The Evolution of Inequality in Education Trajectories and Graduation Outcomes in the US*

Christian Belzil
CREST, Paris Polytechnic Institute and IZA and CIRANO

Jörgen Hansen
Concordia University, CIREQ and IZA

Xingfei Liu
University of Alberta

May 29, 2022

*Financial support from Investissement d'Avenir (ANR-11-IDEX-0003/ Labex Ecodec/ANR-11-LABX-0047) and Social Sciences and Humanities Research Council of Canada (grant 435-2017-0129) is gratefully acknowledged. We thank seminar participants at CEST-IP Paris and at the CIREQ Conference Advances in Microeconometrics (Montreal, May 2022). The usual disclaimer applies.

Abstract

We model the joint distribution of (i) individual education trajectories, defined by the allocation of time (semesters) between various combinations of school enrollment with different labor supply modalities and periods of school interruption devoted either to employment or home production and (ii) actual graduation outcomes using two cohorts of the National Longitudinal Survey of Youths which we follow from 16 to 28. We discuss the evolution of family income and ability effects where the latter are decomposed into an academic (cognitive) and a practical (technical-mechanical) latent ability factor component correlated with family income and background variables. We find that the individual cognitive-technical ability differential prevailing at 16 was increasing with income in the early 80's but much less so in the early 2000's. We find no evidence of any income-based "trajectory inequality" in either cohort, after conditioning on abilities. Among all graduation and enrollment outcomes, college graduation is the only for which the effect of income has increased between the 1980's and the early 2000's but it reached a level no more important than the high school graduation income effect. In both cohorts, cognitive and technical abilities were the dominant factors but they affect most dimensions of individual trajectories and all graduation outcomes in opposite directions. However, the cognitive ability factor lost half of its effect on college graduation while the impact of the technical-mechanical factor has been more stable across cohorts.

JEL Classification: I2, J1, J3

Keywords: *Education, Inequality, Family Income, Multi-dimensional Abilities, Labor Supply.*

1 Introduction and Motivation

This paper is concerned with the evolution of educational inequality in the US between the early 1980's and the early 2000's. We estimate a reduced-form dynamic discrete choice model on two cohorts of the National Longitudinal Survey of Youths (NLSY) which we follow from 16 to 28. The behavioral model has two main components. The first is the distribution of individual trajectories, defined as the allocation of time (accumulated semesters) between various combinations of school enrollment-labor supply modalities and periods of school interruption devoted either to employment or home production. The second is the distribution of graduation outcomes (high school and 4-year college) conditional on realized trajectories.

The model is estimated conditional on two latent factors measuring cognitive (academic) and technical-mechanical abilities but also incorporates additional unobserved heterogeneity affecting both trajectories and graduation outcomes and which may be interpreted as non-cognitive skills such as motivation or confidence. The factors are identified from the Armed Services Vocational Aptitude Battery (ASVAB) scores available in both cohorts. Both of them are allowed to be correlated with individual and family characteristics measured before high school graduation. As we do not impose stationarity of the distribution of the cognitive and technical factors across cohorts, we can allow the mapping from the characteristics (including family income) onto factors to vary across cohorts and infer to what extent changes in the overall effect of income on education may be decomposed into changes in the ability-income relationship as opposed to net income effects.

At the outset, it should be clear that technical-mechanical ability may be regarded as a specific dimension of the general concept of cognitive ability more in line with the notion of "practical" intelligence. However, to be coherent with the existing literature, we use standard terminology and refer to the factor obtained from the ASVAB components, typically used to measure standard academic ability and compute AFQT scores, as the "cognitive" factor and to the second one as the "technical-mechanical" factor.[1]

Our model is therefore more general than those found in the literature

[1] As far as we know, only Prada and Urzua (2017) have used the technical-mechanical measurements to estimate a specific factor within an empirical schooling model. They review the psychometric and industrial psychology literatures in which considering cognitive abilities as multi-dimensional is strongly advocated.

on the effect of family resources on educational outcomes since it ties graduation to individual specific trajectories which depend themselves on individual characteristics and factors and because it treats cognitive abilities as bi-dimensional and incorporates additional unobserved heterogeneity which we may interpret as non-cognitive ability. The dynamic structure of the behavioral model decomposes progress toward college graduation into a high school phase and a post-high school period. Time allocation decisions during high school trigger variations in age at high school graduation, which affects post-high school time allocation which itself impacts stochastically on graduation outcomes.

Contribution

Our interest in the questions is partly motivated by the inherent difficulty to build bridges between the multiplicity of papers that have been concerned with the link between family resources and education inequality and its recent evolution. The existing literature on the role of family resources is too voluminous to warrant an in-depth survey in the main text as there already exists exhaustive reviews.[2] It is however crucial to recognize that the degree of heterogeneity in approaches is so important that it is difficult to synthesize the literature into a small number of coherent findings.

For instance, different papers: (i) focus on different educational outcomes such as enrollment, years of education, graduation; (ii) condition on different sets of characteristics; (iii) estimate different income effect parameters as some condition on ability while others do not; (iv) treat ability measures differently when those are accounted for (some use AFQT scores, some ignore them or use data that do not contain test scores); (v) measure educational outcomes at different ages; (vi) use different modeling strategies as most papers focus on conditional means and ignore dynamic selection while very few model outcomes sequentially; (vii) focus on different time periods since a very small number of papers analyze the evolution of these effects over time and; (viii) measure family income effect differently as some use family income during teen age years while others use lottery outcomes or other windfall income components. This renders comparisons across studies very difficult.

Our main objective is therefore to design a reduced-form dynamic model

[2] In a recent paper, Bulman et al. (2021) present an exhaustive survey of the literature. We also provide a thorough review of the literature in a web appendix.

that complements: (i) the literature on the role of parental resources on access to higher education which ignores differences in time allocation decisions and trajectories; (ii) the structural schooling literature concerned with liquidity constraints, and which because of the curse of dimensionality, must restrict both the importance of heterogeneity and the complexity of the law of motion, and (iii) the more recent literature on the estimation of semi-structural schooling models in which the distribution of cognitive and non-cognitive skills is obtained through latent factor estimation techniques.[3]

There are four main contributions in our paper. First, by linking family income and multidimensional cognitive abilities to various types of trajectories leading (or not) to college graduation, we can analyze forms of family income-based education inequality which have never been investigated before. For instance, a typical question that we can answer is to what extent working while enrolled in college or interrupting school to work have become more or less affected by family income and abilities over time. This issue is particularly interesting in light of the growing interest in the impact of accumulated student debt in the US since employment while in school may be viewed as an alternative method of education financing.[4]

A second contribution is the estimation of the effects of time allocation decisions (trajectories) on the likelihood of graduating from college as well as on age at graduation, which is rendered possible by the rich dynamic structure of our model.

None of these parameters generated by our first two objectives have ever been reported in the empirical literature.[5]

[3]Our capacity to estimate such a rich model capturing the sequential nature of schooling choices and graduation outcomes, and allowing for so many different types of individuals (close to 1,000) is reached at the expense of specifying a fully structural model such as found in Keane and Wolpin (2001) and Johnson (2015). In a recent paper, Ashworth et al. (2021) use a similar approach. For recent examples of papers combining reduced-form models with latent factor estimation techniques, see Carneiro et al. (2003), Heckman, Humphries and Veramendi (2016) and Prada and Urzua (2017).

[4]Lochner, Stinebrickner and Suleymanoglu (2021) discuss many issues related to the optimal design of financial aid repayment regulations and estimate the determinants of student debt repayment.

[5]Some structural schooling models such as Keane and Wolpin (1997, 2001) incorporate the possibility of discontinuous schooling patterns but ignore the distinction between enrollments and graduation. The curse of dimensionality also prevents them to condition on individual specific regressors (let alone latent factors) and forces them to consider only

Our third contribution is the comparison of a wide range of parameters measuring the effect of family income on education in a much more transparent manner than would a comparison of estimates obtained in papers measuring different outcomes and using different methodologies. In order to cover the totality of educational outcomes considered in the literature, we measure the impact of family income and cognitive and technical factors on high school graduation and on the following college outcomes: enrollment, persistence (defined as enrolling in at least 5 semesters of post-high school education), 4-year college graduation and age at graduation (conditional on graduation).

We evaluate the effects on the general population (as done in the empirical literature) as well as on the sub-population of high school graduates and quantify precisely to what extent the overall correlation between family income and educational outcomes is driven by the correlation between family income and abilities (at 16) as opposed to net income effects prevailing after differences in abilities have been accounted for.

Finally, a fourth contribution is the analysis of the evolution of the cognitive-technical ability differential and the role it plays in explaining changes in income effects. To do this, we allow the factors to be correlated with pre-high school graduation background variables and thereby obtain three different types of income effects; one passing through cognitive ability, another one through technical ability and a third one measuring the net income effect. Our approach is at odds with the existing literature in which background variables are often introduced into measurement equations (outside the factor) and therefore focuses on the effects of the factors, net of parents background variables.[6] Our paper is probably the first that estimates the distinct effects of pure cognitive ability vs. technical-mechanical ability within a sequential dynamic discrete choice model of time allocation decisions and graduation outcomes and that can compare the evolution of their relative importance. As far as we know, it is also the first one that documents differences in the mappings from background variables onto cognitive and technical ability latent factors.

Motivation

small number of unobserved types (usually 4).

[6]Prada and Urzua (2017) measured the impact of cognitive and non-cognitive factors on enrollment and wages observed in the 1979 cohort of the NLSY but ignore the sequential nature of individual decisions. Their analysis is also limited to white males only.

The US labor market has been in constant transformation over the last 50 years. Some of the most salient changes have concerned the market for higher education. Between 1970 and 2005, net tuition for a 4-year college degree nearly tripled and average borrowing levels have increased substantially.[7] At the same time, education financing opportunities have also developed and an increasing number of students are working in the market while enrolled in school. Dynarski and Scott-Clayton (2016) show that the emergence of a multiplicity of higher education financial aid programs has translated into a decrease in net prices between 2005 and 2011. Scott-Clayton (2012), Bound, Lovenheim and Turner (2012), Murphy and Topel (2016) and Ashworth et al. (2020), have documented a raise in time needed to graduate from college in conjunction with a clear increase in work experience during college. According to the National Center for Education Statistics, the yearly flow of individuals graduating with a Bachelor's degree, which was below 1 million in the early 1980's, fluctuated between 1.5 and 3.0 million per year between 2005 and 2018.[8]

There is also reasonable evidence that enrollments have increased mostly at lower quality colleges and that average academic standards have decreased.[9] These statistics suggest that both the demand and supply for higher education services may have changed substantially and that the nature of educational selectivity that prevailed in the early 1980's may have changed accordingly.

Although the analysis presented in the paper entails comparing individual decisions taken in the early and mid-1980's with those taken mostly between 2000 and 2007, we believe that documenting changes taking place over this specific period may be of great interest for understanding the present state of the higher education market in the US. As pointed out earlier, many institutional changes (tuition rates, financial aid opportunities, college enrollments) initiated in the early 1980's have continued until very recently (Dynarski and Scott-Clayton, 2016) so that trends documented from the 1979 and 1997 cohorts of the NLSY are likely to persist nowadays.

Main Results

Despite a strong positive correlation between the ASVAB cognitive and

[7] Abel and Deitz (2014).

[8] Source: 120 years of American Education: A Statistical Portrait, National Center for Education Statistics, 1993.

[9] This point is emphasized in Hoxby (2009) and Babcock and Marks (2011).

technical measurements, the correlation between the cognitive and the technical-mechanical latent factors is negative in both cohorts (-0.17 in the 79 cohort and -0.36 in the 97 cohort). The effect of income on cognitive ability has been divided almost by 5 between the early 80's and the early 2000's. A \$30,000 difference in family income (practically equivalent to an increase of one standard deviation in income in the early 1980's) generated a difference of 0.14 standard deviation in the cognitive factor but only a 0.03 standard deviation increase in the early 2000's. At the same time, a \$30,000 difference in family income generated a negative difference of -0.06 standard deviation in the technical factor both in the early 80's and in the early 2000's, thereby indicating that some specialization already takes place before 16 and that those coming from low income families tend to favor activities fostering technical and mechanical skills more than those coming from advantaged backgrounds. In other words, the differential between individual cognitive and technical abilities was increasing with family income in the early 80's but has become less sensitive to family income in the early 2000's.

We find no evidence of any family income-based "trajectory inequality", after conditioning on abilities. In other words, work while in school, school interruptions (to work) and delayed enrollment were all unaffected by family income in both cohorts. Trajectories are mostly explained by cognitive and technical-mechanical abilities but they tend to play in opposite directions. That is high cognitive ability individuals worked less during school, enrolled earlier and graduated earlier, whereas high technical ability individuals worked much more during school, enrolled later and graduated later.

The model implies that, among those enrolled in education and who eventually graduated, a \$30,000 increase translated into an increase of 0.4 hour of employment per week in 79 and a reduction of 0.25 hour in 97. At the same time, a one standard deviation increase in cognitive skill reduced employment while in school by 2 hours per week in 79 and 1 hour in 97 but a one standard deviation increase in the technical-mechanical factor increased labor supply while in school by 6 hours per week both in the 79 and the 97 cohorts.

Among all graduation and enrollment outcomes, college graduation is the only for which the effect of income has increased between the 1980's and the early 2000's. The effects of income on high school graduation, enrollment and persistence have all decreased. However, the college graduation income effect reached a level no more important than the high school graduation income effect as our estimates imply that a \$30,000 real income difference raised high school and college graduation frequencies by 0.02 to 0.03.

The evolution of the impact of the cognitive factor on college graduation goes in the opposite direction of the income effect. We find evidence of a strong decrease in college graduation-cognitive ability selectivity but the effect of the technical factor, which dominates in absolute values, is also more stable across cohort. A one standard deviation increase in cognitive ability raised college graduation by 0.20 among high school graduates in 79 and by 0.09 in 97 (about 3 times a positive \$30,000 difference in 97). However, a one standard deviation in the technical-mechanical factor reduced college graduation by 0.27 in 79 and by 0.23 in 97 (about -10 times a positive \$30,000 difference in 97).[10]

In the early 80's, the total effect of income on high school graduation and college graduation were much higher than the net effects (which were either low or practically null) and indicate that graduation outcomes were sensitive to income mostly because individual abilities at 16 were correlated with parental income and other characteristics and more precisely, because the cognitive-technical ability differential was increasing with income. In the early 2000's, the net income effects capturing financing opportunities were more important than the residual ability income effect partly because the latter was reduced.

The reduced-form mappings from individual characteristics onto age at graduation are fully compatible with the weak dependence of most trajectory characteristics on family income and the strong dependence on cognitive ability as most income effects are negligible. A \$30,000 difference in family income increased age at graduation by less than 1 month in the 79 cohort and reduced it by about 2 months in the 97 cohort.

The cognitive ability factor was the strongest determinant of age at graduation in the 1980's but its decrease in importance noted for graduation outcomes seems to be mirrored in its impact on age at graduation. Precisely, a 1 standard deviation increase in the cognitive factor reduced age at graduation by 1.1 year in the early 80's but its impact was reduced to 0.7 year in the early 2000's. At the opposite, a one standard deviation in the tech-

[10] Recent journalistic evidence (Wall Street Journal, April 8, 2022) documents that many US employers currently favor hiring and promoting individuals with sound "tech" knowledge to those having obtained a college degree.

nical factor increased age at graduation by half a year in 79 but its impact increased over this 20 year period to reach 1.6 years.

We find that omitting the technical factor leads to a serious exaggeration of the role of cognitive abilities. This is likely explained by the existence of individuals who have a reasonable level of cognitive ability but low technical ability and who have no other option than graduating from college. This omission reinforces the effect of measured cognitive abilities but also induces an exaggeration of the college graduation income effect in some cases.

In the next section, we present some of the main features of the NLSY data that we use in our analysis. In Section 3, we present both the latent factor estimation method and the main behavioral model. Section 4 is devoted to identification and estimation issues. In Section 5, we briefly discuss model estimates while Section 6 documents our main findings regarding inequality in post-high school trajectories. The effect of individual trajectories on graduation outcomes is found in Section 7. Finally, in Section 8, we summarize the main results obtained from the reduced-form mappings from family income and abilities onto the main educational outcomes analyzed in the literature. The conclusions are found in Section 9.

2 The NLSY79 and NLSY97 Cohorts

Our analysis is based on data from two cohorts of the National Longitudinal Survey of Youth, NLSY79 and NLSY97. The NLSY is one of the most common data set used in microeconometrics so we limit this section only to details that are needed in order to comprehend the results.

2.1 Cognitive and Technical Ability Measurements

One specificity of the NLSY is the availability of Armed Services Vocational Aptitude Battery (ASVAB) test scores for both cohorts. To measure the cognitive factor, we select 6 components; namely Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Mathematics Knowledge (MK), Coding Speed (CS) and Numerical Operation (NO).

The ASVAB also contains information about mechanical and electronics skills. These measures have been overlooked by most economists who use the

NLSY because of the usual focus on standard cognitive measures. However, and as pointed out in the literature concerned with characterizing cognitive abilities in a multi-dimensional setting, the capacity to comprehend various physical mechanisms may be highly valued in the labor market.[11]

There are 3 measures available in the NLSY; Electronics Information (EI), Automotive and Shop Information (AI) and Mechanical Comprehension (MC). While these measures are often interpreted as measures of blue collar mechanical (and manual) abilities, they also incorporate a strong cognitive dimension. Indeed, as shown in Prada and Urzua (2017), the measures are highly positively correlated with standard cognitive measures such as scientific knowledge and mathematical skills.

As is well known in the literature, ASVAB scores are not comparable across cohorts as the 97 cohort wrote a computer-based exam which differed from the pen-pencil exam administered in the 79 cohort.[12]While efforts have been invested in making AFQT scores comparable, Altonji, Baradwaj and Lange (2012) do not perform a formal factor decomposition. One option is to use their corrected ASVAB measures and assume that the factors have a stationary distribution across cohort. However, stationarity would be difficult to maintain with our objective to decompose the factors into a portion that depends on individual and family characteristics measured at 16. For instance, the distribution of many of those regressors such as mother's education and the fraction of individuals raised with their nuclear family have also changed. This suggest that the mapping from characteristics onto factors may have changed as well and that forcing the relationship to be common across cohort may be erroneous. So instead of estimating stationary factors, we estimate them separately for each cohort and thereby allow the correlation between factors and characteristics to vary across cohorts. More details are provided below.

[11]The interest of industrial psychologists in characterizing traits and abilities fostering success in different occupations dates back to the early 20th century and is surveyed in Blauvelt (2006). Prada and Urzua (2017) use the measures available in the 79 cohort and document their importance for labor market outcomes.

[12]Altonji, Baradwaj and Lange (2012) compute AFQT scores that are meant to be comparable to the 1979 measures using an experiment in which a random set of participants wrote the 79 version and another one the 97 version. However, MaCurdy and Vytlacil (2003) argue that the 97 cohort cannot be assumed to be equivalent to the 79 cohort.

2.2 Sample Selection

Our study is based on the representative samples and does not incorporate over-samples. In order to document both employment and enrollment histories as early as age 16, we need to impose strict age conditions at survey time and thereby focus on a set of younger individuals. Our cohorts are characterized as follows.

- For the 1979 cohort: we keep individuals born between September 1961 and 1964
- For the 1997 cohort: we keep individuals born between 1981 and 1984

Aside from those conditions, we need to exclude those with missing information on included observed characteristics such as family income, ASVAB scores, mother's education, family stability (whether the individual report having been raised within the biological family or not), number of siblings, area of residence (urban vs. rural), age at survey time, gender and ethnic background. Finally, we do not impose any restrictions on minimum number of years in the panel.

2.3 School Enrollment, Work Histories and Family Income

The most demanding step of data collection is the assignment of a school enrollment status and an average number of hours worked (on a per-week basis) for each semester. Potential school enrollment takes place over 2 semesters in each year; fall (September to December) and Spring (January to May). Our definitions of enrollment follow closely Keane and Wolpin (2001) and Johnson (2015). To be enrolled, an individual must declare attending school for at least 3 months between September and December (in the fall) and 4 months between January and May (spring). Hours of work are obtained by taking the sum of weekly hours over each semester and dividing those by the number of weeks for each semester.

To model the initial condition, we make use of information about the highest grade completed by age 16 as well as age at which the ASVAB's were recorded.

To model graduation outcomes, we follow various NLSY official publications which report that the variables measuring diploma outcomes are likely to be more reliable than the yearly information on grade completed, especially in college. This is why the portion of the model concerned with educational outcomes focuses solely on high school graduation and 4-year college graduation probabilities and ignores grade accumulation on a yearly basis.[13]

As our primary objective is to measure changes in family income-based education inequality, we follow most papers concerned with the role of family income and use income information from the years preceding college enrollment. So, for each individual, we use income at ages 16 and 17 if available, and construct an average income measure. If income is only available for one of the years, the average income is replaced by that income. If no income information is available for these ages, we consider income at earlier ages if available in order to minimize the number of individuals dropped because of missing income. For both cohorts, we express income in year 2000 dollars using the CPI for all urban consumers as in Belzil and Hansen (2020).

2.4 Some Descriptive Statistics

Some of the main characteristics of our samples are found in Table 2.1 (schooling attainments and characteristics) and Table 2.2 (parental income). Our sample data point to a higher proportion of 4-year college graduates in the 1997 cohort and a higher proportion of high school graduates in the 1979 cohort but a comparable proportion of high school drop-outs.[14] The evolution of individual characteristics (mother's education, family stability, number of siblings, racial indicators..etc.) is similar to what is reported in many other papers, so we do not discuss it further.

As documented in Table 2.2, mean family income grew by about 19% percent (from \$51,150 to \$61,089) over the period considered. This corresponds to a growth rate slightly below 1% per year, which matches aggregate mea-

[13]An alternative approach would be to model grade progression year by year. However, there is a relatively high incidence of reversions in reported grade. For instance, there are individuals reporting a grade increment in one year and a lower grade level the following year.

[14]The definition of high school drop-outs is not necessarily coherent across studies as many authors include GED's with High school graduates whereas we treat them as drop-outs.

sures provided by the Bureau of Labor Statistics.[15] Not surprisingly, the median family income grew more slowly as it increased by only 8% over the same period.

As is well known, income dispersion has increased even more. In our sample, the standard deviation of family income increased by 58% going from $30,156 to $47,715. As expected, family income is higher than average among college graduates in both cohorts.

To summarize individual enrollment decisions, it is convenient to examine the distribution of age at high school graduation and 4-year college graduation and compare it across cohorts. These distributions are reported in Table 3 and indicate that for both outcomes, there seems to be higher proportions at the right tails in the 97 cohort compared to the 79 cohort.

First, we note that in both cohorts, 18 is the most common age for high school graduation. However, a much larger portion of the 79 cohort high school graduates (19.6%) actually graduated by 17 than in the 97 cohort (12.3%). At the same time, a higher proportion of individuals graduating from high school did it at 19 (or beyond) in the in the 97 cohort (about 14%) than in the 79 cohort (about 10%).

For college graduation, we find a higher proportion of individuals graduating after age 26 (8.7%) in the recent cohort than in the 79 cohort (6.9%). At the other end of the spectrum, a higher proportion (12.7%) of the 79 cohort graduated at (or before) 21 than in the 97 cohort (5.4%).

Despite those changes, the distributions of semesters of enrollment needed to complete college have not changed substantially. First, and contrary to conventional wisdom, only one third of college graduates obtain their degree after enrolling for 8 semesters (the most common way to do it). This is true in both cohorts. About 45% of the 79 college graduates and 40% of the 97 graduates required 11 or more semesters of enrollments and about 15% in the 79 cohort and 10% in the 97 cohort finished college after having enrolled more than 15 semesters. These numbers indicate clearly that progression toward college graduation is far from being as simple as in standard textbook models.

[15] According to the Bureau of Labor Statistics (variable MEFAINUSA672N), median household income grew by 20.6 percent between 1980 and 2000. Other studies using the NLSY, including Kinsler and Pavan (2011), Lovenheim and Reynolds (2011), Castex and Dechter (2014) and Nielsen (2015) report similar family income growth. Castex and Dexter (2014) report changes in the logarithm of income but their sample data also discloses a growth in real income levels which is comparable to ours.

TABLE 2.1
Educational Attainments and Individual Characteristics in the 1979 and 1997 Cohorts

	1979 Cohort	1997 Cohort
Highest Degree		
High School Drop-outs (inc. GED)	25.0%	24.2%
High School Graduates	50.0%	46.6%
Associate Degree	6.6%	5.8%
4-Year College Graduates	18.4%	23.4%
Characteristics		
Male	50.9%	50.7%
Mother's education	11.7	13.1
Intact Family	74.8%	57.7%
Rural	23.1%	26.5%
Number of Siblings	3.1	2.3
Black	12.0%	13.8%
Hispanics	7.7%	11.2%
Age at Survey	17.3	15.2
Grade at 16	10.0	9.9
Sample Size	2013	2219

Table 2.2
Parental Income in the 1979 and 1997 Cohorts

	1979 Cohort	1997 Cohort
	Parental Income (in $2000)	
Mean	$51,150	$61,089
Median	$47,467	$51,598
St. dev.	$30,156	$47,715
High School Drop-Out	$38,435	$47,127
High School Graduate	$50,555	$58,059
Associate Degree	$56,568	$58,872
College Graduates	$67,800	$82,055

Note: Parental Income is measured in dollars of year 2000 between age 14 and 16 (or 17). It is deflated with the CPI for urban consumers.

Table 2.3
Age at High School and College Graduation

	Frequencies (in %)	
	1979	**1997**
Age at High School Graduation		
17 or less	19.6%	12.3%
18	69.8%	73.3%
19	9.1%	13.0%
20 or more	1.5%	1.4%
Age at 4-year College Graduation		
21 or less	12.7%	5.4%
22	34.9%	40.5%
23	23.4%	25.8%
24	13.2%	12.9%
25	8.9%	6.7%
26 or more	6.9%	8.7%
Semesters of Enrollment		
before College Graduation		
8 (or less)	29.5%	33.7%
9	11.4%	13.1%
10	12.4%	14.0%
11-12	18.6%	18.0%
13-14	13.6%	12.3%
15-16	7.1%	6.0%
17 or more	7.4%	3.1%

3 The Econometric Model

Our model is a reduced-form representation of the intertemporal utilities of choosing a particular school attendance-employment-home time combination in each semester from age 16 to 28 (or until college graduation). It incorporates rich dynamic features enabling us to account for the impact of different types of trajectories on college outcomes such as completion and age at graduation. As we estimate the distribution of latent factors non-parametrically and incorporate more than 800 types, we do not adopt a fully structural approach such as in Keane and Wolpin (1997).

3.1 Definitions and Notation

We assume that the decision process starts at age 16 (the legal school leaving age) and that one decision is exerted every semester. Due to the NLSY data collection methods, it is difficult to distinguish between summer enrollment and early fall enrollment. For this reason, we exclude summer attendance as a choice in the model.

The choice variables are indexed by a period (age) indicator denoted a ($a = 16, 17, ..28$) and a semester indicator ($m = 1, 2$). Each period starts in September so that $m = 1$ denotes the fall semester and $m = 2$ denotes the winter semester.

Ignoring individual subscripts for the moment, the choice variables, denoted d_{amj}, are equal to 1 when option j is chosen in semester m of period a and and 0 if not. The notation is detailed in the following panel:

Notation Panel

	Choice Combinations		Accumulated Periods	
	In School	**Hours worked**	**In HS**	**Post HS**
Choice Index (j)				
$d_{a,m,s0}$	yes	0	$S^{hs}_{a,m,s0}$	$S^{phs}_{a,m,s0}$
$d_{a,m,spt}$	yes	1-20	$S^{hs}_{a,m,spt}$	$S^{phs}_{a,m,spt}$
$d_{a,m,sft}$	yes	21 or more	$S^{hs}_{a,m,sft}$	$S^{phs}_{a,m,sft}$
$d_{a,m,pt}$	no	1-20	$W^{hs}_{a,m,pt}$	$W^{phs}_{a,m,pt}$
$d_{a,m,ft}$	no	21 or more	$W^{hs}_{a,m,ft}$	$W^{phs}_{a,m,ft}$
$d_{a,m,h}$	no	0	$H^{hs}_{a,m}$	$H^{phs}_{a,m}$

For each possible choice, we define a variable that records the number of accumulated periods between 16 and the moment at which the decision for semester m of period a is exerted. Those are found in the last two columns of the Notation Panel. To capture more complex forms of dynamics in the model, we distinguish between time allocation decisions undertaken in high school and after high school graduation. The superscript hs refers to choices between 16 and high school graduation while the phs subscript denotes choices after high school graduation. For instance, $S^{hs}_{a,m,sft}$ denotes the number of accumulated semesters of enrollment with full-time work before high school graduation, by semester m at age a.

In parallel to the sequence of individual enrollment and work decisions, we model two potential graduation outcomes; high school graduation and 4-year college graduation. The transitions are assumed to take place in the summer and for this reason, we do not need to use any semester subscript when modeling graduation. In both cases, graduation arising at the end of year a requires attendance for at least one term during year a.

Let's denote age at high school graduation by a^{hsg} and introduce two different vectors representing past choices: one characterizing time allocation during high school (up to a^{hsg} if one graduated from high school or until 21 for someone who has not graduated) and one characterizing decisions after high school graduation (if the person graduated from high school). Those vectors are denoted TR^{hs} and TR^{phs}, respectively. The TR^{hs} and TR^{phs} vectors are as follows:

$$TR^{hs}_{a,m} = \{S^{hs}_{a,m,0}, S^{hs}_{a,m,pt}, S^{hs}_{a,m,ft}, W^{hs}_{a,m,pt}, W^{hs}_{a,m,ft}, H^{hs}_{a,m}\} \text{ for } a < a^{hsg}$$

$$TR^{phs}_{a,m} = \{S^{phs}_{a,m,0}, S^{phs}_{a,m,pt}, S^{phs}_{a,m,ft}, W^{phs}_{a,m,pt}, W^{phs}_{a,m,ft}, H^{phs}_{a,m}\} \text{ for } a \geq a^{hsg}$$

where each element of TR^{phs} records accumulated periods in each state from high school graduation (a^{hsg}) until semester a, m.

3.2 Modeling Utilities and Choice Probabilities

The utilities are represented by linear (in the parameters) functions which depend on family income, family characteristics, two latent factors (cognitive and technical-mechanical abilities) and past choice histories. This sort of reduced-form approach, which is capable of approximating a wide variety of dynamic structures, has been used by Cameron and Heckman (2001) and more recently by Ashworth et al. (2021).

Utilities in High School

The intertemporal utility of choosing option j for individual i in high school entering semester m at period a ($U^{hs}_{a,m,j,i}$) is written as follows:

$$U^{hs}_{a,m,j,i,} = \bar{U}^{hs}_{a,m,j,i} + \varepsilon^{hs}_{a,m,j,i}$$

where $\bar{U}^{hs}_{a,m,j,i}$ is the deterministic part and is a linear (in the parameters) function that depends on the determinants appearing on the right-hand side of the following equation (inside the brackets):

$$\bar{U}^{hs}_{a,m,j,i} = \bar{U}^{hs}\{u^{hs}_{j,i}, X_i, I_i, I_i^2, TR^{hs}_{a,m}\} \text{ for } a < a^{hsg}$$

and where $\varepsilon^{hs}_{a,m,j,i}$ represents a stochastic idiosyncratic shock independent across a, m, j and i, and is assumed to follow an Extreme Value distribution with scale normalized to 1.

The term $u^{hs}_{j,i}$ represents individual-state specific heterogeneity affecting choices. It incorporates two latent factors: one measuring cognitive ability and one measuring technical and mechanical ability but also additional unobserved heterogeneity. We provide more details below. The variable I_i denotes family income and the vector X_i contains individual and family characteristics measured at age 16 and which are described in Section 2. The utilities are not defined beyond age 21, since we do not model subsequent trajectories for high school drop-outs.

Post-High School Utilities

The post-high school utilities are also written as the sum of a deterministic part, $\bar{U}^{phs}_{a,m,j,i}$, and a random shock. As for utilities during high school, $\bar{U}^{phs}_{a,m,j,i}$ is a linear (in the parameter) function. It has completely separate parameters from its high school counterpart. That is

$$U^{phs}_{a,m,j,i} = \bar{U}^{phs}_{a,m,j,i} + \varepsilon^{phs}_{a,m,j,i}$$

$$\bar{U}^{phs}_{a,m,j,i} = \bar{U}^{phs}\{u^{phs}_{j,i}, X_i, I_i, I_i^2, a^{hsg}, I(age = a), TR^{phs}_{a,m}\} \text{ for } a \geq a^{hsg}$$

where $u^{phs}_{a,m,j,i}$ is a heterogeneity term also detailed below, $\varepsilon^{phs}_{a,m,j,i}$ is an Extreme Value random shock with scale normalized to 1 and $I(age = a)$ is an indicator generating a set of binary variables for each age. Note that the post-high school utilities depend indirectly on $TR^{hs}_{a,m}$ evaluated at a^{hsg} since realized choices during high school trigger potential variations in age at high school graduation. The introduction of an aging effect is needed because our model is not fully structural and cannot account for non-stationarity induced by a finite horizon setting.

After conditioning on unobserved heterogeneity, the choice probabilities have the usual Multinomial Logit functional form; that is

$$\begin{aligned}
\Pr(d_{a,m,j,i} &= 1) = \frac{\exp(\bar{U}^{hs}_{a,m,j,i})}{\sum_{k=1}^{6} \exp(\bar{U}^{hs}_{a,m,k,i})} \text{ for } a < a^{hsg} \\
\Pr(d_{a,m,j,i} &= 1) = \frac{\exp(\bar{U}^{phs}_{a,m,j,i})}{\sum_{k=1}^{6} \exp(\bar{U}^{phs}_{a,m,k,i})} \text{ for } a \geq a^{hsg}
\end{aligned}$$

The Graduation Stochastic Process

Define G^{hs}_a and G^{co}_a as the high school and 4-year college graduation indicators when entering semester 1 (beginning of September) of period a and denote age at college graduation by a^{cg}. We also denote the total number of post-high school enrollment periods by $S^{phs}_{a,m}$. Specifically,

$$\begin{aligned}
G^{hs}_a &= 1 \text{ if High School graduate when entering } a \text{ and } 0 \text{ if not} \\
G^{co}_a &= 1 \text{ if College graduate when entering } a \text{ and } 0 \text{ if not} \\
S^{phs}_{a,m} &= S^{phs}_{a,m,0} + S^{phs}_{a,m,pt} + S^{phs}_{a,m,ft}
\end{aligned}$$

For both high school and college, the graduation probability (a discrete time hazard rate) depends on observed and unobserved heterogeneity, and on various trajectory indicators defined below. Given that individuals start graduating from college at least 3 years after high school graduation, we must define the risk set (the individuals that have a non 0 probability of graduating) accordingly. For college graduation, we impose a 0 graduation probability until a minimal enrollment level of 6 semesters is realized. For high school, we impose that in order to be in the risk set, one must have been enrolled for at least one semester over the academic year.

Formally, we have the following

$$\begin{aligned} \Pr(G_{a+1}^{hs} &= 1 \mid (G_a^{hs} = 0) = PHSG_{a,i} \text{ if at risk} \\ \Pr(G_{a+1}^{hs} &= 1 \mid G_a^{hs} = 0) = 0 \text{ if not at risk} \end{aligned}$$

and we parameterize $PHSG$ as a logistic probability which depends on the same determinants as the choice probabilities $(X_i, I_i, I_i^2, TR_{a,m}^{hs}, I(age = a))$ and an heterogeneity term, h_{gi}, discussed below.

The transition probability to college graduation, PCG_i, is also logistic and is defined similarly:

$$\begin{aligned} \Pr(G_{a+1}^{co} &= 1 \mid G_a^{co} = 0) = PCG_{a,i} \text{ if at risk} \\ \Pr(G_{a+1}^{co} &= 1 \mid G_a^{co} = 0) = 0) \text{ if not at risk} \end{aligned}$$

The index defining PCG is also a linear (in the parameters) function depending on $X_i, I_i, I_i^2, a^{hsg}, TR_{a,m}^{phs}$, $S_{a,m}^{phs}$, $(S_{a,m}^{phs})^2$ and a heterogeneity term cg_i. The dependence on total elapsed enrollment semesters takes into account potential forms of duration dependence. For instance, we cannot rule out that abnormally long enrollment periods may induce drop-out behavior.

4 Identification and Estimation

We now discuss some of the most important issues pertaining to the identification and estimation of the model.

4.1 Unobserved Heterogeneity and Latent Factors

The different intercept terms of the utilities of choices ($\bar{U}^{hs}_{a,m,j,i}$ and $\bar{U}^{phs}_{a,m,j,i}$) and graduation probabilities ($PHSG_{a,i}$ and $PCG_{a,i}$) are defined as follows:

$$
\begin{aligned}
u^{hs}_{j,i} &= u^{hs}_{0ji} + u^{hs}_{cj} \cdot C_i + \ u^{hs}_{tj} \cdot T_i \\
hg_i &= hg_{0i} + hg_c \cdot C_i + \ hg_t \cdot T_i \\
u^{phs}_{j,i} &= u^{phs}_{0ji} + u^{phs}_{cj} \cdot C_i + \ u^{phs}_{tj} \cdot T_i \\
cg_i &= cg_{0i} + cg_c \cdot C_i + \ cg_t \cdot T_i
\end{aligned}
$$

where C denotes the cognitive factor, T is the latent factor measuring technical-mechanical skill, and $u^{hs}_{0ji}, u^{phs}_{0ji}, hg_{0i}$ and cg_{0i} play the role of residual unobserved heterogeneity affecting time allocation decisions and graduation outcomes. The parameters u^{hs}_{cj}, u^{hs}_{tj}, u^{phs}_{cj} and u^{phs}_{tj}capture the effect of cognitive and technical skills on the intertemporal utilities of each possible choice. The parameters hg_c, hg_t, cg_c and cg_t play a similar role for the graduation probabilities.

We write both factors as the sum of a component assumed to depend on a vector of observed family and individual characteristics (denoted X_i) and family income (denoted I_i) at age 16, and an orthogonal component denoted $\tilde{C}_i$ and $\tilde{T}_i$:

$$C_i = C_x \cdot X_i + C_I \cdot I_i + \tilde{C}_i$$

$$T_i = T_x \cdot X_i + T_I \cdot I_i + \tilde{T}_i$$

where $\{C_x, C_I\}$ and $\{T_X, T_I\}$ are vectors of parameters measuring the correlation between the factors and observed regressors.

Denote the jth cognitive measure of individual i by $CM_{i,j}$ and the qth technical-mechanical measure of individual i by $TM_{i,q}$ and assume a recursive measurement system.[16] Specifically, we obtain the following equations:

$$
\begin{aligned}
CM_{i,j} &= cm_{0j} + cm_{cj} \cdot C_i + \varepsilon^{cm}_{ij} \\
TM_{i,q} &= tm_{0q} + tm_{t,q} \cdot T_i + tm_{c,q} \cdot C_i + \ \varepsilon^{tm}_{i,q}
\end{aligned}
$$

[16]Prada and Urzua (2017) also make use of various non-cognitive measures such as the Rosenberg self-esteem measures and the Rotter Index measures to define a socio-emotional factor. However, as those are only measured in early ages in the 79 cohort, we cannot make use of them. However, we allow for additional unobserved heterogeneity.

where $j = 1, 2, ..6$, $q = 1, 2, 3$.

The parameters cm_{0j} and tm_{0q} are intercept terms affecting the location of each measure, cm_{cj}, $tm_{t,q}$ and $tm_{c,q}$ are loading parameters and $\varepsilon_{ij}^{cm}, \varepsilon_{i,q}^{tm}$ are measurement error shocks which follow a Normal distribution with mean 0 and standard deviation σ_j^{cm} and σ_q^{tm} respectively.

Test achievements available in the NLSY have been obtained within a practically incentive-free environment and are therefore likely to provide only partial information about the relevant level of skills driving actual choices. This stresses the need for considering additional unobserved heterogeneity affecting actual choices and graduation outcomes. To some extent, our additional heterogeneity terms may be interpreted as representing non-cognitive skills.

4.2 Identification

To estimate the model, various restrictions (or normalizations) need to be imposed. First, we assume a recursive factor system and more precisely that technical measurements are affected by both the cognitive and the technical factors but that cognitive measurements depend only on the cognitive factor. We thereby obtain a clear interpretation of the factors. For the Arithmetic Reasoning measure, we set the loading parameter associated to the cognitive factor to 1, and the intercept to 0. For the Mechanical Comprehension score, we set the loading associated to the technical factor to 1 and the intercept to 0.

When estimating our model, we use a discrete approximation of the joint factor distribution. In line with Bajari, Fox and Ryan (2007) and Train (2008), we adopt a fixed mass points approach by choosing grid points covering the entire range of possible values and estimate all associated frequencies (type probabilities).

To proceed, we normalize the measurements so to help us set up support points for the distribution of each factor. We then assume that the orthogonal part of each factor can take one of 29 values between -4 and 4 and which are equidistant by 0.2 between -2.0 and 2.0 { -2.0, -1.8,.... -0.2, 0.0, 0.2,...,1.8, 2.0} and by 0.5 at both extremes (-4.0, -3.5, -3.0, -2.5) and (2.5, 3.0, 3.5,4.0). This approach generates 841 different combinations (or types). Our task is to estimate the probabilities of all possible combinations. The probability of a given realization is denoted p_r ($r = 1, 2, ...841$), and each p_r is estimated as a Logistic transform.

Finally, identification of the distribution of the additional heterogeneity vector $\{u^{hs}_{0j,i}, u^{phs}_{0j,i}, hg_{0i}, cg_{0i}\}$ is easy to achieve given the large number of observed choices between 16 and 28. To avoid confusion with the index used for combinations of cognitive and non-cognitive skills (type r), we use the letter u to identify different types with respect to residual unobserved heterogeneity. We assume that the distribution of residual unobserved heterogeneity is orthogonal to the factors and set the number of types to 4 in order to reduce the number of parameters as each type has 12 intercept terms (5 for high school utilities, 5 for post-high school utilities, 1 for high school graduation and 1 for college graduation).

4.3 Estimation

As indicated earlier, our model is estimated separately for each cohort. From now on, we use the index k to designate a specific cohort. The total likelihood function has 3 different parts which we now describe but the estimation is achieved in two steps. In the first step, we estimate the distribution of the cognitive and technical factors. For a given individual i belonging to cohort k, and with a set of measurements $\{CM^k_{i,j}, TM^k_{iq}\}$ taking modalities $\{cm_{i,j}, tm_{iq}\}$, the contribution to the likelihood of the measurements, denoted $L^{M,k}_i(type\ r)$, is equal to

$$\begin{aligned} L^{M,k}_i(type\ r) &= \{\prod_{j=1}^{6} \Pr(CM^k_{ij} = cm_{ij} \mid X_i, I_i, \tilde{C}_r, \tilde{T}_r) \cdot \\ \prod_{q=1}^{3} \Pr(TM_{iq} &= tm_{iq} \mid X_i, I_i, \tilde{C}_r, \tilde{T}_r)\} \end{aligned}$$

This part of the likelihood contains 18,837 contributions in cohort 79 and 19,971 in the 97 cohort. From those contributions, we obtain estimates of all parameters of the measurement equations and all type probabilities by maximizing the unconditional log likelihood given by the following expression:

$$L^{M,k}_i = \sum_{m=1}^{841} p_r \cdot L^{M,k}_i(type\ r)$$

The second part of the likelihood is concerned with individual time allocation choices and it is denoted $L_i^{CH,k}(type\ r, type\ u)$. For a given individual i in cohort k, the contribution to the likelihood of observed enrollment and labor supply choices given type (r, u) is equal to

$$L_i^{CH,k}(type\ (r,u)) = \prod_{a=1}^{A_i} \prod_{m=1}^{2} \{\Pr(d_{a,m,j^*,i} = 1 \mid type\ r, u)^{I(d_{a,m,j^*})}\}$$

where j^* denotes the optimal choice of individual i and $I(.)$ denotes the identity function. The individual subscript for the last period where choice is exerted, A, is explained by the fact that we do not model choices beyond college graduation or age 21 if not finishing high school.

Finally, the 3rd component of the likelihood is the distribution of each individual's graduation transition history. To define it, we use the following indicators needed to identify ages at which individuals are at risk of graduating and at which a graduation transition is observed. First, in order to define the risk sets, we have $I_{phsg}(a) = 1$ if individual is a potential high school graduate at a (is in the risk set) and 0 if not, and $I_{pcg}(a) = 1$ if individual is a potential college graduate at a and 0 if not.

To complete, we need actual graduation indicators and define $I_{hsg}(a)$ and $I_{cg}(a)$ such that $I_{hsg}(a) = 1$ when an individual has graduate from high school between a and $a + 1$ and 0 if not and $I_{cg}(a) = 1$ if someone has graduate from college between a and $a + 1$and 0 if not.

Before formulating the probability of a graduation transition at age a, denoted $PG(a)$, we need an indicator denoted $I_{pg}(a) = I_{phsg}(a) + I_{pcg}(a) + I_{hsg}(a)+I_{cg}(a)$, which equals 1 whenever an individual at age a is in a position to graduate or actually graduates and 0 if not. For those at risk, we have

$$\begin{gathered} PG_i(a; type\ (r,u)) = \\ \{(1 - PHG_i(a; type\ (r,u)))^{1-I_{hsg}(a)} \cdot PHG_i(a; type\ (r,u))^{I_{hsg}(a)}\}^{I_{phsg}(a)} \cdot \\ \{(1 - PCG_i(a; type\ (r,u)))^{(1-I_{cg}(a))} \cdot PCG_i(a; type(r,u))^{I_{cg}(a)}\}^{I_{pcg}(a)} \end{gathered}$$

For those who are not at risk at a, we impose that $PG_i() \equiv 0$ and model choices only.

The likelihood of graduation from high school and college in cohort k, is equal to

$$L_i^{G,k}(type\ (r,\ u)) = \{\prod_{a=1}^{A_i} (PG_i(a) \mid type\ (r,\ u))^{I_{pgi}(a)}\}$$

where the indicator, I_{pg}, guarantees that only outcomes taking place when an individual is at risk, contribute to the likelihood.

Taken together, the portion of the likelihood of choices and graduation contains 23,259 contributions in the 79 cohort and 20,913 in the 97 cohort. Recalling that we first estimated the distribution of cognitive and technical factors and obtained $\hat{p}_r$ and $\hat{L}_i^{M,k}(type\ r)$ in the first stage, we then maximize the full likelihood conditional on $\hat{p}_r$ and $\hat{L}_i^{M,k}$. Assuming a form of conditional independence, namely that choices and graduation outcomes are independent after conditioning on unobserved heterogeneity, the likelihood function of individual i, denoted L_i^k, is therefore equal to

$$L_i^k(.) = \sum_{u=1}^{4} p_u \cdot \left\{ \sum_{m=1}^{841} \hat{p}_r \cdot \{ \hat{L}_i^{M,k}(type\ r) \cdot L_i^{CH,k}(type\ (r,u)) \cdot L_i^{G,k}(type\ (r,u)) \right\}$$

Parameters are obtained by maximizing the criterion function obtained after taking the log of the product of each L_i.

5 Estimates and Model Fit

Before answering the questions that motivate our research, we summarize some features of the heterogeneity distributions and document the capacity of our model to fit the data.

5.1 The Distribution of Cognitive and Technical Abilities

As a first step, we estimated the parameters of the measurement equations and the distribution of factors from a total of 38,808 likelihood contributions: 18,117 in cohort 79 and 19,971 in the 97 cohort. The parameters of the measurement equations are found in the supplementary file but the estimates of the mapping of individual and family background variables onto the cognitive and the technical factors, which have been standardized in order to obtain comparable marginal effects, are found in Table 5.1.

The distributions of the cognitive and the technical-mechanical factors are plotted in Figure 1 and Figure 2. The cognitive factor is more dispersed than the technical factor in both cohorts. For instance, it has positive densities as far as -1.8 and 1.6 and it discloses the features of a multi-modal distribution. The technical factor is much less dispersed as a much larger share is found around 0.

The variance decompositions found at the bottom of Table 5.1 indicate that both the cognitive and the technical factors tend to be strongly correlated with background variables but that both factors have become less dependent on background variables. The totality of the regressors account for 44% of the total cognitive factor variance in the 79 cohort and 36% in the 97 cohort. For the technical factor, the regressors account for 55% in 79 and 35% in 97. It is interesting to note that despite a strong positive correlation between cognitive and technical measurements (documented in the supplementary file), the correlation between the cognitive and the technical-mechanical factors is negative in both cohorts (-0.17 in the 79 cohort and -0.36 in the 97 cohort).

As should be obvious upon examination of Table 5.1, the impact of many regressors changed across cohorts. This validates our approach to avoid estimating a unique (stationary) distribution across cohorts. Within both cohorts, we find that females dominate males with respect to the cognitive factor (although the male-female differences equal to -0.06 and -0.02 are insignificant) but males have a significant advantage in technical-mechanical abilities over females. The difference, which was about 1.4 standard deviation in 79 and about 1.0 in 97, is substantial and is likely to affect gender differences in employment and earnings.[17]

Consistent with the investment deficit in early childhood investments often reported in the literature, Blacks and Hispanics have a lower average level of cognitive and technical-mechanical abilities than Whites although the disadvantage in the technical factor is not as pronounced. Not surprisingly, both the cognitive and the technical factors are positively correlated with the highest grade level completed at age 16.

The most striking results are those pertaining to the income effects. The parameter estimate of the effect of income on cognitive ability has been divided almost by 5 from the early 80's to the early 2000's, going from 0.14 to 0.03. This implies that a $30,000 difference in family income generated a difference of 0.14 standard deviation in the early 80's but only a 0.03 standard deviation increase in the early 2000's. This substantial drop is likely to impact our estimates of the evolution of income effects.

Another surprising result is the negative impact of income on the technical-mechanical factor although the effect has been relatively small in both cohorts (a decrease of 0.06 standard deviation). This negative impact most likely indicates that specialization already takes place before 16 and that those coming from low income families tend to favor activities fostering technical and mechanical skills more than do those coming from advantaged backgrounds. Put differently, this finding implies an increasing gap between cognitive and technical abilities as family income increases in the 79 cohort but a weaker relationship in the 97 cohort.

An interesting feature of the projection of the factors onto individual and family characteristics is the difference in the evolution of the impact of the family stability indicator and family income. The effect of being raised with the biological parents (the variable "Intact Family") on the cognitive factor has been multiplied by about 2, going from 0.08 standard deviation to 0.17.

[17] As far as we know, the contribution of the female deficit in the technical-mechanical factor to the total gender wage gap has never been documented formally.

5.2 Model Estimates and Goodness of Fit

The model of individual choices and outcomes (after including unobserved heterogeneity) requires the estimation of 341 parameters for each cohort. As for the factor distribution parameters, those are available in a supplementary file. One efficient way to assess the capacity of the model to fit data is to compare the predicted number of semesters in each state beyond high school graduation with observed frequencies. As is evident upon looking at Table 5.2, our model captures the allocation of time over the post high school period. Our model is also capable of predicting graduation outcomes accurately as indicated by the entries found in Table 5.3.

Table 5.1
Decomposing the Cognitive and Technical Factors
into Individual and Family Background Variables

	Factor			
	Cognitive		**Technical**	
	1979	**1997**	**1979**	**1997**
Characteristics				
Male	-0.0628	-0.0167	1.4234**	0.9610**
Black	-0.8553**	-0.6086**	-0.2607**	-0.3644**
Hispanic	-0.2365**	-0.3536**	-0.3174**	-0.1571**
Intact Family	0.0816**	0.1907**	0.0200**	-0.0905**
Income (30k)	0.1391**	0.0339**	-0.0601**	-0.0669**
Mother's educ.	0.0856**	0.0929**	-0.0181**	-0.0494**
South	-0.1144**	-0.0366**	-0.0586**	-0.0348**
Rural	0.0395*	-0.0705**	0.1308**	0.3011**
# of Siblings	-0.0435**	-0.0090	0.0125**	-0.0153*
Education at 16	0.3672**	0.4112**	-0.1385**	-0.1833**
Age at Test	0.0556**	0.2239**	0.0796**	0.0438**
% explained	44%	36%	55%	35%
Correlations	-0.17	-0.36	-	-

Note: Parameters measure the marginal effects of background variables on standardized cognitive and technical factors.

Note: Estimates with (**) are significant at 1% level. Estimates with (*) are significant at 5% level.

Table 5.2
Model Fit: Predicted and Actual number of Semesters
beyond High School Graduation

	Choices		**Accumulated Periods**			
	In School	**Hours worked**	Predicted		Data	
			1979	1997	1979	1997
$S_{a,m,s0}$	yes	0	1	1	1	1
$S_{a,m,spt}$	yes	1-20	1	1	2	1
$S_{a,m,sft}$	yes	21 or more	2	1	2	2
$W_{a,m,spt}$	no	1-20	1	1	1	1
$W_{a,m,,sft}$	no	21 or more	11	11	12	12
$H_{a,m}$	no	0	2	2	2	2

Note: Number of semesters are rounded to the nearest integer.

Table 5.3
Model Fit:
Predicted and Actual Graduation Outcomes

	Graduation Outcomes			
	Predicted	Data	Predicted	Data
	1979	1979	1997	1997
High School	70.4%	75.0%	78.0%	76.8%
4-year College	16.0%	18.0%	26.0%	23.4%

6 Inequality in Post-High School Trajectories among College Graduates

In this section, we use the model to detect sources of inequality in various types of trajectories. Given the complexity of the model and especially its non-linearity, we need to simulate it in order to uncover easily interpreted marginal effect parameters.

Our objective is now to map individual attributes onto trajectory characteristics, as defined by the allocation of time between various combinations of school enrollment, labor supply and school interruption. For instance, we evaluate to what extent the decisions to work either part-time or full-time while enrolled in college have become more or less affected by family income and abilities. Similarly, we measure the sensitivity of the incidence of school interruptions and delayed college graduation to differences in family income.

To do this, we simulate a large number of individual trajectories and graduation outcomes (about 50,000) and regress the number of accumulated semesters on the exogenous variables of the system (individual and family characteristics at age 16, latent factors and unobserved type indicators) for both cohorts.

To ease presentation, our discussion focuses mostly on 3 of the individual specific characteristics; family income and the cognitive and technical-mechanical factors. On top of this, and in order to provide a comparison basis with characteristics often used in empirical work, we also report the effects of mother's education and the binary indicator for nuclear family. To assess their relative importance, we scale factors and mother's education according to the standard deviations obtained for the 79 cohort. For family income, we choose units of $30,000 as it is very close to the income standard deviation in the 1979 cohort while for mother's education, the unit is 2.6 years of education which is also the standard deviation in the 1979 cohort. While the model incorporates flexibility by introducing income squared in the choice utilities, we restrict the reduced-form regressions to linear income effect so to allow a more direct comparison across cohorts.

One major challenge in evaluating the marginal effects of individual characteristics is to ensure comparability of the time horizon over which trajectory characteristics are measured so that the total number of accumulated semesters in a given state can be compared across individuals. To render the exercise meaningful, it is first necessary to condition on age at high school graduation. We thereby take the set of individuals predicted to graduate at 18 (the most common age at high school graduation in the US) and use a period of 4 years (8 semesters) following high school graduation to characterize individual trajectories. This 4-year period is motivated by the fact that 22 is by far the most common age at college graduation for those graduating from high school at 18.

Another challenge is the separation of individuals who may intend to graduate from college and those who have no intention to do so. While a clear separation cannot be achieved in observational data unless a purely structural model in which individual beliefs are fully specified is estimated, we perform our analysis on the subset of individuals predicted to graduate by age 28.

We raise 4 specific questions. First, we ask which characteristics make individuals more or less likely to work while enrolled in college over the 4 year period after high school graduation? To do this, we compute 3 different measures of the propensity to work while enrolled. Those are the frequencies of i) semesters of enrollment with no work, and ii) semesters with full-time work in reference to the total number of semesters of enrollment over the 4-year period. The third measure is the average number of weekly hours of work while enrolled using the modal number of hours in each labor supply class.

The second question is what makes individual more or less likely to delay enrollment and therefore graduation. To answer this, we construct the frequencies of enrollment periods realized over the first 8 semesters. The third and fourth questions, which are somewhat related, are about what makes an individual more likely to interrupt school i) in order to work (either full-time or part-time) and (ii) in order to involve in household activities. Using simulated employment periods over the 4-year period, we construct similar frequencies and regress them on characteristics and factors.

6.1 The Importance of Family Income

Upon examination of the results found in Table 6.1, one notes that real income increased the proportion of enrollment periods with no work in 79 (column 1) but also that this effect vanished in 97 (column 2). Income increased the proportion of individuals working full-time while enrolled in 79 by 0.03 (column 3) but reduced it in 97. To obtain a clearer picture, it is informative to examine the overall effects of income on hours worked while in school, which are reported in columns 5 and 6. With a $30,000 difference leading to a difference of less than half an hour per week (0.38 hour in 79 and -0.26 hour in 97), one can conclude against the existence of any significant family income-based inequality in employment while in school.

For all three other trajectory characteristics (found in the lower panel of Table 6.1), the impact of income are even smaller thereby obviating the need for commenting on their evolution. We therefore conclude that in both cohorts, family income had neither an impact on the propensity to enroll earlier (within the 8 semesters following high school graduation), nor on the incidence of school interruptions devoted to work.

6.2 The Importance of Cognitive and Technical Abilities

Unlike family income, cognitive and technical abilities constitute important determinants of individual trajectories to reach college graduation. A first finding is the dominance of the technical-mechanical factor.[18] A one standard deviation increase in the technical factor raises weekly employment by 6 hours in both cohorts. On the other hand, a one standard deviation increase in the cognitive factor reduced employment by 2 hours in 79 and 1 hour in 97.

However, the most striking finding is the conflict between cognitive and technical skills as their impact are practically always of opposite signs. As documented in the lower panel of Table 6.1, those with a high level of technical abilities tend to be more likely to interrupt school to work in both cohorts and less likely to enroll early (or more likely to delay enrollment). For those with a high level of cognitive abilities, this is the exact opposite. In other words, high cognitive ability individuals work less during school and enroll earlier, while high technical-mechanical ability individuals work much more during school and enroll later.

Overall, our findings therefore suggest that inequality in trajectories of college graduates is ability-based and not family income-based.

[18]For most regressions, we also ran a specification without unobserved type indicators in order to evaluate the importance of additional unobserved heterogeneity. Overall, it is found to be more important for choices than for graduation outcomes. For instance, for some choices such as hometime and periods of work with no enrollment, unobserved heterogeneity accounts for more than the cognitive and the technical factors combined. The regressions are found in the supplmentary file.

Table 6.1
Labor Supply while in School, School Interruptions to Work
and Early Enrollments

	Dependent Variables					
	Enroll-No work		Enroll-Work FT		Hours per week	
Cohort	1979	1997	1979	1997	1979	1997
Income (30k)	0.0201	0.0009	0.0292**	-0.0124**	0.3826**	-0.2574**
Cognitive	0.0168	-0.0007	-0.0899**	-0.0547**	-1.9656**	-1.0262**
Technical	-0.2060**	-0.1498**	0.2144**	0.2235**	6.3493**	5.9778**
Mother's Educ.	0.0032	0.0264**	0.0082	-0.0484**	0.1320	-1.2329**
Intact Family	-0.0358	0.0498**	-0.0165	-0.0786**	0.0279	-0.0786**
R square	0.14	0.47	0.30	0.47	0.27	0.48

	Dependent Variables					
	Home Time		Work PT or FT		Enrollment	
	1979	1997	1979	1997	1979	1997
Income (30k)	-0.0012	-0.0076**	0.0011	0.0033**	0.0001	0.0043**
Cognitive	-0.0462**	-0.0273**	-0.0765**	-0.0440**	0.1227**	0.0714**
Technical	0.0545**	0.0539**	0.1649**	0.1101**	-0.2195**	-0.1640**
Mother's Educ	0.0014	-0.0065**	0.0104**	-0.0126**	-0.0118**	0.0191**
Intact Family	0.0023	-0.0101**	0.0217**	-0.0317	-0.0240**	0.0417**
R square	0.18	0.42	0.43	0.21	0.47	0.43

Note: The "Enroll-no work" and "enroll-work FT" variables are measured by the number of relevant semesters divided by the total number of semesters of enrollment over the 4-year period beyond high school graduation. All other dependent variables (Home time, Work PT or FT and Enrollment) are measured by the number of relevant semesters divided by 8.

Note: Family Income is measured in \$30,000. Mother's education is divided by 2.6 (the standard deviation in the 79 cohort) and the cognitive and technical factors are standardized.

Note: $**$: significant at 1% level, $*$: significant at 5% level

7 The Effect of Individual Trajectories on College Graduation and Age at Graduation

We now investigate the effect of trajectories on college graduation and age at college graduation. Measuring the impact of different types of trajectories on college graduation is a natural question to ask, although it has been ignored in the literature. With it, we can gain further insight about the dynamics of college graduation.

For instance, comparing the effect of past enrollment periods during which individuals either work full-time or part-time with the effects of periods of employment and periods of home production (with no school enrollment) is particularly informative as periods of non-enrollment may be more easily coined as school interruption periods if we condition on subsequent college graduation. It is also informative to compare outcomes of those who have worked while enrolled to those who enrolled in school and did not work since this comparison delivers an estimate of the effect of working among a set of individuals more likely to intend to graduate.

For both college graduation and age at graduation, we use the same simulated outcomes used in the previous section (simulated trajectories for the 4 year period following high school graduation at 18) and compute the marginal effects similarly. In order to estimate the causal effect of past trajectories on college graduation, we do not restrict our sample to those who have graduated. We however do so for age at graduation. We record both outcomes (graduation and age at graduation) over the period starting after the 8th semester and lasting until age 28. We then measure the impact of accumulated enrollment periods with full-time or part-time work on both outcomes using OLS regressions where we exclude one of the choices because the sum over all options equals 8 semesters. To simplify the presentation, we report the effects of individual trajectories only but the specification incorporates all exogenous regressors and factors and the reduced-form regressions are in the supplementary file.

7.1 College Graduation

The parameters reported in Table 7.1 are obtained after excluding the number of periods of enrollment without working, so that negative effects associated to any given option indicate that, other things equal, an additional period in that state reduces the outcome of interest (the left-hand side variable) relative to school-no work. The results indicate that a semester of enrollment accompanied with part-time work does not harm college graduation as the impacts (-0.0006 for 79 and -0.0080 for 97 per semester) are practically equal to 0. However, working full-time while enrolled appears to be more harmful as the estimates, around -0.06 in both cohorts, imply that a full year of enrollment (2 semesters) with full-time work reduces graduation probabilities by 0.12.

The comparability of graduation probabilities between those enrolled-not working and those enrolled-working part-time may be interpreted as an indication that the education financing function of part-time employment dominates the potential negative impact of working while in school while the stronger negative effect of full-time work while in school may indicate that with a high volume of market work per week, academic effort becomes more diluted and thereby reduces subsequent graduation probability.

Not surprisingly, periods of non-enrollment either devoted to work or to household activities have much stronger negative impacts on graduation as the estimates imply effects between –0.15 and -0.20 for each additional year of employment or household activities.

These estimates must be interpreted with caution as the negative impact of non-enrollment periods (and to some extent the negative impacts of enrollment periods with full-time work) may reflect the absence of any intention to attend college (let alone graduation) among a subset of high school graduates. They may also capture the fact that some individuals are indifferent between enrollment and non-enrollment and decide to diversify their activities. To say more about the impact of working full-time while enrolled and working in the labor market after high school, it is crucial to examine their impact on age at graduation.

7.2 Age at Graduation

For age at graduation, the sort of interpretation problem noted when analyzing a population of high school graduates is avoided as the results found in columns 3 and 4 have been obtained for a sub-population of college graduates. The results are therefore easier to interpret. One similarity with results obtained for college graduation is that we also find that working part-time while in school has no real impact on age at graduation. Indeed, the negative estimates (-0.0330 for 79 and -0.0744 for 97) indicate that those working part-time while enrolled would tend to graduate faster than those who don't, other things equal, after conditioning on all other characteristics and factors.

In the early 80's, working full-time while enrolled for two semesters increased age at graduation by 0.2 year (2 times 0.103) but interestingly, the effect became negative in the early 2000's and points to a small reduction in age at graduation of about 0.1 year (2 times -0.049).

To summarize, and to the extent that conditioning on subsequent graduation outcomes allows us to capture individuals who have college graduation intentions, our estimates indicate that as of the early 2000's, working while in school (including full-time work) did not delay graduation, although working while not in school significantly delays age at graduation. This is generally coherent with the education financing function played by labor supply while in school.

Table 7.1
The Effect of Past Trajectories
on the Probability Graduating from College
and Age at College Graduation

	Dependent Variables			
	College Grad.		**Age at Coll Grad.**	
Trajectories	**1979**	**1997**	**1979**	**1997**
School-no work	-	-	-	-
School-Part-time work	-0.0007	-0.0080*	-0.0330*	-0.0747*
School-Full-Time work	-0.0600*	-0.0512*	0.1035*	-0.0493*
Work Part-time	-0.0946*	-0.1197*	0.0124	0.2848*
Work-Full-time	-0.0804*	-0.1165*	0.2279*	0.2877*
Home Time	-0.0721*	-0.1073*	-0.0101*	0.3109*
R square	0.37	0.47	0.08	0.29

Note: The dependent variables are binary indicators for college graduation in columns 1 and 2, and age at graduation in 3 and 4

Note: all right-hand side variables are measured in semesters

Note: $**$: significant at 1% level, $*$: significant at 5% level

Note: The reduced-form regressions also control for all individual characteristics and factors.

8 The Evolution of the Relationship between Graduation Outcomes, Family Income and Abilities

We now use our model to investigate the evolution of the impact of key individual characteristics on a range of educational outcomes found in the literature. As noted earlier, comparisons across studies are difficult to make but our model provides us with the opportunity to analyze different outcomes within a unique model structure and to compare different parameters in a much more transparent manner.

As a first step, we use the reduced-form relationships induced by our estimates to evaluate changes in the mapping from individual characteristics onto graduation outcomes (both high school and college), enrollment and persistence and age at college graduation between the 1980's and the early 2000's. Enrollment is defined as having enrolled for at least one semester beyond high school graduation, and college persistence is defined as having enrolled for at least 5 semesters beyond high school.

In the second step, we evaluate the sensitivity of both income and cognitive factor effects on graduation outcomes, enrollment and persistence. For each cohort, we compare the estimates obtained with the full specification (presented in Tables 8.1 and 8.2) with those obtained when no factor is included in the reduced-form regression and also when only the cognitive factor is incorporated. Comparing the full specification with the former allows us to evaluate to what extent the total effect of family income on educational outcomes is driven by the correlation between income and abilities before college decisions (at 16) as opposed to net income effects prevailing after differences in abilities have been accounted for. Comparing the full specification with the one controlling only for the cognitive factor allows us to evaluate the impact of treating cognitive skill as a unidimensional object. This exercise is particular meaningful as technical ability is virtually always ignored in the existing literature except in Prada and Urzua (2017). Table 8.3 incorporates both the new estimates as well as those already reported either in Table 8.1 and Table 8.2 so to ease comparison.

8.1 Reduced-Form Mappings

8.1.1 High school Graduation

Usually ignored in the literature, the effect of income on high school graduation implies a 0.036 increase in the frequency of graduation for a \$30,000 difference in family income. In the early 2000's, the same income difference generated a smaller effect (about 0.021). These estimates will be particularly interesting to compare with those obtained for college graduation.

While the cognitive factor is a more important determinant of high school graduation, its marginal effect has almost been divided by 2 over the period covered in our analysis. In the 79 cohort, a one standard deviation increase in cognitive ability raised high school graduation by 0.18 while in the early 2000's the corresponding increase was only 0.11. This means essentially that with respect to high school graduation, a one standard deviation in cognitive skill was worth about 5 times one standard deviation in income in the early 2000's.

Taken in absolute values, the technical-mechanical factor effect was the most important in the early 80's and as important as the cognitive factor effect measured in the early 2000. As noted above when modeling trajectories, the technical factor tends to play in opposite directions than the cognitive factor as it reduces high school graduation. However, as was the case with the cognitive factor, its importance has also been divided by 2 over the period that we consider, as indicated by the movement in the marginal effects going from -0.19 (in 79) to -0.08 (in 97).

It is interesting to note that the variable indicating if one has been raised by biological parents at age 14 has become more important. Although we are already conditioning on cognitive and technical abiities, it may be difficult to interpret its effect as being causal because the frequency of family dissolution has increased over that period and the population composition of those raised within the biological family may therefore have changed. However, and as a comparison, the high school graduation differential induced by a stable family environment is about the same as the effect of one standard deviation in the cognitive factor in the 97 cohort.

8.1.2 Enrollment and Persistence in College

In many studies, the impact of family income on education outcomes is measured from college attendance (regardless of graduation) and is meant to measure the effect of family resources on access to college. The estimates found in columns 3 to 6 of Table 8.1 have been obtained without conditioning on high school graduation to ease comparison with many estimates reported in the literature.

As was the case with high school graduation, the income effects on enrollment have decreased over the 1980-2000 period (going from 0.015 to -0.007) but basically indicate that either in the early 80's or in the early 2000's, income had no significant impact on the probability of enrolling in college. The cognitive and technical factors, which are about 10 times as important as income, have again opposite signs. It is important to note that the effect of the cognitive factor has been divided by 2 (going from 0.103 to 0.047) while the strong negative effect of the technical factor measured in the 79 cohort (-0.096) has practically disappeared in the early 2000's (to reach -0.007).

These results may be reconciled with the rapid expansion of 2 year colleges and lower-quality 4 year colleges (Hoxby, 2009, Babcock and Marks, 2011) and also by the very low real cost of 2-year colleges (Dynarski and Scott-Clayton, 2013).

On the other hand, it is interesting to note that the importance of being raised in a stable family (even after controlling for cognitive and technical abilities) has increased to reach practically the same level of impact as a one standard-deviation in the cognitive ability factor (0.042).

As our measure of college persistence is defined as the probability of having attended (enrolled in) college for a minimum of 5 semesters, it is less likely to be affected by those individuals going to 2-year institutions. Still, our estimates also point toward the irrelevance of family income and to the decrease in the effects of both the cognitive and the technical factors, which are again of opposite signs. The evolution of income and ability effects on college persistence are therefore similar to those obtained for enrollment.

8.1.3 4-Year College Graduation

As noted earlier, the literature focusing on family income and higher education usually avoids modeling high school graduation and therefore estimates income effects over the general population. With our model, it is possible to measure both a conditional (on high school graduation) effect and an unconditional one (for the overall population). The income effects among high school graduates (in columns 1 and 2) and among the entire population (columns 3 and 4) of Table 8.2 indicate that the distinction may sometimes be relevant.

First, in the 1980's, income had no impact on college graduation after conditioning on abilities. However, unlike what was found for high school graduation, enrollment and persistence, our findings suggest that college graduation has become more income sensitive in the early 2000's. Precisely, it moved from values practically equal to 0 (even negative) to 0.023 (conditional on high school graduation) and 0.027 (without conditioning). This illustrates that inference about the evolution of family income-based education inequality may be highly dependent on the educational outcome considered, as we just saw that both enrollment and persistence income effects appear to be negligible in the early 2000's. It is however interesting to note that the college graduation income effect of the 97 cohort is virtually equal to the high school graduation income effect. So, family income had no more impact on college graduation than on high school graduation in the early 2000's and had less impact on college graduation than high school graduation in the early 1980's.

The evolution of the impact of the cognitive factor goes in the opposite direction of the income effect. In the 80's, a 1 standard deviation difference raised graduation frequencies by 0.20 among high school graduates and 0.15 among the overall population but the corresponding effects for the 97 cohort, which were equal to 0.08 and 0.04, indicate that the cognitive factor had lost half of its predictive power. This tendency had been noticed also for high school graduation and for college persistence. This provides evidence in favor of the hypothesis that college selectivity, as defined by the mapping from cognitive ability onto graduation outcomes (holding other factors constant) may have dropped substantially over the 1980-2000 period (Hoxby, 2009 and Babcock and Marks, 2011).

Finally, the last striking result is the strong negative effect of the technical factor on college graduation. Indeed, the technical factor effect not only dominates the impact of the cognitive factor in absolute values but is also more stable across cohort. Among high school graduates, the effect of a 1 standard deviation difference went from -0.27 (in 79) to -0.23 (in 97) while it went from -0.19 (in 79) to -0.17 (in 97) in the general population. The strength of the technical factor is most likely explained by the fact that it provides access to better job market opportunities for non-college graduates than does pure cognitive skills.

8.1.4 Age at College Graduation

As argued earlier, differences in age at graduation may be a different way to coin education inequality. It is a simple way of summarizing trajectories that are defined by multi-dimensional characteristics and may capture fundamental differences in access to educational financing if those with low family background are forced to work and/or to interrupt college in order to finance their studies.

The reduced-form mappings summarized in columns 5 and 6 of Table 8.2 are fully coherent with the weak dependence of most trajectory characteristics on family income and the strong dependence on cognitive ability established in Section 6 as most income effects are negligible. A $30,000 difference in family income increased age at graduation by less than 1 month in the 79 cohort and reduced it about 2 months in the 97 cohort.

The cognitive ability factor was the strongest determinant of age at graduation in the 1980's but its decrease in importance noted for graduation outcomes seems to be mirrored in its impact on age at graduation. Precisely, a 1 standard deviation in the cognitive factor reduced age at graduation by 1.1 year in the early 80's but its impact was reduced to 0.7 year in the early 2000's. At the opposite, a one standard deviation in the technical factor increased age at graduation by half a year in 79 but its impact increased over this 20 year period to reach 1.6 years.

To summarize, there is no evidence of any significant form of income driven inequality in age at graduation for any cohort. Differences in age at college graduation are driven mostly by differences in cognitive and technical abilities. Indeed, by the early 2000's, the technical-mechanical factor became the most important determinant of both college graduation and age at graduation.

8.2 Net vs. Total Income Effects and Sensitivity Analysis

There are two striking findings that emerge from Table 8.3. The first one is easily captured by comparing estimates of column 1 with column 2 for the 79 cohort and those of column 4 with column 5 for the 97 cohort. Essentially, we find that in the early 80's, the total effects of income on high school graduation, college graduation, enrollment and persistence (found in column 2) were higher than the net effects (in column 1) and indicated that educational outcomes were sensitive to income mostly because individual abilities at 16 were correlated with parental income and more precisely, because the cognitive-technical ability differential was increasing with income.

As an example, the total high school graduation income effect, equal to 0.072, is well above the net effect equal to 0.030 and the unconditional college graduation effect (in the population), equal to 0.031, is to be compared with the net effect which is practically 0 (-0.005). The total enrollment and persistence income effects are also three times the net effects.

In the early 2000's, the relative importance of the net income effect and the ability income affect was reversed as the net income effects absorbed most of the total effects. For instance, the college graduation net income effect conditional on high school graduation (0.023) accounted for more than half the total effect (0.032). This is due largely (but not only) to the increasing level of orthogonality between both cognitive and mechanical abilities on one hand and parental income on the other hand.

The second finding, even more striking, arises from comparing estimates of column 1 and 3 for the 79 cohort and 4 and 6 for the 97 cohort. The results point to an extreme sensitivity of cognitive ability effects on high school and college graduation to the presence of controls for technical abilities.

In the early 80's, omitting the technical factor automatically inflates the effect of cognitive abilities on high school graduation and college graduation as indicated by differences of the magnitude of 0.04 to 0.05. In the early 2000's, the differences are even higher, especially for college graduation. The omission raises the impact on college graduation among high school graduates from 0.088 to 0.155 and from 0.045 to 0.135. Omitting the technical factor therefore leads to a serious exaggeration of the role of cognitive abilities. This is also the case for enrollment and persistence although the differences are not as pronounced.

This finding is likely explained by the existence of an important mass of high technical ability-low academic ability individuals who have good job opportunities without graduating from college and would be classified only as low cognitive ability if it is the only factor incorporated. This omission therefore reinforces artificially the effect of cognitive ability.

Omitting the technical factor may also impact on the net income effect. This is notable with the college graduation net income effects measured in the early 2000's and which would be equal to 0.032 and 0.036 in absence of the technical factor but are actually estimated to be equal to 0.023 and 0.027 when both factors are taken into account. The overstatement of the net income effect is most likely due to the fact that some lower income individuals who tend to have high mechanical ability do not graduate from college simply because they respond to opportunities that do not require college. Taken as such our estimates indicate that studies concerned with educational outcomes, and which control only for standard cognitive measures such as AFQT scores or extract a single cognitive factor from a subset of ASVAB measures, will tend to seriously over-estimate the importance of cognitive abilities on college graduation outcomes.

Table 8.1
High School Graduation and College Enrollment

	Dependent Variable					
	HS. Grad.		Enrollment		College Persistence	
Population	All		HS graduates		HS graduates	
	1979	1997	1979	1997	1979	1997
Inc.($30 K)	0.0306**	0.0210**	0.0158**	-0.0078**	0.0099**	-0.0104**
Cognitive	0.1811**	0.1097**	0.1031**	0.0471**	0.0637**	0.0407**
Technical	-0.1941**	-0.0838**	-0.0956**	-0.0073**	-0.1287**	-0.0410**
Mother's Educ	-0.0281**	0.0280**	-0.0136**	-0.0121**	-0.0098**	-0.0073**
Intact Family	0.0348**	0.1042**	0.0297**	0.0415**	0.0022	0.0169**
Age HSG.	-	-	-	-	-	-
R square	0.36	0.28	0.09	0.09	0.12	0.07

Note: The high school graduation indicator is equal to 1 for those graduating (before 21) and 0 if not.

Note: The enrollment indicator is equal to 1 if someone has ever enrolled beyond high school graduation.

Note: The persistence indicator is equal to 1 if someone has enrolled at least 5 semesters beyond high school graduation

Note: Family Income is measured in $30,000 . Mother's education is divided by 2.6 (the standard deviation in the 79 cohort) and the cognitive and technical factors are standardized.

Table 8.2
College Graduation and Age at Graduation

	Dependent Variable					
	College Grad.		College Grad.		Age at Coll. Grad.	
Population	HS graduates		All		College Graduates	
	1979	1997	1979	1997	1979	1997
Inc. ($30 K)	-0.0087	0.0232**	-0.0049**	0.0269**	0.1204**	-0.1701**
Cognitive	0.2029**	0.0884**	0.1497**	0.0455**	-1.0722**	-0.6626**
Technical	-0.2666**	-0.2311**	-0.1892**	-0.1695**	0.4145**	1.6150**
Mother's Educ	-0.0317**	0.0536**	-0.0316**	0.0456**	0.1746**	-0.2970**
Intact Family	-0.0722**	0.0868**	-0.0546**	0.0714**	0.3041**	-0.4292**
Age at HSG	-0.0517**	0.0331**	-	-	1.0550**	0.7234**
R square	0.44	0.49	0.35	0.47	0.28	0.47

Note: The college graduation indicator is equal to 1 for those graduating and 0 if not.

Note: Age at college graduation is measured in years

Note: Family Income is measured in $30,000. Mother's education is divided by 2.6 (the standard deviation in the 79 cohort) and the cognitive and technical factors are standardized.

Table 8.3
Net vs Total Income Effects:
The Importance of Controlling for the Technical Factors

	1979				1997	
Cognitive	yes	no	yes	yes	no	yes
Technical	yes	no	no	yes	no	no
HS Grad.						
Income (30K)	0.0306	0.0721	0.0358	0.0210	0.0325	0.0254
Cognitive	0.1811	-	0.2246	0.1097	-	0.1414
Enrollment						
Income (30K)	0.0158	0.0385	0.0184	-0.0078	-0.0050	-0.0074
Cognitive	0.1031	-	0.1245	0.0471	-	0.0500
Persistence						
Income (30K)	0.0099	0.0284	0.0134	-0.0104	-0.0054	-0.0082
Cognitive	0.0637	-	0.0926	0.0407	-	0.0561
Coll. Grad. (cond.)						
Income (30K)	-0.0087	0.0155	-0.0035	0.0232	0.0337	0.0323
Cognitive	0.2029	-	0.2063	0.0884		0.1549
Coll. Grad. (uncond.)						
Income	-0.0049	0.0312	0.0002	0.0269	0.0427	0.0360
Cognitive	0.1497	-	0.1921	0.0455		0.1350
Age at Coll. Grad.						
Income (30K)	0.1204	0.0223	0.1048	-0.1701	-0.1987	-0.1997
Cognitive	-1.0722	-	-0.9091	-0.6626		-0.8949

9 Conclusion

This paper was motivated by the research for a novel way to characterize education inequalities and its evolution in the US over the period covering the early 1980's until the early 2000's. We modeled the joint distribution of time allocation decisions leading (or not) to graduation as well as graduation outcomes (conditional on realized trajectories) within a dynamic model. The distinction between a latent cognitive factor and another one measuring technical and mechanical abilities, and the allowance for their potential correlation with individual and family characteristics at 16, turned out to be crucial as we found that the gap between one's cognitive and technical ability levels was increasing with family income in the early 80's but much less in the early 2000's.

Despite the existence of a wide degree of dispersion in time allocation decisions between various combinations of school enrollment, hours worked when in school and school interruptions (either for work or hometime), the results suggest that heterogeneity in family income played practically no role in the 80's and even less in the early 2000's. Trajectories are explained by cognitive and technical-mechanical abilities: high cognitive ability individuals work less during school, enroll earlier and graduate earlier, high technical and mechanical ability individuals work much more during school, enroll later and graduate later.

Among all graduation and enrollment outcomes, college graduation is the only for which the effect of income has increased between the 1980's and the early 2000's. The effects of income on high school graduation, enrollment and persistence have all decreased. However, the college graduation income effect reached a level no more important than the high school graduation income effect as our estimates imply that a $30,000 real income difference raised high school and college graduation frequencies by 0.02 to 0.03.

The total effect of income (incorporating the effects of income on cognitive and technical abilities at age 16) on high school graduation and college graduation were higher than the net effects in the 80's mostly because individual abilities were correlated with parental income and other characteristics.

In the early 2000's, the relative importance of the net income effect and the ability income affect was reversed as net income effects capturing financing opportunities became relatively more important than the residual ability income effects.

However, income effects remain very small compared with ability effects as a one standard deviation increase in cognitive or technical abilities can be between 5 to 15 times as large as a $30,000 increase depending on the cohort and the factor considered but our estimates indicate that over the 1980-2000 period cognitive ability has lost about half of its predictive power on 4-year college graduation. This provides strong evidence in favor of the existence of a strong decrease in college graduation selectivity.

We also found no evidence of any significant form of income driven inequality in age at graduation for any cohort. Differences in age at college graduation are driven mostly by differences in cognitive and technical abilities. Indeed, by the early 2000's, the technical-mechanical factor became the most important determinant of both college graduation and age at graduation.

Finally, both income and ability effects have been assumed to be identical across racial groups. Over the period considered, institutional features such as Affirmative Action, financial aid policies (biased toward low income groups) as well as labor market discrimination might have affected time allocation decisions differently for different groups. As those features of the higher education environment are likely to translate into different income and/or ability effects, it would be interesting to focus on racial differences and examine if the evolution of income and ability effects may differ between Blacks, Whites and Hispanics and to what extent the distinction between academic and practical intelligence may be relevant.

References

[1] Abel J.R. and R. Deitz (2014) “Do the Benefits of College Still Outweigh the Costs? *Current Issues in Economics and Finance, NY Federal Reserve Board,* 20 (3)

[2] Altonji, J., Bharadwaj P. and F. Lange (2012) “Changes in the Characteristics of American Youth: Implications for Adult Outcomes, “*Journal of Labor Economics,* 30 (4), 783-828.

[3] Ashworth, Jared, V. Joseph Hotz, Arnaud Maurel, and Tyler Ransom (2020), “Changes across Cohorts in Wage Returns to Schooling and Early Work Experiences”, Journal of Labor Economics, 39 (4)

[4] Arcidiacono Peter, Esteban Aucejo, Arnaud Maurel and Tyler Ransom (2019) “College Attrition and the Dynamics of Information Revelation” NBER Working Paper.

[5] Babcock P. and Marks M. (2011) “The Falling Time Cost of College: Evidence from a Century of Time Use Data “*The Review of Economics and Statistics”*, May 2011, 93(2): 468–478

[6] Bajari, Patrick, Fox, Jeremy and Steve Ryan (2007) “Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients” *American Economic Review.* Apr 2007, Vol. 97, No. 2: Pages 459-463

[7] Belley P. and L. Lochner (2007) “The Changing Role of Family Income and Ability in Determining Educational Achievement” NBER Working Paper 13527

[8] Beaudry, Paul, David A. Green, and Benjamin M. Sand. 2014. The declining fortunes of the young since 2000. *American Economic Review* 104, no. 5:381-386.

[9] Belzil, Christian and Jorgen Hansen (2020) “The Evolution of the US Family Income-Schooling Relationship and Educational Selectivity” *Journal of Applied Econometrics,* 35 (7): 841-859

[10] Blauvelt, Glenn Raymond. 2006. Machine shop: A design environment for supporting children's construction of mechanical reasoning and spatial cognition. PhD dissertation, University of Colorado at Boulder

[11] Bound, John, Michael F. Lovenheim, and Sarah Turner (2012) "Increasing time to baccalaureate degree in the United States", *Education Finance and Policy* 7, no. 4:375-424.

[12] Bulman, G, Fairly R., Goodman S. and A. Isen (2021) "Parental Resources and College Attendance: Evidence from Lottery Wins, *American Economic Review* 111 (4): 1201-1240

[13] Cameron S. and James J. Heckman (2001) "The Dynamics of Educational Attainment for Black, Hispanic, and White Males" *Journal of Political Economy,* Vol. 109, No. 3 (June 2001), pp. 455-499 (45 pages)

[14] Carneiro, Pedro, Karsten Hansen, and James J. Heckman. 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* 44, no. 2:361–422.

[15] Carneiro, Pedro and Sokbae Lee. 2011. Trends in quality-adjusted skill premia in the united states, 1960-2000. *American Economic Review* 101, no. 6:2309-2349.

[16] Cunha, Flavio, Fatih Karahan, and Ilton Soares. 2011 "Returns to skills and the college premium" *Journal of Money, Credit and Banking* 43, no. 5:39-86.

[17] Castex, G. and E. Dechter (2014) "The Changing Role of Education and Ability in Wage Determination" *Journal of Labor Economics",*Vol. 32, No. 4, pp. 685-710

[18] Dynarski, Susan M. and Judith Scott-Clayton (2016), Financial Aid Policy: Lessons from Research, Future of Children, 23, 67-91.

[19] Gneezy U, List J.A., Livingston J.A., Qin X, Sadoff S., and Xu Yang (2019)"Measuring Success in Education: The Role of Effort on the Test Itself, *American Economic Review, Insights*, 1 (3), 291-308

[20] Heckman, J.J. and T. Kautz (2014) “Achievements Tests and the Role of Character in American Life” in *The Myth of Achievement Tests: The GED and the Role of Character in American Life,* Chapter 1, 3-56, University of Chicago Press.

[21] Heckman, James J., John E. Humphries, and Gregory Veramendi (2017). “Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking.” *Journal of Political Economy*

[22] Hoxby, Caroline, The Changing Selectivity of American Colleges, *Journal of Economic Perspective*, 23 (2009), 95-118.

[23] Johnson, M.T. “Borrowing Constraints, College Enrollment, and Delayed Entry” *Journal of Labor Economics*, 2013, vol. 31, no. 4]

[24] Juhn, C., K.M. Murphy and B. Pierce (1993) “Wage Inequality and the Rise in Returns to Skill”, *Journal of Political Economy,* 101 (3), 410-442.

[25] Keane, Michael P. and Kenneth I. Wolpin (1997), The Career Decisions of Young Men, *Journal of Political Economy*, 105, 473-522

[26] Kinsler, Josh and Ronni Pavan, Family Income and Higher Education Choices: The Importance of Accounting for College Quality, Journal of Human Capital, 5 (2011), 453-477.

[27] Landerso, Rasmus and James J. Heckman (2017) The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US, *Scandinavian Journal. of Economics* 119(1), 178–230, DOI: 10.1111/sjoe.12219

[28] Lochner, Lance J., Todd Stinebrickner, and Utku Suleymanoglu. 2021. Parental support, savings and student loan repayment, American Economic Journal: Economic Policy, 13, no. 1:329–71.

[29] Lochner, Lance J. and A. Monge-Naranjo (2011), The nature of credit constraints and human capital, *American Economic Review* 101(6), 2487-2529.

[30] Lovenheim, Michael and C. Lockwood Reynolds, Changes in Postsecondary Choices by Ability and Income: Evidence from the National Longitudinal Surveys of Youth, Journal of Human Capital, 5 (2011), 70-109.

[31] Lovenheim, Michael F. and C. Lockwood Reynolds, The E ect of Housing Wealth on College Choice: Evidence from the Housing Boom, Journal of Human Resources 48 (2013), 1-35.

[32] Murphy, Keven M. and Robert H. Topel (2016) "Human Capital Investment, Inequality, and Economic Growth," *Journal of Labor Economics*, vol 34(S2), pages S99-S127.

[33] Nielsen, Eric R., The Income-Achievement Gap and Adult Outcome Inequality, Finance and Economics Discussion Series 2015-041. Washington: Board of Governors of the Federal Reserve System, http://dx.doi.org/10.17016/FEDS.2015.041, 2015.

[34] Page, Lindsay C. and Judith Scott-Clayton, Improving College Access in the United States: Barriers and Policy Responses, NBER Working Paper No. w21781, 2015.

[35] Prada, M and S.Urzua (2017) "One Size Does Not Fit All: Multiple Dimensions of Ability, College Attendance, and Earnings", *Journal of Labor Economics*, 2017, vol. 35, no. 4]

[36] Scott-Clayton, J., 2012, "What explains trends in labor supply among US undergraduates?," *National Tax Journal*, 65 (1), 181-210.

[37] Train, Kenneth (2008) EM Algorithms for nonparametric estimation of mixing distributions, *Journal of Choice Modelling* Vol 1, Issue 1, Pages 40-69.

Figure 1. Distribution of factors for the 1979 cohort.

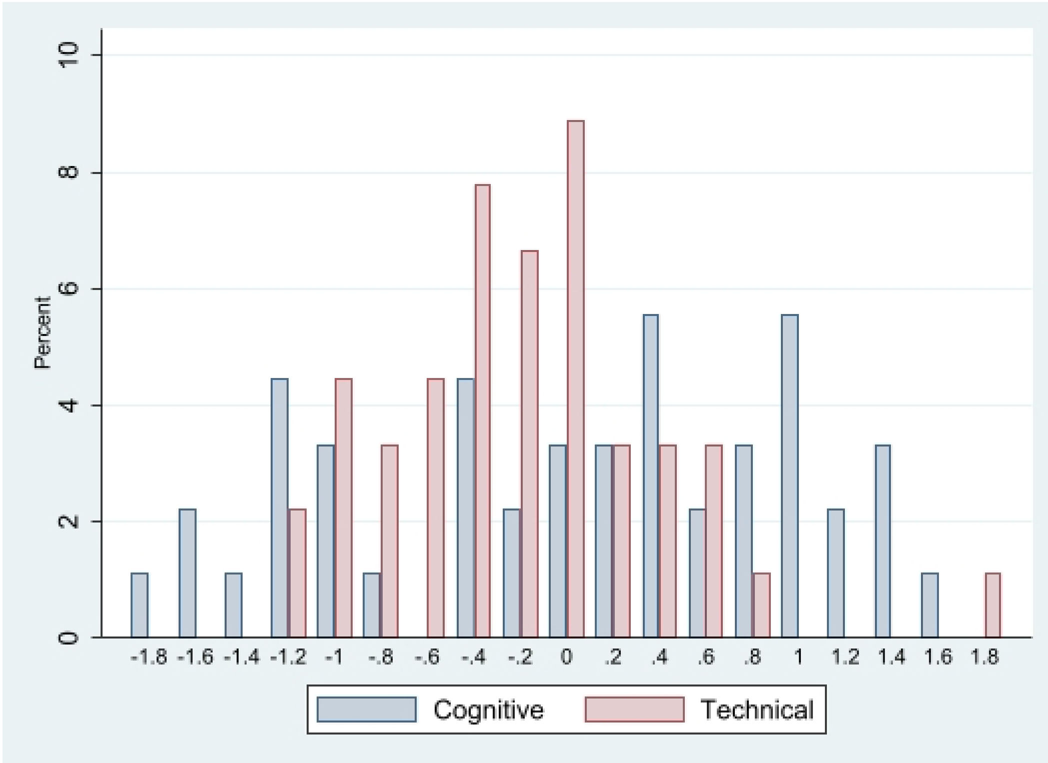

Figure 2. Distribution of factors for the 1997 cohort.