Early bird caught the worm?

The effect of a student aid reform on time-to-degree

Susanna Sten Gahmberg*

December 15, 2014

Abstract

Delayed graduation from higher education is an issue that has received much attention in recent years. This paper studies how students responded to a student aid reform that aimed at increasing the share of students who graduated on time in Norway in the early 1990's. The reform gave students in some study programs financial incentives to graduate on time by offering a restitution of their study loan. Using a difference in difference strategy we find that one additional year of treatment, defined as the number of years spent studying in the reform period, increased the probability of graduating on time on average by 1.3-1.5 percentage points compared to a baseline probability timely graduation of 14 percent. The reform was also successful in decreasing average delay on the treated students: one year of treatment reduced average delay by 0.065 years. There was, however, considerable heterogeneity in the take-up of treatment. The treatment effect was largely driven by high ability students and students with highly educated parents. Further, the treatment effect varied by study program. This casts some doubt on the efficiency of the reform design and highlights the importance of considering student heterogeneity when designing policies to affect student behavior.

JEL-Classification: I23, J24

Keywords: higher education, study progression, time-to-degree, financial incentives, student aid

*Norwegian School of Economics, Department of Economics, Helleveien 30, NO-5045 Bergen, Norway. e-mail: susanna.sten@nhh.no.

1 Introduction

Delayed graduation in higher education has received increasing attention from both policy makers and researchers in recent years.[1] For instance, Brunello and Winter-Ebmer (2003) report that the share of students who expect to delay graduation ranges from close to zero in the UK and Ireland to about 31 % in Sweden and Italy. In Norway, only 29 % of the graduates from 5-year graduate programs and 44 % of the graduated from 3-year undergraduate programs completed on expected time in 2011-2012 (Statistics Norway, 2013). In addition, delays are often quite long (Häkkinen and Uusitalo, 2003; Deschryvere, 2009; Statistics Norway, 2013).[2] As a consequence, policies to reduce delays have been implemented in a number of countries. For example, Norway implemented a progression dependent student aid system in 2002 (Ministry of Education and Research, 2002; St.meld. nr.7, 2008) and Finland is currently looking to reform their student aid system to more efficiently promote progression (Ministry of Education and Culture, 2010, 2012). In Germany and Italy higher tuition fees for those who are delayed have been introduced (Heineck et al., 2006; Garibaldi et al., 2012).

The aim of this study is to analyze one of the earliest attempts to influence the study progression that took place in Norway in 1990 when a student aid reform, often referred to as the *turbo reform*, was implemented. The reform entitled students in certain graduate programs to an extra restitution of their study loan if they graduated on stipulated time. This created a sharp discontinuity in the incentives to graduate on time and offers a good opportunity to estimate how students respond to financial incentives. We focus especially on heterogeneity in the treatment effect by studying how the take-up varies by parental background, student ability and field of study. Further, we discuss efficiency issues related to the design and implementation of the reform.

The study uses rich and accurate register data on education and student characteristics and we link ability test scores from the military draft to the male students in the sample to study how the treatment effect varies over the ability distribution.

[1] See for example (Brunello and Winter-Ebmer, 2003; Häkkinen and Uusitalo, 2003; Garibaldi, Giavazzi, Ichino, and Rettore, 2012; US Department of Education, 2003; Budsjett-instilling, 1990; NOU, 1999).

[2] The OECD has repeatedly pointed out that graduating at old age is costly both from the society's point of view in terms of increased costs of education and delayed tax revenues, and from the individual's point of view through postponed labor market entry (OECD, 2011, 2013)

Within a difference in difference framework we also assess the robustness of the results by constructing alternative control groups, using synthetic control groups as suggested by Abadie, Diamond, and Hainmueller (2010).

We find that on average one additional year of treatment, defined as the number of years enrolled in higher education during the reform period, resulted in a 1.3-1.5 percentage point increase in the probability of graduating on time. This effect is robust to using different control groups, but slightly smaller than what Gunnes et al. (2013) found in a recent paper.[3] Given that the probability of graduating on time was 14 % percent in in the treatment group in the pre-reform period, this translates into an increase of about 50 % in the probability of graduating on time for a student who was treated for five years. The reform also reduced average delay; one additional year of treatment resulted in a 0.065 year reduction in delay.

The turbo reform was much debated because it was feared to reward good students who would have graduated on time anyway and punish weaker students who would accumulate more debt both because of the longer study duration and because of the foregone restitution. Partly because of these concerns, the turbo grant was abolished only five years after implementation. The worries of the opponents of the reform are in part confirmed as we find a considerable heterogeneity in the take-up of treatment, and show that the students who responded to the reform came from the upper tail of the ability distribution and from stronger socioeconomic backgrounds. Thus, the reform was likely to benefit students who would have performed well irrespective of the extra restitution.

Further, we find that there were significant differences in take-up by study program even though these estimates are rather imprecise because of small sample sizes. Some possible explanation for the this heterogeneity is offered by Berg (1994) who conducted a survey to find out why students delay graduation. She found that the reasons for delaying graduation were often correlated with study program, and that students in some programs were less able to affect their study progression than others, e.g. because the delays were caused by structural problems in the study program or because courses were very demanding. This points to another inefficiency of the

[3]Gunnes et al. (2013) evaluated the effect of the reform on the number of semesters delayed, the share of students who graduated on time and non-completion. The authors focus mainly on average effects and only briefly mention heterogeneity in the treatment effect

reform, namely that study progression was outside the control of some students, leaving them with little chances of complying.

These results show that designing a policy that affects study progression is challenging. Students are a highly heterogenous group and their motives to delay graduation might also vary. Some delay graduation voluntarily, for example by taking extra courses and improving grades. This is the kind of behavior policy makers often want to get rid of. But many delay graduation involuntarily e.g. because they are struggling with the curriculum or face problem related to the structure of the study program. These students might not be able to do much to about their progression, and stricter requirements on progression might be discouraging and lead to increased drop out behavior.

The data only offer limited possibilities to study the channels through which the reform worked. However, we present some results suggesting that the reform partly worked through a reduction of work activities. While there is no evidence that fewer treated students worked in the reform period, their earnings decreased substantially compared to the non-treated students, suggesting that the treated students cut back on hours worked in the reform period.

2 Related literature

Student aid is a topic that has been debated frequently in the last years, and for many different reasons. In countries like the US where education is costly student aid has been promoted as a means of improving access to higher education for students from less privileged backgrounds. In countries where education is free, such as the Scandinavian countries, the debate has been concerned with the optimal level of financial aid. Many countries struggle with long study durations and student aid systems are currently being re-designed to incentivize students to increase their study effort and reduce delays.

The importance of this issue is reflected in the number of studies that tries to estimate the effect of financial incentives on student behavior. A substantial part of this literature has been concerned with the effect of student aid on the extensive margin, that is increasing enrollment or access to higher education. For example, Dynarski (2003) finds substantial effects of student aid on college enrollment in the US. Similarly, Bettinger (2004) finds positive effects of student aid on persistence in higher education.

In more recent years, researchers have also shown interest in the effect of financial aid or rewards on the intensive margin. One strand of this literature is concerned with academic performance and effort[4], while the line of research most in line with this study has focused on study duration and completion. The evidence on duration and completion is less conclusive than the evidence on enrollment and persistence. What is also problematic is that many of the studies cannot convincingly control for confounding factors as the effect is often identified from comparing students who graduated before and after the policy intervention.

One of the earliest contributions to this literature is a study by Häkkinen and Uusitalo (2003) who evaluate a Finnish student aid reform that aimed at reducing study duration by increasing the total support to facilitate full-time studies. They find only limited effects of the reform and conclude that part of their findings are explained by increasing unemployment rates that reduced student employment possibilities. Heineck, Kifmann, and Lorenz (2006) found that the introduction of tuition fees for delayed students at a German university affected student behavior but that the effect varied by major. While tuition made students in some majors graduate faster, average duration increased in others. Tuition was also found to increase the dropout rate in some majors, which was an unintended and undesirable effect of the reform. Glocker (2011) uses German panel data to study the relationship between student aid and duration and graduation probabilities. She finds that while higher levels of financial aid has no effect on study duration, it is positively correlated with the probability of graduating with a degree. Students who receive financial aid are also found to graduate faster than students who fully rely on private funds.

Garibaldi, Giavazzi, Ichino, and Rettore (2012) manage to work around the problem of confounding factors using discontinuities in the tuition fees at a private university in Italy as they estimate the effect of increased tuition on the probability of graduating on time. They find that students who face the threat of having to pay higher tuition fees after their expected graduation year are more likely to graduate

[4] See e.g. Angrist and Lavy (2009); Angrist, Lang, and Oreopoulos (2009); Leuven, Oosterbeek, and van der Klaauw (2010).

on time than students who do not face the same threat of higher tuition fees. This study offers the most credible evidence of student responses to financial incentives to date.

The main message of this literature is that financial incentives has some impact on study duration and college completion, but that the effect on completion is not as strong as that on enrollment. Also there seems to be heterogeneity in the responses. Both Häkkinen and Uusitalo (2003) and Heineck et al. (2006) find that students in different majors react differently to incentives and the latter study also find increased dropout behavior following the introduction of tuition fees from delayed students. However, the evidence is still too scarce to draw any definitive conclusions on the effect of student aid on duration and completion but there seem to be a need for more research on this topic, and especially research that takes student heterogeneity into account.

3 Institutional settings

The higher education system and the student aid system in Norway were both restructured in 2002/2003 as a part of the Bologna process, and what follows is therefore a description of the old systems that were relevant for this study.

3.1 The Norwegian higher education system

The Norwegian higher education system consisted of universities, specialized universities and regional university colleges. All types of institutions offered both undergraduate and graduate courses, but the regional university colleges mostly provided shorter (two to three year long) vocationally oriented programs. Most undergraduate programmes at universities lasted for three or four years. The graduate programs were structured either as integrated study programs with a total duration of five to six years, such as medicine or law, or as a combination of an undergraduate and a related graduate program, also with a combined duration of five to six years. The education system was similar to that in the U.S. in that most students would leave university after the undergraduate level. Both the undergraduate and the graduate degrees in the 1990's were more comprehensive than the post-Bologna Bachelor's and Master's degrees.

The majority of the students were enrolled in public institutions.[5] Tuition fees, which were only paid in the private higher education sector, were low, making the private costs of higher education very low.[6]

3.2 The Norwegian State Loan fund

The Norwegian State Loan Fund (NSLF) is the main provider of student financial aid in Norway. The purpose of the NSLF is to promote equality in society by enabling students to participate in education irrespective of age, sex, geographical, economic and social conditions and to ensure a satisfactory work environment for students (Lånekassen, 2012). This means that practically all citizens were entitled to financial support from the NSLF if enrolled in higher education. Since the costs of higher education are virtually zero, the NSLF provided loans and grants to cover the students' living expenses during the academic year. The student support was not tested against parental income, but dependent on students own income and wealth. Students were allowed to work during the academic year, but the allowed earnings were restricted to NOK 5,200 per month. Earnings in the summer months were not included in the calculations.

In the time period of the study 87 % of the financial support was distributed as a loan, and 13 % as a grant. The total support during an academic year was decided upon every year by the Parliament and it ranged from 52,000 to 60,000 NOK in nominal value in 1991-1995.[7] If a student did not make any progress the support was cut, but there were special arrangements in the case of sickness, maternity leaves etc (Lånekassen, 2012).

The loans provided by NSLF were very favourable. Interest was not calculated while the student was enrolled, and repayments only started about ten months after graduation (or after dropping out). The interest rate was usually lower than the market interest rate. If a person could not repay her debt, for example because of illness or unemployment, the loan could be fully or partly cancelled.

The uptake of loan was high. Among the students who graduated with a long

[5] In 2011 the figure was 87 % (Kunnskapsdepartementet, 2012)

[6] Students in private institutions can apply for a loan from The Norwegian State Loan fund to cover their tuition fees.

[7] Corresponds to 13,000 to 15,000 USD in 2011.

graduate degree in 1990, 97 % had taken up some loan, but only 28 % had taken up the full amount (Berg, 1997). In 1995 51 % of the 5th semester students had taken full loans, while 31 % had taken some loan and 18 % had not taken any loan or only the grant part for the support (Berg, 1997). Too high labor income, unwillingness to accumulate debt and living for free at home with parents were the most common explanations for not lifting full support (Berg, 1997).

Enrollment in graduate programs was low in the 1980's and to stimulate enrollment in these programs all students who graduated from a graduate program got a reduction of their loan after graduation, irrespective of time-to-degree. Until 1989/1990 this amount was fixed for all study programs (NOK 27,300 in 1989/1990), but from 1990/1991 it was differentiated by the length of the program, ranging from NOK 28,400 for 10 semesters to NOK 43,400 for 13 semesters in 1990/1991. All the programs in this study were affected similarly by this scheme, and we don't expect it to influence the results.

3.3 The turbo reform

The history of the *turbo grant reform* (Turbostipendreformen) was rocky from the start and until its termination. The turbo scheme was made public on October 4th 1990 as a part of the National budget for 1991. It stipulated that students in certain graduate study programmes who completed their degree on stipulated time were entitled to a restitution of their student loan of about 18,000 NOK from the NSLF.[8] The new rules applied to students who graduated after August 15th 1990, and thus some students received the grant retrospectively.

Not all students were eligible for the restitution. Students in undergraduate programs were exempted, as well as students in certain fields of study. Delayed graduation was a widespread problem in higher education, and the reform was targeted specifically at programs where delays were common. These were mostly loosely structured study programmes taught at universities, such as humanities, social sciences and science. The largest groups not covered by the reform were engineering and medicine students. However, when the reform was first announced it was not clear which study programs were covered by the reform and the first official guide-

[8] This corresponded to about 35 % of one years study loan or 9 % of the total study loan of a student who had followed normal study progression and taken up the full loan.

lines from the NSLF were not published before July 1991. Therefore there was quite some uncertainty about the reform in the first year after its implementation. There is also no record of this reform having been discussed in the media prior to the date it was announced, and therefore it is very unlikely that students could anticipate the reform.

The reform was debated from the start. The main arguments for the discontinuation of the turbo scheme were that the rules were difficult to administer and that the restitution was likely to be given to students who would have graduated on time anyway, thus rewarding the good students and punishing the weaker students. Therefore, the grant was abolished, and students who graduated after August 14th 1995 were not eligible for the turbo grant.

Even though the termination of the turbo scheme had been discussed by policy makers already in the spring of 1994 (St. Meld. nr. 14, 3 94), the students were unlikely to anticipate its ending. The changes were announced in the spring/summer 1995, but at that time the plan was to replace the turbo scheme with a similar but more general scheme that would cover all students. Only later in the fall 1995 was it announced that the turbo grant would not be replaced after all.

4 Data

The study uses register data from Statistics Norway covering all students enrolled in higher education in 1974–2010. The data is reported directly from the educational institutions to Statistics Norway and is therefore considered to be very accurate. The data contains enrollment and graduation dates, completed degree, institution from which the degree is obtained, duration of study program, as well as data on whether the student completed her degree on stipulated time and if not, and the number of semesters delayed. The data also contains information on demographic characteristics, as well as parental education and income.

We focus on the students who completed a degree, although dropouts could be included to study whether the reform had an effect of dropout behavior.[9] We restrict the sample to students whose expected graduation year is 1986 or later mainly because of data restrictions. Students who enrolled in higher education in the fall

[9] Gunnes et al. (2013) find no effect of the reform on non-completion.

1991 or later are also excluded to avoid selection into treatment.[10]

To make the sample more homogenous and to ensure that we focus on students who follow a relatively normal study progression before enrollment in higher education we restrict the sample to students aged 18-21 at high school graduation (90.5 % of the remaining sample). Older students are less likely rely on student aid and more likely to work and study part time, and therefore less likely to be affected by the incentives offered by the turbo reform. For this reason we also restrict the sample to students who are aged 18-25 at first enrollment in higher education and no older than 40 years at graduation (1 % of the remaining sample).

The treatment status of a student depends on two factors. First, not all study programs were covered by the reform. Eligibility was decided on the university-study program level, but in practice this often coincided with the study program level. The treated and non-treated study programs are listed in Table 1. Humanities, social sciences, science and law were the largest of the treated programs. Among those not treated, which we refer to as the control programs, were engineering, medicine and agriculture.[11] [12]

Second, the eligibility for the restitution depended crucially on the expected rather than the actual graduation date. Because delays were common, many students who graduated in the reform period had already passed their expected graduation date when the reform was implemented and were thereby not eligible for the restitution. Date of expected graduation is not recorded in the data, but we combine the date of first enrollment in higher education and the stipulated duration (Column 3 in Table 1) to impute this date. This adds some uncertainty to the data because it was relatively common to switch majors or to take leaves from the university so that total enrollment does not necessarily reflect enrollment in the study program from which the student enrolled. However, the NSLF followed very strict rules when counting

[10] The last cohort included in the study enrolled in the spring 1991, which means that they submitted their applications in the fall 1990. Thus, these students could have been aware of the turbo grant but since there was no information about the treatment status of the study programs at this point, it should not have affected the student's choice of study program.

[11] The degree obtained by the agriculture students is called *Cand.agric.*, which signals that is it related to agriculture and they graduated from what is now called the Norwegian University of Life Sciences. In practice the students had a number of different majors including engineering, business administration, resource management and biology. Thus, the agriculture students studied many of the same majors represented in the treatment group.

[12] The reform status of some study programs was unclear, most commonly because the status of the program changed during the reform period, or because NSLF could not determine the duration of the study program. Students in these programs (who make up for about 4 % of the remaining student population) are excluded from the sample.

the number of semesters enrolled and would normally not cancel extra semesters obtained by program switchers. Therefore we do not expect the imputation of the expected graduation date to be a problem.

Conditional on being delayed at graduation, the average delay in the treated group was 2.1 years. This has important implications for the expected take-up of the reform. For many students who were approaching their stipulated graduation date when the reform was implemented, it was very hard to comply with the new rules even if they wanted to. Therefore, no big jump in the share of students graduating on time is expected at the time of implementation. Instead, we expect a gradual increase in this share for later cohorts who had more time to adapt to the new rules.

Put differently, students in different cohorts were treated at different intensities. A simple parametrization of treatment intensity is presented in Table 2 using information on the expected graduation date and duration of study program. We define treatment as the number of years the student was studying in the reform period up to her expected graduation date. Given that the reform was implemented in 1990, we define students who were expected to graduate in 1991 to be treated for one year, 1992 graduates to be treated for two years and so on. Students who were expected to graduate in the pre-reform period, and students who were expected to graduate in the reform period, but graduated before the reform was implemented are not treated. Students who were expected to graduate in the fall of 1990 were strictly speaking treated for two months, but their possibilities to comply with the reform were very limited and we treat these students as not treated. If there was a positive reform effect on these students, the estimates are downward biased. Students who were expected to graduate after the reform period ended were also partly treated, although not at the end of their studies. We include these students in the analysis.

The main goal of the turbo reform was to increase the share of students who graduated on stipulated time and this is also the main outcome variable of the analysis. The outcome variable is a dummy variable that indicates whether the student graduated on stipulated time on not and it is derived by Statistics Norway by combining data on the stipulated length on a study program and the number of semesters a students was enrolled in higher education before graduation. Thus, breaks and gap years are not counted. The turbo grant might also have had an impact on other dimensions of student behavior. In Appendix C the analysis is repeated using delay measured in years as the outcome variable. The duration of the delay is also policy relevant because reductions in delay are associated with reductions in public spending on education.

The fraction of students graduating on time by graduation year and treatment status is shown in Figure 1. The turbo reform was targeted at study programs where delays were common, and as a result the share who graduated on time was much lower in the treatment group than in the control group throughout the period. As expected there is no immediate jump in the outcome variables for the treated group straight after the implementation in 1990, but rather a gradual change over the reform period. The treatment and control group actually follow the same pattern in the first two years of the reform period, but then the trends part as the reform starts to kick in.

5 Empirical strategy

The fact that there are treated and non-treated study programs and a given implementation date makes this a suitable application for a difference in difference strategy. Given that there are ten treated study programs we can study program specific treatment effects. In the main analysis we will, however, only estimate an average treatment effect controlling for study program fixed effects. Study program specific treatment effects are briefly investigated in Section 6.3 but the small sample sizes makes the estimates imprecise.

Just as there are several potential treatment groups, there are six non-treated study programs in the sample. Any combination of these can be used as a control group. The baseline results are estimated using a control group that is an unweighted average of all the six non-treated study programs. The choice of control group could, however, be motivated in many ways, and the robustness of the results is tested by using a number of different control groups in Section 6.4.1.

In the simplest estimation equation we ignore the differences in treatment intensity and simply estimate an average treatment effect for the reform and post-reform periods:

$$y_i = \alpha + studyprogram + cohort + \beta_1 T_i + \beta_2 PT_i + \delta X_i + \epsilon_i \quad (1)$$

where y_i is an indicator variable for whether student i graduated on time or not. $studyprogram$ and $cohort$ are study program and expected graduation cohort dummies, and T_i and PT_i are the difference in difference variables indicating whether the student was expected to graduate from a treated program in the reform period (spring 1991 to spring 1995) or the post-reform period (fall 1995 to fall 1997). X_i are control variables including demographic and family characteristics and ϵ_i is an error term. The parameters of interest are β_1 and β_2, which measure the average treatment effect in the reform and post-reform period respectively.

In the main specification we use the measure of treatment intensity introduced in the previous section:

$$y_i = \alpha + studyprogram + cohort + \beta treatment_i + \delta X_i + \epsilon_i \quad (2)$$

where $treatment_i$ measures treatment intensity (equal to zero in the control group and ranging from zero to five years in the treatment group). The rest of the variables are defined as before. The coefficient β measures the effect of one additional year of treatment. Because of the short reform period most of the students were only partially treated. By multiplying β with the total study duration, it is possible to extrapolate the effect to a fully treated student.

When using difference in difference estimators unadjusted standard errors will often understate the true standard errors of the estimated coefficients due to the presence of unobserved group-level effects and/or serial correlation in the error term (Moulton, 1990; Wooldridge, 2003; Bertrand, Duflo, and Mullainathan, 2004; Donald and Lang, 2007). While there is consensus that the standard errors need to be adjusted when applying difference in difference estimators, there is less agreement on the best way to adjust them. In cases where there are many groups or clusters, the most straightforward approach is to cluster the standard errors at the group level.[13] When the number of clusters is small the problem is harder to solve. Donald and Lang (2007) offer a solution but it is only applicable in the case where the treatment

[13]In practice this can be done by using the cluster option in STATA. This procedure allows for general within-group covariance and heteroscedasticity (Wooldridge, 2003; Donald and Lang, 2007).

variable is binary.

In the preferred specification we cluster the standard errors. Even if the group variable used in the analysis is study program, we cluster the standard errors at the university-study program level for three reasons.[14] First, this is the level at which treatment status is decided even if, in most cases, it coincides with the study program level. Second, one can easily argue that if there are common group effects or shocks, these are most likely to appear at the university-study program level. Think about economics students at two different universities. They study the same major, but the course structure of the program, the labor market and other factors that might affect study progression might differ between the universities (and cities). A third motivation for clustering at the university study program level is to increase the number of clusters from 16 study programs, to 61 university-study program clusters. This improves the reliability of the standard error adjustment, as the method is only consistent as the number of clusters grows large. The minimum number of clusters required to obtain reliable standard errors is often said to be 50 (Bertrand et al., 2004; Donald and Lang, 2007).

5.1 Validity of the DD estimator

The validity of the difference in difference estimator relies on a number of assumptions. First, identification is threatened if students can manipulate their treatment status either through changing their expected graduation date or by switching between the treatment and the control group. The fact that the reform was retrospectively implemented is comforting because there was no way a student could manipulate her graduation date relative to the implementation date. It is also very unlikely that a student would move from the control group to the treatment group. When the reform was announced all the students in the sample were already enrolled or in the process of enrolling in the study they later graduated from. Changing their treatment status would mean enrolling in a different study program and start over again, which would in itself make them non-eligible for the restitution because it would cause them to spend too long obtaining their degree. Thus, we are confident

[14] Standard errors have also been calculated using clustering at study program level and study program times cohort level. The results are presented in Table 11 in the Appendix. Clustering on university-study program level is the most restrictive alternative in terms of statistical significance.

that the selection into treatment is not a serious issue.

A potential worry is that even though students could not manipulate their treatment status, assignment into study program which in turn determines treatment status, is not random. But even if choice of study program is not random, the choice was made before treatment status of the study programs was assigned and the choice should be uncorrelated to the take-up of the reform.

The identifying assumption of the difference in difference estimator says that the reform effect can be estimated if the time trend of the outcome variable in the treatment and control group would be would have been the same, had it not been for the reform. The difference in difference estimator automatically deals with any differences in levels of the outcome variables and therefore it is not a problem that the treatment and control group differ in their likelihood of graduating on time as long as their time trends are parallel.

The assumption of parallel trends is ultimately not testable, but there are some ways of assessing its plausibility. A first step towards ensuring that the control group is good is to graphically compare the pre-reform trends of the treatment and control group. The pre-reform trends in on-time graduation in the full sample are shown in Figure 1, which shows that the pre-reform trends are fairly parallel. The parallel trend assumption is studied further in Table 10 in Appendix A, where the difference in pre-reform trends is estimated using both a linear time trend and year dummies. The assumption of parallel time trends in the pre-reform period cannot be rejected in either case. [15]

The goodness of the baseline control group is further investigated in Table 3 by performing a balancing test of pre-determined characteristics. As one could expect based on the fact that the treatment and control group students study different majors, there are some differences in background characteristics. The control group students are less likely to be female and have higher IQ score (available only for male students), which is very likely to be explained by the study program mix. The control group students also come from families with slightly higher income and higher parental educational attainment.

The last column of Table 3 reveals whether there are differential trends in the

[15]The parallel trend assumption is also robust to changes in the control group. Results are not presented here for brevity, but are available on request.

pre-determined characteristics that could explain the reform effect. The sample is balanced on family background and ability score, but the share of female students increased at a significantly higher rate in the treatment group in the reform period. Increasing female educational attainment is a well-documented phenomenon over this period, and it is therefore not surprising that the share of female students increased rapidly in the treatment group. Many of the control group study programs are known to be male dominated, and it is not surprising that the female share did not increase as much in these programs over time. The gender imbalance is driven by the fact that men and women choose to study different majors, and it remains no matter how we define the control group. The sample is also unbalanced on age at high school graduation. However, the difference of 0.04 years (15 days), which is unlikely to make a big difference in practice. In the main analysis we control for pre-determined variables to make sure they are not driving the results.

6 Results

In this section the main analysis in presented. The baseline results using the unweighted control group are presented in Section 6.1. In Section 6.2 we study whether the reform effect differs by student characteristics such as gender, parental background and ability. In Section 6.3 we split the sample by study program and estimate study program specific treatment effects. The robustness of the results are investigated in Section 6.4 where we use alternative control groups and test for other confounding factors such as the unemployment rate and increasing enrollment in higher education.

6.1 Effect on probability of graduating on time

The estimated effect of the turbo reform is presented in Table 4 where Equation 1 is estimated in the first two and Equation 2 is estimated in the last two columns. We find a positive effect on the probability of graduating on time in the reform period of about 3.8 percentage points but it is only significant on the 13 % level. Given that only 14 % of the students in the treated programs graduated on time in the pre-reform period, it must still be seen as a relatively large effect. There is also a positive effect of the reform in the post-reform years but this is even less precisely estimated.

The measure of treatment intensity is introduced in Columns 3 and 4 and the estimates suggests that one additional year of treatment is associated with a 1.5 percentage point increase in the probability of graduating on time. If the effect extrapolated to a student who was treated for five years (which was the duration of a typical study program) the accumulated effect corresponds to a 7.3 percentage point increase in the probability of graduating on time. This corresponds to a 52 % increase in the probability of graduating on time. The inclusion of control variables does not significantly change the estimates. This is very comforting since it suggests that selection into treatment and student characteristics is not driving the results.

6.2 Heterogenous treatment effects

The results in Table 4 indicate that the average treatment effect is not driven by background characteristics but that study progression varies with the characteristics of the student. Female students and students with highly educated parents are less likely to graduate on time. In this section we exploit the data further to see whether the treatment effect also varies with students characteristics.

For comparison the baseline estimates with background controls are presented in Column 1 of Table 5. The estimates in Column 2 suggest that the treatment effect was not significantly different for men and women even though women are on average less likely to graduate on time. The importance of family background is investigated in Columns 3 and 4. Although students with at least one parent with tertiary education are on average less likely to graduate on time, these students are driving the treatment effect. The treatment effect is increasing in parental education and ranging from close to zero and insignificant for students with parents who have less than high school education to 1.6 percentage points and significant at the 5 % level for students of parents with higher education.

Many Norwegian students receive financial support from their families and both the probability of receipt and amount of support is positively correlated with family background (Berg, 1997; Anders Barstad and Thorsen, 2012). If students with highly educated parents receive more support from their families, they can focus on their studies and be less dependent on labor income. This could again affect study progression. There is no information about support from parents in the data, so this hypothesis cannot be tested directly. In Column 4 we instead test whether the treatment effect is increasing in family income. The students are divided into quartiles based on family income at age 16 and we interact the quartiles with the treatment intensity variable. There is no relationship between treatment and family income which suggests that the positive effect of parental education on the take-up of treatment works through other channels than income. Examples of such mechanisms are attitudes or skills that are more often taught in homes of highly educated parents, such as motivation, determination or effort.

It is not far fetched to assume that students differ in their prerequisites to follow academic courses, study preparedness or motivation (qualities that may or may not be correlated with parental background). However, these are variables that are very rarely available in register data and their effect on treatment is therefore hard to investigate. In Norway all men do a cognitive ability test at age 18 as a part of the military draft. We have the unique possibility to link the test score data to the sample of male students to study the relationship between treatment and ability. In Column 7 we divide the sample of male students into quartiles based on their cognitive test scores and interact them with the treatment variable. There does not seem to exist any clear relationship between IQ and the probability of graduating on time as such. The interactions terms, on the other hand, suggests that the treatment effect is driven mainly by the students in the highest quartile of the ability distribution. One additional year of treatment increases the probability that a student in the top quartile of the ability distribution graduates on time by 2 percentage points relative to a student in the students in the lowest quartile. This is a very interesting finding that casts doubt on the efficiency of the reform. It is not clear that the students in the top quartile of the ability distribution are the ones that policy makers want to help in the form of a reduced study loan.

6.3 Study program

So far we have focused on the average effect across study programs, but since there is evidence students in different majors react differently to other types of changes in financial aid (Häkkinen and Uusitalo, 2003; Heineck et al., 2006) it is also interesting to study whether the effect of the turbo reform differed between study programs. In Table 6 we estimate study program specific treatment effects using the unweighted control group. The unweighted control group might not be optimal for all of the treated study programs so the estimates should be interpreted with care. Smaller sample sizes also make some of the estimates quite imprecise. Still, the estimates carry interesting information.

We find that there is substantial differences in take-up between study programs. The biggest effects are found in the Humanities, Social Sciences, Sciences, Economics and Psychology. On the other hand, there seems to be no effect of the reform in Law and Dentistry. The estimates are robust to the inclusion of control variables, suggesting that the differences are not driven by differences in student composition.

One can get some insights into some possible explanations for these differences from a survey by Berg (1994) where students were asked about factors that had affected their study progression. In addition to identifying the most common reasons for delay (taken extra credits, working, failed exams, need for more time to understand the curriculum etc.), Berg finds that the reasons for delay vary by study program. For example, she finds that students in Economics and Law are particularly active in retaking exams to improve grades, and that law students fail exams more often than others. Psychology students often struggle with collecting enough credits (due to poorly structured programs) while students in the Humanities often take too many credits.

What is important to note is that students differ in their possibilities to change their study pace. Law is known to be one of the most demanding study programs, and if students are struggling with the curriculum and pressure to get good grades, they might not be able to graduate any faster even if they are encouraged to do so. Students who retake exams to improve their grades or take extra courses, on the other hand, can more easily change their study pace. Berg (1994) also expresses a concern that increased competition in the labor market due to increasing unemployment rates and supply of graduates led to hoarding of education in some programs. This point is very important in the discussion of how students respond to financial incentives. A rational student would compare the costs and benefits of graduating sooner when faced with the possibility of getting the turbo grant. If it was common practice to delay graduation in order to improve ones transcript of records, the turbo restitution might not weigh up for the competitive advantage lost by not improving a grade or taking an extra course.

These results all indicate that there was heterogeneity in the treatment effect. Already at the implementation there was a worry that the turbo reform would benefit only the good students. This worry is confirmed in this analysis since the students who responded the most were those in the highest quartiles in the ability distribution and those with the highest educated parents. The finding that the reform effect varies between study programs can also be symptomatic of structural problems in some programs that might make it hard for students in certain programs to change their behavior.

6.4 Sensitivity analysis

In this section we further investigate the robustness of the results. First we show that the results are relatively robust to the use of different control groups. Later we discuss how the economic condition in Norway might have had an impact on the effect of the reform.

6.4.1 Alternative control groups

While most control groups are chosen based on economic reasoning, Abadie, Diamond, and Hainmueller (2010, 2012) have developed a data driven method for constructing a control group. The synthetic control method was originally developed for case studies where there is one treated unit[16] and a number of possible control units, but where none of the control units serve as good control group either on their own or in an unweighted combination. The idea of Abadie and co-authors is to create a synthetic control group based on a weighted combination of the possible control units based on matching of pre-intervention values of the outcome variable and/or pre-intervention values of predictors of the outcome variable.

The matching procedure is explained in more detail in Section B in the Appendix. In short, we use the group specific share of students graduating on time in each of

[16] In this case there are of course more than one treated unit but we pool all the treated programs into one unit.

the years 1986-1989 as predictor variables. Because the matching algorithm matches on levels rather than trends, and because of the large differences in levels between the treated and the control units, we normalize the outcome variable to have mean zero in 1990. The result of the matching is presented in Figure 2 where the trends of the unweighted and the synthetic control group are compared.

The difference in difference estimates using the synthetic control group are presented in Column 2 of Table 7. The estimated reform effects are slightly smaller than the baseline specification in Column 1 predicts: one additional year of treatment is predicted to increase the probability of graduating on time by 1.3 percentage points and the estimate is only statistically significant on the 14 % level.

Ideally, one would like the treated and control group students to study the same majors. Although this is not feasible, it is possible to make the treatment and control group more similar by including only the programs that are the most similar in the analysis. Most of the treated study programs were loosely structured programs with little organized teaching and a large degree of freedom of choice (and responsibility for one's progression) for the students, while the (health related) non-treated programs were tightly structured programs with classroom-like teaching. The treated programs had no or very low admission criteria, while competition for study places was fierce in most of the non-treated programs. The future labor markets of the two groups also differed. While the students in the treated programs were likely to be employed in both the private and the public sector, the majority of the students in the health related non-treated programs were likely to be employed in the public sector. In general, students in treated programs faced more volatile labor markets and were more sensitive to fluctuations in the unemployment rate. All of these factors might in turn affect study progression.

Following these lines of argument one can exclude three programs from the control group: medicine, veterinary medicine and pharmaceutical science. Of the remaining three programs we also exclude architecture because it had some features which makes it hard to compare to the other programs. That leaves engineering and agriculture in the control group. Studies in engineering are more tightly structured that some of the treated programs, but all in all the program is relatively similar. Also, the engineering students face very similar labor market conditions as the teated students in terms of sensitivity for recessions and unemployment rates. The same is holds for the agriculture students.

In Column 3 of Table 7 we estimate the reform effect using this control group. Compared to the baseline estimates in Column 1, the estimated reform effect is now marginally smaller than in the baseline case and only borderline significant. The F-statistic and the R-squared are slightly higher, suggesting that the control group with only engineering and agriculture students might be better than the baseline control group.

Following the same line of argument, we estimate the reform effect for health related study programs only in Column 4 of Table 7. More specifically, the treatment group consists only of students in dentistry and the control group consists of students in medicine, veterinary medicine and pharmaceutical science. In this case, the reform did not seem to have any significant effect on the share of students who graduated on time. One explanation for this could be that Dentistry was a strictly structured study program, and that dentistry students might have limited possibilities to affect their study progression.

Taken together, the results are fairly robust to changes in the control group. The estimates are all in the same ball park and suggest that the one additional year of treatment increased the probability of graduating on time by 1.3-1.5 percentage points.

6.4.2 Unemployment and increasing enrollment

Despite seemingly parallel trends in the pre-reform period the validity of the difference in difference estimator can be threatened if there is some confounding factor that coincided with the implementation of the reform and that had a different impact on the treatment and control group over time. In this section we study the presence of such factors.

The most serious potential confounder is the recession that hit Norway in the late 1980's and that led to increasing unemployment rates in 1987-1993. This was likely to affect the employment possibilities of fresh graduates and students looking for a part time job in the relevant period. As pointed out above, increasing unemployment rates is only a threat to identification if the treated and non-treated students are affected differently. Unfortunately, there is reason to believe that this might be the case for a number of reasons discussed below.

In Section 6.4.1 we argued that the treated students probably faced more volatile labor markets than the control students. Linking this to the literature that suggests that there are severe and persistent negative effects from graduating in a recession on labor market outcomes (Kahn, 2010; Oreopoulos, von Wachter, and Heisz, 2012; Liu, Salvanes, and Sørensen, 2012) one can assume that treated students might have stronger incentives to strategically delay graduation in bad times. If treated student strategically delayed graduation in response to the increasing unemployment rates, the reform effect would be underestimated.

Häkkinen and Uusitalo (2003) offer an alternative hypothesis relating unemployment and graduation behavior. They found that higher unemployment rates in the 1990's in Finland lead to fewer student jobs, forcing students to study full-time and to graduate faster despite the bad economy. Working while studying is and has historically been very common in Norway. In a Norwegian student survey conducted in the mid 1990's about 60 % of the respondents agreed at least to some extent with the claim that they had to work to manage financially (Berg, 1997), and 80 % of the respondents in another survey answered that work activities had affected their study progression (Berg, 1994). About 59 % of the students in the sample of this study worked at some point while enrolled in higher education, and working was more common among treated students (66 vs 49 %). If there were fewer students jobs available to students during the recession, this probably hit the treated students harder, which would lead to an overestimation of the reform effect.

In Column 2 in Table 8 we control for the effect of increased unemployment by interacting the study program dummies with the national unemployment rate in the year of expected graduation for individuals aged 25-54.[17] The interaction terms (not reported for brevity) are all mostly small and insignificant, but there seem to be some variation between study programs. The estimated reform effect is 1.4 percentage points and significant at the 13 % level, suggesting that if anything, not controlling for unemployment rates actually lead to overestimating the reform effect slightly, which is in line with what Häkkinen and Uusitalo (2003) find.

[17] Adding the unemployment rate 1 to 2 years before expected graduation no big impact on the estimates.

The hypothesis of Häkkinen and Uusitalo (2003) is investigated in more detail in Table 9, as we study whether treated students changed their working behavior during enrollment in the reform period. We study both the probability of having a job and the total earnings of the students. To study the impact of the reform on the probability of working while studying we use two measures, none of which is perfect. First we generate a dummy for having a job in a given year based on job spell data, but many missing stop dates adds some uncertainty to the variable. Second we generate a dummy for having any pension qualifying earnings in the tax record data. This data also include other types of income than labor income, and therefore the share of students with some income is higher than the share of students who actually work. Finally, we also investigate whether the reform lead to a reduction in total pension qualifying earnings. The variables are measured in every calender year from enrollment up to the year prior to expected graduation.[18]

We use a difference in difference strategy, but the year fixed effects now refer to the year the work activity was observed. The difference in difference variable of interest measures the average effect over the reform and post-reform periods. Study program dummies and background characteristics are included as before. We also control for the national unemployment rate (interacted with study program) since changes in the unemployment rate might affect the employment possibilities of students.

The reform effect on the probability of having a job is investigated in Column 1 of Table 9. We find a small positive but insignificant effect on the probability of working in the reform years. Column 2 suggests that the reform had no impact on the probability of having some income. Taken together, there is no evidence that the reform affected the probability of working. In Column 3 we study the effect of the reform on total earnings. Treated students reduced their earnings by on average about 13 % in the reform period and the estimates are highly significant in both statistical and economical terms.

The results in Table 9 indicate that the incentives offered by the turbo reform could have worked is through reducing time spent working while studying. The results are consistent with a story where students don't stop working completely, but where they reduce their working hours and spend more time on their studies.

[18]In the expected graduation year many of the students graduate and enter the labor market which of course affects earnings in that year.

The recession also increased youth unemployment and in order to keep youth unemployment rates at acceptable levels the government increased the number of study places at the universities. This led to an enormous increase in the number of students enrolled in higher education from 1987-1994.[19] Further, it is documented that the expansion was mainly concentrated to the treated programs simply because it was easiest to expand capacity in the loosely structured study programs (Try, 2000). One might worry that this expansion affected study progression for a number of reasons, but it is not clear in what direction. The supply increase could lead to increases in delay if there was congestion in the education system due to failure to provide resources necessary to meet the needs of the bigger cohorts. Furthermore, as more students enrolled in the universities, the composition of the student mass probably also changed. In particular, average quality of students in terms of ability, study preparedness and motivation might have decreased, which could impact academic performance and progression negatively. On the other hand, increased competition for jobs between the students might have put more pressure on the students to graduate faster to signal ability and motivation to future employers.

We control for cohort size in Columns 3 to 5 in Table 8. In Column 3 we include a control for the logarithm of the total graduation cohort size by year of expected graduation, which we interact with study program to allow students in different programs to differ in their sensitivity to competition. In this specification students in different programs are assumed to be substitutes at least to some degree. In Column 4 we instead include a control for study program specific cohort size. In this specification students are only facing competition from their peers in the same study program. Both variables are included in Column 5.

Once cohort size is taken into account, one additional year of treatment is estimated to increase the probability of graduating on time by 1-1.2 percentage points, which is lower than the baseline estimate. This finding is in line with the idea that students increase their study progression in response not only to the reform, but also in response to the increases in supply, which overestimates the reform effect.

[19]Between 1986 and 1994 the number of study places increased from 101,000 to 169,000 mostly as ad hoc solutions to the unemployment problem. Despite wishes that the new places would mainly be created in the the college sector where most study programs were short and vocationally oriented, the university sector grew faster than the college sector in this period Try (2000).

7 Conclusion

Long study durations is a problem in surprisingly many industrialized countries. It is considered inefficient both because it consumes public resources that could be used differently, and because it delays labor market entry of the students, shortening their working career and lowering their lifetime income, resulting also in lower tax revenues.

For this reason a number of policies have recently been designed to shorten time-to-degree. The aim of this study was to evaluate one such policy, the turbo grant reform that was implemented in the 1990's in Norway. Students who graduated on stipulated time from certain study programs were entitled to a restitution of their student loan that amounted to about 35 per cent of one year's loan. The analysis suggests that the share of students who graduated on time increased and that average delay decreased following the reform. One extra year of treatment, defined as the number of years studying in the reform period, resulted in an increase in the probability of graduating on time of about 1.3-1.5 percentage points and reduced delay by 0.065 years. Recalculated as the effect of studying under the new regime for five years the effects correspond to a 7.3 percentage point increase in the probability of timely graduation (relative to a baseline of 14 %) and a reduction in delay of about 1/3 of a school year (compared to a mean of 2.5 years). The estimated average effect is non-negligible, but the fact remains that most of the turbo recipients would have graduated on time irrespective of the restitution. It is also important to note that even if the reform had a significant effect on graduation behavior in the treatment group, it was far from enough to close the gap between the treatment and control group.

One important finding of this study is that the treatment effect was driven by a rather selected group of students. The students who responded to the reform are those in the upper tail of the ability distribution and those with highly educated parents. This casts some doubt on the efficiency of this reform, as this might not be the group of students that one would like to help. In addition, there was only significant treatment effects in some study programs. Survey evidence from Berg (1994) highlight the possibility that in some cases it might lie outside the influence of the students to impact their study progression.

Taken together these results have important implications for policy makers that are looking to affect study progression in higher education. While reducing delays is an important policy goal, not all means of getting there might be equally good. It is important to keep in mind that not all students delay graduation voluntarily and that punishing these students for slow progression might have adverse effects on student behavior.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of californias tobacco control program. *Journal of the American Statistical Association 105*(490), 493–505.

Abadie, A., A. Diamond, and J. Hainmueller (2012, June). Comparative politics and the synthetic control method. Mit political science department research paper no. 2011-25, MIT Political Science Department.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review 93*(1), 113–132.

Anders Barstad, T. L. and L. R. Thorsen (2012). Studenters inntekt, økonomi og boutgifter. levekår blant studenter 2010. Report 38/2012, Statistics Norway.

Angrist, J., D. Lang, and P. Oreopoulos (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics 1*(1), 136–63.

Angrist, J. and V. Lavy (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *American Economic Review 99*(4), 1384–1414.

Berg, L. (1994). Hvilke forhold forstyrrer studieprogresjonen? In N. Vibe and T. Nygaard (Eds.), *Utdanning og arbeidsmarked 1994*, pp. 58–69. Utredningsinstituttet for forskning og høyere utdanning.

Berg, L. (1997). Leve på lån. studie- og låneatferd fem semestre etter studiestart (delrapport 1 fra studiefinansieringsprosjektet). Rapport 11/97, Norsk institutt for studier av forskning of utdanning.

Bertrand, M., E. Duflo, and S. Mullainathan (2004, February). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Bettinger, E. (2004). How financial aid affects persistence. In *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It*, NBER Chapters, pp. 207–238. National Bureau of Economic Research, Inc.

Brunello, G. and R. Winter-Ebmer (2003). Why do students expect to stay longer in college? evidence from europe. *Economics Letters 80*(2), 247 – 253.

Budsjett-instilling (1989-1990). Budsjett-instilling s. nr. 12 1989-90.

Deschryvere, M. (2009). A comparative survey of structural characteristics of finnish university departments. Discussion papers 1195, The Research Institute of the Finnish Economy.

Donald, S. G. and K. Lang (2007, May). Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics 89*(2), 221–233.

Dynarski, S. M. (2003). Does aid matter? measuring the effect of student aid on college attendance and completion. *The American Economic Review 93*(1), pp. 279–288.

Garibaldi, P., F. Giavazzi, A. Ichino, and E. Rettore (2012, August). College cost and time to complete a degree: Evidence from tuition discontinuities. *The Review of Economics and Statistics 94*(3), 699–711.

Glocker, D. (2011). The effect of student aid on the duration of study. *Economics of Education Review 30*(1), 177 – 190.

Gunnes, T., L. J. Kirkebøen, and M. Rønning (2013). Financial incentives and study duration in higher education. *Labour Economics 25*, 1 – 11.

Heineck, M., M. Kifmann, and N. Lorenz (2006, January). A duration analysis of the effects of tuition fees for long-term students in germany. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik) 226*(1), 82–109.

Häkkinen, I. and R. Uusitalo (2003). The effect of a student aid reform on graduation: A duration analysis. Working Paper Series 2003:8, Uppsala University, Department of Economics.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics 17*(2), 303 – 316.

Kunnskapsdepartementet (2012). Høyere utdanning 2012. Tilstandsrapport, Kunnskapsdepartementet.

Leuven, E., H. Oosterbeek, and B. van der Klaauw (2010, December). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association 8*(6), 1243–1265.

Liu, K., K. G. Salvanes, and E. . Sørensen (2012). Good skills in bad times: Cyclical skill mismatch and the long-term effects of graduating in a recession. Technical report, NHH Norwegian School of Economics, Department of Economics.

Lånekassen (2012, May). Lånekassen www.lanekassen.no.

Ministry of Education and Culture (2010). Ei paikoillanne, vaan valmiit, hep! koulutukseen siirtymistä ja tutkinnon suorittamista pohtineen työryhmän muistio. Opetusministeriön työryhmämuistioita ja selvityksiä 2010:11.

Ministry of Education and Culture (2012). Opintotuen rakenteen kehittäminen 2012. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2012:29.

Ministry of Education and Research (2002). Kvalitetsreformen. Veiledninger og brosjyrer 14.08.2002.

Moulton, B. R. (1990, May). An illustration of a pitfall in estimating the effects of aggregate variables on micro unit. *The Review of Economics and Statistics 72*(2), 334–38.

NOU (1999). Nyttige lærepenger - om utdanningsfinansieringen gjennom Lånekassen. Norges offentlige utredninger 1999:33.

OECD (2011). Education at a glance 2011: Oecd indicators.

OECD (2013). Education at a glance 2013: Oecd indicators.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012, January). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

St. Meld. nr. 14 (1993-94). Studiefinansiering og studentvelferd.

Statistics Norway (2013, September). Throughput of students in tertiary education, 2011/2012.

St.meld. nr.7 (2007-2008). Statusrapport for kvalitetsreformen i høgre utdanning.

Try, S. (2000). Veksten i høyere utdanning: Et vellykket arbeidsmarkedstiltak? Rapport 2/2000, Norsk institutt for studier av forskning of utdanning.

US Department of Education (2003). The condition of education 2003. NCES 2003-067, Institute of Education Sciences.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review 93*(2), 133–138.

Figure 1: Probability of graduating on time by treatment status

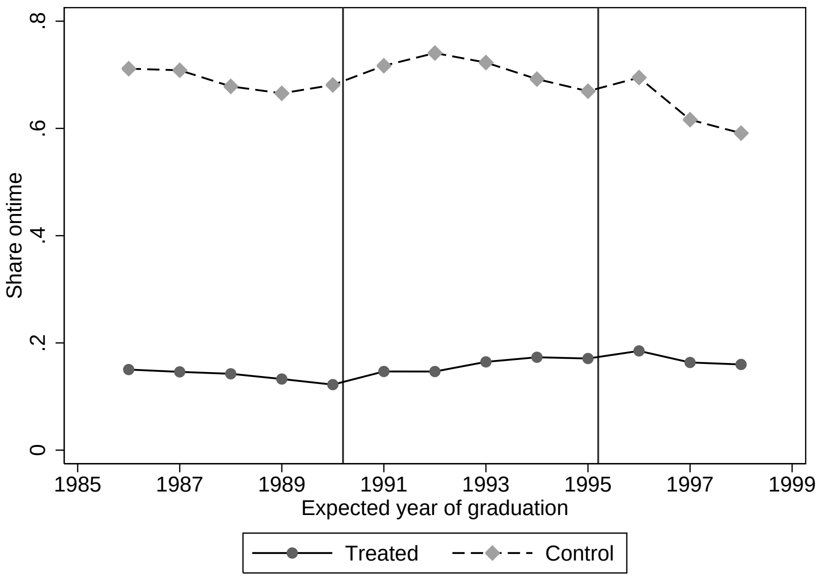
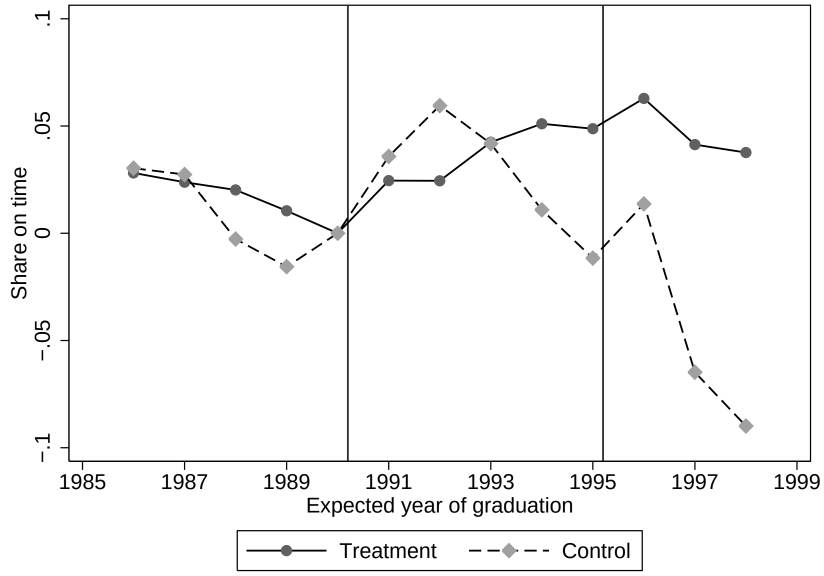
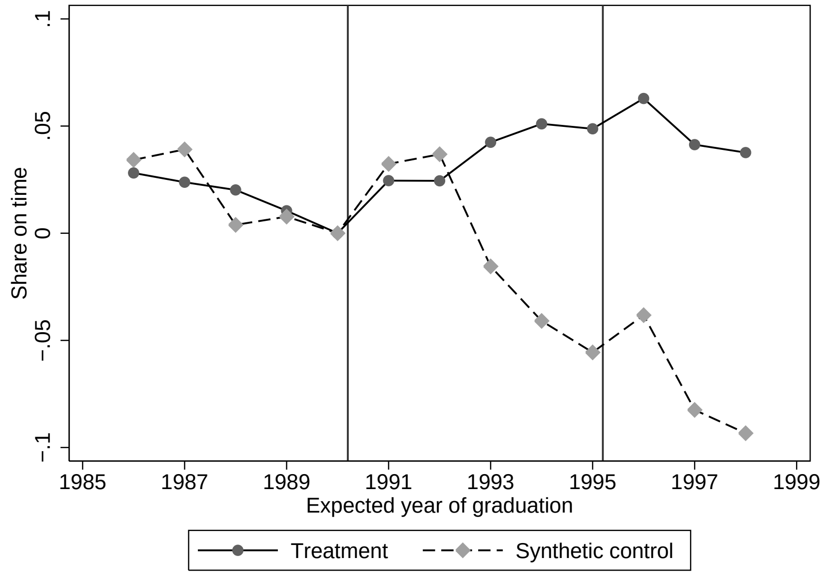


Figure 2: Demeaned probability of graduating on time by treatment status



(a) Baseline control group



(b) Synthetic control group

Table 1: Sample characteristics

Treated	No of students	Percentage	Expected duration	Delay in years	Share on time
Humanities	2,805	8.23	6	2.665 (1.748)	0.086 (0.280)
Social Sciences	3,185	9.35	6	2.162 (1.862)	0.187 (0.390)
Science	4,637	13.61	5 (6)	2.453 (1.544)	0.071 (0.257)
Law	5,174	15.15	(6) 6.5	1.705 (1.566)	0.206 (0.404)
Arts	158	0.46	6	1.101 (1.293)	0.373 (0.485)
Theology	493	1.45	6 (6.5)	1.294 (1.498)	0.294 (0.456)
Economics	519	1.52	5.5 (6)	1.362 (1.510)	0.291 (0.455)
Psychology	1,143	3.36	6.5	2.986 (1.511)	0.034 (0.181)
Dentistry	622	1.83	5	0.675 (0.827)	0.497 (0.500)
Fishery	87	0.26	5	2.161 (1.337)	0.057 (0.234)
	18,823	0.553		2.130 (1.698)	0.156 (0.363)
Control					
Medicine	2,702	7.93	6 (6.5)	1.039 (1.063)	0.262 (0.440)
Agriculture	1,974	5.79	5	-0.102 (1.056)	0.805 (0.396)
Engineering	9,418	27.65	5	-0.156 (1.040)	0.818 (0.386)
Pharmaceutical Science	307	0.90	5	-0.022 (1.103)	0.762 (0.426)
Veterinary Medicine	368	1.08	5.5 (6)	0.165 (0.594)	0.810 (0.393)
Architecture	475	1.39	5.5	1.914 (1.372)	0.051 (0.219)
	15,244	0.447		0.139 (1.187)	0.692 (0.462)
N	34,067	100.00		1.240 (1.789)	0.396 (0.489)

Notes: Sample distribution by study program. The expected length of a study program might vary by educational institution. The second most common duration listed in parenthesis if variation in duration. Standard deviations of delay and share on-time in parenthesis.

Table 2: Parametrization of treatment intensity

	Stipulated duration of study in years			
Expected graduation year	5	5.5	6	6.5
1986-1990	0	0	0	0
1991	0/1	1	1	1
1992	1/2	2	2	2
1993	2/3	3	3	3
1994	3/4	4	4	4
1995	4/5	5	5	5
1996	4	4/5	5	5
1997	3	3/4	4	4/5

Notes: The intensity of treatment is determined by expected graduation year and in some cases also by duration of the study program. Science students (5 year program) became eligible in 1991, and their treatment variable therefore takes on the value t-1 where t is the treatment intensity of other 5 year programs.

Table 3: Means of predetermined variables by treatment status

	Treated			Control			Difference-in-
	Pre-reform	Post-reform	Difference	Pre-reform	Post-reform	Difference	Difference
	(1)	(2)	(2)-(1)	(3)	(4)	(4)-(3)	[(2)-(1)] -[(4)-(3)]
Female	.462	.5336	0.071***	.3350	.3359	0.001	.0703***
	[.006]	[.004]	[.007]	[.006]	[.005]	[.008]	[.011]
Age at	19.096	19.076	-0.020***	19.044	19.067	0.0224***	-.0421***
HS graduation	[.006]	[.004]	[.007]	[.005]	[.005]	[.007]	[.010]
Ability (males)	6.981	6.939	-0.133***	7.617	7.497	-0.120***	-0.0128
	[.021]	[.022]	[.029]	[.020]	[.018]	[.027]	[.041]
Immigrant	0.036	0.045	0.006*	0.041	0.041	0.0006	0.005
	[.002]	[.002]	[.003]	[.002]	[.002]	[.003]	[.004]
Mother's years	11.32	11.65	0.327***	11.39	11.72	0.330***	-.0026
of education	[.038]	[.028]	[.047]	[.037]	[.033]	[.049]	[.068]
Father's years	13.06	13.13	0.077	13.11	13.26	0.158**	-.0796
of education	[.046]	[.034]	[.058]	[.047]	[.041]	[.062]	[.085]
Family income	518946	562324	43378***	525186	561123	35937***	7441
at age 16	[2926]	[2392]	[3889]	[2851]	[2675]	[3934]	[5999]

Notes: Means and differences in means of predetermined variables. Ability score available only for male students.
Standard errors in square brackets. $*p < 0.10, **p < 0.05, ***p < 0.01$

Table 4: The effect of the turbo reform on the probability of graduating on time

	(1)	(2)	(3)	(4)
Reform period*Treated	0.0369	0.0390		
	(0.0250)	(0.0245)		
Post-reform period*Treated	0.0314	0.0380		
	(0.0317)	(0.0315)		
Years in treatment			0.0144*	0.0149*
			(0.00813)	(0.00802)
Female		-0.0382***		-0.0381***
		(0.00554)		(0.00551)
High school graduation age		-0.00310		-0.00283
		(0.00524)		(0.00527)
Mother's Education Middle		0.00197		0.00212
		(0.00889)		(0.00890)
Mother's Education High		-0.0168		-0.0167
		(0.0127)		(0.0127)
Father's Education Middle		-0.0104		-0.0107
		(0.00667)		(0.00657)
Father's Education High		-0.0255**		-0.0258**
		(0.0115)		(0.0113)
log Family income at age 16		-0.0121***		-0.0120***
		(0.00336)		(0.00339)
Immigrant		0.00926		0.00922
		(0.0112)		(0.0112)
Constant	0.282***	0.528***	0.286***	0.525***
	(0.0236)	(0.0986)	(0.0239)	(0.0993)
R^2	0.428	0.432	0.428	0.432
Observations	34067	34067	34067	34067

Notes: All specifications include study program and cohort fixed effects. Columns 2 and 4 also contain dummies for region of residence at age 16 and unknown parental education. Parental education relative to low education. Educational levels defined as low=less than high school, middle=high school, high=post secondary education. Standard in parentheses errors clustered at study program-school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Estimated effect on the probability of graduating on time by student characteristics

	(1) Baseline	(2) Female	(3) Par educ	(4) Fam inc	(5) IQ
Years in treatment	0.0149*	0.0137	0.00269	0.0161*	0.0101
	(0.00802)	(0.00821)	(0.0101)	(0.00869)	(0.0142)
Female	-0.0381***	-0.0412***	-0.0378***	-0.0379***	
	(0.00551)	(0.00682)	(0.00548)	(0.00551)	
Female*Years treated		0.00242			
		(0.00255)			
Intermediate parental education			-0.0120		
			(0.00814)		
High parental education			-0.0402**		
			(0.0170)		
Intermediate parental education*Years treated			0.0104**		
			(0.00400)		
High parental education*Years treated			0.0140**		
			(0.00590)		
Family income 2nd quartile				0.000620	
				(0.00579)	
Family income 3rd quartile				-0.00384	
				(0.00662)	
Family income 4th quartile				-0.0350***	
				(0.00716)	
Family income 2nd quartile*Years treated				-0.00585*	
				(0.00320)	
Family income 3rd quartile*Years treated				-0.00278	
				(0.00387)	
Family income 4th quartile*Years treated				0.00256	
				(0.00475)	
Ability 2nd quartile					-0.000850
					(0.0183)
Ability 3rd quartile					0.00124
					(0.0262)
Ability 4th quartile					-0.0195
					(0.0233)
Ability 2nd quartile*Years treated					0.00295
					(0.00574)
Ability 3rd quartile*Years treated					0.00964
					(0.00700)
Ability 4th quartile*Years treated					0.0206***
					(0.00746)
Constant	0.525***	0.528***	0.538***	0.381***	0.605***
	(0.0993)	(0.100)	(0.0963)	(0.0854)	(0.142)
R^2	0.432	0.432	0.432	0.433	0.442
Observations	34067	34067	34067	34067	18652

Notes: Difference in difference estimates of one additional year of treatment, defined as studying in the reform period. All specifications include study program and cohort fixed effects and controls for age at high school graduation, region of residence and log family income and parental education at age 16. Standard errors are clustered at study program-school level and are shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Estimated effect on the probability of graduating on time by study program

	(1) Humanities	(2) Social Sciences	(3) Science	(4) Law	(5) Arts
Years in treatment	0.0187**	0.0177**	0.0175	0.00732	0.0295***
	(0.00823)	(0.00811)	(0.0107)	(0.00752)	(0.00693)
Constant	0.654***	0.493***	0.567***	0.687***	0.583***
	(0.171)	(0.133)	(0.142)	(0.171)	(0.178)
R^2	0.412	0.360	0.463	0.367	0.282
Observations	18049	18429	19881	20418	15402

	(6) Theology	(7) Economics	(8) Psychology	(9) Dentistry	(10) Fishery
Years in treatment	0.0126	0.0659**	0.0163*	-0.00762	0.0354*
	(0.0167)	(0.0274)	(0.00792)	(0.0121)	(0.0185)
Constant	0.589***	0.650***	0.605***	0.592***	0.632***
	(0.177)	(0.205)	(0.173)	(0.172)	(0.199)
R^2	0.289	0.292	0.367	0.273	0.289
Observations	15737	15763	16387	15866	15331

Notes: Study program specific difference in difference estimates using the unweighted control group. Control variables included, see Notes in Table 5. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Estimated effect on the probability of graduating on time using alternative control groups

	(1) Baseline	(2) Synth	(3) E+A	(4) Health
Years in treatment	0.0149*	0.0128	0.0141	-0.00103
	(0.00802)	(0.00856)	(0.0107)	(0.00883)
Constant	0.525***	0.451***	0.951***	1.087***
	(0.0993)	(0.111)	(0.0933)	(0.269)
R^2	0.432	0.364	0.454	0.188
Observations	34067	33285	30215	3999

Notes: All specifications include study program, cohort fixed effects and six dummies for region of residence at age 16. Parental education relative to low education. Educational levels defined as low=less than high school, middle=high school, high=post secondary. Standard in parentheses errors clustered at study program-school level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Sensitivity checks: the effect of unemployment and cohort size

	(1) Baseline	(2) Unemployment	(3) log(Cohort)	(4) log(Cohort)	(5) log(Cohort)
Years in treatment	0.0149*	0.0139	0.0126	0.0121**	0.0104*
	(0.00802)	(0.00906)	(0.00867)	(0.00513)	(0.00618)
Constant	0.525***	0.548***	0.396**	0.535***	0.466***
	(0.0993)	(0.0930)	(0.175)	(0.0923)	(0.156)
R^2	0.432	0.433	0.432	0.433	0.433
Observations	34067	34067	34067	34067	34067

Notes: Difference in difference estimates of one additional year of treatment, defined as studying in the reform period. All specifications include study program and cohort fixed effects and controls for age at high school graduation, region of residence and log family income and parental education at age 16. Unemployment rate is national average unemployment rate for ages 25-54 measured at expected year of graduation. Cohort size refers to the number of students in the sample expected to graduate in a given year. Standard errors are clustered at study program-school level and are shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Reform effect on the probability of working and earnings while studying

	(1) P(Work)	(2) P(Earnings)	(3) log(Earnings)
Reform period*Treated	0.0257	-0.0000262	-0.133***
	(0.0177)	(0.00363)	(0.0222)
Post-reform period*Treated	-0.0123	-0.00731	-0.303***
	(0.0233)	(0.00636)	(0.0328)
Constant	-0.358***	0.417***	10.03***
	(0.0914)	(0.0443)	(0.167)
R^2	0.123	0.031	0.035
Observations	185270	185270	172558

Notes: Estimated average effect of the reform on the probability of working (1) or having pension qualifying earnings (2) and total earnings while studying. All specifications include study program and year fixed effects as well as control variables. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix

A Additional tables

Table 10: Pre-reform trends

	(1) Ontime	(2) Ontime	(3) Delay	(4) Delay
Treatment group*Year	0.00736		-0.00947	
	(0.00818)		(0.0225)	
Year	-0.0115		0.00660	
	(0.00697)		(0.0199)	
Treatment group*1987		-0.00855		-0.00128
		(0.0380)		(0.113)
Treatment group*1988		0.0291		-0.0820
		(0.102)		(0.0851)
Treatment group*1989		0.0241		-0.0286
		(0.0265)		(0.0752)
Treatment group*1990		0.0219		-0.0334
		(0.0275)		(0.0960)
1987		0.00597		-0.0170
		(0.0364)		(0.0834)
1988		-0.0278		0.0794
		(0.101)		(0.0509)
1989		-0.0359		0.0544
		(0.0229)		(0.0403)
1990		-0.0374*		0.00150
		(0.0218)		(0.0744)
R^2	0.465	0.465	0.448	0.449
Observations	13297	13297	13297	13297

Notes: All specifications include study program fixed effects and control variables. Standard errors are clustered at study program-school level and are shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Comparison of standard errors using different levels of clustering

	(1) Unadjusted	(2) Study program	(3) Study*Year	(4) Study*School
Outcome variable: On time				
Years in treatment	0.0149***	0.0149**	0.0149***	0.0149*
	(0.00222)	(0.00514)	(0.00298)	(0.00802)
Constant	0.310***	0.310**	0.310***	0.310***
	(0.102)	(0.132)	(0.108)	(0.0974)
No. of clusters	0	16	174	61
R^2	0.432	0.432	0.432	0.432
Observations	34067	34067	34067	34067
Outcome variable: Delay				
Years in treatment	-0.0650***	-0.0650	-0.0650***	-0.0650**
	(0.00829)	(0.0377)	(0.0156)	(0.0304)
Constant	1.314***	1.314	1.314***	1.314
	(0.380)	(0.823)	(0.446)	(0.809)
No. of clusters	0	16	174	61
R^2	0.407	0.407	0.407	0.407
Observations	34067	34067	34067	34067

Notes: Cohort and study program fixed effects included, as well as a full set of control variables. Column 1 uses unadjusted OLS standard errors. In Column 2 standard errors are clustered on study program. Column 3 uses standard errors clustered at the study program times expected graduation year level, and Column 4 uses standard errors clustered at the study program by university level. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B Synthetic control method

The synthetic control method was developed for aggregate data where there is one treated unit and several possible control units, typically states in the US (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2012). It is, however, possible to apply the method to individual level data with some simple modifications. In order to perform the synthetic control matching we aggregate the individual level data to study program by expected graduation year level and since the synthetic control method only allows for one treated unit we further aggregate all the treated programs into one unit.

The matching algorithm aims at creating a control group that is as close as possible to the treated group in terms of the level of the outcome variable. Because the levels of the outcome variable in the treatment and control group are very different, we use the demeaned values of the outcome variables when performing the matching.[20]

The matching is then performed separately for both outcome variables. The predictor variables that are used to construct the synthetic control group are simply the group average of the outcome variable in each of the pre-intervention years 1986-1989.

The results from the matching is presented in Table 12. For both outcome variables, the root mean squared prediction error (RMSPE) is very low, which suggests that the fit of the synthetic control group is good. The weights of the control units are presented in the Panel B. In both cases, the most weight is given to engineering students The predictor variable means of the unweighted and the synthetic control groups are displayed in Panel C, and it is clear that the sample means of the synthetic control groups are closer to the mean of the treated group than the mean of the unweighted control group. These results, in combination with Figure 2, suggest that the synthetic control method was successful in generating a synthetic control group from the treated study programs.

[20]This is done by normalizing the level of the outcome variable to zero in 1990.

Table 12: Comparison of baseline and synthetic control groups

	Treated	Unweighted control	Synthetic control On time	Synthetic control Delay
Panel A: Root Mean Squared Prediction Error				
RMSPE			.011674	.0250628
Panel B: Weights				
Medicine			.240	.066
Agriculture			.368	0
Engineering			.388	.825
Pharmaceutical science			0	.016
Veterinary science			.004	.093
Architecture			0	0
Panel C: Predictor balance based on aggregate data				
Demeaned ontime 1986	.0281167	.05008947	.0342356	
Demeaned ontime 1987	.0238058	.0492494	.0391092	
Demeaned ontime 1988	.0202041	.01147067	.0038917	
Demeaned ontime 1989	.0104794	.00490785	.0077573	
Demeaned delay 1986	.0041058	-.0629625		-.0094867
Demeaned delay 1987	-.0062487	-.05099232		-.0237934
Demeaned delay 1988	.0229188	.05708111		.0678294
Demeaned delay 1989	.0338595	.02322695		.0356111

C Delay

The turbo grant reform aimed at increasing the share of students who graduated on time, but the data also offers an opportunity to study the effect on another related variable, namely delay. This variable is of at least as big interest as the share of students graduating on time since it is informative of the resources that are spent on each delayed student. If the goal of the government is to decrease education spending, it should be concerned with reducing delay as this affects both spending at the educational institutions and student aid.

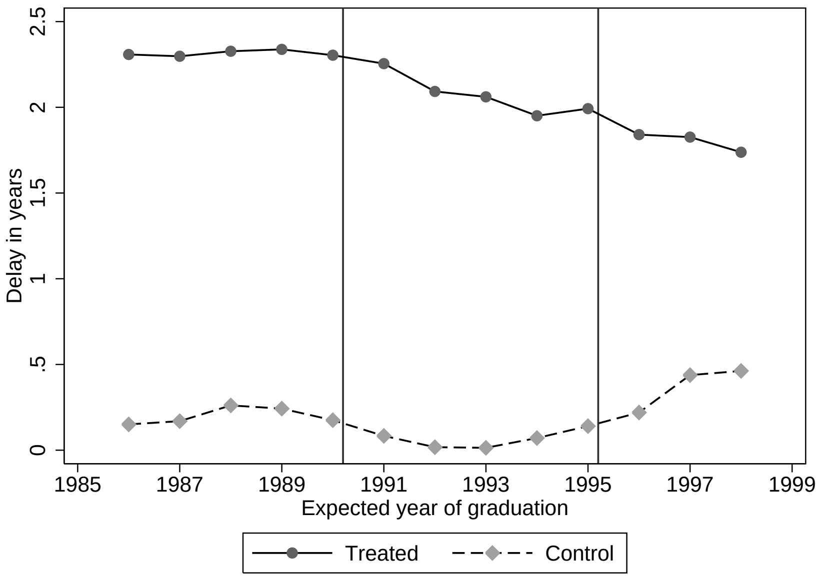
In the data reported delay varies between -5 and 16 years, and we suspect that some of the extreme values are due to reporting error. We drop observations below the 5th and above the 95th percentile in the delay distribution, but this does not largely affect the results. The results are also robust to excluding 1 and 10 percent in the tails.

Figures 3 and 4 and Tables 13 to 17 are identical to those in Section 6 only the outcome variable is different. The average treatment effect on delay is studied in Table 13. Average delay was reduced by 0.22 years during the reform period, and this result is significant on the 5 % level. The post-reform period treatment effect is not significant even though it is of considerable size. Columns 3 and 4 suggest that one additional year of treatment reduced delay by about .065 years, which is equivalent to 24 days. If we extrapolate the result to a treatment of five years the accumulated effect is a reduction of 0.33 years, or four months. This corresponds to a 14 % reduction compared to the average pre-reform delay in the treatment group of 2.3 years. As in the main analysis, the results are not driven by differences in the student composition.

In Table 14 we study the effect on delay by student characteristics and we find the same patterns as in the main analysis. Again, the treatment effect is driven by high ability student and students of highly educated parents. The remaining results are also very similar to the main analysis and will not be commented on any further.

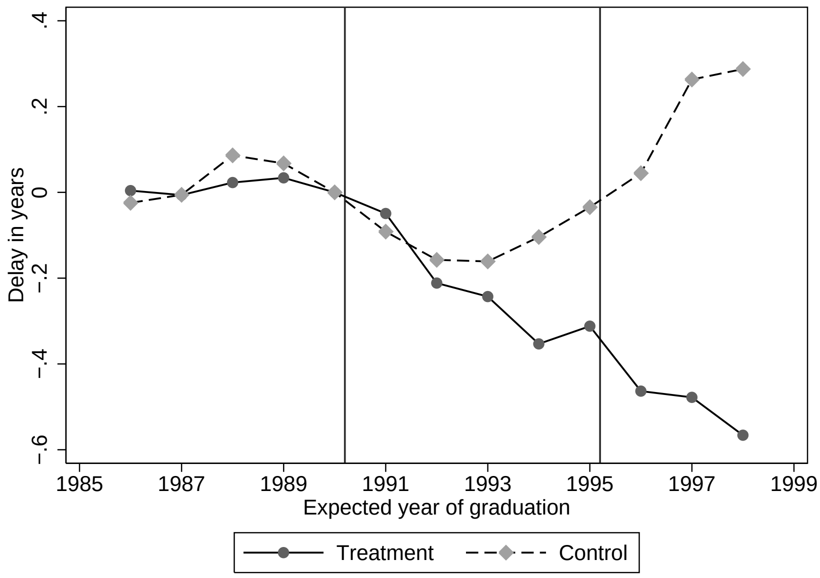
The effect on average delay is perhaps not as striking as the one on the probability to graduate on time, but this is expected. Average delay was a little more than two years in the treatment group, but a considerable share of the students were delayed by a lot more than that. These students were not, however, likely to respond to the

Figure 3: Delay by treatment status

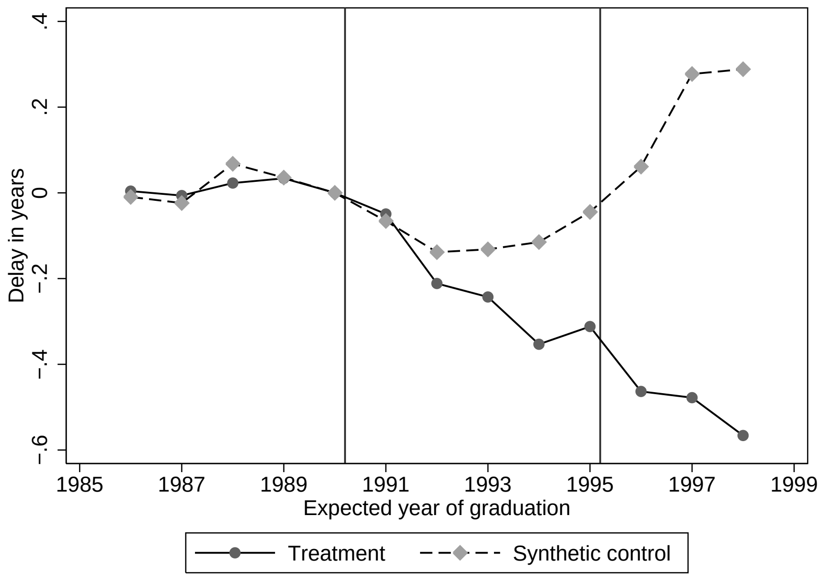


incentives and they would therefore keep the average delay high. The compliers are likely to be students who would otherwise have been only a little delayed, and they therefore contribute to a smaller response in average delay.

Figure 4: Demeaned delay by treatment status



(a) Baseline control group



(b) Synthetic control group

Table 13: The effect of the turbo reform on delay

	(1)	(2)	(3)	(4)
Reform period*Treated	-0.214**	-0.222**		
	(0.106)	(0.102)		
Post-reform period*Treated	-0.153	-0.179		
	(0.145)	(0.143)		
Years in treatment			-0.0634**	-0.0650**
			(0.0311)	(0.0304)
Female		0.172***		0.171***
		(0.0260)		(0.0263)
High school graduation age		0.0536		0.0531
		(0.0460)		(0.0467)
Mother's Education Middle		-0.0147		-0.0155
		(0.0404)		(0.0403)
Mother's Education High		0.0493		0.0489
		(0.0595)		(0.0593)
Father's Education Middle		-0.0166		-0.0153
		(0.0276)		(0.0277)
Father's Education High		0.0487		0.0498
		(0.0468)		(0.0466)
log Family income at age 16		0.0401*		0.0398*
		(0.0214)		(0.0216)
Immigrant status		-0.0448		-0.0455
		(0.0538)		(0.0540)
Constant	1.074***	-0.516	1.067***	-0.508
	(0.0553)	(0.767)	(0.0511)	(0.778)
R^2	0.401	0.406	0.401	0.407
Observations	34067	34067	34067	34067

Notes: All specifications include study program and cohort fixed effects. Columns 2 and 4 also contain dummies for region of residence at age 16 and unknown parental education. Parental education relative to low education. Educational levels defined as low=less than high school, middle=high school, high=post secondary education. Standard in parentheses errors clustered at study program-school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Estimated effect on delay by student characteristics

	(1) Baseline	(2) Female	(3) Par educ	(4) Fam inc	(5) IQ
Years in treatment	-0.0650**	-0.0563*	-0.0184	-0.0632*	-0.0280
	(0.0304)	(0.0298)	(0.0434)	(0.0346)	(0.0458)
Female	0.171***	0.193***	0.170***	0.171***	
	(0.0263)	(0.0334)	(0.0260)	(0.0264)	
Female*Years treated		-0.0166			
		(0.0103)			
Intermediate parental education			-0.0295		
			(0.0449)		
High parental education			0.0871		
			(0.0760)		
Intermediate parental education*Years treated			-0.0404*		
			(0.0221)		
High parental education*Years treated			-0.0533*		
			(0.0276)		
Family income 2nd quartile				0.0479*	
				(0.0266)	
Family income 3rd quartile				0.0392	
				(0.0376)	
Family income 4th quartile				0.104*	
				(0.0531)	
Family income 2nd quartile*Years treated				-0.00437	
				(0.0103)	
Family income 3rd quartile*Years treated				-0.00250	
				(0.0145)	
Family income 4th quartile*Years treated				-0.000531	
				(0.0205)	
Ability 2nd quartile					-0.000747
					(0.0896)
Ability 3rd quartile					0.0374
					(0.150)
Ability 4th quartile					0.148
					(0.172)
Ability 2nd quartile*Years treated					-0.00834
					(0.0274)
Ability 3rd quartile*Years treated					-0.0649
					(0.0407)
Ability 4th quartile*Years treated					-0.109**
					(0.0463)
Constant	-0.508	-0.528	-0.425	-0.0452	-0.389
	(0.778)	(0.785)	(0.737)	(0.810)	(0.731)
R^2	0.407	0.407	0.407	0.407	0.426
Observations	34067	34067	34067	34067	18652

Notes: Difference in difference estimates of one additional year of treatment, defined as studying in the reform period. All specifications include study program and cohort fixed effects and controls for age at high school graduation, region of residence and log family income and parental education at age 16. Standard errors are clustered at study program-school level and are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Estimated effect on delay by study program

	(1) Humanities	(2) Social Sciences	(3) Science	(4) Law	(5) Arts
Years in treatment	-0.189***	-0.0893***	-0.106***	0.00803	-0.0606**
	(0.0399)	(0.0186)	(0.0360)	(0.0182)	(0.0214)
Constant	0.189	0.434	-0.0940	-0.648	0.565
	(0.892)	(0.786)	(0.988)	(1.387)	(0.923)
R^2	0.460	0.372	0.471	0.337	0.242
Observations	18049	18429	19881	20418	15402

	(6) Theology	(7) Economics	(8) Psychology	(9) Dentistry	(10) Fishery
Years in treatment	-0.0175	-0.160***	0.00490	0.00291	-0.274
	(0.0244)	(0.0522)	(0.0324)	(0.0255)	(0.168)
Constant	0.539	0.507	0.596	0.479	0.423
	(0.909)	(0.939)	(0.843)	(0.926)	(0.993)
R^2	0.249	0.254	0.421	0.240	0.251
Observations	15737	15763	16387	15866	15331

Notes: Study program specific difference in difference estimates using the unweighted control group. Control variables included, see Notes in Table 5. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Estimated effect on delay using alternative control groups

	(1) Baseline	(2) Synth	(3) E+A	(4) Health
Years in treatment	-0.0650**	-0.0382	-0.0576*	-0.0352
	(0.0304)	(0.0299)	(0.0335)	(0.0212)
Constant	-0.508	-0.390	-1.514*	-0.864
	(0.778)	(0.854)	(0.881)	(0.501)
R^2	0.407	0.382	0.416	0.153
Observations	34067	31618	30215	3999

Notes: All specifications include study program, cohort fixed effects and six dummies for region of residence at age 16. Parental education relative to low education. Educational levels defined as low=less than high school, middle=high school, high=post secondary. Standard in parentheses errors clustered at study program-school level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Sensitivity checks: the effect of unemployment and cohort size

	(1) Baseline	(2) Unemployment	(3) log(Cohort)	(4) log(Cohort)	(5) log(Cohort)
Years in treatment	-0.0650**	-0.0628**	-0.0353	-0.0571**	-0.0159
	(0.0304)	(0.0287)	(0.0265)	(0.0269)	(0.0235)
Constant	-0.508	-0.679	1.110	-0.595	1.051
	(0.778)	(0.758)	(1.156)	(0.756)	(1.039)
R^2	0.407	0.409	0.408	0.409	0.410
Observations	34067	34067	34067	34067	34067

Notes: Difference in difference estimates of one additional year of treatment, defined as studying in the reform period. All specifications include study program and cohort fixed effects and controls for age at high school graduation, region of residence and log family income and parental education at age 16. Unemployment rate is national average unemployment rate for ages 25-54 measured at expected year of graduation. Cohort size refers to the number of students in the sample expected to graduate in a given year. Standard errors are clustered at study program-school level and are shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$