \quad (18)$$

where $k = \text{Math}$ at $t = 1$, $k = \text{College}_t$ for $t \in \{2, \dots, \tilde{t}\}$, and $k = 0$ for $t \in \{\tilde{t}, \dots, T\}$.

The model has to be numerically solved, since an analytical solution is not feasible. The model is solved by backward recursion. In the third stage ($s = 3$) all individuals work full-time on the labor market. The indirect utility of entering the labor market in period $t \geq \tilde{t}$ equals the expected present discounted value of life-time earnings (plus a random component). Predicted life-time earnings are calculated based on individual earnings up until 19 years after high school. It is assumed that from the 20th year after high school and 25 years onwards (approximately until age 60) individuals are employed full-time, i.e. $Exp_{t+1} = Exp_t + 1$, $t \in \{20, \dots, 45\}$. Hence, the optimization problem becomes trivial, since the other state variables are fixed: $G_{t+1} = G_t$ and $Math_{t+1} = \text{Math}$. The value function simplifies to:

$$V_t^0(S_t) = U_t^0(S_t) + \sum_{\tau=t+1}^T \delta^{\tau-t} E[W(X_\tau^w, \varepsilon_\tau^w)] \quad (19)$$

where wages depend on attained education and experience. When calculating the expected value of lifetime earnings, wages are assumed to be constant after 25 years of labor market experience. Given that wages are log-normally distributed, but enter in levels in the optimization problem, the wage expectation in (19) includes the variance of the idiosyncratic earnings ability shock: $\exp(\alpha X_\tau^w + \sigma_w^2)$. Note that year effects cannot be included in the calculation, since these cannot be forecasted beyond the sample period.

In the second stage ($s = 2$), students choose whether to attend college or not. The conditional value functions are given by:

$$V_t^k(S_t) = U_t^k(S_t) + \delta \sum_{g \in \{0, 2, 3, 4, 5\}} P(\text{Grad}_t = g | S_t, \text{College}_t = k) \cdot \max_{\kappa \in \{0, 1\}} V_{t+1}^\kappa(S_{t+1}) \quad (20)$$

where $P(\text{Grad}_t = g | S_t, \text{College}_t = k)$ is the probability of graduating at level $g \in \{0, 2, 3, 4, 5\}$ between time t and $t+1$, and follows an ordered logit model since ε_t^g is logis-

tically distributed.¹¹ In each period $t \in \{2, \dots, \tilde{t}\}$ individuals choose the college option giving them maximum expected value, i.e. $V_t(S_t) = \max\{V_t^0(S_t), V_t^1(S_t)\}$. The conditional probability of college enrollment, $P(\text{College}_t = 1|S_t) = P(V_t^1(S_t) > V_t^0(S_t))$, where the choice specific value functions are given in (20) corresponds to a logit model of current and future expected value, since ε_t^c is logistically distributed.¹²

Finally, while still in high school ($s = 1$) students either choose advanced Math or not. The expected value of choosing the advanced Math level is given by:

$$V_1^1(S_1) = b^m(PS) + \varepsilon_t^m + \delta E[V_2(S_2) | \text{Math} = 1]$$

while the value of not choosing advanced Math is given by:

$$V_1^0(S_1) = \delta E[V_2(S_2) | \text{Math} = 0]$$

Individuals choose the Math level they expect will give them the maximum value, i.e. $V_1(S_1) = \max\{V_1^0(S_1), V_1^1(S_1)\}$. Hence, the probability of choosing advanced Math follows the logit model:

$$P(\text{Math} = 1|S_t) = \frac{1}{1 + \exp(b^m(PS) + \delta(E[V_2(S_2) | \text{Math} = 1] - E[V_2(S_2) | \text{Math} = 0]))}$$

¹¹In other words, the value of working is given by:

$$V_t^0(S_t) = W(X_t, \varepsilon_t^w) + \delta \max_{\kappa \in \{0,1\}} V_{t+1}^\kappa(S_{t+1})$$

where $Exp_{t+1} = Exp_t + 1$ and $G_{t+1} = G_t$ if working, since $P(\text{Grad}_t = 0|S_t, \text{College}_t = 0) = 1$. If enrolled in college then $Exp_{t+1} = Exp_t$ and $G_{t+1} = \max\{G_t, \text{Grad}_t\}$. The value of college enrollment is:

$$V_t^1(S_t) = b^c(X_t) + \varepsilon_t^c + \delta \sum_{g \in \{0,2,3,4,5\}} P(\text{Grad}_t = g|S_t, \text{College}_t = 1) \max_{\kappa \in \{0,1\}} V_{t+1}^\kappa(S_{t+1})$$

¹²More specifically:

$$\begin{aligned} & P(\text{College}_t = 1|S_t) \\ &= P(V_t^1(S_t) > V_t^0(S_t)) \\ &= P\left(b^c(X_t) - W(X_t, \varepsilon_t^w) + \delta \left(\frac{\left(\sum_{g \in \{0,2,3,4,5\}} P(\text{Grad}_t = g|S_t, \text{College}_t = 1) \max_{\kappa \in \{0,1\}} V_{t+1}^\kappa(S_{t+1}) \right)}{-P(\text{Grad}_t = 0|S_t, \text{College}_t = 0) \max_{\kappa \in \{0,1\}} V_{t+1}^\kappa(S_{t+1})} \right) > -\varepsilon_t^c \right) \\ &= \frac{1}{1 + \exp\left(b^c(X_t) - W(X_t, \varepsilon_t^w) + \delta \left(\frac{\left(\sum_{g \in \{0,2,3,4,5\}} P(\text{Grad}_t = g|S_t, \text{College}_t = 1) \max_{\kappa \in \{0,1\}} V_{t+1}^\kappa(S_{t+1}) \right)}{-P(\text{Grad}_t = 0|S_t, \text{College}_t = 0) \max_{\kappa \in \{0,1\}} V_{t+1}^\kappa(S_{t+1})} \right) \right)} \end{aligned}$$

We now have simple analytical forms for all the conditional choice probabilities. The main difficulty in calculating the current choice probabilities arises because they depend on future expected utilities; hence the computation requires that the utility of all potential state-choice combinations must be determined.

In the estimation, we solve this dimensionality problem using the method developed in Hotz and Miller (1993) that relies on a representation of the value function in which future conditional choice probabilities (CCPs) are treated as data rather than functions of the underlying structural parameters. Since data is available on future choices, these probabilities can be calculated from the sample proportions. First note that one great simplification provided by the distributional assumptions is that the $E \max$ in (18) becomes a closed form expression:¹³

$$\begin{aligned} E[V_{t+1}(S_{t+1})|S_t, College_t = k] &\equiv E\left[\max_{\kappa \in \{0,1\}} V_{t+1}^{\kappa}(S_{t+1})|S_t, College_t = k\right] \\ &= \gamma + E\left[\ln\left(\sum_{\kappa=0}^1 \exp(V_{t+1}^{\kappa}(X_{t+1}))\right)|X_t, College_t = k\right] \end{aligned}$$

where γ is Euler's constant and $V_{t+1}^{\kappa}(X_{t+1})$ is the expectation of the alternative κ specific value function given current observed state, X_t , and current alternative, k . Consequently, these assumptions obviate the necessity of numerically computing multivariate integrals and greatly reduce the computational burden. Second note that equation (21) can be used in the calculation of the probability statements in the second stage which are necessary to evaluate the likelihood function. The CCPs are then treated as nuisance parameters in the estimation. Hotz and Miller (1993) show that Bellman's equation (18) can always be written as a function of current utilities and future CCPs. Note that given the model assumptions, the term inside the $\ln(\cdot)$ in (21) is the denominator of the CCP of choosing any of the alternatives k given the state, i.e. $\xi_k(X_{t+1}) = P(College_t = k|X_{t+1}) = \frac{\exp(V_{t+1}^k(X_{t+1}))}{\sum_{\kappa=0}^1 \exp(V_{t+1}^{\kappa}(X_{t+1}))}$. Particularly, this implies that $\ln(\xi_0(X_{t+1})) = V_{t+1}^0(X_{t+1}) - \ln(\sum_{\kappa=0}^1 \exp(V_{t+1}^{\kappa}(X_{t+1})))$ and that the one-period

¹³Consult McFadden (1978, 1981) and Rust (1987) for details and a derivation of this result.

ahead value function conditional on alternative k chosen this period (21) can be written:

$$\begin{aligned}
& E[V_{t+1}(X_{t+1}) | X_t, k] \\
&= \gamma + \sum_{g \in \{0,2,3,4,5\}} P(Grad_t = g | X_t, k) \ln \left(\sum_{\kappa} \exp(V_{t+1}^{\kappa}(X_{t+1})) \right) \\
&= \gamma + \sum_{g \in \{0,2,3,4,5\}} P(Grad_t = g | X_t, k) (V_{t+1}^0(X_{t+1}) - \ln(\xi_0(X_{t+1})))
\end{aligned} \tag{22}$$

Consequently, the value function is solely a function of the flow utility, $U_t^k(S_t)$, the one-period ahead expected value of exiting the university, $V_{t+1}^0(X_{t+1})$, and the one-period ahead CCP of choosing to exit the university.

The graduation probability, $P(Grad_t = g | X_t, k)$, controls student expectation about the one-period ahead state transition and together with the wage equation it also controls the expectation about the one-period ahead value of the full-time labor market alternative. Furthermore, in the third stage $V_{t+1}^0(X_{t+1})$ simplifies to (19). Finally, this implies that the conditional probability of choosing alternative k that enters the likelihood function is given by substituting (19) and (22) into the college choice probabilities. In the terminology of Arcidiacono and Miller (2008), the model is said to exhibit the one period dependence (OPD) property, since the current value function only depends on the one-period ahead value of university exit and the probability of choosing to exit the university and start working full time on the labor market.

3.2 Estimation

The parameters of the structural model are estimated by a maximum likelihood based procedure. The model basically requires two types of parameters to be estimated: utility function (preference) parameters: β 's, μ 's (and α 's), and transition parameters: γ 's and α 's.¹⁴ The transition parameters are used in forming expectations about uncertain future events; these include the γ parameters in the graduation process (15) through which students learn about their academic abilities, as well as the α parameters of the wage process (16) that form student expectations about future wages. Note that one important feature of the model is that the wage equation (16) is both part of the law

¹⁴When estimated as an ordered logit model, the basic model has 38 parameters to be estimate: $\mu = (\mu_0, \mu_1)$, $\beta = (\beta_0, \beta_1, \dots, \beta_7)$, $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_7)$, $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_7, \alpha_{12}, \alpha_{13}, \dots, \alpha_{17}, \alpha_{23}, \alpha_{24}, \dots, \alpha_{27})$, and σ_w .

of motion and an important part of utility.¹⁵

Let $O_{it} = (PS_i, Math_i, College_{it}, Grad_{it}, W_{it})$ denote the vector of observed choices and outcomes for individual i at time t . At $t = 0$, we observe pilot school status, PS_i , at $t = 1$, we observe the Math choice, $d_{i1} = Math_i$, and from $t = 2$ onwards we observe a sequence of college enrollment choices, $d_{it} = College_{it}$. The observed outcomes are accepted wages, W_{it} , and acquired college degrees, G_{it} . The likelihood function for the sample of individuals $i = 1, \dots, N$ observed from period $t = 0, 1, \dots, T_i$ is given by the product over the individual likelihood functions, which is the density for the sequence of observables conditional on the model parameters. Because of the additive separability and conditional independence assumptions, the individual likelihood contribution, $\mathcal{L}_i(\theta)$, can be decomposed into a product of conditional and marginal densities for each transition. With independent errors across each of the outcomes, the likelihood function factors into:

$$\begin{aligned} \mathcal{L}_w(\alpha) &- \text{the likelihood contribution of wages} \\ \mathcal{L}_m(\mu) &- \text{the likelihood contribution high school Math choices} \\ \mathcal{L}_g(\gamma) &- \text{the likelihood contribution of college degree progression} \\ \mathcal{L}_c(\theta) &- \text{the likelihood contribution of college-work choices (utility)} \end{aligned}$$

where $\theta = (\mu, \alpha, \gamma, \beta)$. The sample log likelihood function is then the sum of these three components:

$$\ln \mathcal{L}(\theta) = \ln \prod_{i=1}^N (\mathcal{L}_{mi}(\mu) \times \mathcal{L}_{wi}(\alpha) \times \mathcal{L}_{gi}(\gamma) \times \mathcal{L}_{ci}(\theta))$$

Note that the entire set of model parameters enters the likelihood through the college choice probabilities and that subsets of the parameters enter through the other structural relationships as well - μ through the Math choice probabilities, α through the wage equation, and γ through the college graduation probabilities. Given the additivity of $\ln \mathcal{L}(\theta)$, estimation could be carried out by fast sequential maximum likelihood. Since we have data on Math and college choices, accumulated college degrees, and wages, we can consistently estimate the μ , α and γ parameter vectors by maximizing $\mathcal{L}_m$, $\mathcal{L}_w$ and $\mathcal{L}_g$ separately. Then using the $\hat{\mu}$, $\hat{\alpha}$ and $\hat{\gamma}$ parameter estimates, we can consistently

¹⁵This is a standard feature of structural dynamic discrete choice schooling models, see e.g. Belzil (2007) for further discussion and a review of the literature.

estimate the preference parameter vector β by maximizing $\mathcal{L}_c(\hat{\mu}, \hat{\alpha}, \hat{\gamma}, \beta)$. Estimating the parameters stepwise rather than jointly saves significant computational time. The resulting inconsistent standard errors for the preference parameters, due to the estimation error, could be corrected with one Newton step over the whole likelihood, cf. Rust (1994). The rate of time preference is fixed at $\delta = 0.95$ in all the estimations.

Note that additive separability and conditional independence imply that if there is no unobserved individual heterogeneity in the model, the final estimation step reduces to estimating a logit of current choices on current flow utility and the discounted one-period ahead expected value function.

3.2.1 Unobserved Heterogeneity

Given the diversity of high school graduates' observed characteristics, it is unlikely that they have the same preferences for college, as well as unobserved work and college abilities. Hence, it seems important to account for persistent unobserved heterogeneity in multiple traits that may themselves be related. A common approach in the literature is to treat these initial traits as unmeasured and drawn from a mixture distribution; see e.g. Keane and Wolpin (1997), Eckstein and Wolpin (1999), and Arcidiacono (2004). We assume there is a fixed number of discrete types of individuals who differ in the parameters that describe their preferences, their academic ability and motivation, and their labor market ability. We adopt this nonparametric approach introduced by Heckman and Singer (1984) and allow for a finite mixture of M types. Each type comprises a fixed proportion, π_m , $m \in \{1, \dots, M\}$, of the population. This way of accounting for unobserved heterogeneity allows for flexible correlation of the errors across the choices as well as correlation over time.

In the estimation, wage offers are allowed to differ by unobserved type reflecting persistent differential labor market skills, $\alpha_0 = \sum_{m=1}^M \alpha_{0m} \mathbf{1}[type = m]$, in equation (16). Introducing, $\mu_0 = \sum_{m=1}^M \mu_{0m} \mathbf{1}[type = m]$, in equation (13) allows persistent Math abilities to differ by type. Persistent college abilities are also allowed to differ by type by introducing, $\gamma_0 = \sum_{m=1}^M \gamma_{0m} \mathbf{1}[type = m]$, in equation (15). Likewise, the consumption value of college in equation (14) is allowed to differ by unobserved type, $\beta_0 = \sum_{m=1}^M \beta_{0m} \mathbf{1}[type = m]$.¹⁶ The likelihood function becomes a finite mixture (or weighted average) of the type-specific likelihoods. Hence, every type is described by a

¹⁶Hence, the model with unobserved heterogeneity has $37 + 4M$ parameters to estimate.

vector of parameters that are given to them in high school, corresponding to their labor market skills, Math abilities, college abilities, and their preferences for college.

To conserve on parameters and avoid identification issues, we consistently only allow for first-order heterogeneity effects. This approach is common in the literature.

3.3 Potential Behavioral Effects of Math

The model includes several channels through which advanced Math can affect educational choices and outcomes, as well as labor market outcomes. First, Math is allowed to affect labor market outcomes directly (through shifting wages up by α_1). Second, Math can also affect the wage return to college enrollment and graduation (α_{12} through α_{15}) as well as the wage return to experience (α_{16} and α_{17}). Third, advanced Math may change preferences for college attendance directly through changing the consumption value b^C (through β_1). Finally, Math can increase the college graduation probabilities P_G (through γ_1) if the skills required in the Math course are more valuable than skills required in other courses as to being better able to successfully complete a college education.

3.4 Extended Model: Field Choice

This section presents the extended model, where students in the second stage also choose a field of college education. This amounts to include more choices for $College_{it} = 1$ in (10) and would give us direct estimates of transferability of academic skills across fields, as well as the impacts of advanced high school Math on these switching costs.

Furthermore, we can explicitly include college entry requirements in the model. These extensions of the model would also allow us to evaluate policies concerning field specific entry requirements.

3.5 Identification

TBW

4 Data

For our empirical analysis we use a panel data set comprising the population of individuals starting high school from 1984–87 in Denmark. The data are administered by Statistics Denmark, which has gathered the data from administrative registers. For each individual, we have data from complete detailed educational histories, including detailed codes for the type of education followed (level, subject, and educational institution) and the dates for entering and exiting the education, along with an indication of whether the individual completed the education successfully, dropped out or is still enrolled as a student. Furthermore, we have information on the branch choice in high school and on high school GPAs. The GPA is a weighted average of final exam grades for each course. Both the quality of the courses and the GPA are comparable across high schools since monitoring high schools is centralized at the Ministry of Education. Furthermore, all high school students within each high school cohort take identical written exams, while oral exams and major written assignments are evaluated both by the student’s own teacher and an external examiner assigned by the Ministry of Education. Note that there are no tuition fees for education in Denmark, and all students 18 years and older receive a study grant that suffices to cover living expenses. The grant is independent of parental income (after age 19), educational effort, and achievement as long as the student is less than one year behind the prescribed norm. We have yearly observations on labor income (earnings), gross income, and net income for 1997–2000. All incomes are observed at year-end and deflated to real values measured in DKK in 2000 using the average wage index for the private sector. Other individual background variables used in our estimations are gender and actual labor market experience (including a squared term). The parental background variables used include: A set of mutually exclusive indicator variables for the level of highest completed education of the mother and father, respectively, and their income as observed at the end of the year before the individual started high school.

For the empirical analysis, the central dependent variables are college enrollment, graduation and wages. We regard the individual as being enrolled in college ($College_i = 1$), if the individual enrolls within 5 years after high school graduation. We regard the individual as having graduated ($Grad_i = 1$), if the individual has graduated from college within 10 years after high school graduation. We record wages (W_{it}) 12, 13, and 14 years after high school enrollment, which means that the mentioned definition will allow

us to capture the dominant part of enrollment and graduation as well as observe each individual in the labor market at least once. College may be either short-cycle (e.g. diplomas in health assistance, computer programming), medium-cycle (e.g. BA and BSc obtained at nursery college, teachers' college) or long-cycle higher education (MA and MSc degrees).

4.1 Sample Selection

Among the gross population of high school entrants for 1984–87, only high school graduates who finished in three years are selected. Furthermore, we exclude individuals with missing labor market income 12, 13, and 14 years after starting high school; hence, individuals who have left the country, died, are unemployed full-time year-round, or out of the labor force all three years are excluded. Since students with high-level Math are less likely to be unemployed or nonparticipants than others (see the next section), excluding individuals with missing incomes potentially introduces a negative bias on the parameter of main interest, meaning that our conclusion becomes conservative. After the restrictions, the sample contains observations on 61,140 high school graduates, which makes about 15,000 graduates per cohort coming from about 140 different high schools; see Table 1. Due to concerns about validity of the experiment, we disregard cohort 1987 and continue the empirical analysis with 45,936 observations from cohort 1984–86.

4.2 Data Description

The descriptive statistics for the individuals in the sample are shown in Tables 2 and 3. Table 2 shows summary statistics of all background variables as well as the difference in these variables across school types, while Table 3 shows summary statistics of various outcome variables across school type as well as the differences by Math level. Tables 2 and 3 show that students who attended a pilot school have more favorable individual characteristics (apart from family background) than the students who attended nonpilot schools. The most favorable characteristics exist for students at schools who have advertised their pilot status ($PilotSchool = 1$ & $PilotIntro = 0$), while the least favorable characteristics are found for individuals at nonpilot schools. Students who were unexpectedly exposed to the pilot scheme ($PilotSchool = 1$ & $PilotIntro = 1$)

had 0.10 years more education and 3 percent higher earnings, on average, than the average graduate from a nonpilot school, while students who were expectedly exposed to the pilot scheme ($PilotSchool = 1$ & $PilotIntro = 0$) had an additional 0.23 years of education and an additional 4 percent higher earnings. Table 3 reveals that the group of individuals choosing advanced Math have more favorable characteristics no matter what the school type and that this group of students stands out even more at the nonpilot schools, where the students are compelled to take advanced physics in order to take advanced Math than they do at any of the two school types offering the pilot program. High-level Math students have higher high school GPAs and more high-level Math students attend and complete a higher education at any level. Aside from having higher completion rates, high-level Math students also complete a given educational level at a faster rate. Hence, high-level Math students seem to be more effective in the higher educational system. This is corroborated by their transitions through the higher educational system displayed in Figure 1 and 2. Figure 2 reveals that more than 50% of high-level Math high school graduates go directly to college, compared to less than 30% of the others. High-level Math students are also more persistent as almost all of them stay on after the first college year, compared to only around 80% among those with a lower Math level. In addition, Table 3 shows that they are more successful after entering the labor market as they are unemployed less and earn more. High-level Math students' log earnings are 0.29 higher than the earnings of other high school students. As a point of reference, the Math log earnings gap is more than five times larger than the gender log earnings gap for these high school graduates. Hence, we set out to find out whether this huge earnings gap arises because high-level Math students are more likely to enroll in higher education, because they are more productive in completing a higher education, or because they become more productive on the labor market, while accounting for potentially favorable unobservable characteristics among those who selected the high-level Math course.

5 Results

5.1 Results from Estimation of Reduced Form Models

We commence our attempt to separately identify the effect of advanced math on college enrollment, college graduation and earnings with estimation of simple reduced form

models. In this context, the reduced form models consist of a set of IV estimations where Math is treated as an endogenous variable which is instrumented by the indicator variable measuring whether individuals are unexpectedly exposed to the pilot scheme, $PS_i = 1$. The three outcome variables are: an indicator for college enrollment within 3 years after high school graduation, an indicator for college graduation within 8 years after high school graduation, and log earnings as measured by the log of average wages in the 12th -14th year after enrollment.

Looking at the first stage, we find that the instrument, PS_i , has a strong effect on the probability of choosing the high level math course (z-stat around 10). Furthermore, the magnitude of the effect is large: being unexpectedly exposed to the pilot program increases the probability of choosing high level math by 10 pct points.

Turning to the second stage, we find that high level math has a significantly positive effect on all three outcome variables when gender is controlled for. In the specification with no control variables, the effects of math as estimated from OLS is significantly positive for all three outcome variables, while the effect is only significant for earnings in the IV estimations (columns (1) and (6)). As soon as gender is added as a control variable, the estimated effects of math are significantly positive, and they are robust to adding more control variables (columns (2)-(4) and (7)-(9)). In the last specification (columns (5) and (10)), we add post-graduation controls in terms of experience and indicators for educational level and subject grouping. It only makes sense to estimate this specification when the outcome variable is earnings, and here we find that the effect of math is halved in the OLS regression while it becomes insignificant in the IV estimation.

The reduced form results shows that the effect of high level math on earnings is in the range 15-20 pct points. Furthermore, they indicate that math influences both college enrollment and graduation strongly, and that the effect of math on earnings is entirely driven by the indirect influences through these two indicators of college behaviour. In the next sub-section, we explore the underlying channels through which math causes more favorable outcomes in more detail by estimation of our structural model presented in Section 3.

5.2 Results from Estimation of Structural Models

TBW

6 Conclusion

TBW

References

- [1] Altonji, J. (1995), “The Effect of High School Curriculum on Education and Labor Market Outcomes.” *Journal of Human Resources* 30(3):409-38.
- [2] Arcidiacono, P. (2004), “Ability Sorting and the Returns to College Major”, *Journal of Econometrics*, 121(1-2): 343-375.
- [3] Arcidiacono, P. and J. B. Jones (2003), “Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm”, *Econometrica*, 71(3): 933-946.
- [4] Arcidiacono, P. and R. A. Miller (2008), “CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity”, manuscript, Duke University.
- [5] Attanasio, O., C. Meghir and A. Santiago (2005), “Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progreso”, University College London, IFS Working Paper EWP05/01.
- [6] Belzil, C. (2007), “The Return to Schooling in Structural Dynamic Models: A Survey”, *European Economic Review*, 51(5): 1059-1105.
- [7] Card, D. and D. R. Hyslop (2005), “Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers”, *Econometrica*, 73(6): 1723-1770.
- [8] Eckstein, Z. and K. I. Wolpin (1999), “Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities”, *Econometrica*, 67(6): 1295-1339.
- [9] Hanushek, E. and D. Kimko (2000) “Schooling, Labor-Force Quality and the Growth of Nations”, *American Economic Review* 90(5):1184–1208.
- [10] Heckman, J. J. and B. Singer (1984); “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data”, *Econometrica*, 52 (2): 271-300.
- [11] Imbens, G. W. and J. D. Angrist (1994), “Identification and Estimation of Local Average Treatment Effects”, *Econometrica*, 62(2): 467-475.
- [12] Jamison, E., D. Jamison, and E. Hanushek (2007), “The Effects of Education Quality on Income Growth and Mortality Decline.” *Economics of Education Review*, 26(6): 771–88.

- [13] Joensen, J. S. and H. S. Nielsen (2009), “Is there a Causal Effect of High School Math on Labor Market Outcomes?”, *Journal of Human Resources*, 44(1): 171-198.
- [14] Keane, M. P. (2009), “Structural vs. Atheoretic Approaches to Econometrics”, forthcoming in the *Journal of Econometrics*.
- [15] Keane, M. and K. I. Wolpin (1997), “The Career Decisions of Young Men”, *Journal of Political Economy*, 105(3): 473-522.
- [16] Keane, M. P. and K. I. Wolpin (2007), “Exploring the Usefulness of a NonRandom Holdout Sample for Model Validation: Welfare Effects on Female Behavior”, *International Economic Review*, 48(4): 1351-1378.
- [17] Meghir, C. (2006), “Dynamic Models for Policy Evaluation”, University College London, IFS Working Paper EWP06/08.
- [18] Rose, H. and J. Betts (2004), “The Effect of High School Courses on Earnings”, *Review of Economics and Statistics* 86(2): 497-513.
- [19] Rust, J. (1987), “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher”, *Econometrica*, 55(5): 999-1033.
- [20] Rust, J. (1994), “Structural Estimation of Markov Decision Processes”, in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, North-Holland, Amsterdam.
- [21] Todd, P. and K. Wolpin (2006), “Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility”, *American Economic Review*, 96(5): 1384-1417.
- [22] Willis, R. J. and S. Rosen (1979), “Education and Self-Selection”, *Journal of Political Economy*, 87(5): 7-36.

Table 1. Overview of the Introduction of the Pilot Scheme in Danish High Schools.

Cohort starting in high school in year	PilotSchool=0		PilotSchool=1		PilotSchool=1		All
			PilotIntro=0		PilotIntro=1		
	#schools	#students	#schools	#students	#schools	#students	#students
1984	121	12880	0	0	22	3260	16140
1985	105	10561	22	2977	15	1741	15279
1986	90	8392	37	4514	15	1611	14517
All		31833		7491		6612	45936

Table 2. Descriptive Statistics.

Variable	Means and (standard deviations)		
	Overall mean mean	Mean difference between <i>PilotSchool=1</i> <i>PilotIntro=1</i> and <i>PilotSchool=0</i>	Mean difference between <i>PilotSchool=1</i> <i>PilotIntro=1</i> and <i>PilotSchool=1</i> <i>PilotIntro=0</i>
Educational variables:			
Attended primary school 10th grade	0.31	0.00	0.01
High School:			
Age at high school entry	16.25 (0.61)	-0.01	0.01
Math track students	0.68	0.02	-0.02
High level Math	0.27	0.10	-0.02
GPA	8.25 (0.97)	0.00	-0.04
Highest completed education:			
Length of education (years)	15.16 (2.39)	0.10	-0.23
Educational level grouping:			
Only high school	0.18	-0.02	0.01
Vocational training	0.11	0.01	0.02
Short higher education (SHE)	0.07	0.01	0.00
Medium higher education (MHE)	0.33	-0.02	-0.02
Long higher education (LHE)	0.25	0.02	-0.03
Educational level-subject grouping:			
SHE Technical	0.02	0.00	0.00
SHE Other	0.05	0.01	0.00
MHE Teacher	0.08	-0.01	0.00
MHE Humanities	0.04	-0.01	0.00
MHE Health	0.07	0.00	0.00
MHE SocSci	0.05	0.00	0.00
MHE TechSci	0.07	0.00	0.00
MHE Other	0.02	0.00	0.00
LHE Health	0.02	0.00	0.00
LHE NatSci	0.03	0.00	0.00
LHE Technical	0.05	0.01	-0.01
LHE Humanities	0.03	0.00	-0.01
LHE SocSci	0.10	0.02	-0.01
LHE Other	0.02	0.00	0.00
Income:			
Log Labor Income (2000 DKK)	12.03 (0.96)	0.03	-0.04
Other individual variables:			
Female	0.57	-0.01	0.01
Labor Market Experience (years)	4.94 (2.80)	-0.02	0.18
Parental background variables:			
Income:			
Fathers' log income (2000 DKK)	8.85 (5.51)	-0.03	0.05
Mothers' log income (2000 DKK)	8.18 (5.19)	-0.30	-0.17
Highest completed education:			
Father basic school	0.16	-0.02	0.01
Father high school	0.01	0.00	0.00
Father vocational training	0.28	0.01	0.02
Father short higher education	0.03	0.00	0.00
Father medium higher education	0.13	0.01	0.00
Father long higher education	0.11	0.00	-0.01
Mother basic school	0.23	-0.01	0.04
Mother high school	0.01	0.00	0.00
Mother vocational training	0.29	0.00	-0.01
Mother short higher education	0.04	0.01	0.00
Mother medium higher education	0.16	-0.01	-0.01
Mother long higher education	0.03	0.00	-0.01
Number of Individuals	45936		

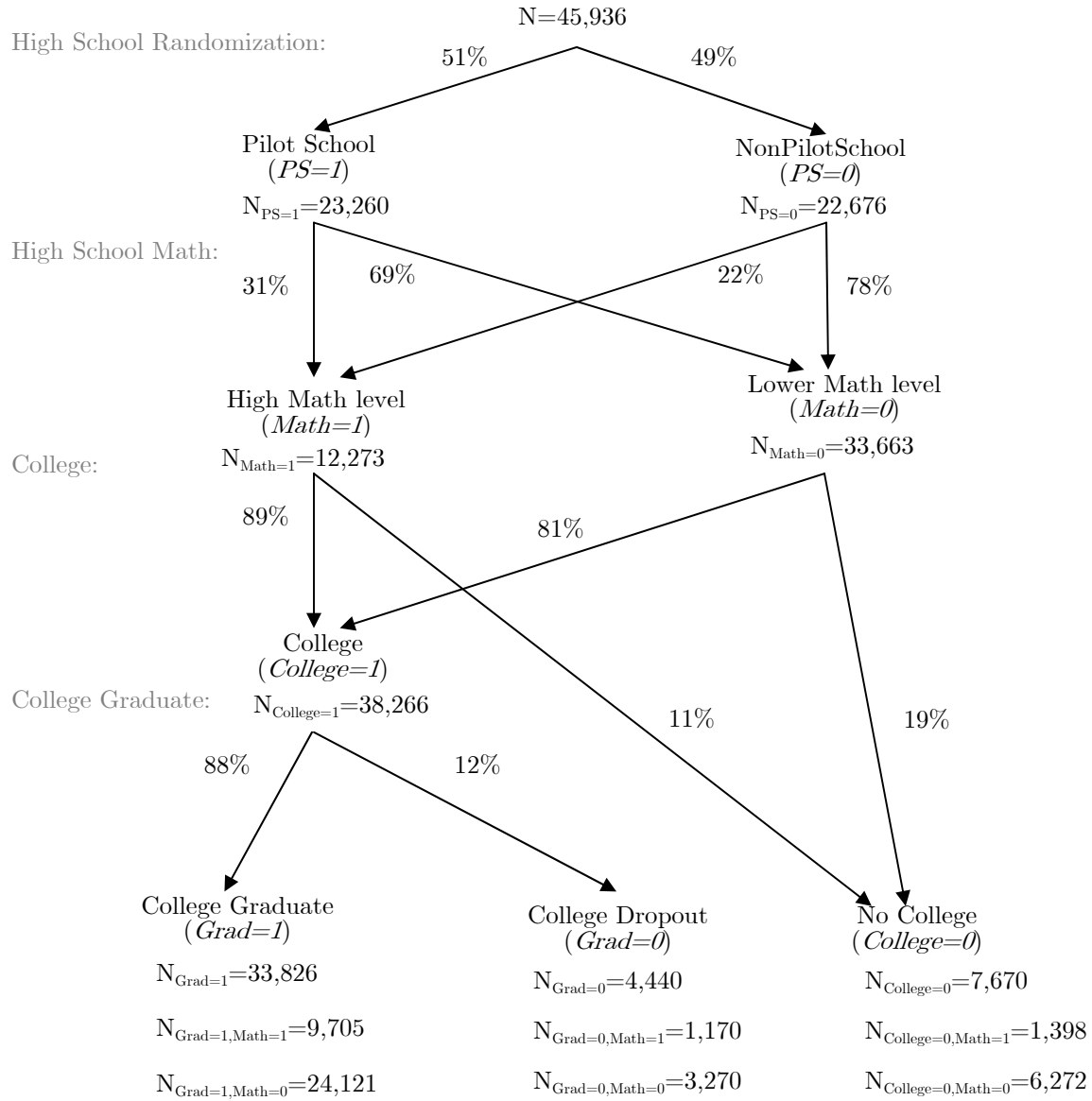
Table 3. Descriptive Statistics of Labor Market and Educational Outcomes.

	Sample means and (standard deviations)							
	All		<i>PilotSchool</i> = 0		<i>PilotSchool</i> = 1 <i>PilotIntro</i> = 0		<i>PilotSchool</i> = 1 <i>PilotIntro</i> = 1	
	Overall mean	Mean difference by <i>Math A</i>	Overall mean	Mean difference by <i>Math A</i>	Overall mean	Mean difference by <i>Math A</i>	Overall mean	Mean difference by <i>Math A</i>
Outcomes								
Educational Outcomes:								
GPA in High School	8.25 (0.97)	0.24	8.24 (0.97)	0.25	8.28 (0.97)	0.23	8.24 (0.96)	0.19
<i>College Enrollment:</i>								
Enrolled in Higher Education	0.83	0.07	0.83	0.08	0.86	0.06	0.83	0.05
Enrolled in Long Higher Education	0.46 (0.49)	0.18	0.45 (0.49)	0.19	0.50 (0.49)	0.17	0.46 (0.49)	0.18
<i>College Graduation:</i>								
College Degree	0.74	0.07	0.73	0.08	0.77	0.05	0.73	0.06
Master's Degree	0.35	0.17	0.34	0.16	0.39	0.16	0.36	0.16
Length of Highest Completed Education (years)	15.16 (2.36)	0.85	15.09 (2.36)	0.87	15.42 (2.33)	0.74	15.19 (2.38)	0.80
<i>Time to College Graduation:</i>								
Years from High School Graduation to Master's Graduation	8.73 (2.46)	-1.16	8.81 (2.50)	-1.20	8.45 (2.28)	-1.01	8.69 (2.48)	-1.09
Labor Market Outcomes:								
Labor Income (2000 DKK)	12.09 (0.94)	0.29	12.07 (0.95)	0.30	12.15 (0.90)	0.27	12.12 (0.91)	0.24
Degree of Unemployment (scale 0-1000)	50.84 (146.88)	-19.66	52.41 (149.90)	-18.61	45.22 (135.69)	-20.81	49.64 (144.37)	-20.01
Number of Individuals	45936		31833		7491		6612	

Table 4. Estimates of Causal Effects of High level Math on Labor Market and Educational Outcomes.

	Parameter estimates (standard errors) [marginal effects]									
	OLS					IV				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Effect of High level Math on Outcome:										
College Enrollment	0.07 *** (0.00)	0.07 ** (0.00)	0.06 ** (0.00)	0.06 ** (0.00)		-0.05 (0.05)	0.07 * (0.04)	0.09 ** (0.04)	0.07 * (0.04)	
College Graduation	0.07 *** (0.00)	0.08 *** (0.00)	0.08 *** (0.00)	0.08 *** (0.00)		0.03 (0.06)	0.10 ** (0.05)	0.12 *** (0.05)	0.12 ** (0.05)	
Earnings	0.32 *** (0.01)	0.24 *** (0.01)	0.24 *** (0.01)	0.24 *** (0.01)	0.13 *** (0.01)	0.23 *** (0.13)	0.18 *** (0.10)	0.17 *** (0.10)	0.16 *** (0.10)	-0.05 (0.07)
High level Math First-Stage:						0.29 *** (0.02)	0.31 *** (0.02)	0.30 *** (0.02)	0.28 *** (0.02)	0.30 *** (0.02)
Pilot School Intro						[0.10]	[0.11]	[0.10]	[0.09]	[0.10]
Additional control variables:										
<i>Individual variables:</i>										
Gender		+	+	+	+		+	+	+	+
Experience (quadratic)					+					+
Educational level grouping					+					+
<i>Parental variables (for mother and father):</i>										
Highest completed education and income			+	+	+			+	+	+
<i>Regional controls:</i>										
County indicators			+	+	+			+	+	+
<i>Cohort controls:</i>										
Entry cohort fixed effects			+	+	+			+	+	+
<i>High School specific controls:</i>										
Average parental background in 1983				+	+				+	+

Figure 1. Overview over Individual Choices in Basic Model.



Unconditional Probabilities:

$$P(Grad = 1|Math = 1) = .79, \quad P(Grad = 0|Math = 1) = .10, \quad P(College = 0|Math = 1) = .11$$

$$P(Grad = 1|Math = 0) = .72, \quad P(Grad = 0|Math = 0) = .09, \quad P(College = 0|Math = 0) = .19$$

Figure 2. College Enrollment over time after High School Graduation, by MathA.

