ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Albæk, Karsten et al.

Working Paper

Better Late Than Never? How Late Completion Affects the Early Careers of Dropouts

IZA Discussion Papers, No. 12560

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Albæk, Karsten et al. (2019) : Better Late Than Never? How Late Completion Affects the Early Careers of Dropouts, IZA Discussion Papers, No. 12560, Institute of Labor Economics (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/207386

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12560

Better Late Than Never? How Late Completion Affects the Early Careers of Dropouts

Karsten Albæk
Rita Asplund
Erling Barth
Lena Lindahl
Marte Strøm
Pekka Vanhala

AUGUST 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12560

Better Late Than Never? How Late Completion Affects the Early Careers of Dropouts

Karsten Albæk
Danish National Centre for Social Research

Rita Asplund
Research Institute of the Finnish Economy

Erling Barth
Institute for Social Research, Oslo and IZA

Lena Lindahl
Swedish Institute for Social Research and IZA

Marte Strøm
Institute for Social Research, Oslo

Pekka Vanhala
Research Institute of the Finnish Economy

AUGUST 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Better Late Than Never? How Late Completion Affects the Early Careers of Dropouts[1]

Across the OECD countries, dropouts from upper secondary schooling fare worse in the labor market, with higher NEET rates more spells of unemployment and lower earnings. Among the dropouts, there are however significant shares who complete at a later age. In this paper, we thus ask the question: Does it pay for young adults who do not complete upper secondary schooling by the age of 21, to do so at some point during the subsequent 7 years, that is, before turning 28? In all four Nordic countries under scrutiny, we find that late completion lowers the probability of being outside employment, education or training (NEET) at age 28. Moreover, the exact age of completion does not seem to matter. Our estimates are robust to the inclusion of extensive controls for socioeconomic background and early schooling paths, and similar to the ones produced by event history analysis with individual fixed effects. This indicates that late completion of upper secondary schooling plays an important role for the labor market inclusion of young dropouts.

JEL Classification: I21, J24, J64
Keywords: upper secondary schooling, dropouts, NEET rates

Corresponding author:
Erling Barth
Institute for Social Research
Munthesgate 31
0208 Oslo
Norway
E-mail: eba@socialresearch.no

[1] We acknowledge funding from the VAM program of the Norwegian Research Council, grant # 247996. The paper has been presented at the ESPE conference in Antwerpen, June 2018.

1. Introduction

A lot of attention has recently been devoted, both by the EU and across the OECD, to dropouts from upper secondary schooling. Non-completion of upper secondary schooling is a serious risk factor later in life, in particular due to higher probabilities of being NEET (Not in Education, Employment or Training) and of low earnings. With technological change and increased international competition in the products market (trade) and the labor market (immigration), a large literature has demonstrated that those who lose from these developments are low skilled, manual labor in Western countries (see e.g. Autor and Katz (1999), Autor et al. (2006), Goos et al. (2009)). Would it help to raise the educational level of those who have the lowest skills to be better able to adapt to these developments?

The literature on high school dropouts has focused on how to prevent young students at risk from dropping out, and find positive effects on upper secondary school completion and later earnings of e.g. raising the compulsory school leaving age (Angrist and Krueger (1991), Harmon and Walker (1995), Oreopoulos (2007)) or giving a means-tested conditional cash transfer to 16- to 18-year-olds who continue in full-time education (Dearden et al. (2007)). Little attention has been paid to the quite substantial share of youth that completes upper secondary school at a later than normal age. As pointed out by Albæk et al. (2015), the question of high school dropouts, in particular as defined by OECD standards as youth not having completed upper secondary schooling by the age of 21, is in many cases factually a question of late completion rather than non-completion.[2]

In this paper, we estimate returns to late completion. We ask: Does it pay for young adults who have not completed upper secondary schooling by the age of 21, to do so at some point during the subsequent 7 years? And if it pays, does it matter if they complete early rather than late? Specifically, we investigate the potential gains of late completion, between 21 and 27 years of age, for the probability of being a NEET at the age of 28. So far there is very little evidence on this question to rely on.

[2] OECD (2014) shows that more than 10% of upper secondary graduates in Denmark, Finland, the Netherlands and Norway are 25 or older, and as many as 20% in Iceland.

Of course, there is likely to be a non-negligible selection of dropouts into those who complete at a later point in time, which may bias the estimated returns to late completion. For example, those who are more able or more motivated, or have parents that push them more, may be more prone to complete upper secondary schooling, and less likely to become a NEET for the same reasons. Since we do not perform a natural experiment, we rely on two methods for identification. One is to include an unusually rich set of observable characteristics to control for selection by observables, and the other is utilizing event history analysis.

Our rich set of observable control variables include the sequence of events occurring in each youth's life between age 16 and 20 (schooling, employment, unemployment, benefit recipiency and unknown status by year) eventually leading up to a "dropout status", as obtained from using cluster analysis. Included is also parental background as measured separately by the father's and the mother's education, as well as their earnings level when the child was 16. These controls potentially pick up both the direct relationship between early paths and family background (labor market experience, pushy parents) and the more indirect relationship (ability).

We also study late completion in an event-study setup, exploring changes in NEET status before and after the year of completion. The event-study provides, for each young adult, the whole trajectory of changes in NEET status before and after completion (we include individual fixed effects to control for all selection on fixed characteristics). If the individual NEET status trajectories significantly changes around the year of completion, the change may be interpreted as causal effects of completion under the identifying assumption that the exact timing of completion (age of completion) is random (conditional on all fixed characteristics of the individual).

We find that completing upper secondary schooling at a later than normal age is associated with significantly better labor market prospects of young dropouts. The probability of being a NEET when aged 28 is 12 to 15 percentage points lower for late completers than for non-completers. We show that our control variables are relevant to potential selection: early paths and family background do predict late completion and also the timing of late completion. However, including this highly detailed information as controls does not change the estimated returns to completion much. Event history analysis measures the effect to be in the range of 8 to 14 percentage points. Although there is clear selection on observable characteristics into late completion, a large part of the returns are orthogonal to this selection.

There are several indications that the estimated returns to late completion of upper secondary schooling are not driven by selection: the inclusion of detailed controls does not change the coefficients much, there is no difference between completing early (at e.g. age 22) or late (at e.g. age 28), the within-individual (event-study) estimates are consistent with the cross-sectional evidence and the estimates are similar and robust across four different countries and educational systems (Norway, Sweden, Denmark and Finland). We also use Oster's (2019) method to test the sensitivity of our results to selection on unobservables, and to provide a bound for the true effect of completion.[3]

The paper is organized as follows. First we describe the data and provide a graphical description of key facts regarding the labor market trajectories of young dropouts according to their age of (late) completion between 21 and 27 years of age. Next, we provide a description of the methodology, including a brief outline of the Oster (2016) bounds. The subsequent section presents the key results, and the last section concludes.

2. Data

The starting point of our analysis is all Danes, Finns, Norwegians and Swedes who were born in the years 1977 to 1989. Individual-level information from different official registers are merged for each country, and all analyses are run on country-specific datasets. We use information from residential registers, education registers, tax-registers, employer-employee registers and registers on welfare payments. Equally good registers on employment across all countries start in 1992, and all labor market outcomes are therefore measured in the period 1992-2010.

The variable defining the dropout group uses information from the educational register about the highest completed education in each year. We pick the first year the individual is observed with a completed upper secondary diploma, and the person is an upper secondary dropout if the age of upper secondary completion is higher than 21, or missing (OECD definition of dropout).

[3] Oster (2019) develops a method to provide an assessment of the severity of the selection on unobservables, using assumptions on the relation between the unobservables and completion relative to the relation to early paths and socioeconomic background, as well as on the hypothetical total explanatory power of the model if unobservables were also included.

As background information we use birth cohort, residency, and information about parents' education and income. We condition on residency in the year we measure the outcome, but also on residency in the years when the individual is 16-20 years old. First, this is because we want to include the youth that had a real opportunity to be in the respective country's educational system before they complete or drop out. A second reason is that we include individual educational/labor market histories at the ages 16-20 as important explanatory variables in the analysis.

The outcome variables that we study are indicators of labor market status at the ages 21-31. We stop at 31 to get enough cohorts having reached this age in the observation window (cohorts 1977-1979 have reached age 31 in 2010). The different statuses that we focus on are "Student", "Employed" and "NEET". We use the status registered in October every year. If an individual is observed in both employment and education, we set the status to "Student". The NEET status covers individuals registered as unemployed or receiving welfare pensions.

The labor market histories of the individuals at the ages 16-20 (pre-paths) are more detailed than the outcome variables. In calculating these histories, we use a total of five different statuses: "Student", "Employed", "Unemployed", "Welfare Pensioner" and "Other". We use two different ways of combining the status information for the five years at the ages 16-20. First, we use cluster analysis to group the different sequences into 16 "Ideal paths". Second, we use the actual sequence as a control variable (one such sequence could for example be "student, student, unemployed, employed, employed"). These pre-paths are more carefully discussed in section 4.1 below.

3. Institutional background and dropout rates in the Nordic countries

The educational systems in the four Nordic countries under scrutiny are overall similar. For the cohorts that we study, children started school at age 6 or 7. Compulsory schooling lasts for 9 years. Upper secondary schooling is voluntary, but almost everyone continues in upper secondary schooling. There are two upper secondary schooling tracks: an academic track to prepare for further studies, and different vocational tracks ending, after an obligatory on-the-job training period, with a certificate as skilled worker.

Conspicuously large shares of those who start an upper secondary education have dropped out by age 21: the dropout rate is smallest in Sweden at 16.7 percent and highest in Denmark at 34.2 percent (first line, Table 1). In the dropout group, there are however significant shares who complete only after the age of 21. Most of them complete in their early twenties while very few complete in their late twenties. Various routes to late completion are available in each country: ordinary youth education or various modes of adult education.

Table 1. Dropout rates and late completion, percentages: Norway, Sweden, Denmark and Finland

	Norway	Sweden	Denmark	Finland
Dropout measured at age 21 (percent of all)	29,6	16,7	34,2	18,1
Late completers (percent of dropouts at 21):				
At 22	9,6	6,0	19,0	8,6
At 23	4,8	4,1	9,2	6,5
At 24	3,0	3,0	5,5	4,8
At 25	2,0	2,3	4,0	3,7
At 26	1,4	2,0	2,9	2,8
At 27	1,1	1,6	2,3	2,1
At 28	0,8	1,2	1,7	1,6
At 29	0,6	1,0	1,3	1,1
At 30	0,5	0,8	0,9	0,8
At 31	0,3	0,6	0,5	0,5
Still not completed at 31 (percent of all)	22,5	12,9	18,0	12,2

Note: Own calculations on register data. The first and last line shows the percentage of all youth who have not completed upper secondary schooling by the age of 21 and 31 respectively. The middle lines shows how many percent in the dropout group (measured at age 21) who complete at a later age. E.g. for Norway, 29,6 percent of all youth have not completed upper secondary schooling at age 21. In this group, 9,6 percent completes at age 22, 4,8 percent completes at age 23 etc. Measured at age 31, 22,5 percent of all youth have still not completed.

The last row in Table 1 shows what percentage is still dropouts at age 31. Dropout rates measured at the age of 31 are substantially lower than dropout rates measured at age 21 (they would be much lower also if they were measured at e.g. age 25). Denmark is an extreme case in that a large share of the youth population completes upper secondary schooling later than the cut off at 21 years of age, when the share of non-completers is 34.2 percent; by age 31, the share is down at 18 percent. The overall pattern is however similar in all four Nordic countries (see also Albæk et al. (2015) and Education at a Glance (2017, Table A2.1)).

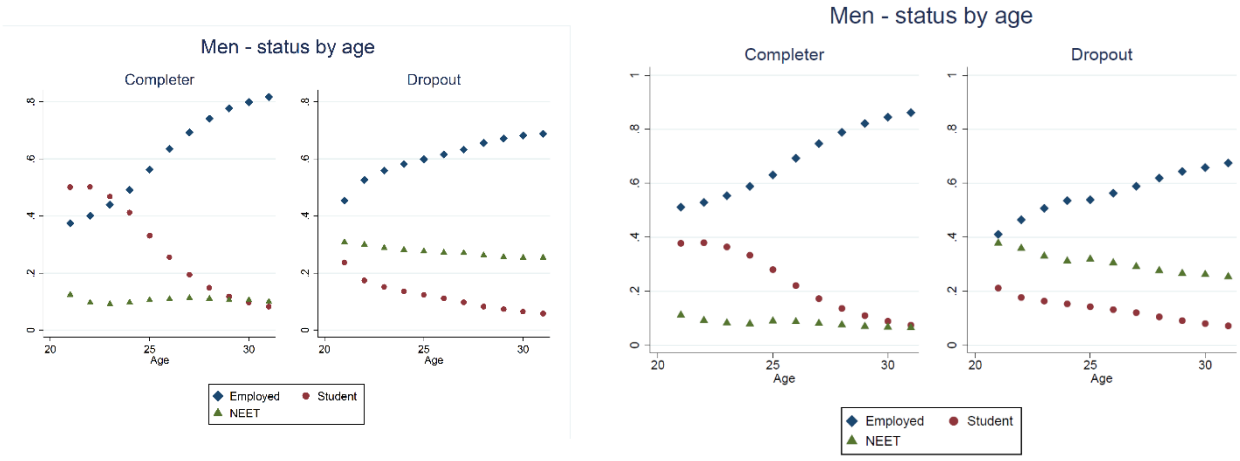
4. Early career trajectories of dropouts and completers in four Nordic countries

Albæk et al. (2015) showed that dropout rates in the Nordic countries are high, similar to those of the US and other OECD countries, and stable over time. As in the rest of OECD countries, dropouts fare worse on the Nordic labor market, with higher NEET rates, more spells of unemployment and lower earnings.

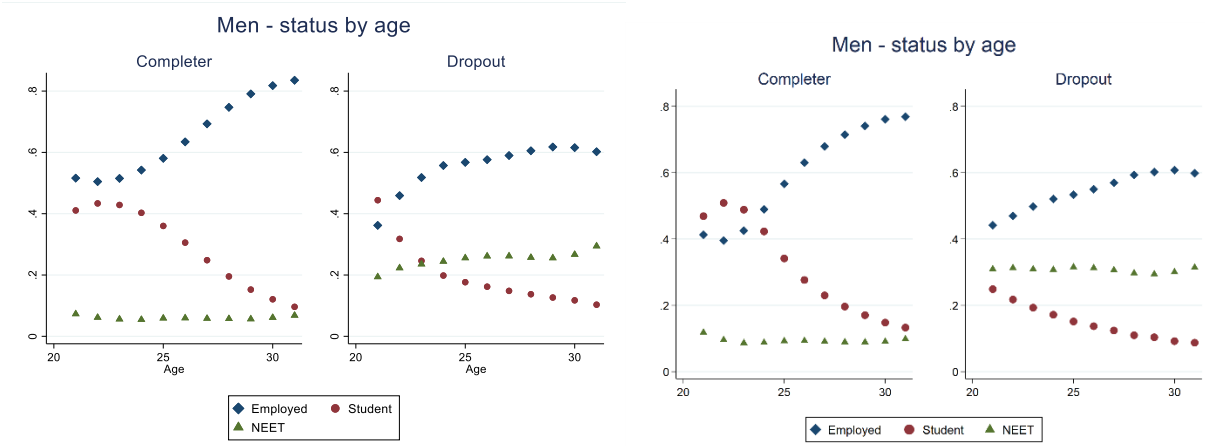
Figure 1 shows average early career labor market trajectories of dropouts versus completers in the Nordic countries. The patterns are highly similar across the four countries: Those who complete upper secondary by age 21 often continue in education or training also after upper secondary schooling. At age 32, around 80 percent of them are employed and less than 10 percent belong to the category of NEETs. Dropouts, on the other hand, have employment rates around 20 percentage points below those of completers while their NEET rate is 20-30 percent.

Figure 1. Early career trajectories, upper secondary dropouts versus completers

Norway *Sweden*



Denmark *Finland*



Note: Status as "Completer" versus "Dropout" is defined by completion status at 21 years of age.

In Figures A.1-A.4 in the appendix, we show early career trajectories of male and female dropouts by completion age. In all four countries, the employment rate is clearly higher after upper secondary school completion as compared to before completion, regardless of the age of completion. Employment rates are even similar after completion to the employment rates for completers in Figure 1. NEET-rates are also lower after completion compared to the first years after 21.

Moreover, the patterns are strikingly similar in all four countries. It is also striking that there does not seem to be a clear and visible connection between the age of completion and the positive jump in employment probabilities or the drop in NEET probabilities. Completing upper secondary school seems to matter, but when you do it, does not seem to matter.

4.1 Early paths of youth who drop out

Let us now move one step back and see what the dropouts did before age 21, i.e. during the years when they were expected to complete upper-secondary schooling. We observe early school-to-work transition paths for everyone in our sample. As indicated earlier, we use these early paths as control variables in the estimations. To form meaningful control variables and sample splits, we perform cluster analysis to allocate individuals into "early path groups" that group together young persons who are highly similar with respect to their early post-compulsory-school educational

and labor market experiences. This method is more closely described in Albæk et al. (2015), from where some of the below descriptions are taken.

We use five different education/labor market statuses to categorize a young person's activity at each age from 16 up to 20: (1) "Student", (2) "Employed", (3) "Unemployed", (4) "Welfare Pensioner" and (5) "Other". The alternative trajectories (sequences of main activities) that a young person may follow are huge in numbers. From one age to the next – e.g. from age 16 to age 17 – the young person can in principle make five different moves: stay in the same activity or move to one of the other four main activities. Accordingly, the total number of possible combinations of main activities over five ages, from 16 to 20, is 5^5 or in total 3.125 .

The real-life number of alternative trajectories from age 16 up to age 20 is smaller, though, less than one third of all the possible combinations, but still the real-life numbers of alternative trajectories are so high that it is impossible to describe and analyze each of them separately. A major aim of the cluster analysis method is to obtain an overview of transition patterns by reducing large numbers of empirical sequences to a smaller number of classificatory categories of sequences.

Cluster analysis minimizes the distance between sequences within each cluster. One way of measuring the distance between two sequences is to count the number of substitutions required for changing one sequence into the other sequence, and then use the number of substitutions needed as a measure of the distance between the two sequences. This is what we do. However, in this paper, instead of using an algorithm to define clusters, we form a total of 16 reference or "ideal" clusters, as described below, and allocate all individual sequences in the data to the cluster where the distance of the sequence to the reference sequence is smallest. We then apply these same reference sequences in each of the four Nordic countries, and allocate young people's early school-to-work trajectories across them. Clustering by reference sequences allows us to form clusters in the same way across all four countries[4]: a particular sequence observed in two or more of the Nordic countries is, in each country, allocated into exactly the same reference sequence cluster. This also implies that we may straightforward compare the magnitude of distinct pathways across the four countries.

[4] As opposed to e.g. clustering based on optimal matching.

Table 2 shows our 16 reference paths, shortly described in the first column and with the specific sequence set out in column 2. We see that some paths are more common than others – and the pattern is often the same across the countries. The two most common paths are "Late dropout, employment" and "Student". "Late dropout, employment" are those who are registered as students for the full three years that upper secondary schooling takes, but then for some reason dropped out/did not pass. They then leave school and start work. Indeed, other sequences that start with studies and then shift to employment after dropping out are also common. "Student", in turn, are those who are registered as students at all ages between 16-20. The least common sequences are those starting with studies and ending up on some welfare pension arrangement. Hence, most of the dropouts drop out to work, although there are also quite large shares ending up in the "other" category which is the one most similar to NEET status (only taking out the registered unemployed and welfare pensioners).

Table 2: Shares of dropouts on different early paths (percentages) based on cluster analysis. Dropout is measured at age 21.

		Norway		Sweden		Denmark		Finland	
Description of path	Ideal Sequence	Men	Women	Men	Women	Men	Women	Men	Women
Late dropout, other	(1 1 1 5 5)	10,0	8,1	3,6	3,3	5,0	5,5	4,5	5,8
Late dropout, employment	(1 1 1 2 2)	25,5	29,7	27,4	25,2	9,8	10,5	25,5	14,9
Late dropout, unemployment	(1 1 1 3 3)	4,2	4,1	12,7	10,7	2,9	4,6	3,7	3,4
Late dropout, pensioner	(1 1 1 4 4)	0,7	1,1	1,1	1,2	0,3	0,3	0,5	0,6
2nd year dropout, other	(1 1 5 5 5)	5,6	6,1	2,3	2,3	3,2	4,0	2,3	3,4
2nd year dropout, employment	(1 1 2 2 2)	10,3	9,1	7,0	8,4	11,2	10,9	11,1	7,7
2nd year dropout, unemployment	(1 1 3 3 3)	2,5	2,2	5,2	5,4	2,6	4,3	3,8	2,9
2nd year dropout, pensioner	(1 1 4 4 4)	0,2	0,3	0,6	0,6	0,6	0,7	0,4	0,5
Early dropout, other	(1 5 5 5 5)	2,2	2,9	1,7	1,8	3,1	4,8	1,4	2,2
Early dropout, employment	(1 2 2 2 2)	2,7	2,2	1,7	1,9	9,3	8,6	3,8	3,2
Early dropout, unemployment	(1 3 3 3 3)	1,0	0,9	0,9	1,1	1,0	2,0	2,8	2,7
Early dropout, pensioner	(1 4 4 4 4)	0,0	0,0	3,2	3,0	0,7	0,7	1,9	2,0
Student	(1 1 1 1 1)	25,7	23,3	16,5	20,7	39,3	32,3	12,6	21,1
Other	(5 5 5 5 5)	2,0	3,1	7,0	6,4	2,6	2,8	7,5	10,5
Employment	(5 2 2 2 2)	1,5	1,3	2,0	1,5	4,1	3,3	5,8	4,6
Late starter	(5 1 1 1 1)	5,9	5,5	7,2	6,7	4,5	4,6	12,6	14,6
Total number of observations		127294	88612	119287	90310	161978	114480	90969	61982

Footnote: The cluster analysis uses the number of substitutions required to change one sequence into the "ideal path" to match each sequence into one path, see text for discussion. The "ideal paths" are showed in the column "Ideal Sequence". Activity indicators: 1 = Student, 2 = Employed, 3 = Unemployed, 4 = Disability pension, 5 = Other

4.2. Who are the late completers?

Table 3 shows that our background variables indeed predict late completion: Children with high-earning parents have a higher probability of late completion of upper secondary schooling. Some early paths are also more common among late completers: in all four countries, continuing as a student or starting upper secondary schooling later than normal are both associated with a higher probability of completing upper secondary (later than age 21). The groups with the lowest probabilities are those who drop out and receive a welfare pension afterward (and are as such probably the groups with the most severe problems, e.g. with a disability or a serious health problem). Those who drop out to employment have higher probabilities of completing upper secondary, albeit late, compared to those who end up in unemployment. The heterogeneity between groups in the probability of late completion shows that these group indicators do pick up characteristics (observed and/or unobserved) of the youth that affect their selection into completion.

Table 3: The impact of socioeconomic background and early paths on late completion (t-values)

	Norway	Sweden	Denmark	Finland
Mother high education	-0,01	0,13	0,03	0,08
	(-2,55)	(35,81)	(11,91)	(10,46)
Father high education	-0,01	0,12	0,04	0,07
	(-3,20)	(30,59)	(14,01)	(9,66)
Mother high earnings	0,02	0,01	0,05	0,03
	(5,84)	(3,99)	(18,18)	(8,18)
Father high earnings	0,01	0,01	0,07	0,03
	(4,91)	(4,91)	(28,31)	(9,11)
Woman	0,01	0,12	0,06	0,12
	(2,34)	(49,75)	(23,05)	(33,39)
Early paths (reference group Late dropout, other):				
Late dropout, employment	0,03	0,02	-0,04	0,09
	(6,21)	(2,23)	(-5,38)	(10,92)
Late dropout, unemployment	-0,05	-0,02	-0,13	0,04
	(-5,57)	(-2,49)	(-15,19)	(3,09)
Late dropout, pensioner	-0,12	-0,10	-0,33	-0,13
	(-6,40)	(-5,17)	(-11,76)	(-4,99)
2nd year dropout, other	-0,04	0,02	-0,06	-0,06
	(-5,01)	(1,71)	(-7,20)	(-4,25)

2nd year dropout, employment	-0,03	-0,02	-0,07	0,01
	(-4,11)	(-2,60)	(-10,18)	(1,10)
2nd year dropout, unemployment	-0,07	-0,04	-0,21	-0,03
	(-6,80)	(-4,04)	(-22,97)	(-2,51)
2nd year dropout, pensioner	-0,14	-0,10	-0,40	-0,14
	(-3,70)	(-5,52)	(-22,94)	(-3,44)
Early dropout, other	-0,07	0,02	-0,14	-0,03
	(-6,12)	(1,59)	(-17,02)	(-1,59)
Early dropout, employment	-0,05	-0,03	-0,05	0,01
	(-5,67)	(-2,34)	(-7,55)	(1,06)
Early dropout, unemployment	-0,08	-0,05	-0,23	-0,03
	(-4,96)	(-3,46)	(-19,30)	(-2,54)
Early dropout, pensioner	-0,16	-0,14	-0,40	-0,21
	(-2,75)	(-14,50)	(-19,33)	(-15,70)
Student	0,24	0,21	0,35	0,25
	(40,94)	(27,07)	(57,14)	(28,47)
Other	-0,11	0,00	-0,23	-0,08
	(-9,42)	(0,43)	(-22,48)	(-8,40)
Employment	-0,12	-0,04	-0,22	-0,06
	(-9,16)	(-3,74)	(-24,37)	(-5,70)
Late starter	0,09	0,11	0,23	0,14
	(10,95)	(13,73)	(24,47)	(15,07)
Constant	0,19	0,10	0,32	0,17
	(29,12)	(11,66)	(47,92)	(19,32)
Observations	85,433	11,0299	12,2875	67,132

Sample: Non-completers at age 21. Dependent variable: Completion between age 22 and 28. Reference group for Early paths: Late drop out, other.

5 Methodology

Our main challenge is selection into late completion among those who did not complete upper secondary by the age of 21. As a starting point we *compare the NEET rates at 28 years of age* of late completers and non-completers (those who have not completed by the age of 28). We add a comprehensive set of controls reflecting socioeconomic backgrounds and behavior between 16 and 21 years of age. Secondly, we use an *event history approach*, comparing NEET rates every year before and after completion among the late completers. Thirdly, we use a *difference in difference* approach to compare changes in NEET status over time for late completers and non-completers. Finally, we compare the estimates obtained and use the Oster (2019) approach to

discuss the likely impact of selection on unobservables. The following outlines our methodology in some detail.

5.1 Comparing NEET rates at age 28 for late completers and non-completers

Our empirical analysis departs from the following simple equation, estimated on the sample of individuals who had not completed upper secondary schooling by the age of 21:

$$(1)\ y_{i28} = X_i\boldsymbol{\beta} + C_i\boldsymbol{\rho} + \gamma_t + u_i$$

Where y_{i28} is NEET status at 28 years of age for individual i[5], X_i is a row vector of individual characteristics, C_i is a row vector of dummy variables reflecting age of completion, one for each year of age between 21 and 28, taking the value 1 if the individual completed upper secondary at year τ; $C_i = [C_{21,i}\ C_{22,i}\ C_{23,i}\ C_{24,i}\ C_{25,i}\ C_{26,i}\ C_{27,i}]$, and **β** and **ρ** are column vectors of regression coefficients. The reference person has not completed upper secondary by the age of 28 and has a zero on all elements of C. γ is a year dummy and u is the error term. In our first specification, we run this model separately for each country, including gender and year dummies only.

Our key parameters are the ρ's, describing the change in probability of being a NEET at 28 years of age, depending on the year of completion of upper secondary. Below, we find very few significant differences between the ρ's, so in the following we will often discuss a restricted model where $C_i\boldsymbol{\rho}$ is replaced by a simple dummy and the common coefficient: ρC, where C is a dummy if the individual has completed upper secondary between 21 and 28 years of age, and ρ gives the change in NEET probability associated with completion.

Our main worry is selection into late completion of upper secondary schooling among those who did not complete before the age of 21. We have an unusually rich set of observable characteristics to control for selection by observables. Firstly, we include parental background by controlling for both the father's and the mother's education, as well as for their earnings when the young person was 16 years of age. We label these controls as "Socioeconomic Background".

Secondly, we constructed a sequence of events describing each young adult's early path after compulsory school (when aged 16 to 20) by classifying each year by their main activity as

[5] We have chosen the age of 28 to have enough cohorts in the sample having reached this age (these are the cohorts born 1977-1989).

described above in Section 4.1. We control for these sequences in two ways, one is grouping sequences into 10 groups using cluster analysis as described above, and the other is to add a fixed effect for each and every possible (and realized) combination of the 5 categories over the 5 years from 16 to 20. We label these controls as "Early Paths".

Socioeconomic background and early paths capture key ingredients in what we would regard as the main selection issues, such as individual ability, motivation, and health status. Socioeconomic background controls for factors that may influence all such aspects while early paths may provide strong indicators of individual capabilities through their effect on successes and failures prior to the starting age for our analysis. Adding these background factors as controls may thus be viewed as adding to the equation proxies for the underlying selection mechanisms; they capture the direct effect of the variables themselves, in addition to any factor left in the error term that is correlated with socioeconomic background and early paths. In Section 4 above we showed that socioeconomic background and early paths are significant predictors of late completion.

We are left with two estimates of the effects of late completion where we define ρ^0 as the estimate without controls for socioeconomic background and early paths, and $\tilde{\rho}$ is the estimate conditional on these controls. The way in which these parameters change between specifications provides information on both selection on observables and possibly also on selection on unobservables. We apply Oster's (2019) method to assess the impact of selection on unobservables. Using assumptions on the R square of the hypothetical regression of NEET status on completion and the full set of both observed and unobserved covariates, as well as on the relative degree of selection on observed and unobserved variables, we may derive a consistent estimator of the effect of late completion, and furthermore provide an assessment of the magnitude of potential selection on unobservables necessary to eliminate the estimated effect of completion.

5.2 Event history trajectories

Event history models estimate the full trajectory of within-individual changes in NEET status before and after upper secondary school completion. This approach allows us both to effectively control for all observed and unobserved fixed individual characteristics, and to investigate in

detail the time pattern of NEET rates before and after completion. We estimate the following, with event time g measuring distance from completion in years:

$$(2)\ y_{itg} = \sum_{j\neq -6} \beta_j \times \boldsymbol{I}[j = g] + FE_i + u_{it}$$

where y_{itg} is the probability of being a NEET at a specific age, t, for individual i, who completed upper secondary schooling at event time g=0. $\sum_{j\neq -6} \beta_j \times \boldsymbol{I}[j = g]$ include the full set of event time dummies from 9 years before upper secondary school completion to 10 years after j∈[-9,10].

The key identifying assumption is that, conditional on the individual fixed effect, the completion year is random[6]. This allows for a correlation between the individual fixed effects and the completion year, for instance that more "able" youth complete earlier, and identifies the effect of completion conditional on this relation. A no-anticipation assumption provides a necessary exclusion restriction (Athey and Imbens (2018)). A no-anticipation assumption is, however, not likely to be valid for the three to four years preceding completion, simply because the student has to apply to programs and undertake whatever is left of such a program in order to achieve completion in a given year. Indeed, the event history graphs shown below reveal that the NEET rates are considerably smaller in the 3-4 years preceding completion, predictably due to higher student rates preceding the year of completion (see Appendix figures for details). We consequently use a "long difference" with a reference year of -6, well below this dip[7]. In addition, we include individual fixed effects to control for all fixed observed and unobserved heterogeneity.

We also include a specification where we use a full set of control variables instead of the individual fixed effects. This specification does not suffer from the identification issue around the linear components of the pre-trend, as discussed by Borusyak and Jaravel (2017), but does not allow for a full control on unobservables. Comparing the coefficients from the two models allows for an assessment of the impact of unobservables on the estimated coefficients.

The β_j coefficients provide the event history estimates for the effects of late completion on NEET rates, when -6 is chosen as the reference year. Since they are allowed to vary according

[6] See Athey and Imbens (2018) for a discussion of estimation and inference in situations with "staggered adoption".
[7] We could use any year including and between j=-9 and j=-5 as reference year, as the estimated differences between j=-6 and these years are only barely significant for some periods (but still economically insignificant).

to years since completion, they are more flexible in that dimension than the ρ coefficient discussed above. If the β's for j>0 are not significantly different, β_1 or an average of the β's may be compared directly with the coefficient $\tilde{\rho}$ above. In this case we have used the within-individual variation. In the next section we compare the pre-completion trajectory of late completers to that of non-completers.

5.3 Comparison with non-completers

Next we use the non-completers as a control group in a difference in difference framework in order to sweep out a potential common trend[8]. Because the non-completers (naturally) do not have a year of completion, we set their event time g=0 to the year they were 27 years old. This is to provide enough observations backwards to event time j=-6. As shown in the event figures, this choice is not very consequential, since the time paths of both non-completers and pre-completers[9] are quite flat. Assuming that the pre-trends are identical for late completers and non-completers, we may use the non-completers' change in NEET rates with age as a counterfactual change in NEET probabilities for the late completers. With a reference time of j=-6, we effectively compare the difference in NEET rates at age 21 and 28 for non-completers to the difference in NEET rates between the year after completion (j=1) and six years before completion (j=-6) for the late completers. We thus run the following specification of the event history analysis for non-completers:

$$(2)\ y_{itg} = \sum_{j\neq -6} {\beta_j}^{nc} \times \boldsymbol{I}[j = g] + FE_i + u_{it}$$

Where g is set to zero at the age of 27. The difference in difference estimator for the effect of late completion will be given by the difference $\beta_j - {\beta_j}^{nc}$. In this case we again identify the effect using pre and post NEET rates for the same individuals, but now by using the change for non-completers as a control.

[8] See Borusyak and Jaravel (2017) who show that a linear component of the pre-trend is not identifiable in the presence of unit and time effects in event history analyses. This problem can be solved by using the never treated to identify the pre-trends. This strategy is only valid under the assumption of a common pre-trend for completers and non-completers.

[9] "Pre-completers" denotes late completers in the period before completion

5.4 Comparing results

From the event history analysis, we may derive an estimator comparable to the estimator from the analysis of late completion on NEET rates at age 28. We complete our analysis with a discussion of the magnitude of selection using the Oster method on the difference between the estimates obtained from the different methods.

6 Regression results

6.1 Baseline results

In this paragraph, we report results from the estimation of NEET rates at age 28 on late completion (equation 1 above) with different specifications of the control vector X_i. In the Baseline model, we have only included year and gender indicators. In the next specification "w/ses", we have included both parents' level of education and place in the earnings distribution. After that, we add controls for the cluster groups that we have estimated based on early paths, and in the last column, we add controls for the full actual sequence of early paths.

Table 4: Changes in NEET-rates after upper secondary school completion by age of completion, measured at age 28.

	Norway				**Sweden**			
Completion	Baseline	w/SES	w/cl groups	w/sequence	Baseline	w/SES	w/cl groups	w/sequence
By age 22	-0,162	-0,151	-0,141	-0,133	-0,109	-0,115	-0,118	-0,118
	(-32,37)	(-29,3)	(-26,84)	(-24,81)	(-18,24)	(-19,34)	(-20,60)	(-20,39)
By age 23	-0,15	-0,138	-0,129	-0,118	-0,136	-0,135	-0,118	-0,117
	(-22,13)	(-19,76)	(-18,66)	(-17,06)	(-20,01)	(-19,79)	(-18,53)	(-18,48)
By age 24	-0,155	-0,145	-0,134	-0,123	-0,147	-0,144	-0,115	-0,113
	(-18,59)	(-16,93)	(-15,91)	(-14,71)	(-18,42)	(-17,98)	(-15,47)	(-15,31)
By age 25	-0,133	-0,121	-0,108	-0,099	-0,144	-0,14	-0,111	-0,107
	(-13,46)	(-11,87)	(-10,81)	(-10,01)	(-16,19)	(-15,78)	(-13,53)	(-13,09)
By age 26	-0,138	-0,134	-0,121	-0,112	-0,166	-0,162	-0,126	-0,122
	(-12,38)	(-11,66)	(-10,77)	(-10,03)	(-17,63)	(-17,25)	(-14,49)	(-14,18)
By age 27	-0,164	-0,157	-0,146	-0,137	-0,187	-0,182	-0,145	-0,14
	(-13,63)	(-12,67)	(-11,97)	(-11,33)	(-18,36)	(-17,98)	(-15,40)	(-15,02)
By age 28	-0,14	-0,132	-0,117	-0,107	-0,191	-0,188	-0,155	-14,8
	(-10,63)	(-9,62)	(-8,72)	(-8,09)	(-17,22)	(-17,01)	(-15,08)	(-14,58)
Year, gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

SES background		Yes	Yes	Yes		Yes	Yes	Yes
Early paths			Yes	Yes			Yes	Yes
	Denmark				**Finland**			
Completion	Baseline	w/SES	w/cl groups	w/sequence	Baseline	w/SES	w/cl groups	w/sequence
By age 22	-0,220	-0,203	-0,167	-0,154	-0,188	-0,178	-0,156	-0,146
	(-79,48)	(-71,94)	(-52,31)	(-44,04)	(-28,79)	(-27,34)	(-24,62)	(-22,69)
By age 23	-0,209	-0,194	-0,157	-0,145	-0,186	-0,176	-0,149	-0,142
	(-56,62)	(-52,43)	(-41,34)	(-37,12)	(-25,86)	(-24,5)	(-21,73)	(-20,51)
By age 24	-0,210	-0,198	-0,163	-0,152	-0,195	-0,182	-0,151	-0,139
	(-46,12)	(-43,42)	(-35,85)	(-33,26)	(-24,26)	(-22,74)	(-19,76)	(-18,45)
By age 25	-0,201	-0,189	-0,157	-0,144	-0,183	-0,173	-0,144	-0,137
	(-39,18)	(-36,77)	(-30,73)	(-28,54)	(-20,55)	(-19,53)	(-17,12)	(-16,5)
By age 26	-0,202	-0,192	-0,161	-0,147	-0,183	-0,177	-0,148	-0,137
	(-35,14)	(-33,42)	(-28,11)	(-25,98)	(-18,67)	(-18,11)	(-15,92)	(-14,96)
By age 27	-0,202	-0,192	-0,164	-0,154	-0,176	-0,170	-0,137	-0,128
	(-33,11)	(-31,53)	(-27,02)	(-25,55)	(-16,52)	(-16)	(-13,61)	(-12,88)
By age 28	-0,177	-0,168	-0,141	-0,130	-0,150	-0,143	-0,114	-0,102
	(-26,38)	(-25,20)	(-21,16)	(-19,73)	(-13,06)	(-12,53)	(-10,5)	(-9,53)
Year, gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SES background		Yes	Yes	Yes		Yes	Yes	Yes
Early paths			Yes	Yes			Yes	Yes

T-values, corrected by individual clustering, are in parentheses.

In all four countries, completing upper secondary schooling heavily lowers the probability of being a NEET when aged 28. The probabilities are from 11 to 20 percentage points lower but settle mostly around 16 percentage points. The coefficients are largest in Denmark and smallest in Sweden. Age of completion does not seem to matter, however. The lower NEET rates of those who complete are of about the same size and are not significantly different from each other in any of the four countries.

The estimates are surprisingly robust to the inclusion of extensive controls for socioeconomic background and early paths. Moreover, the pattern is quite similar across the four countries: Adding socioeconomic background reduces the estimated coefficients by around 1 percentage point. Adding careful controls of early paths adds 2-3 percentage points to the attenuation of the estimated effects. This result indicates that also late completion of upper secondary schooling plays an important role for the labor market inclusion of young dropouts. We are still left with the question of selection on unobservables, though.

6.2. Selection on unobservables?

To evaluate the importance of selection on unobservables, we apply the Oster (2019) procedure. This procedure uses movements in both the coefficient of interest and in R-squared when adding covariates to assess how large the selection on unobservables must be for the relationship between upper secondary school completion and lower NEET rates to fall to 0.

Table 5: Test of selection on unobservables using Oster (2019). The effect of having completed upper secondary schooling on the probability of being a NEET, measured at age 28.

	(1) Baseline effect	(2) Controlled effect	(3) Identified set	(4) δ for β=0 given R_max	(5) R_max
Norway	-0,150	-0,122	[-0,122, -0,110]	5,33	0,143
Sweden	-0,149	-0,124	[-0,124, -0,114]	6,88	0,267
Denmark	-0,207	-0,148	[-0,148, -0,093]	1,74	0,222
Finland	-0,184	-0,137	[-0,137, -0,115]	4,26	0,270

Table 5 reports the results from using the Oster procedure. The baseline effects in column (1) are similar to the Baseline specification in Table 1, controlling only for year and gender, but this time including an indicator for whether the individual has completed upper secondary schooling instead of estimating a separate coefficient for each age of completion. The controlled effect in column (2) is similar to the last columns in Table 1, where we included the most detailed controls for socioeconomic background and early paths (fixed effects for each full sequence).

Comparing the baseline effect and the controlled effect, we see that the estimated effect is attenuated by 17 percent in Sweden, 19 percent in Norway, 26 percent in Finland and 29 percent in Denmark. The question now is how much more the effect would be attenuated (or in general changed), if we could add the remaining unobservables to the equation.

Oster starts out by setting an a-priori limit on how much more of the variation in NEET rates that we expect can be explained by unobservable factors (i.e. a limit on R-squared). Oster recommends setting the limit on R-squared to 1,3 times as much as the specification with the full control vector can explain. This gives us a maximum R-squared ranging from 0,143 in Norway to

0,27 in Finland. The estimates arising from the Oster procedure in this case are given by the highest number in the "identified set" as reported in colum (3). These numbers are remarkably similar across countries: 11.0, 11,4, 9.3, and 11.5.

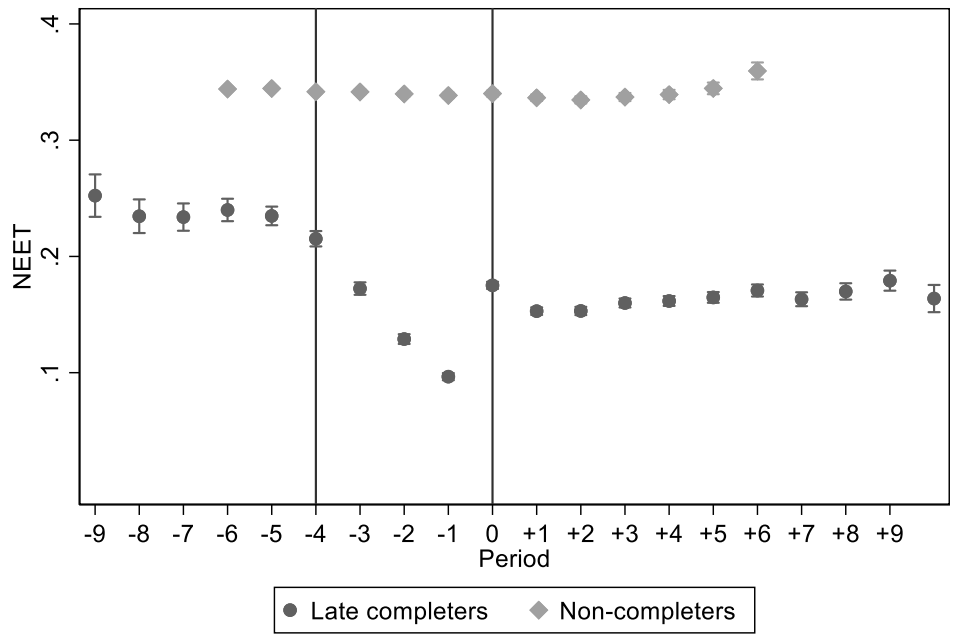
Column (4) reports how large the selection on unobservables must be for the effect of upper secondary school completion to be zero. We see that the selection must at least be from 1,7 to 5,3 times larger than the selection that we pick up in the observable variables. Considering the extensive set of relevant controls that are included in our specification, this result indicates that our results are quite robust.

6.3. Event history analysis

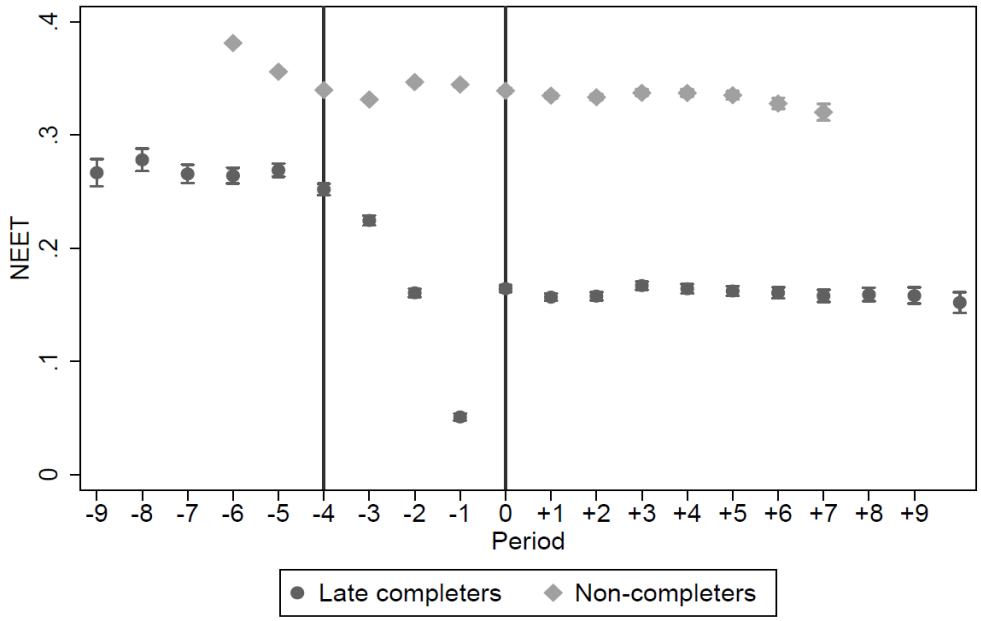
To probe more into the selection issue, we now turn to an event history analysis. An event history analysis uses variation over time, comparing outcomes measured at different years both before and after completion. We also include individual fixed effects to control for all fixed observed and unobserved characteristics. The results of the event history analysis are best illustrated using figures. Figure 2 shows the probability of being a NEET in the years preceding completion and in the years following completion with a 95 percent confidence interval. We may think of the event history analysis as an average of the patterns revealed in Figures A2 to A4 in the Appendix.

Figure 2. The probability of being a NEET before and after completion. Late- and non-completers

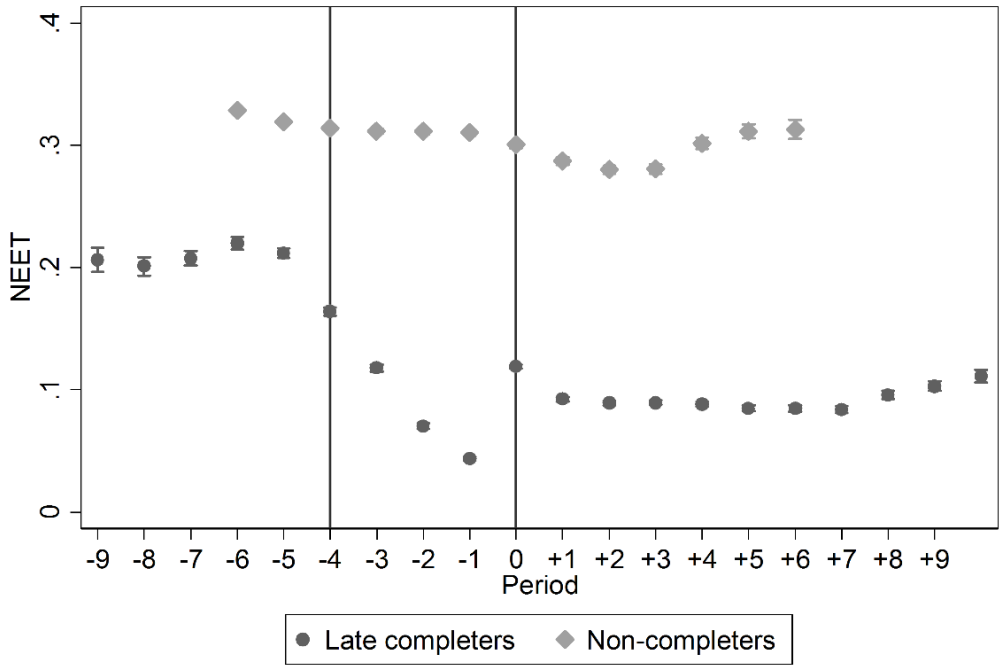
Norway



Sweden



Denmark



Finland

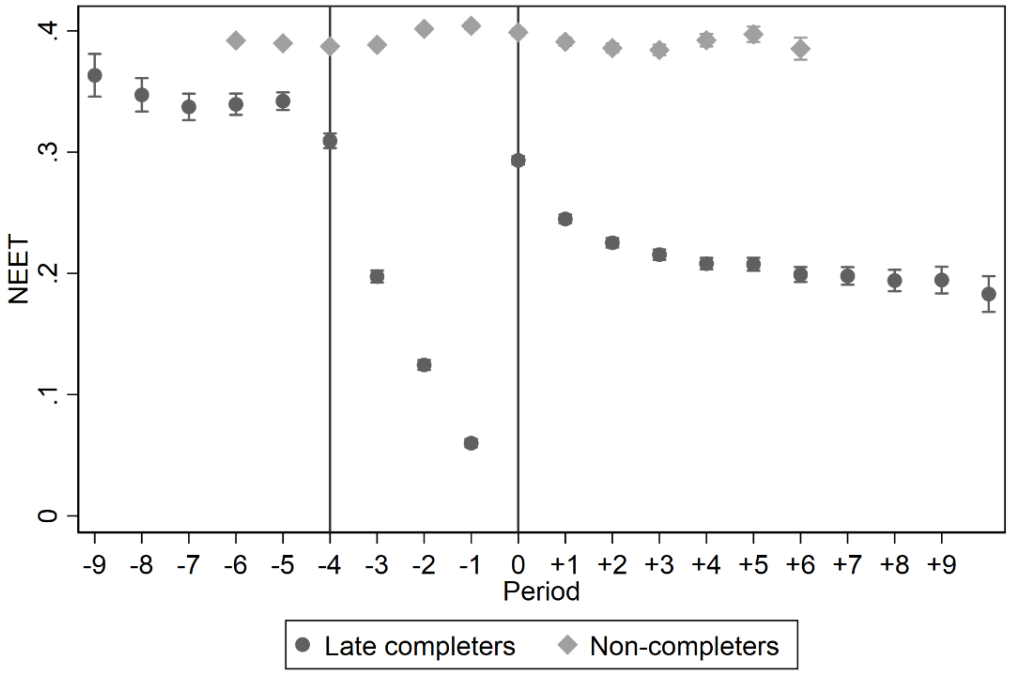


Figure 2 and the second set of specifications in Table 6 show the results when we estimate the probability of being a NEET, including only years before and after completion, and individual fixed effects in the model.[10] Consider first the black markers showing the event history trajectories for late completers. We find that the NEET rates among the late completers are quite stable, for example around 24 percent for Norway, up through five years before completion. Then they drop considerably as we approach the year of completion. This decline reflects that these young people are students in the years preceding completion. Indeed, figures A5-A8 in the appendix showing both employment and student rates in a similar diagram display a notable spike in the probability of being a student in the years immediately preceding completion. Because of this pattern, we compare the estimates for the outcome 6 years before completion[11] to the outcomes in all the years following completion. Because the average age of late completion is around 23 years, the estimate 5 years after completion is a comparable estimate to the previous OLS estimates at age 28. We will therefore comment specifically upon the estimate in event time j=5[12]. We note that there is a marked drop in NEET rates after completion, with NEET rates approximately halved compared to the pre-completion situation.

A distinct feature of our results also is that the change in NEET rates is reasonably flat in both the period before and after late completion. Comparing the NEET rate 6 years before completion to the NEET rate 5 years after completion gives us an estimated decline of 0.08 or 8 percentage points (pp) in Norway, of 10 pp in Sweden, 14 pp in Denmark and 13 pp in Finland. These coefficients are very similar to the estimates obtained using cross-sectional comparisons of NEET rates at age 28 in specification (1) of Table 1. Note that now we use the late completers in the period before completion as a comparison group.

[10] We have tested several specifications of the control set, and note that the event history effects are only mildly smaller in absolute terms when we include individual fixed effects, compared to including parental background variables, a complete set of sequence dummies to describe the early path from 16 to 20 years of age, and age and year dummies. We have thus chosen to show the FE specifications even though they do not identify the event history conditional on year and age, which are not separately identifiable from the event history within individuals for both late completers and non-completers (see e.g. Borusyak and Jaravel (2019)).

[11] The estimates for event time j=-9 to j=-5 are, however, economically and for the most part also statistically insignificant, so our results are not sensitive to the choice of reference period.

[12] The estimates in event time j=1 to j=10 are quite stable in size, and so j=5 is quite representative of an average estimate.

In the first set of specifications of Table 6 we have included specifications with control for the full set of observables, including gender, socioeconomic background, and early paths with the full set of dummies for each sequence, instead of the individual fixed effects. In this specification, year dummies are included as well. Comparing the results from the specifications with and without controls for unobservables, we find that adding unobservable fixed effects does not alter the overall picture much. We find a 1 pp attenuation of the estimated effects for Denmark and Norway; the NEET rate 6 years before completion compared to the NEET rate 5 years after completion gives us an estimated decline of 8 vs 9 pp in Norway, and 14 vs 15 pp in Denmark when introducing individual fixed effects. For Sweden and Finland, we actually find somewhat larger effects when controlling for unobservables; 10 vs 7 pp in Sweden, and 13 vs 12 pp in Finland.

Table 6: Event-study estimates (t-values)

Event time	Full set of control variables				Individual fixed effects			
	Norway	Sweden	Denmark	Finland	Norway	Sweden	Denmark	Finland
-9	0,02	0,00	0,01	0,03	0,01	0,00	-0,01	0,02
	(2,03)	(-0,62)	(1,99)	(2,72)	(1,20)	(0,38)	(-2,44)	(2,43)
-8	0,00	0,01	0,00	0,01	-0,01	0,01	-0,02	0,01
	(-0,09)	(1,89)	(-0,51)	(1,58)	(-0,63)	(2,36)	(-4,24)	(0,96)
-7	0,00	0,00	-0,01	0,00	-0,01	0,00	-0,01	0,00
	(0,47)	(0,25)	(-1,83)	(0,23)	(-0,83)	(0,30)	(-3,23)	(-0,32)
-5	-0,01	0,01	-0,01	0,00	-0,01	0,00	-0,01	0,00
	(-1,18)	(1,60)	(-2,77)	(0,25)	(-0,84)	(1,08)	(-2,58)	(0,46)
-4	-0,02	0,00	-0,05	-0,03	-0,02	-0,01	-0,06	-0,03
	(-3,72)	(-0,79)	(-17,15)	(-4,87)	(-4,31)	(-2,87)	(-18,86)	(-5,71)
-3	-0,07	-0,03	-0,10	-0,14	-0,07	-0,04	-0,10	-0,14
	(-10,90)	(-6,17)	(-32,51)	(-24,56)	(-12,31)	(-9,73)	(-35,95)	(-27,99)
-2	-0,11	-0,09	-0,14	-0,21	-0,11	-0,10	-0,15	-0,21
	(-19,17)	(-21,60)	(-50,49)	(-38,44)	(-20,91)	(-26,21)	(-54,51)	(-43,54)
-1	-0,14	-0,20	-0,17	-0,27	-0,14	-0,21	-0,18	-0,28
	(-25,68)	(-47,18)	(-63,01)	(-51,59)	(-27,68)	(-55,31)	(-65,92)	(-57,69)
0	-0,07	-0,08	-0,10	-0,04	-0,06	-0,10	-0,10	-0,05
	(-12,65)	(-19,54)	(-37,68)	(-8,23)	(-12,52)	(-25,89)	(-37,77)	(-9,56)
1	-0,10	-0,08	-0,14	-0,09	-0,09	-0,11	-0,13	-0,09
	(-16,72)	(-18,91)	(-48,24)	(-16,76)	(-16,64)	(-27,38)	(-47,33)	(-19,26)
2	-0,10	-0,08	-0,14	-0,11	-0,09	-0,11	-0,13	-0,11

	(-16,67)	(-17,92)	(-49,90)	(-19,97)	(-16,43)	(-26,64)	(-47,91)	(-22,89)
3	-0,09	-0,07	-0,15	-0,12	-0,08	-0,10	-0,13	-0,12
	(-15,32)	(-15,58)	(-49,65)	(-20,98)	(-14,93)	(-23,85)	(-47,27)	(-24,40)
4	-0,09	-0,07	-0,15	-0,12	-0,08	-0,10	-0,13	-0,13
	(-14,84)	(-14,74)	(-48,18)	(-20,97)	(-14,38)	(-24,07)	(-47,00)	(-25,25)
5	-0,09	-0,07	-0,15	-0,12	-0,08	-0,10	-0,14	-0,13
	(-13,96)	(-13,96)	(-46,66)	(-19,38)	(-13,57)	(-23,95)	(-47,48)	(-24,57)
6	-0,08	-0,07	-0,15	-0,12	-0,07	-0,10	-0,14	-0,14
	(-12,38)	(-12,97)	(-44,39)	(-18,81)	(-12,20)	(-23,64)	(-46,38)	(-25,01)
7	-0,09	-0,07	-0,15	-0,12	-0,08	-0,11	-0,14	-0,14
	(-13,79)	(-12,29)	(-42,79)	(-16,71)	(-13,03)	(-23,33)	(-45,44)	(-23,76)
8	-0,08	-0,06	-0,14	-0,12	-0,07	-0,11	-0,12	-0,15
	(-12,28)	(-10,36)	(-39,46)	(-15,70)	(-11,30)	(-22,01)	(-39,91)	(-22,41)
9	-0,08	-0,06	-0,15	-0,12	-0,06	-0,11	-0,12	-0,14
	(-11,03)	(-9,73)	(-37,10)	(-13,79)	(-9,06)	(-20,39)	(-35,19)	(-19,80)
10	-0,11	-0,07	-0,15	-0,12	-0,08	-0,11	-0,11	-0,16
	(-12,18)	(-10,16)	(-32,70)	(-11,48)	(-9,74)	(-18,98)	(-28,84)	(-17,67)
Observations	354504	384526	913073	345181	354504	384526	913073	345181

Note: In the first set of specifications, the full set of control variables includes gender, dummies for year of observation, indicators for socioeconomic background and dummies for the full set of sequences of early paths, all of which are absorbed by the individual fixed effects in the second set of specifications.

There may be a change in NEET rates with age that we cannot account for in the within-completer estimation. As a comparison group to account for such life-cycle variation in NEET rates, we may use the group of non-completers. We have estimated the probability of being a NEET for non-completers by age, and drawn them in as if their potential completion year was at age 26 (grey markers in Figure 2). Since the estimated NEET rates are quite flat by age, this choice of age is not very consequential, and is basically made to give us a sufficient number of pre-completion years. We note that the NEET rate of non-completers of, for example, around 34 percent for Norway is higher than the pre-event NEET rates of the completers (as shown in Figure 2 by the difference in the level of NEET rates in event time j=-6), reflecting selection on both observables and unobservables.

Table A1 in the Appendix shows the estimated coefficients for non-completers. We find that the coefficients are close to zero, and move somewhat up or down around zero with age. On average, however, they are close to zero (an exception is Sweden, but this may be because the NEET rates of Swedish non-completers are especially high in event-time j=-6 and j=-5; for the rest of the period, the development is quite flat as in the other countries). We conclude, therefore,

that the within-completer group estimates in Figure 2 and Table 6 are robust to a control group comparison.

To sum up the results from the event history analysis: the probability of being a NEET significantly falls in all four countries after completion, and the estimates are of similar size to the cross-sectional evidence from model (1).

We may use the event history results to provide another Oster (2019) inspired assessment of the importance of selection for the average controlled effect at age 28, as estimated in Table 4. Table 8 first provides the controlled effects as estimated with the full set of controls in Table 4. Next we include in the table the estimate from the event study, using the comparison between years -6 and +5. If we take the event history results as the true effect, since it is based on comparing outcomes over time within the same individuals, the reported delta suggests that selection on unobservables has to have 2 to 3 times larger impact than the observables in Norway and Sweden, but of only minor size in Denmark and Finland.

Table 8 How large must the selection on unobservables be for the Event-study estimates in t=5 to be the true effects.

	(0) Controlled effect	(1) Event-study (ES) effect	(2) δ for β=ES given R_max	(3) R_max
Norway	-0,122	-0,08	2,93	0,143
Sweden	-0,124	-0,10	2,18	0,267
Denmark	-0,148	-0,14	0,18	0,222
Finland	-0,137	-0,13	0,32	0,270

Note: Controlled effect is the estimated effect using the specification with the full set of controls (column 2, Table 4). ES-effect is the average effect measured at 5 years after late completion.

6. Conclusion

Visual inspection of the descriptive paths between 21 and 30 years of age for those who had not completed upper secondary schooling by the age of 21 shows the following patterns in all four Nordic countries under scrutiny:

1 There seems to be a positive jump in employment/education status even after late completion. This jump is both economically and statistically significant.

2 There are few discernable differences between the positive gains depending on the age of late completion.

3 Pre-graduation patterns do not look very different from the outcomes for those who do not complete, at least not before turning 31.

The two latter parts of this pattern are surprising; they suggest that it does not matter at what age you complete upper secondary schooling, and that there is little evidence of selection in our data.

A more careful study of regression results confirms all three observations. There is no evidence of different coefficients from different years of late completion. There is very little evidence of selection: the results do not change when adding an unusually rich set of controls, including a cluster analysis and complete control for the sequences from 16-20 years of age leading up to a "dropout status" at age 21. An analysis following the Oster (2019) assumptions supports a limited role of selection on unobservables, as do event history analyses, both when relying on within-individual changes before and after completion, and when comparing to the non-completers. Since we do not have a natural experiment, we are still reluctant to conclude that we have pin pointed a causal effect. However, it seems clear that our data does not give us any reason to suspect selection to be a severe problem.

The results are strikingly similar across all four countries and, thus, are not likely to be due to idiosyncrasies related to a country's particular system of education or labor market policies. Our results indicate that policies to promote young dropouts to re-enter education at a later age may be an attractive way to raise educational levels and improve employment prospects. Even if it always pays to finish early, just because there are more years to reap the gains, we show that there are large gains to completing upper secondary schooling also at a later age.

The opportunity to complete late could also compensate young people for some of the negative effects of growing up in a less resourceful family. Low socioeconomic status influences the probability of dropping out. Our results indicate, however, that dropouts have no lesser

abilities to reap the gains from completing upper secondary schooling at a later age. Policies promoting them to complete – even after the normal age – can therefore be expected to help disadvantaged youth to a more equal chance in the labor market.

References

Albæk, K., Asplund, R., Barth, E., Lindahl, L., Simson, von, K., Vanhala, P. (2015). *Youth Unemployment and Inactivity: A Comparison of School-to-Work Transitions and Labour Market Outcomes in Four Nordic Countries* (TemaNord 2015:548). Copenhagen: Nordic Council of Ministers.

Angrist, Joshua D., and Alan B. Keueger. "Does compulsory school attendance affect schooling and earnings?." *The Quarterly Journal of Economics* 106.4 (1991): 979-1014.

Athey, Susan and Guido W. Imbens (2018) "Design-based Analysis in Difference-in-Difference Settings with Staggered Adoption" manuscript, Stanford University.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the US Labor Market." *American Economic Review*, 96(2): 189–94.

Autor, David H., and Lawrence F. Katz. 1999. "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics,*

Borusyak, Kirill and Xavier Jaravel (2017) "Revising Event Study Designs" manuscript, Harvard University

Dearden, Lorraine, et al. "Conditional cash transfers and school dropout rates." *Journal of Human Resources* 44.4 (2009): 827-857.

Goos, Maarten, Alan Manning, and Anna Salomons. "Job polarization in Europe." *American economic review* 99.2 (2009): 58-63.

Harmon, Colm, and Ian Walker. "Estimates of the economic return to schooling for the United Kingdom." *The American Economic Review* 85.5 (1995): 1278-1286.

Murnane, Richard J. 2013. "U.S. High School Graduation Rates: Patterns and Explanations." *Journal of Economic Literature,* 51 (2): 370-422.

Philip Oreopoulos,(2007). "Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling", *Journal of Public Economics*, 91 (11-12): 2213-2229

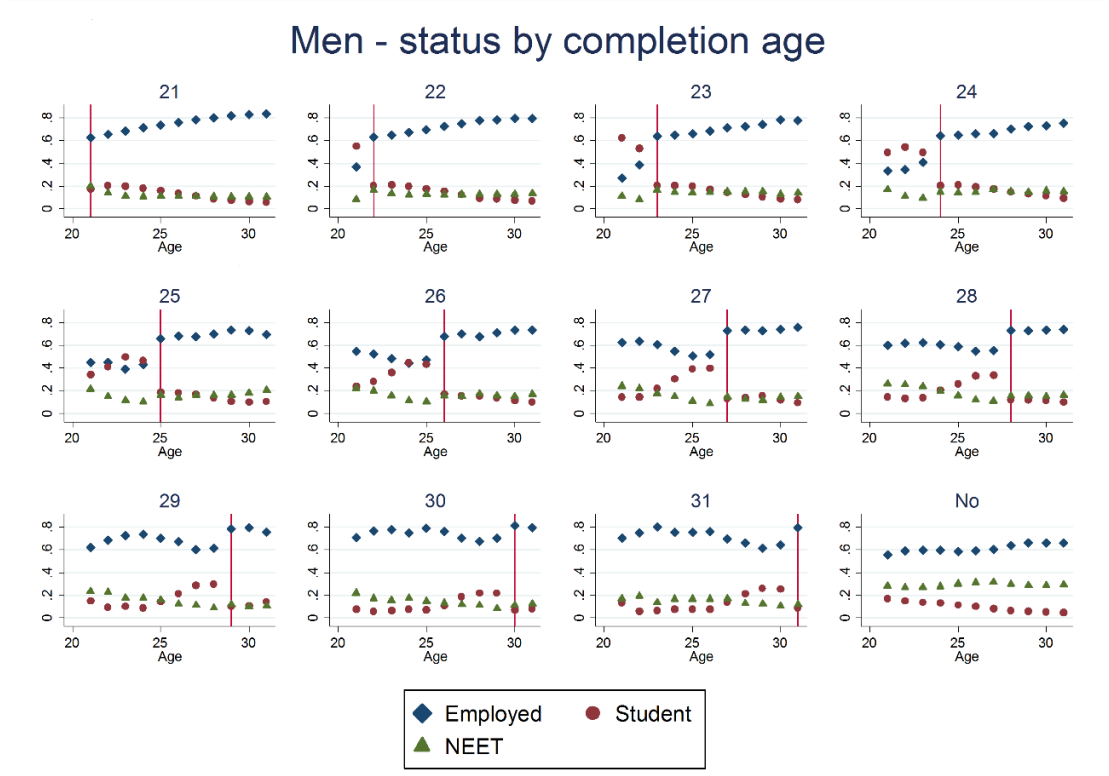
Oster, Emily (2019). “Unobservable Selection and Coefficient Stability: Theory and Evidence”, *Journal of Business & Economic Statistics*, 37:2, 187-204

OECD (2014), “*Indicator A2: How many students are expected to complete upper secondary education?*”, in Education at a Glance 2014: OECD Indicators, OECD Publishing. http://dx.doi.org/10.1787/888933115122

OECD (2017), *Education at a Glance 2017: OECD Indicators*, OECD Publishing, Paris. http://dx.doi.org/10.1787/eag-2017-en

Appendix

Figure A.1 Early career trajectories, late completers versus non-completers. Norway



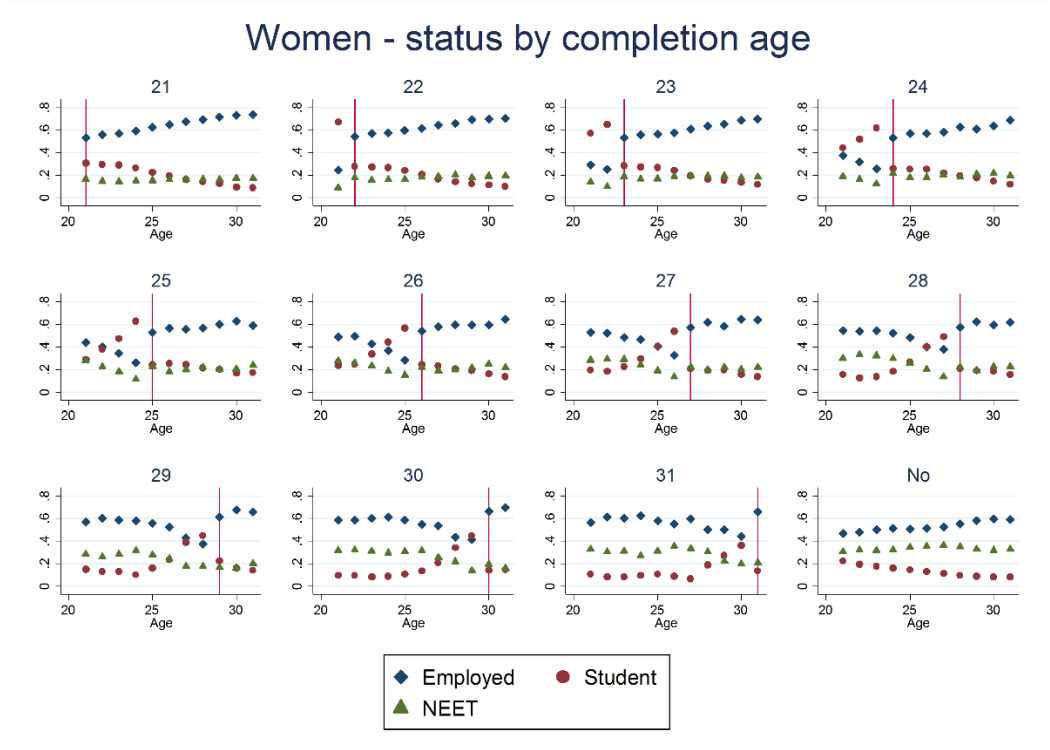
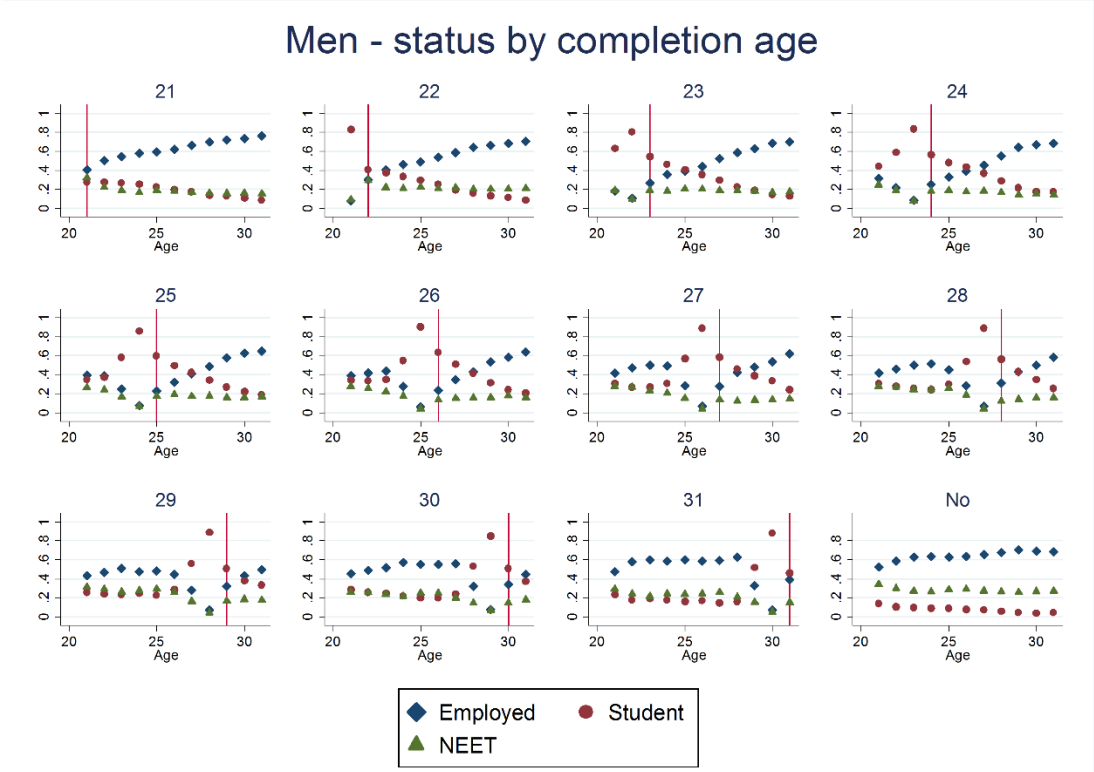


Figure A.2 Early career trajectories, late completers versus non-completers. Sweden



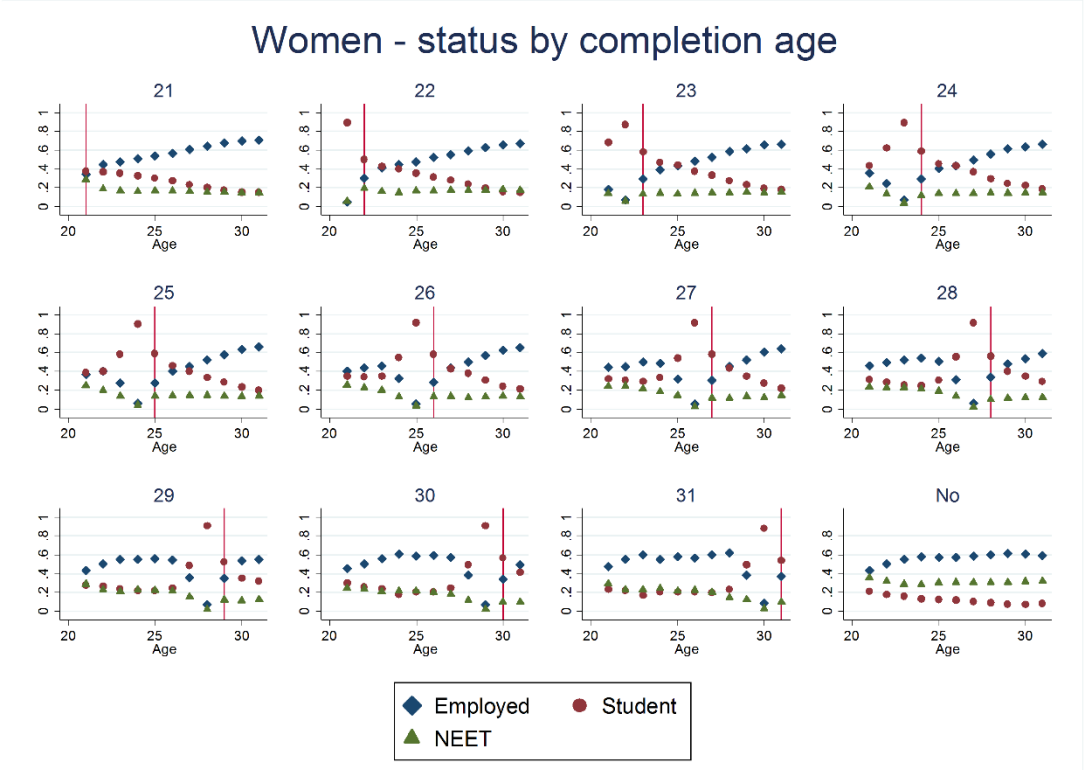


Figure A.3 Early career trajectories, late completers versus non-completers. Denmark

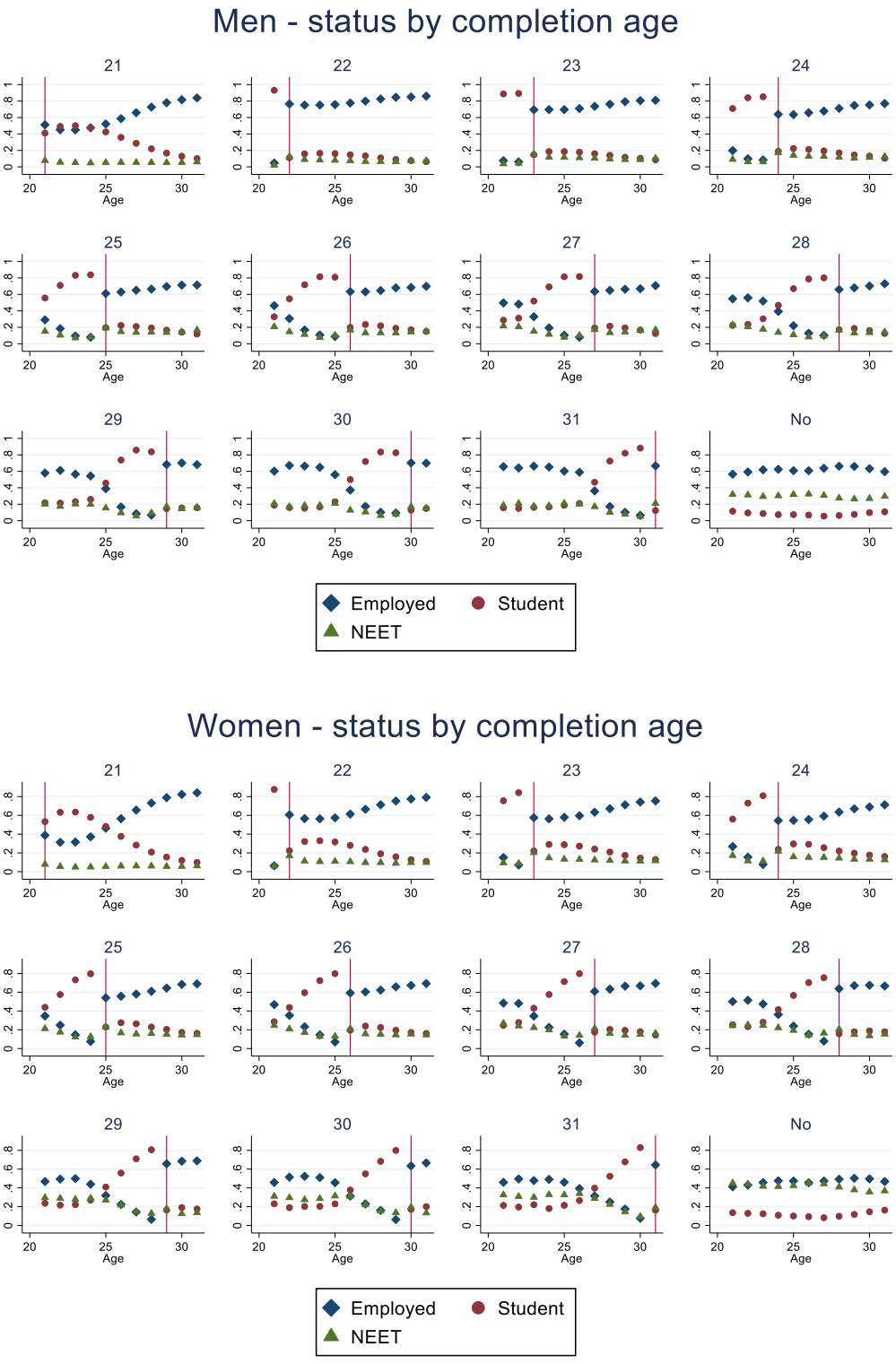
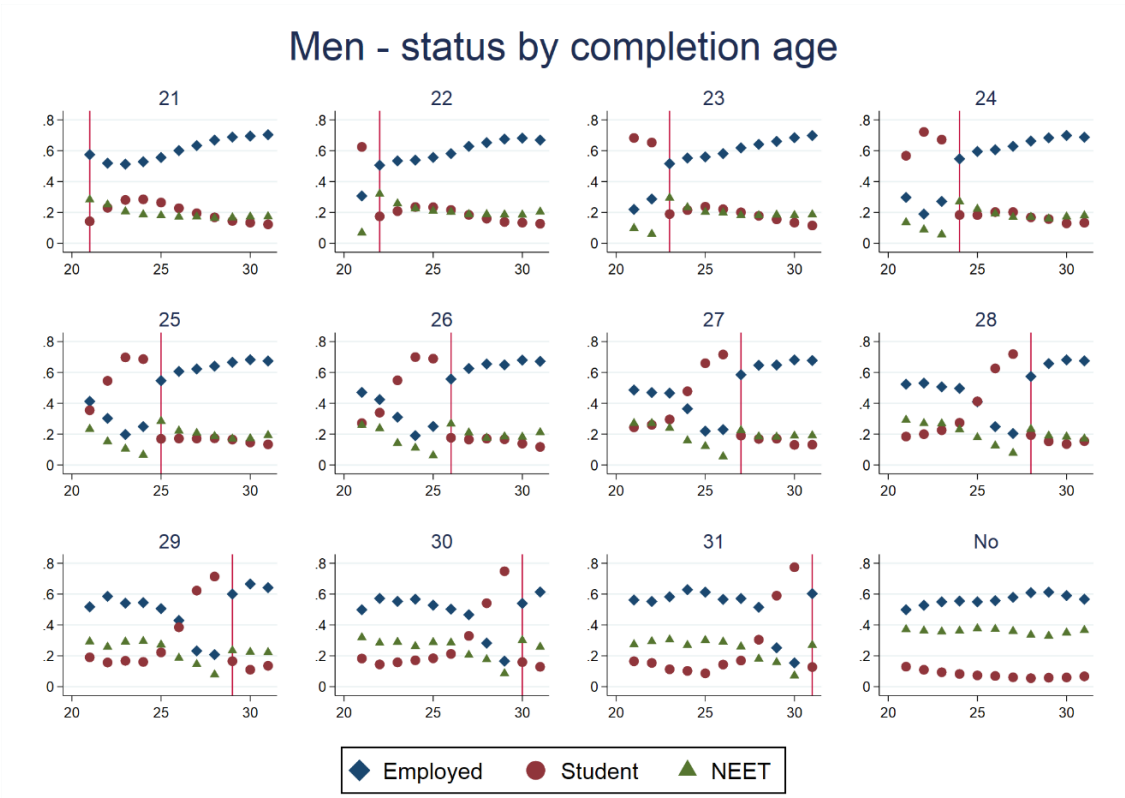


Figure A.4 Early career trajectories, late completers versus non-completers. Finland



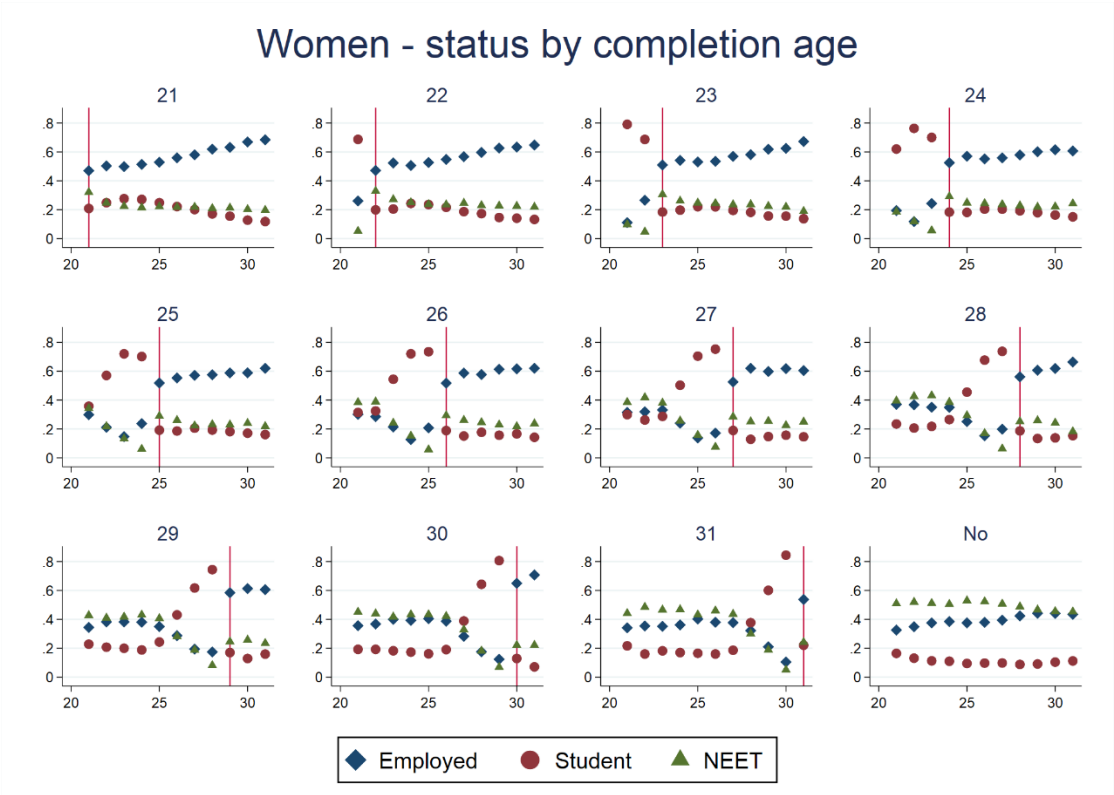
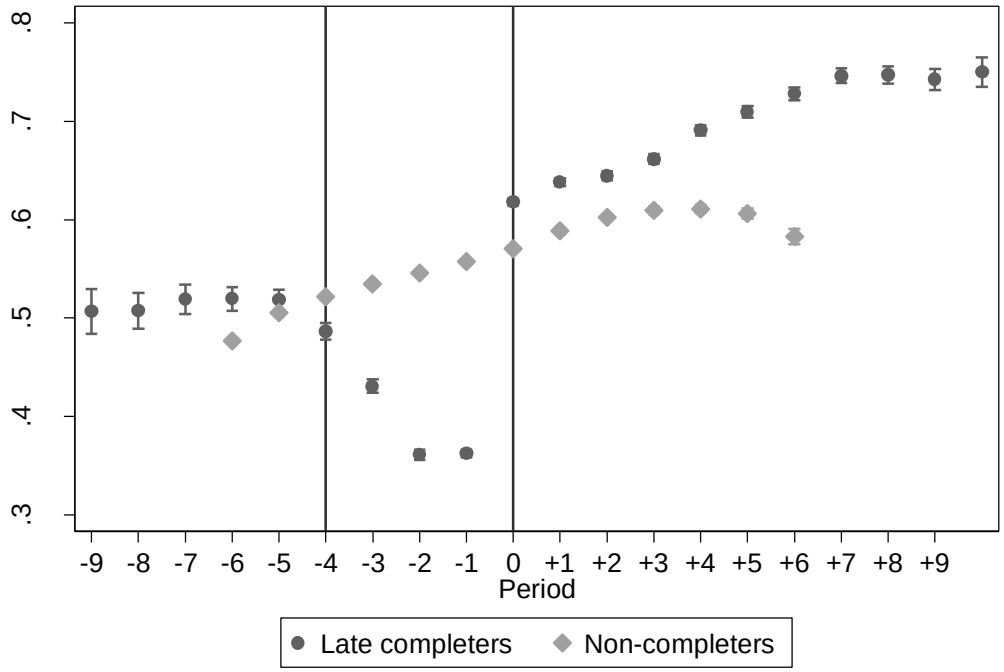


Figure A.5 Event history trajectories for employment and study: Norway



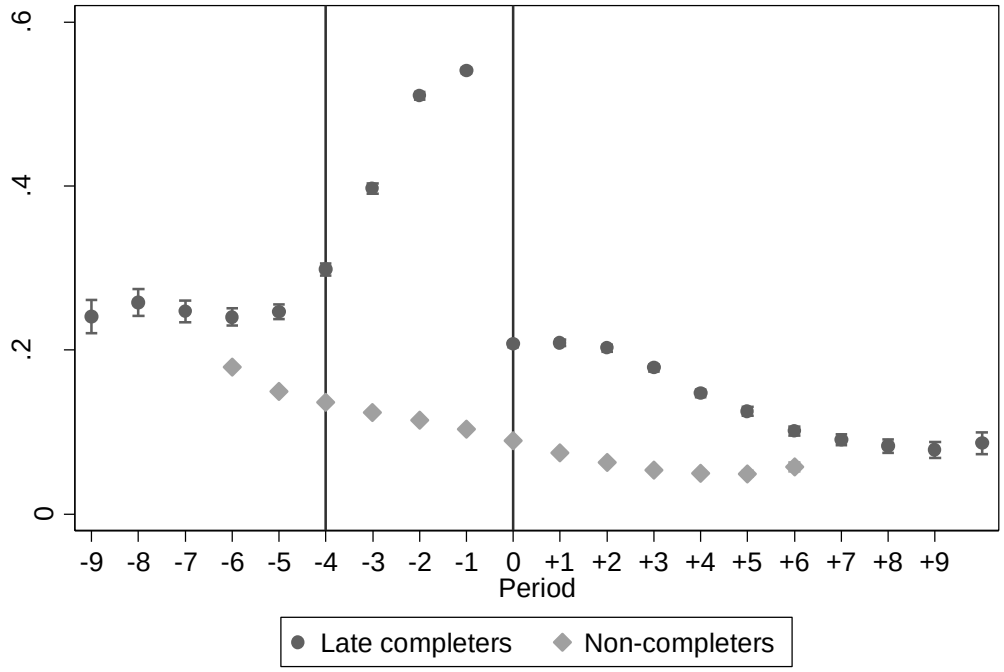
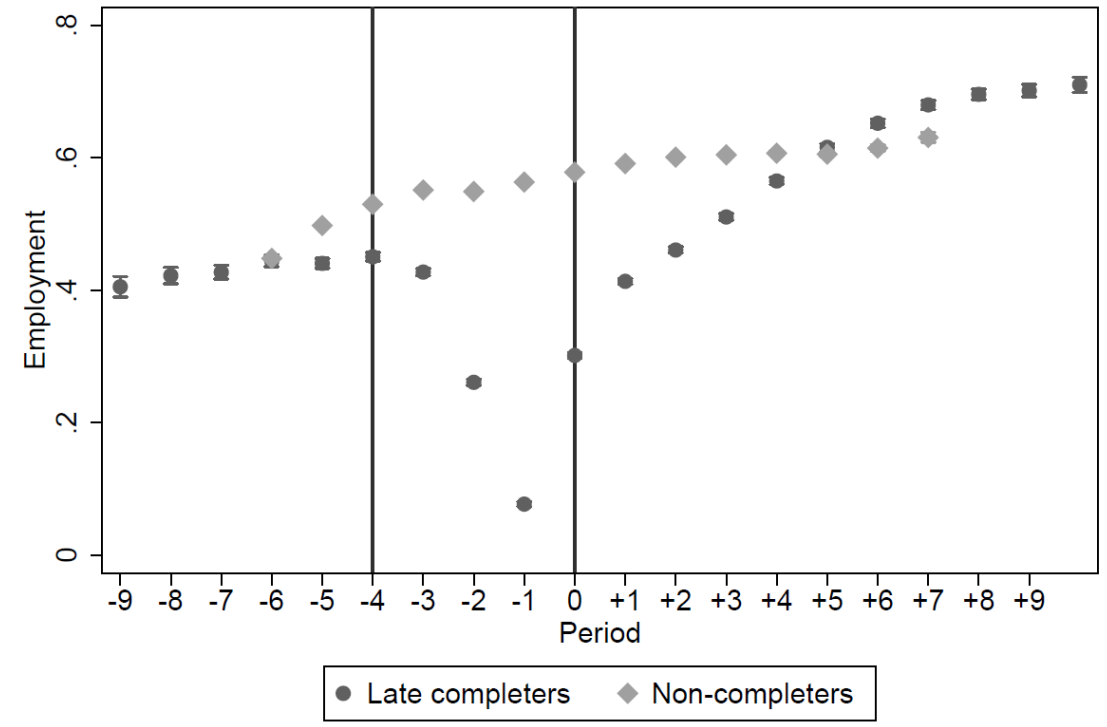


Figure A.6 Event history trajectories for employment and study: Sweden



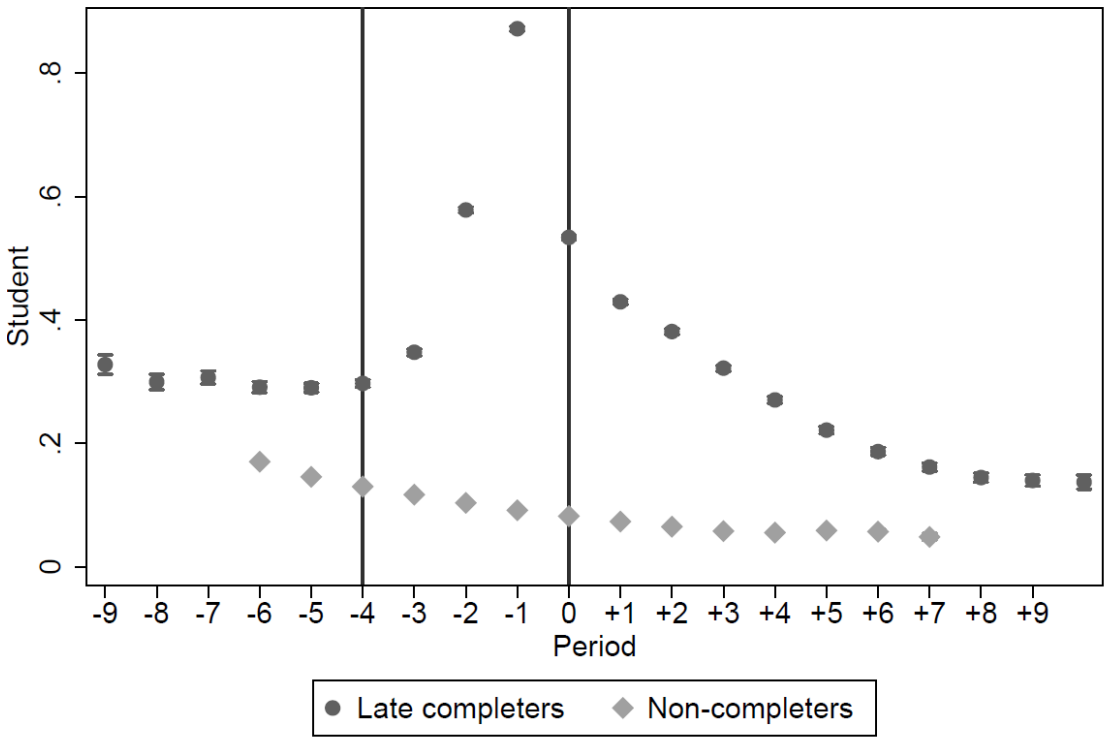
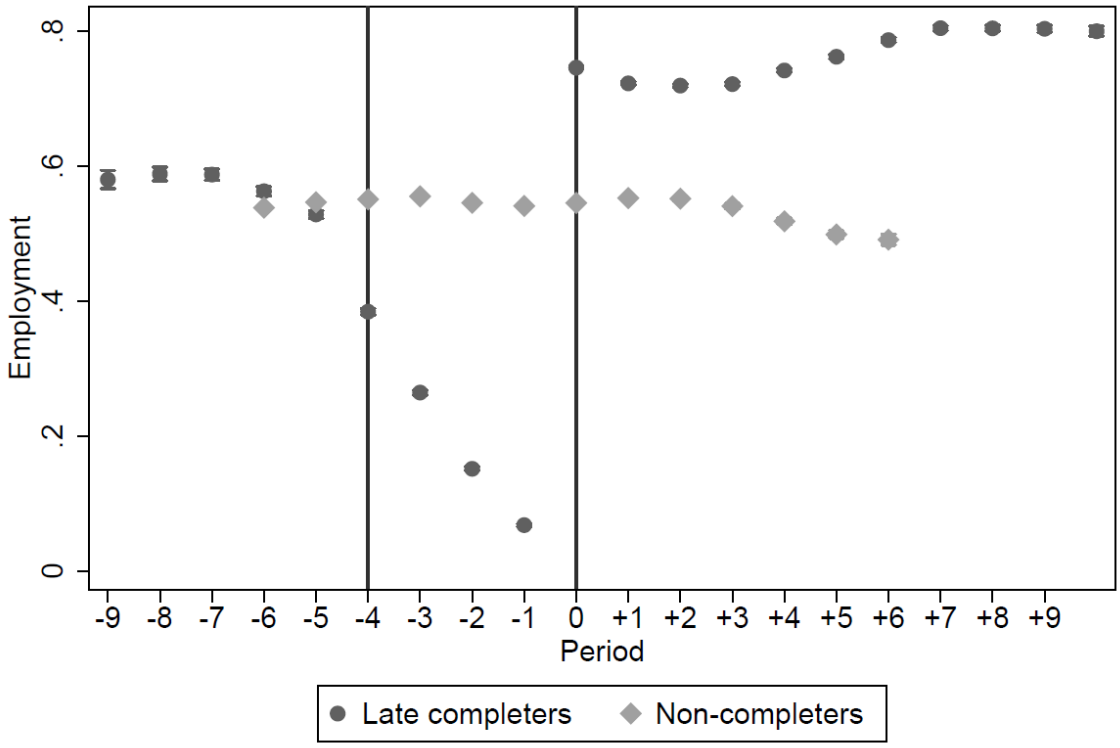


Figure A.7 Event history trajectories for employment and study: Denmark



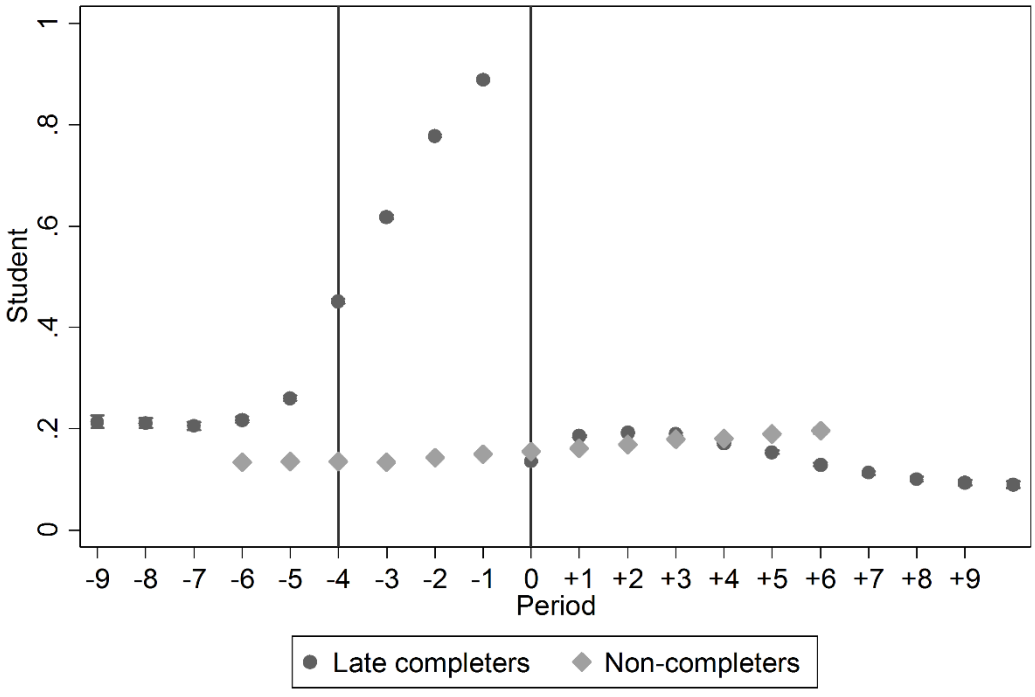
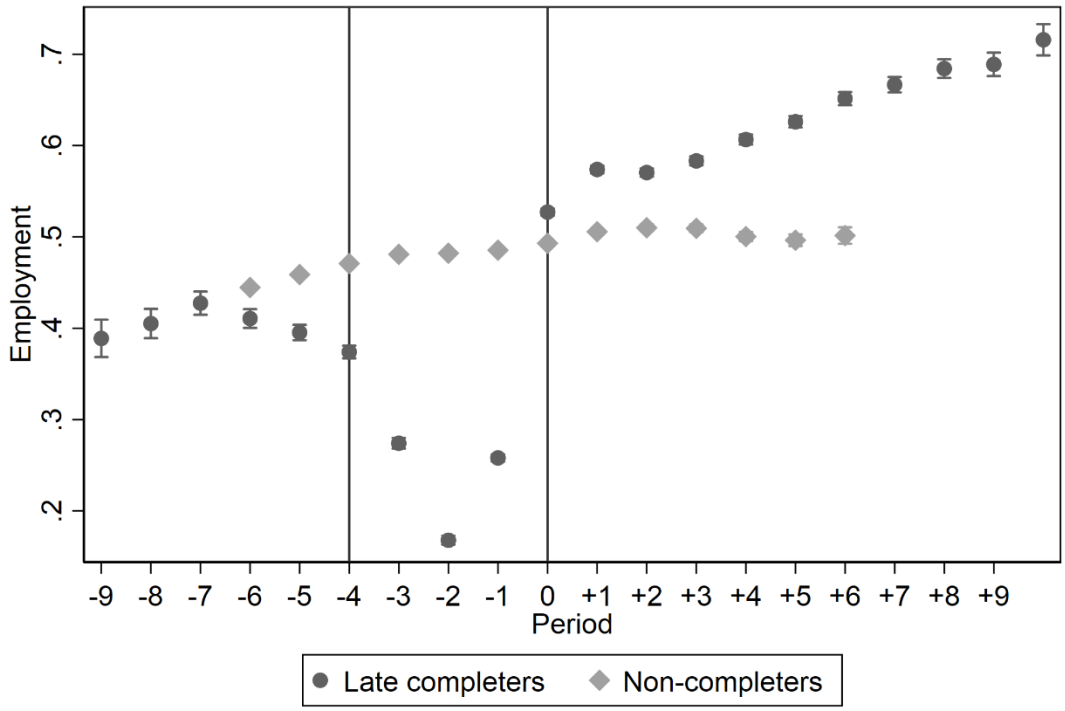


Figure A.8 Event history trajectories for employment and study: Finland



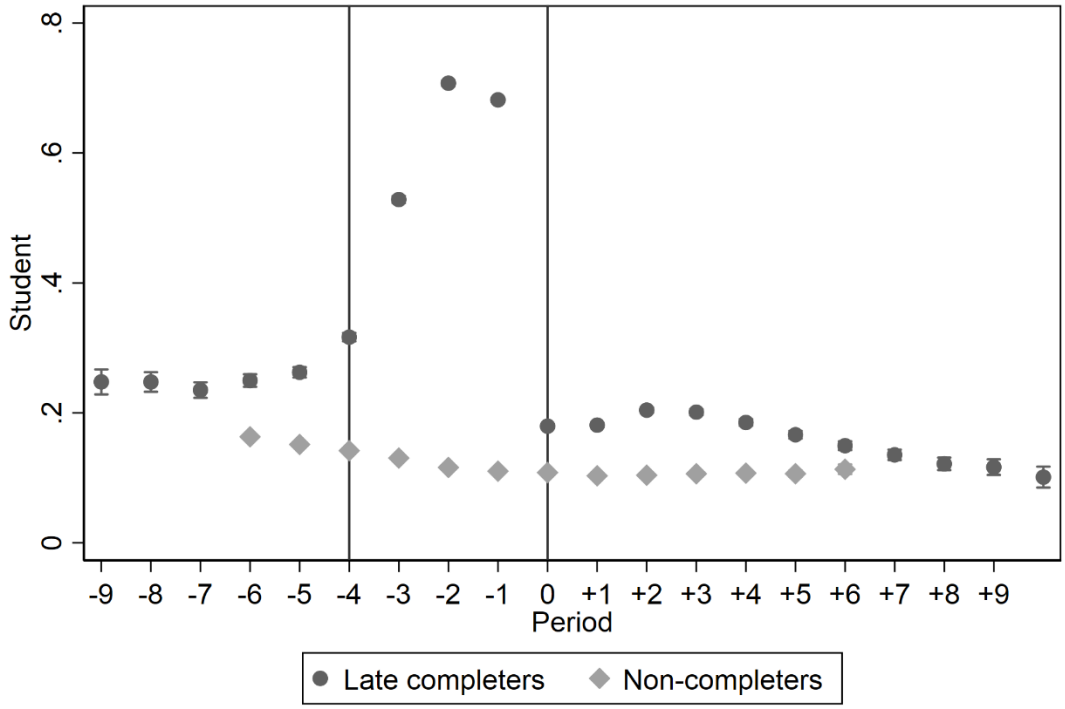


Table A1 Event history coefficients (t-values), non-completers

	Norway	Sweden	Denmark	Finland
-5	0,00	-0,03	-0,01	0,00
	(0,36)	(-21,41)	(-6,81)	(-1,48)
-4	0,00	-0,04	-0,01	0,00
	(-1,67)	(-33,81)	(-9,82)	(-2,96)
-3	0,00	-0,05	-0,02	0,00
	(-1,72)	(-39,05)	(-11,23)	(-2,09)
-2	0,00	-0,03	-0,02	0,01
	(-2,77)	(-25,93)	(-10,71)	(5,41)
-1	-0,01	-0,04	-0,02	0,01
	(-3,48)	(-26,45)	(-10,73)	(6,49)
0	0,00	-0,04	-0,03	0,01
	(-2,35)	(-29,06)	(-15,64)	(3,39)
1	-0,01	-0,05	-0,04	0,00
	(-4,25)	(-30,49)	(-21,90)	(-0,49)
2	-0,01	-0,05	-0,05	-0,01
	(-4,97)	(-29,77)	(-23,63)	(-2,72)
3	-0,01	-0,04	-0,05	-0,01
	(-3,37)	(-25,46)	(-21,37)	(-3,08)
4	0,00	-0,04	-0,03	0,00
	(-2,08)	(-23,53)	(-10,65)	(0,08)
5	0,00	-0,05	-0,02	0,01
	(0,19)	(-22,01)	(-5,65)	(1,47)
6	0,02	-0,05	-0,02	-0,01
	(4,02)	(-20,76)	(-3,69)	(-1,42)
Observations	1097245	1260489	960706	694118

The coefficients measure the difference in probability of being NEET in any period from the probability of being NEET 6 years before age 27 (t=0).