UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Master's Thesis:

*Gender peer effects in doctoral education: Evidence from Sweden**

Simon Lundin

University of Gothenburg

Septemeber 2018

Abstract

By understanding how we are affected by the other members of groups, interactive forums can be more optimally organised, giving rise to welfare increases. This thesis addresses two key research questions: (1) Does the gender composition of a doctoral student's cohort affect said student's academic performance and (2) if gender composition has an effect, does this effect differ between men and women? To address these questions, I use unique individual registry data on all individuals who been enrolled in a Swedish doctoral education from 1971 to 2010. I exploit the within program across cohorts variation in gender composition to obtain exogeneity. The results suggest a negative impact on male academic performance of a greater share of females in the cohort, while the results indicate that there is no overall effect on female performance. However, when examining the effect in different research fields separately, I find a statistically significant positive effect on female performance from a higher share of females within Engineering Sciences.

1 Introduction

Throughout life, we are part of different groups: at school, the workplace, the neighbourhood and so forth. The other individuals we are grouped with, our peers can potentially affect us in a number of ways, which is commonly known as peer effects. Understanding peer effects is important because knowledge about such effects can create opportunities to better organise different kinds of forums, which can then lead to social welfare-enhancements (Hoxby, 2000). In this master thesis, I study gender peer effects within education, i.e., the effects of classroom gender composition. The existence of gender peer effects in schooling has since long been of concern to educators, policymakers, and researchers. However, the scientific evidence is inconclusive (Lavy & Schlosser, 2011). I shed light on this topic by studying the effect of gender composition within doctoral classes on the likelihood of finishing the education by using rich, novel data on Swedish doctoral students.

Specifically, my primary research questions are as follows: (1) Does the gender composition of a doctoral student's cohort affect said student's academic performance and (2) if gender composition has an effect, does this effect differ between men and women? In this thesis, academic performance is measured in terms of (1) the likelihood of graduating within 7 years of entering the education, (2) net time to graduation and (3) gross time to graduation. To address the research questions, I use a new and rich dataset containing individual registry data over all individuals who have been enrolled in a Swedish doctoral education from 1971 until 2010. I examine these questions both overall and for the three main educational fields within my sample (Social Sciences, Engineering Sciences, and Natural Sciences) separately. The data is provided by Statistics Sweden and contains information on the doctoral students and the university where they attended the doctoral program. The main empirical challenge when estimating peer effects is self-selection of students into schools and programmes based on unobservable characteristics (Hoxby, 2000). I address this problem by only examining variation within each program across years while controlling for general time fixed-effect for all programs and program specific time trends, in line with Lavy and Schlosser (2013). The remaining variation between cohorts can plausibly be seen as exogenous, enabling the identification of a causal relationship.

In previous literature, different types of gender peer effects within education have been studied. For example, the effect on students' direction of study has been examined (see e.g. Schneeweis & Zweimueller, 2012; Black et al., 2013; Anelli & Peri, 2016; Brenøe,

2016; Feld & Zölitz, 2016; Hill, 2017)[1] and the effect on long-term outcomes (see e.g. Black et al., 2013; Feld & Zölitz, 2016)[2]. Of particular relevance for this thesis is the literature on the effect of gender composition on student achievements. A number of studies have examined this effect at primary to university (undergraduate) level (e.g. Hoxby, 2000; Whitmore, 2005; Lavy & Schlosser, 2011; De Giorgi et al., 2012; Jackson et al., 2012; Black et al., 2013; Park et al., 2013; Choi et al., 2014; Lee et al., 2014; Oosterbeek & Ewijk, 2014; Feld & Zöliz, 2016; Hill, 2017). This literature is discussed in greater depth in Section 2. To the best of my knowledge, no previous research has been conducted on any gender peer effect within doctoral education. Thus, my key contribution is to extend the research on gender peer effects to doctoral educations by exploiting within program across year variation in gender composition.

There are many mechanisms through which peer gender could affect the educational performance and the likelihood of completing an education. Hoxby (2000) described a few potential channels. According to her, peer effects can occur through students helping each other. Possibly, an increased share of women might have a positive impact on the achievements of both men and women since adult women have been found to be more altruistic than adult men (see the meta-analysis by Engel, 2011), which can imply that women help their fellow students more often. Hoxby (2000) also argued that peer effects could work through the students' influence on the classroom environment. For example, a student exercising disruptive behaviour can affect the other students' learning negatively. Another possible channel can be teacher behaviour. Possibly, teachers behave differently in classes with a larger share of males, which then affects the learning of the class. One study has empirically investigated the mechanisms behind gender peer effects in school, namely Lavy and Schlosser (2011). They found that the positive effect of a greater proportion of females on male's and female's cognitive outcomes are mediated

---

[1]Some studies have found evidence that females exposed to a higher fraction of female classmates are less likely to attain a STEM (science, technology, engineering, and mathematics) degree (Brenøe, 2016; Feld & Zölitz, 2016; Hill, 2017). Anelli and Peri (2016) found that males with a higher fraction of males are more likely to choose male-dominated majors. Schneeweis and Zweimueller (2012) found that female preschoolers with a higher fraction of female peers are less likely to choose female-dominated school types. Black et al. (2013) found that males are less likely to choose an academic track when the fraction of females is higher, while no effect was observed for women.

[2]Black et al. (2013) found a positive effect of a higher fraction of females on both female earnings and female full-time participation in the labour market, while the evidence indicated a negative effect for males, albeit no statistically significant effect could be found for males. Feld and Zölitz (2016) found that a higher fraction of females is associated with lower future earnings for females, while no effect on male labour market outcomes could be found.

through (1) less disruptive and violent behaviour, (2) better interpupil and pupil-teacher relationships, and (3) less teacher tiredness. However, disruptive behaviour might not be as relevant for older students (Oosterbeek & Van Ewijk, 2014).

Instead, for adult doctoral students, the behavioural effect from being in groups with more or less same-sex individuals could be of greater importance. Some recent studies point out that woman's behaviour could change in a male-dominated environment. For instance, women are less likely to take risks when they are surrounded by men (Sjögren Lindquist & Säve-Söderbergh, 2011; Booth & Nolen, 2012; Booth et al., 2014), and they chose to compete to a lesser extent in a male-dominated environment (Booth & Nolen, 2012; Hogarth et al., 2012; Burow et al., 2017). A strand of literature also investigates how well women and men performance in a competitive environment depending on the gender of their opponent(s). Some studies have also found evidence that women perform worse in a male-dominated environment, and that the effect is stronger when women compete with men (Gneezy et al., 2003; Antonovics et al., 2009; Kuhnen & Tymula, 2011; Sousa & Hollard, 2015; Grover et al., 2017). However, some papers have found evidence contradicting the hypothesis that women are less competitive (Dreber et al., 2011; Cárdenas et al., 2012; Dreber et al., 2014). I would argue that the environment a doctoral student face in many ways could be seen as competitive. Before the doctoral education reform from 1998 [3] institutions were not required to have a full financing plan for all admitted student. This could have led to a perceived or real competition between students to get renewed financing during their education if the institution offered such. Further, an important part of the doctoral education can be to apply for both scholarships and to present papers on conferences. In many cases, you might directly or indirectly compete with your peers in these application processes. A third potential factor to why a doctoral education could be viewed as competitive is because of the quite small groups of individuals who in some cases will compete for the same jobs after their education is finished.

Another potential channel could be that women get less recognition for their performances in an environment dominated by men compared to women. Recent research has found that even though women are generally undervalued compared to men, they are even more so by men (see, e.g., Grunspan et al., 2016; Boring, 2017; Mengel, et al., 2018). If this is true in doctoral cohorts as well a woman in a male-dominated cohort would get less recognition compared in a female dominated one, which could affect the

---

[3]Often referred to as the Tham-reform after the Minister of Education at the time - Carl Tham. For a more detailed description of the reform see section 5

academic performance.

A final channel can be the psychological theory of stereotype threats, as suggested by Oosterbeek and Van Ewijk (2014). The theory, developed by Steele and Aronson (1995), states that when an individual is confronted with a stereotype with which said individual has characteristics in common, and this stereotype is associated with poor performance, the performance of the individual is negatively affected. Oosterbeek and Van Ewijk (2014) argued that if stereotypes regarding female mathematical abilities are intensified in cohorts with lower proportions of women, females are predicted to perform better as the proportion of females increases. This could be especially important in STEM (Science, Technology, Engineering, and Mathematics) doctoral programs which are math intensive.

My results show that a larger share of female peers in a doctoral cohort has an overall negative impact on male performance within the same cohort, while no effect on female performance can be detected. Specifically, when the effect on the likelihood of graduating within 7 years is studied, no statistically significant effect is found for males, but the point estimates indicate a negative effect. The effect is estimated to be smaller in absolute terms for females, but no statistically significant gender difference is found either. However, when net and gross to graduation are used as dependent variables a negative effect for males is found. I estimate that a 10 percentage point increase in the share of women increases the net and gross time to graduation among males by approximately one twenty-third and one-tenth of a semester, respectively. For females, the effects are estimated to be close to zero for both net and gross time to graduation overall. However, when studying the main fields of education separately, a highly statistically significant and positive impact on female performance in engineering science is found for all three performance measures. When controlling how robust my main results are to the underlining assumptions made I find some inconclusive results, indicating that the interpretation of the results should be made with caution and that more research is needed.

The remainder of this thesis is structured as follows. In Section 2, an overview of the literature is provided. Section 3 outlines the hypotheses to be tested. Section 4 presents the empirical strategy. Section 5 describes the data. Section 6 presents first the main results and then the results for each of the three main fields of education. Section 7 includes robustness checks. Finally, Section 8 presents a discussion of the results and concludes.

2 Literature overview

The previous literature on the effect of gender composition on student achievements can be divided into two strands. The first focuses on contrasting the academic achievements in single-sex schooling as opposed to mixed-sex schooling (e.g., Jackson, 2012; Park et al., 2013; Choi et al., 2014; Lee et al., 2014)[4], while the second focuses on the effect on achievements of changing the proportion of females within mixed-sex schooling (Hoxby, 2000; Whitmore, 2005; Lavy & Schlosser, 2011; de Giorgi et al., 2012; Black et al., 2013; Oosterbeek & Ewijk, 2014; Feld & Zöliz, 2016; Hill, 2017). The latter strand is most relevant for this thesis. The findings within this strand generally indicate a positive effect of an increased share of females. However, the effect appears to be less apparent at a higher level compared to a lower educational level. Hoxby (2000), Whitmore (2005) and Lavy and Schlosser (2011) have all found that a larger share of females affects the performance of both males and females positively at primary and secondary education level. Only Black et al. (2013), who studied ninth graders in Norway, have found a negative effect of a larger fraction of females on years of education for males, while an indication of a positive effect for females was found.

At the university level, de Giorgi et al. (2012) found a U-shaped effect of increasing the share of women within the class on average grades, meaning that a larger share of females has a positive effect on academic performance only up to a certain threshold. Using data on Dutch university students, Feld and Zöliz (2016) found that an increased proportion of females is associated with increased female performance in non-mathematical courses, but not in mathematical courses. The reverse is true for males, i.e., a higher proportion of females leads to higher male performance in mathematical courses but not in non-mathematical courses. Overall, a positive effect on female performance is found, but hardly any effect on male performance. Hill (2017), on the other hand, found a positive impact of an increased share of females on the share of males graduating, but no effect on females graduating when using data from US colleges. Oosterbeek and Ewijk (2014) found no substantial effect of gender composition on academic achievement at the University of Amsterdam.

---

[4]The findings of the literature focused on single-sex schooling is mixed. Some studies have found a positive effect of single-sex schooling on male's academic performance (Park et al., 2013; Choi et al., 2014). Jackson (2012), however, did not find that men benefit from single-sex schooling, while women did. Lee et al. (2014) found evidence that male performance in single-sex classes in coeducation schools is lower than male performance in mixed-gender classes, while no effect is found for female students.

As discussed by Hoxby (2000), the primary challenge when estimating peer effects in education is to account for the self-selection of individuals into schools and classes. Previous studies have most often used one of two strategies to deal with this endogenous sorting. Firstly, one strand of the literature has used an experimental or quasi-experimental approach where students are randomly assigned to study groups, college roommates, or classes (see e.g. Sacerdote, 2001; Whitmore, 2005; Zimmerman, 2003; Carrell et al., 2009; Duflo et al., 2011; Oosterbeek & van Ewijk, 2014; Feld & Zöliz, 2016). The second commonly used strategy is to exploit idiosyncratic variation in group composition within the same school or university across different cohorts (e.g., Hoxby, 2000; Lavy & Schlosser, 2011; Black, et al., 2013; Hill, 2017). In my thesis, the latter empirical strategy is employed.

The first study to use the latter strategy is Hoxby (2000). She used the variation in gender and racial composition within Texas schools at primary education level across time to estimate the effect on educational outcomes. To ensure that the variation is not driven by time trends, she used specification tests, for example, one in which the order of the years was randomised. She found that if the proportion of girls increases by 10 per cent in a cohort, the test results in math and reading are raised by 8 and 4 per cent of a standard deviation, respectively. The effect is present for both boys and girls in the 3rd to 6th grade. As opposed to Hoxby (2000), I include a programme specific time trend to capture variation generated by time trends. This is in line with Lavy and Schlosser (2011), who used within school variations in gender composition across adjacent cohorts in Israel at primary and secondary education level. They examined the students' scholastic achievements in a number of ways, such as average score, number of credit units, test score in mathematics, etc. They concluded that an increased proportion of girls in a class has an overall positive effect on the performance of both boys and girls.

To summaries, previous research conducted on the effect of an increased share of females on student achievements generally indicates a positive impact for both male and female performance for lower levels of education, while the effect is less apparent for higher levels of education. I am aware of no study investigating this topic for doctoral education. Thus, I aim to fill this gap in the literature.

3   Hypotheses

The theories outlined in Section 1 generally indicate that a higher share of females leads to improved academic performance of both males and females. The empirical findings

presented in Section 2 also supports the notion that a larger share of women has a positive impact, albeit whether a positive effect is present for both males and females is less clear for education at higher level. Thus, I formulate Hypothesis 1, 2 and 3 as follows.

H1: Increasing the share of females in a doctoral cohort leads to an increase in the within 7 year graduation rate for a) males and b) females within the same cohort.

H2: Increasing the share of females in a doctoral cohort leads to a decrease in the net time to graduation for a) males and b) females within the same cohort.

H3: Increasing the share of females in a doctoral cohort leads to a decrease in the gross time to graduation for a) males and b) females within the same cohort.

Both the channel centred on competition and the channel centred on stereotypes presented in Section 1 provide a theoretical indication that the educational outcomes of females are affected to a larger extent than those of males. Hence, Hypothesis 4 is formulated as follows.

H4: Increasing the share of females in a doctoral cohort leads to a a) greater increase in the within 7 year graduation rate, b) greater decrease in net time to graduation and c) greater decrease in gross time to graduation for females than males.

4 Empirical Strategy

Students are most likely to systematically differ in unobservable characteristics that correlate with both the academic performance and gender among universities and academic fields. To correctly identify the peer gender effect, this selection problem needs to be addressed. I, in large, follow the method used by multiple other researchers (see, e.g., Hoxby, 2000; Lavy & Schlosser, 2011; Black, et al., 2013; Hill, 2017). Specifically, I use repeated cross-sectional observations where the within program across time variation in gender composition is exploited. This approach allows for heterogeneity among students in different doctoral programs. By also controlling for general time-specific effects and program-specific time trends I assume that students can be considered to be, as if, randomly assigned to a program. Based on this, I suggest the following regression specification:

$$y_{ipt} = \pi_1 P_{pt} + \pi_2 P_{pt} F_i + \alpha_p + \beta_t + \gamma_p * t + X_{ipt} \delta' + \epsilon_{ipt} \tag{1}$$

where, y_{ipt} is a measure of the outcome for individual i at the doctoral program p who started studying during semester t. Students who started the same doctoral program during the same semester are henceforth referred to as a cohort. α_p measures the program-fixed effect, β_t the time-fixed effects and $\gamma_p * t$ is program specific time trends. X_{ipt} is a row vector including covariates on the individual and program level[5]. P_{pt} is the proportion of women in the concerned doctoral program and cohort. In order to be able to detect heterogeneous effects of gender composition on male and female students, P_{pt} is also interacted with a dummy variable for female, F_i. Using this specification, π_1 measures the average effect on male students' outcome from an increase in the proportion of women in their cohort. Whereas π_2 shows if and how this effect differs for female students.

For the regression to produce unbiased estimates of the peer gender effect the strict exogeneity assumption conditional on the control variables has to be satisfied. The program fixed-effect α_p accounts for endogenous sorting into different programs by students. Further, the time fixed-effect, β_t, controls for potential trends or shocks among all doctoral programs over time that otherwise could lead to the observation of a spurious relationship. For instance, β_t captures business cycle effects on the job market which independently may affect both the proportion of women and the average ability among the admitted doctoral students in a specific year.

The program specific time-trend, $\gamma_p * t$, aim to control for two potential confounding factors. Firstly, some doctoral program may implement policies to increase both the number of women in their program and the graduation rates simultaneously. If this is unaddressed a spurious relationship between gender composition and graduation rates could be observed. Secondly, compared with younger pupils in primary education, doctoral students take a more active decision regarding if and in such case which doctoral program to apply for. Therefore, a doctoral student can make predictions regarding the gender composition of the future peers at different doctoral programs and potentially make their decision regarding which program to apply to based on these predictions. A program specific time-trend would control for the expected gender composition in each cohort. The residual variation in gender composition would then consist of unexpected shocks that an incoming doctoral student cannot foresee. In other words, the variation exploited when including $\gamma_p * t$ is the deviation from each program's long-term linear trend in gender composition[6]

---

[5]The covariates are: age, age^2, gender, an indicator for Swedish born and the cohort size

[6]See section 5 for discussion regarding the functional form of the program specific trend.

Based on the discussion above I would argue that conditional on the fixed effects and the other control variables, the variation in gender composition within each doctoral program can be regarded as quasi-random. Therefore, estimation of equation (1) is predicted to produce unbiased estimates of π_1 and π_2 which represent the casual relationships. To answer my first research question, if the gender composition of one's peers affects the probability of completing the doctoral education, I estimate equation (1) using a binary outcome variable measuring for each student if they had completed their doctoral program 7 years after starting it. The outcome variable takes the value 1 if so is the case and 0 otherwise. As the outcome variable is binary and the relationship might be non-linear, I use Probit maximum likelihood estimator where the standard errors are clustered at the program level in my analysis. If I find that the variable π_1 is statistically different from zero I can conclude that the peer gender effect does exist in doctoral educations. Also, I present the size of the effect by calculating marginal effects when all covariates are held at their respective means. To answer my second research question, if there is a difference in peer gender effect between men and women, π_2 is included. If I find that the variable is statistically different from zero, I can conclude that there is a gender difference.

For the other measures of performance, the gross and net time to graduation I will instead estimate equation 1 using OLS with the standard error clustered at the doctoral program level.

5   Data

The data source I use is from individual registry data maintained by Statistics Sweden. The dataset consists of linked administrative data from the Total Population Register and the University and Higher Education Register for all individuals who have been registered as a doctoral student at a Swedish university or university college between 1971 and 2010. The Higher Education Register is based on information reported each semester from the universities and university colleges. For each doctoral student they report, among else, which faculty the student was registered at and a local code indicating which research topic they were studying. However, the local code for the same research topic does not perfectly correspond among different universities nor within the same university across time. Therefore, Statistics Sweden also includes a national research topic code as a complement [7]. For example *economics* constitutes one research topic

---

[7]The coding system, National register over research topics (Nationell förteckning över forskningsämnen) in 1996, and it is the most specific level of research topic distribution constructed of approximately

code and *engineering physics* another.

I use three different performance measure for the students in the analysis; whether or not the student has obtained a doctoral degree after 7 years from the start of the education, the gross- as well as the net time to graduation. The limit of 7 years for the first variable is chosen out of two reasons. Firstly, as I have information regarding everyone who has been awarded a doctoral degree until 2017 and the last cohorts in my sample began their studies in 2010, it is the longest period that can be used without excluding students. Secondly, it is the median number of years until graduation among the students who obtained a doctoral degree in the estimation sample [8]

In this thesis, I define a doctoral program as a combination of a university or university college, a faculty, and a research topic code. Further, a cohort of students are a group of individuals registered as beginner doctoral student[9] in the same doctoral program and during the same semester. One drawback that I cannot control for, when only including beginner doctoral students in the cohorts is that individual who begin one doctoral program first, drop out from it and then begin to study at a new program will only be counted as a participant in the first program. However, it is probable that it is a small minority of the individuals in the sample who begun more than one doctoral education and argues that my definition yields the best possible proxy for the composition of incoming students to a program.

Both the locally given and the centrally given research topic codes are used independently to create two different sets of cohorts. Then, only the cases where the two methods yield the same cohort of doctoral students are used in the analysis. By doing so, approximately 48 per cent of all students are dropped[10]. However, this method has the advantage that I can utilise the finest distribution of research topics assigned to the student by their universities or university colleges, so the risk of wrongly classifying two

---

240 research topics. When the new national research topic codes were introduced in 1996 Statistic Sweden also retroactively complemented the information for all students since 1971 making it comparable for the whole period I am analysing. A complete list of all the research topic codes in Swedish can be found here: www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-indelning-av-forskningsamnen/

[8]My main results are robust if any number of years between 4 and 8 are used.

[9]I follow Statistics Sweden definition of a beginner doctoral student as an individual who for the first time is registered at a doctoral education with at least 10 per cent degree of activity

[10]In total 50,665 out of 105,696 identified doctoral beginners are dropped. However, there exist large differences among research fields for how well the nationally and centrally given research topic codes corresponds. Among the medical sciences as many as approximately 73 per cent of all students are dropped whereas the corresponding share in ontology is about 3 per cent.

students as peers is minimised. Additionally, it allows me to create longer time series when the local code for a specific research topic is changed from one semester to the next. Even though I assume that these research topic codes correspond well to a doctoral program, they are not designed for this purpose but rather to measure what doctoral student study. Therefore some codes might be to narrow, not include all students who are part of the same peer group, or too broad and include too many. Moreover, some of the codes may correspond well to a peer group in the 1970s but due to an enlargement of that entire research field, today is too broad. Much of these potential problems are dealt with by using both the locally and centrally given code but certainly not all. These limitations if further discussed in Section 8

Furthermore, all cohorts consisting of either less than three students or more than 25 are dropped from the sample [11]. How peer effects influence the success of a doctoral student is likely to be different if the peer group consists of only one other individual compared to multiple. Therefore, I limit the sample to only explore the peer gender effect in groups of at least three students, and drop all other cohorts. When using the method described above to define the cohorts 8 of them or less than 1 per cent of the students, are assigned to cohorts consisting of 25 or more students. In most such cases it is evident that it is due to mistakes done when reporting the information[12]. Therefore, I excluded all cohorts with more participants than 25.

I also exclude all students in a doctoral program where my definition of cohorts yields less than 5 cohort-observations. The variation exploited in my analysis is within program across years and to obtain sufficient variation to identify a relationship the same program need to be observed a couple of years[13]. Furthermore, I would argue that programs where my definition only yields 4 or fewer cohorts are less likely to be designed as a program where students interact on a regular basis.

Finally, the students in 8 doctoral programs where either all enrolled student or none of the enrolled student obtained their degree within 7 years are dropped due to lack of variation in the outcome variable. After these restrictions, the final sample consists of an unbalanced panel of with 21,961 students distributed over 4,387 cohorts in 400 doctoral programs. The cohort characteristics of the final sample are presented in Table 1.

---

[11]The main results presented in section 5 are only affected slightly if the lower limit is increased to 8 students, the upper one decreased to 15 or if both these changes are implemented simultaneously.

[12]the largest cohort according to my definition would consist of 183 students.

[13]In section 7 I present how the main result changes when 7 or 10 cohort-observations instead of 5.

Table 1: Summary statistics of the cohorts in estimation sample

Characteristic	Obs	Mean	Std. Dev.	Min	Max
# Students in cohort	4,387	5.1	2.8	3	25
# Female in cohort	4,387	1.9	1.8	0	14
Share female in cohort	4,387	0.37	0.29	0	1

As we can see, the average cohort consists of approximately 5.1 students and the average percentage of women in a cohort is 37 per cent. Furthermore, table 1 shows that there exists a quite considerable variation in the share of women between different cohorts with a standard deviation of about 29 per cent.

Summary statistics of the individual characteristics for the doctoral students are presented in table 3.

Table 2: Summary statistics of the students in estimation sample

Characteristic	Obs	Mean	Std. Dev.	Min	Max
Gross time to graduation	12,901	15.1	8.1	1	88
Net time to graduation	12,901	9.7	3.7	0.2	36
Ever obtained a degree	21,961	0.59	0.49	0	1
Degree within 7 years	21,961	0.38	0.48	0	1
Indicator if Swedish born	21,961	0.77	0.42	0	1
Female	21,961	0.37	0.48	0	1
Age	21,961	31	7.5	16	77

Among all individuals in the sample, approximately 59 per cent has obtained a doctoral degree before 2017. However, only about 38 per cent of the students managed to obtain the degree within 7 years from the year they first were registered as doctoral students, suggest differences in time to graduation.

Among the sub-sample of students who ever obtained a degree, the average gross time to graduation was just around 15.1 semester, or equivalently, 7.5 years. However, the time varies a lot in the sample from just 1 to 88 semesters with a standard deviation of approximately 8 semesters. The net time to graduation was on average about 9.7 semesters with a relatively smaller standard deviation of around 3.7 semesters. Around 37 per cent of the students in the sample are women, and the average age of all the

students is about 31 years. For both of these variables, there has been a clear trend over time. During the 5 first years, between 1971-1975, around 25 per cent of the beginner doctoral students were women, and the average age was approximately 25.5 years which can compare to 45 per cent women and 30.8 years for their counterparts between 2006-2010.

Figure 1 and 2 present the distribution of beginner doctoral students over time and research fields[14] respectively.
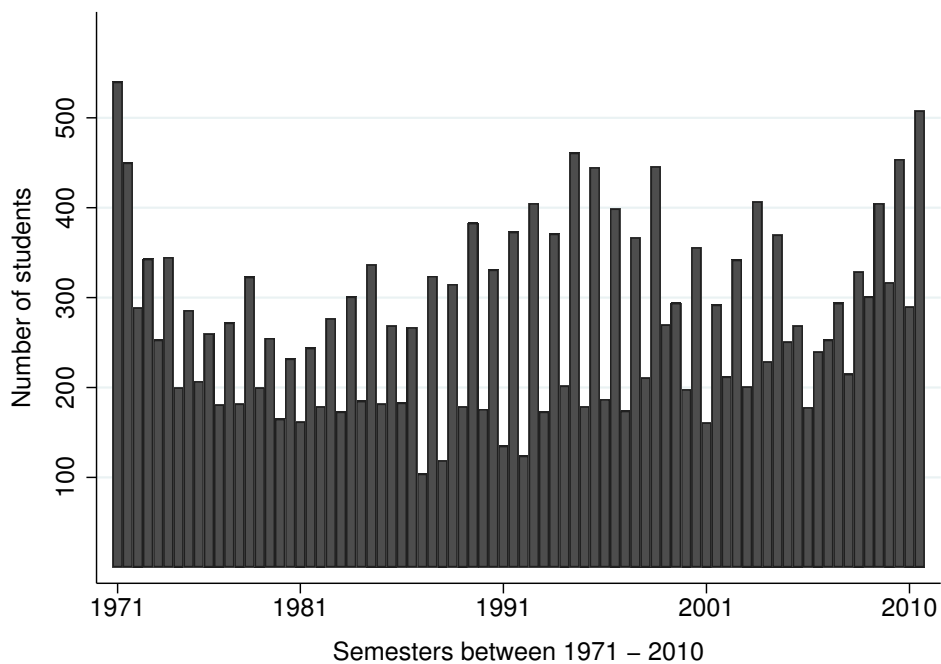

Figure 1: Number of admitted doctoral students for each semester in the estimation sample.

As we can see the pattern that more doctoral students in my sample begin during the fall compared to in the spring semester is quite clear. Also, we can observe a upwards trend in the number of admitted students from around 1975 until a few years before 2000. This deviation from the trend is most likely driven by the implementation of the doctoral education reform in 1998 [15]. The purpose of the reform was, among else, to enhance the efficiency of the education, shorten the time to graduation for the student

---

[14]The 11 research fields is defined by Statistic Sweden.

[15]Often referred to as the Tham-reform.

and tightening the regulation regarding financing of a doctoral education (Haraldsson, 2010). In practice, this significantly affected the possibilities for some intuitions to admit postgraduate students, as they now needed to ensure full financing of the student until graduation. Furthermore, in many institutions the structure of the whole doctoral program was also changed in connection with the reform (Haraldsson, 2010)[16]
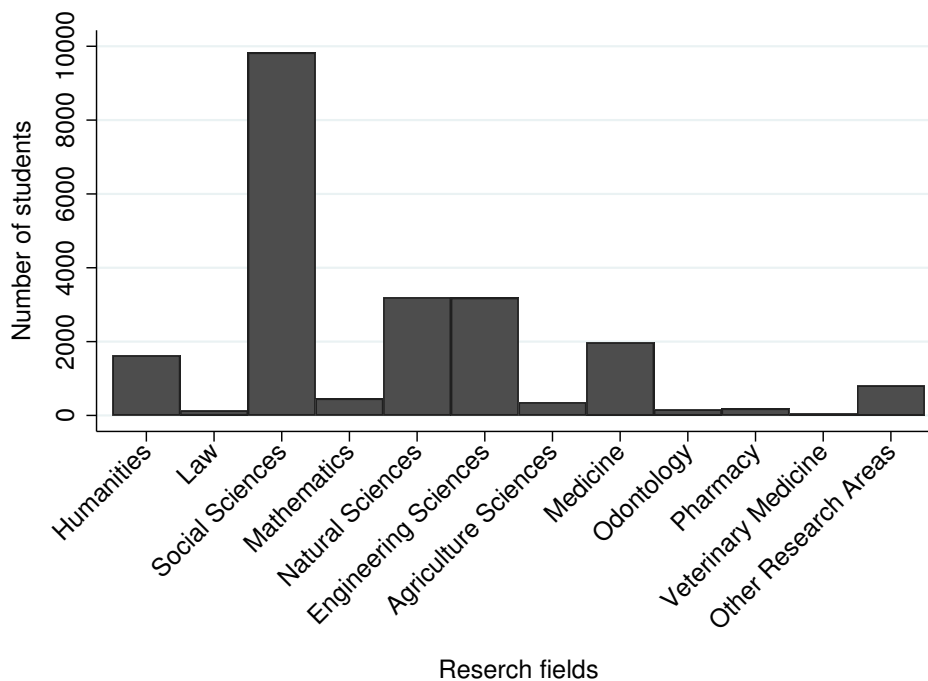

Figure 2: Number of doctoral students for each research fields in the estimation sample.

The distribution among the different research fields is uneven. This is quite natural as some fields are broadly defined, e.g., Social sciences whereas others are more narrowly defined as veterinary medicine. In the analysis, I will estimate equation 1 separately for each of the three largest research fields to investigate if there are differences among the doctoral program in different research fields.

Finally, in figure 3 I present how the share of women among the beginner doctoral students has changed over time. We can see a clear positive trend in the share of women over time. However, there also exists a notable systematic difference in the share of female beginner doctoral students during the fall and spring semesters[17]. In subsection

---

[16]I control for how robust my results are when taking this reform into account in section 7.

[17]A t-test conclude that the difference is statistically significant at the 1 per cent significant level.

5.1 I investigate the variation in gender composition on the program level in more detail.
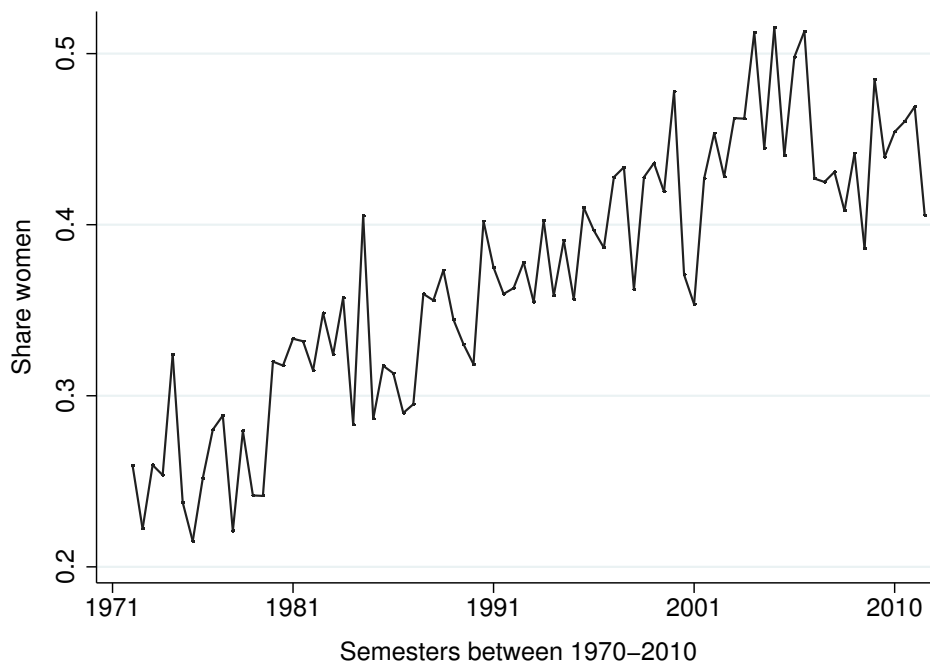

Figure 3: Average number of women among the beginner doctoral students for each semester

5.1   Variation in gender composition

A key concern when using my proposed model is that the variation in gender composition within doctoral programs is sufficiently large. As was evident in figure 3 and table 1 there exists a substantial variation in the share of women overall in my sample. However, the variation of interest is the residual variation in gender composition when I have accounted for that which be explained by the sorting of doctoral students into different programs, the overall changes among all program over time and a program-specific time trend. In table 3 I present the share of the total variation in the gender composition that can be explained by these factors. I do so by using the R^2 calculated from a regression where the share of women in a cohort is regressed on different combination of control variables. Further, in column (3) and (4) I examine how much of the residual variation, when both the time- and program fixed-effects are included that a linear or quadratic

15

program-specific time trend can explain [18]

Table 3: Variation in gender composition explained by fixed effects and program-specific trends.

	(1)	(2)	(3)	(4)
R^2	0.45	0.50	0.56	0.61
Adjusted-R^2	0.39	0.43	0.45	0.45
Program-fixed effects	✓	✓	✓	✓
Time-fixed effects		✓	✓	✓
Linear program-specific time trends			✓	
Quadratic program-specific time trends				✓
Observations	4,387	4,387	4,387	4,387

Note: Calculated R^2 and Adjusted-R^2 values from four regression where the *Sharewomen* in the cohorts is used as dependent variable.

Expectantly the program-fixed effects explain a large share, approximately 45 per cent, of the total variation in gender composition. When combined with time-fixed effects the model explains half of the total variation in gender composition. With the linear and quadratic programs-specific time trends also included in the regression the calculated R^2 increases to about 56 and 61 per cent respectively. These findings suggest that after controlling for this, a substantial amount of the variation still exist. We can also include that the Adjusted-R^2 does not increase when a quadratic program-specific time trend is used compared to a linear one. Based on these findings I will use a linear functional form for the program-specific time trends in my analysis.

5.2 Evidence on the Validity of the identification strategy

My identification strategy hinges on the assumption that the variation in gender composition, conditional on the control variable, can be treated as idiosyncratic within a program over time. Following Lavy and Schlosser (2009) I will try to examine this key assumption. A telling evidence that the variation is non-random would be if an association between the share of women in a cohort with the other covariates could be detected. To investigate this, I regress the covariates; age, an indicator if Swedish born and cohort

---

[18]This is done to find the most suitable functional form for $\gamma_p * t$ in equation 1.

size on the share of women in each students cohort. I also include both program- and time-fixed effect in the regressions. The estimated coefficients and standard errors from these regressions for the variable *ShareFemale* is presented in column 1 below in table 4. In column 2 the corresponding coefficients and standard errors when also a linear program-specific time trend is included are shown.

Table 4: Balancing Tests for the Share of Female Students

Dependent variable	(1)	(2)
Age	0.73*	0.56
	(0.33)	(0.29)
Swedish born	0.016	0.0074
	(0.17)	(0.018)
Cohort size	-0.26	-0.32
	(0.27)	(0.24)
Program-specific time trends		✓
Observations	21,961	21,961

Note: The table report the estimated coefficients from six separate regressions for the variable *Sharefemale*. Note that Age, Swedish born and Cohort size are the dependent variables in these regressions. All regressions also include program- and time-fixed effects. Robust standard errors clustered on doctoral program are reported in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For the covariates Swedish born and cohort size, I find no association with the share of women in the cohort. In the case for the age of the student, the estimated coefficient for share female is statistically significant at the 10 per cent level when I do not include linear program specific trends in the regression. However, when these trends are included the coefficient become non-statically significant. Therefore, I conclude that conditional on fixed effects and program specific trends I cannot find evidence against the assumption that the variation in gender composition within the same program over time is idiosyncratically distributed.

6 Results

6.1 Main results

My main results are presented in table 5. The six columns show the estimated coefficients for equation (1), described in Section 4, both with and without the linear program-specific time trends. The dependent variables are the three performance measures; graduation within 7 years (Model 1-2), net time to graduation (Model 3-4) and gross time to graduation (Model 5-6). The variable *Share female* is the first variable of interest and equal to the proportion of women in the concerned cohort. The estimated coefficient for *Share female* measures the average effect from changes in gender composition on the performance of male students. Similarly, the interaction term *Share female * Female* correspond the average effect from changes in gender composition on the performance of female students.

Table 5: Main results

	Degree in 7 yrs		Net time to graduation		Gross time to gradation	
	(1)	(2)	(3)	(4)	(5)	(6)
	Probit	Probit	OLS	OLS	OLS	OLS
Share female	−0.08	−0.10	0.44**	0.43**	0.87*	1.10**
	(0.07)	(0.08)	(0.22)	(0.21)	(0.47)	(0.48)
*Share female * female*	0.13	0.16	−0.21	−0.33	−0.77	−1.09*
	(0.11)	(0.11)	(0.31)	(0.31)	(0.66)	(0.61)
Program-specific time trends		✓		✓		✓
Observations	21,961	21,961	12,901	12,901	12,901	12,901

Note: All regression specification also include program-fixed effects, time-fixed effects and the following individual and cohort control variables: Age, age^2, indicator if Swedish born, indicator if female and cohort size. Robust standard errors clustered on doctoral program are reported in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

In Model 1 and 2, equation (1) is estimated with a Probit maximum likelihood estimator and a binary outcome variable indicating whether a student graduated within 7

years (*Degree in 7 yrs*) is used. In Model 1, program-specific time trends are excluded while they are included in Model 2. I find no statistically significant effect on the likelihood of either a male or female student obtaining a degree within 7 years in neither of the models. Moreover, the estimated average marginal effects are very close to zero in all cases.

In Model 3 and 4, the dependent variable is *Net time to graduation* while in Model 5 and 6, *Gross time to graduation* is used as the dependent variable instead. In these all four models, I restrict the sample to include only the 12,901 observations of students who have obtained a doctoral degree. The coefficients in these models are estimated with an OLS regression. For *Share female*, the estimated coefficients are consistently negative in all four models and statistically significant at a 5 per cent level across model 3, 4 and 6. In model 5 the estimated coefficient is marginally significant at the 10 per cent level. This suggests a negative effect of a higher share of females on male time to graduation. Specifically, the estimates imply that a 10 percentage point increase in the share of women in a cohort increases the net and gross time to graduation among males from that cohort by approximately one twenty-third and one-tenth of a semester, respectively. The coefficient for the interaction term is consistently estimated to be negative but only marginally significant in Model 6 at the 10 per cent level. This indicates that the adverse effect on time to graduation of a higher share of females is non-existing or at least less prominent for women compared to men. Based on the estimates, the effect on both female net and gross time to graduation is very close to zero.

6.2 Heterogeneous effects among scientific fields

To investigate if the gender composition of one's cohort affects students differently across scientific fields I estimate equation (1) separately for the three largest scientific fields in my sample namely; Social Sciences, Engineering Sciences, and Natural Sciences[19]. Table 6.2 shows the results for each of the three performance measures when investigating the three scientific fields separately.

For both Social and Natural Sciences the estimated coefficients for *Share female* and *Share female * Female* are close to zero when *degree in 7 years* is used as outcome measure. Furthermore, I can find no statistically significant effect on neither of the time to graduation measures used for doctoral programs in Social Sciences. In Natural

---

[19]Mathematics is classified as it's own scientific field by Statistic Sweden and is therefore not included in Natural Sciences.

Sciences, I estimate a negative and statically significant effect of the gender composition of one's cohort on the net time to graduation among male graduates, which is in line with the main results.

Within Engineering Sciences, I find a substantial and statistically significant effect of the proportion of women in a cohort on the performance of women while male performance seems to be unaffected. The coefficient for $Share female * Female$ in Model 7 where *degree in 7 years* is used as an outcome variable is positive and highly statistically significant, meaning that women are more likely to graduate within 7 years as the share of women in their cohort increase. The estimated coefficients correspond to an average marginal effect of an increase in $Share female$ of 0.12 for women. In other words, on average is a woman in an engineering doctoral program 1.2 percentage points more likely to obtain a doctoral degree within 7 years if the share of women in her cohort increases by 10 percentage points. Further, in Model 8 and 9 we can see that the time to graduation is affected for women in Engineering Sciences. A 10 percentage point increase in the share of women in one's cohort is estimated to decrease the time to graduation for a woman by about one-sixth and one-fourth of a semester for the net and gross time, respectively. These results indicate that the effect of gender composition on, especially womens, performance might be heterogeneous among different scientific fields, a possibility that I further discuss below in section 8.

Table 6: Estimates of the effect of gender composition on academic performance for the three largest scientific fields separately

	Social Sciences			Natural Science			Engineering Sciences		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Probit	OLS	OLS	Probit	OLS	OLS	Probit	OLS	OLS
	Degree in 7 yrs	Net-time	Gross-time	Degree in 7 yrs	Net-time	Gross-time	Degree in 7 yrs	Net-time	Gross-time
$Share\,female$	0.10	0.52	1.15	−0.19	0.61**	0.02	−0.25	0.57	0.33
	(0.13)	(0.41)	(0.99)	(0.18)	(0.28)	(0.46)	(0.23)	(0.39)	(0.66)
$Share\,female * female$	−0.05	−0.30	−1.52	0.18	−0.63	−0.04	0.62**	−2.39***	−2.87***
	(0.19)	(0.59)	(1.14)	(0.26)	(0.53)	(0.79)	(0.32)	(0.61)	(1.08)
Program-specific time trends	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	9,826	4,867	4,867	3,200	2,461	2,461	3,121	1,934	1,934

Note: All regression specification also include program-fixed effects, time-fixed effects, and the following individual and cohort control variables:

Age, age^2, indicator if Swedish born, indicator if female and cohort size. Robust standard errors clustered on doctoral program are reported in

parentheses.

*** p<0.01, ** p<0.05, * p<0.1

7 Robustness checks

Before looking at the data, I made up my mind about some choices related to the analysis like cohort definitions, the minimum number of cohort-observations to include a program and how to treat cohorts beginning during the spring semester compared to the ones starting during the fall. In this section, I will check what happens if I change some of these decisions, as well as a few other possible problems with the data. First I will see how my results change if students admitted during the fall and spring not are seen as comparable. Secondly, I will make the analysis separately for cohorts beginning before and after the doctoral education reform from 1998[20]. Thirdly, I will see how robust my results are to changes in the minim number of cohort-observations required.

7.1 Spring and fall cohorts

One concern with the analysis might be that cohorts starting during the fall and spring semesters are considered to be perfectly comparable to one another within programs. However, similarly, as I argued that students might differ between programs in characteristics, they might also do so within program between the spring and fall semesters. To check if this assumption threatens my identification, I estimate equation 1 where I treat cohorts in a program beginning during the fall and spring as actually belonging to two different programs. Henceforth, I refer to these newly defined groups of cohorts as *pseudo-programs*. I apply the same constraint as in the main analysis namely to only include the pseudo-programs with at least five cohort-observations. The estimated results from the regressions are presented below in table 7

When comparing these results with the main results presented in table 5 we can see that the signs of all except one estimated coefficients are the same between the two tables. The exception to the rule is the estimated coefficient for $Share female * female$ in Model 3 where I here estimate a positive effect compared to a negative one in the main results. Furthermore, multiple of the estimated coefficient for $Share female$ that were statistically significant in the main results here become non-significant. Although, the point estimates for the gross time to graduation are very similar and for the net time to graduation, they decrease by about 40 per cent. One possible explanation to why the estimated coefficients are not statistically significant in this specification is that the observations are decreased by around 20 per cent and the coefficients needed to be estimated in the

---

[20]A more detailed description of the reform and its purpose is presented in section 5

Table 7: Estimates of the effect of gender composition on academic performance for cohorts staring during the fall and spring separately

	Degree in 7 yrs		Net time to graduation		Gross time to gradation	
	(1)	(2)	(3)	(4)	(5)	(6)
	Probit	Probit	OLS	OLS	OLS	OLS
$Share\,female$	−0.05	−0.05	0.27	0.25	0.78	1.05*
	(0.08)	(0.10)	(0.24)	(0.24)	(0.59)	(0.60)
$Share\,female * female$	0.09	0.08	0.03	−0.10	−0.13	−0.50
	(0.12)	(0.12)	(0.38)	(0.38)	(0.82)	(0.82)
Pseudo-program-specific time trends		✓		✓		✓
Observations	17,002	17,002	9,781	9,781	9,781	9,781

Note: All regression specification also include pseudo-program-fixed effects, time-fixed effects and the following individual and cohort control variables: Age, age^2, indicator if Swedish born, indicator if female and cohort size. Robust standard errors clustered on pseudo-doctoral program are reported in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

models approximately doubles when controlling for the pseudo-programs instead for the actual programs. Combined these two factors decrease the number of freedoms in the models which could explain that the statistical significance levels changes in three of the models.

7.2 Heterogeneous effects over time

In my main analysis, I assume that two students enrolled in the same doctoral program at different points in time are comparable to one another. I argue that as they have chosen to study the same research topic at the same university, the institutional environment will be the same for the two students. However, as my sample stretches over a quite a long period, 40 years, it is possible that this assumption does not hold for all programs over the whole period. Moreover, in April 1998 a reform of the doctoral educational system took place in Sweden. The reform had a great influence on the regulations regarding

the doctoral education financing of doctoral students which changed the structure of many doctoral programs (Haraldsson, 2010). Below in table 8 I present the estimated coefficient when my sample is restricted to only include either observations before or after the reform in 1998.

Table 8: Estimates of the effect of gender composition on academic performance for cohorts staring before and after the reform in 1998 separately

	Before reform in 1998			After reform in 1998		
	(1)	(2)	(3)	(4)	(5)	(6)
	Probit	OLS	OLS	Probit	OLS	OLS
	Degree	Net-	Gross-	Degree	Net-	Gross-
	in 7 yrs	time	time	in 7 yrs	time	time
$Share\,female$	−0.04	0.39	0.96	−0.17	0.26	0.68*
	(0.10)	(0.29)	(0.71)	(0.14)	(0.24)	(0.35)
$Share\,female * female$	0.12	−0.13	−0.92	0.19	−0.29	−0.64
	(0.16)	(0.48)	(0.96)	(0.16)	(0.28)	(0.45)
Program-specific time trends	✓	✓	✓	✓	✓	✓
Observations	14,595	7,659	7,659	7,366	5,242	5,242

Note: All regression specification also include program-fixed effects, time-fixed effects and the following individual and cohort control variables: Age, age^2, indicator if Swedish born, indicator if female and cohort size. Robust standard errors clustered on doctoral program are reported in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

The signs of the estimated coefficients in table 8 all correspond to their counterparts in the main results. However, none of the coefficients are more than marginally significant in any of the models. We can also see that the estimated effect sizes generally are stronger for the period after 1998 compared to before.

7.3 Minimum number of cohort-observations per program

The final assumption I test in this section is regarding the minimum number of cohort-observation needed for a doctoral program to be included in the analysis. The reason for

having this assumption is twofold. Firstly, to be able to estimate the probit regression when a binary outcome variable is used at enough variation within a program across time in gender composition is needed. Secondly, a program with fewer cohort observations is less likely to be designed as a program where students regularly interact with each other. I assume that if an institution continuously admits more than three students to study the same research topic over time the likelihood that students follow a similar syllabus increase.

Below in table 9 I present how my main results changes if I instead require seven or ten cohort-observations for a doctoral program to be included in the analysis.

Table 9: Estimates of the effect of gender composition on academic performance when only including programs with at least seven or ten cohort-observations

	Min. 7 cohort-observations			Min. 10 cohort-observations		
	(1)	(2)	(3)	(4)	(5)	(6)
	Probit	OLS	OLS	Probit	OLS	OLS
	Degree	Net-	Gross-	Degree	Net-	Gross-
	in 7 yrs	time	time	in 7 yrs	time	time
$Share_{female}$	−0.05	0.27	0.80	−0.08	0.17	0.83
	(0.09)	(0.24)	(0.55)	(0.10)	(0.26)	(0.66)
$Share_{female} * female$	0.11	−0.14	−0.59	0.18	0.07	−0.75
	(0.13)	(0.37)	(0.71)	(0.15)	(0.44)	(0.84)
Program-specific time trends	✓	✓	✓	✓	✓	✓
Observations	17,685	10,050	10,050	14,169	7,820	7,820

Note: All regression specification also include program-fixed effects, time-fixed effects and the following individual and cohort control variables: Age, age^2, indicator if Swedish born, indicator if female and cohort size. Robust standard errors clustered on doctoral program are reported in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Again, the estimated coefficients lose their statistical significance compared to the main results. Given that the change from five cohort-observation to seven can be seen as rather negligible the difference in statistical significance level is noticeable. I discuss this fact more in detail in section 8

8 Discussion and conclusions

This master thesis concerns gender peer effects within education. Specifically, I study the effect of a larger share of female peers within a doctoral cohort on academic performance. I also examine whether this effect, if it exists, differs between men and women. I measure academic performance as (1) the likelihood of graduating within 7 years, (2) net time to graduation and (3) gross time to graduation. To fulfill these aims, I use a large, unique data set on Swedish doctoral students, spanning the years 1971-2010. It is individual registry data maintained by Statistics Sweden. After imposing restrictions, my total sample when studying the effect on the graduation rate within 7 years consists of 21,961 observations, while the corresponding figure is 12,901 when studying the effect on time to graduation.

A key difficulty when estimating gender peer effects within education is nonrandom self-selection of students into schools and programmes (Hoxby, 2000). This difficulty is overcome in my thesis by controlling for time fixed-effects for all programmes and program specific time trends, in line with Lavy and Schlosser (2009). This should provide exogenous variation between cohorts.

I estimate a negative effect of a larger share of women on the likelihood of males graduating within 7 years, but the effect is not statistically significant. The negative impact on the graduation rate is estimated to be smaller for females, but no statistically significant difference between males and females is detected. Thus, no support can be provided for Hypothesis 1. Moreover, the results indicate that a larger share of females in a doctoral cohort increases both the net and gross time to graduation of males. This effect is consistently statistically significant at either the 5 or 10 percent level. For women, this adverse effect is estimated to be smaller, but the difference is only statistically significant at the 10 percent level for the model in which gross time to graduation is the dependent variable and program-specific time trends are included. The estimated effect on female time to graduation is close to zero across models. These findings are contrary to Hypothesis 2a and 3a, while no support is found for Hypothesis 2b nor 3b. As the effect for females is not estimated to be positive, no support for Hypothesis 4 is provided either. Conclusively, the results suggest a negative impact on men's academic performance of a larger share of female peers within doctoral education, while no effect on female performance is found. This negative impact for males is contrary to what I hypothesised and can be considered surprising given the theory presented in Section 1 and the previous empirical findings discussed in Section 2.

Some previous studies have found a positive effect of a larger share of females on the academic performance of either males, females or both (Hoxby, 2000; Whitmore, 2005; Lavy & Schlosser, 2011; de Giorgi et al., 2013; Feld & Zöliz, 2016; Hill, 2017). An explanation for why I do not find a positive effect can be differences in the educational environment. The most convincing empirical evidence supporting the notion that more female peers affect both male and female performance positively are provided at primary and secondary education level (see Hoxby, 2000; Whitmore, 2005; Lavy & Schlosser, 2011). Possibly, the most important channel for the peer gender effect at this level is less disruptive behaviour, a channel discussed by both Hoxby (2000), and Oosterbeek and Ewijk (2014). Empirical evidence for the presence of this channel has also been found (Lavy & Schlosser, 2011). Since this channel reasonably is not relevant at the doctoral level given the age of the students, the positive effect of more female peers might disappear.

Nevertheless, a statistically significant albeit adverse effect on male net and gross time to graduation is found. To the best of my awareness, only one previous study has detected a negative effect of a larger share of females on male performance, namely Black et al. (2013). Interestingly, this is the only other study I am aware of conducted in a Scandinavian country. Possibly, Scandinavian males are affected differently by a larger share of females than for example American males, which is a more common pool of subjects. This might be due to for example cultural differences or differences in the educational organisation. Moreover, it is interesting to note that the point estimates are twice as high when gross time to graduation is used as the dependent variable as opposed to net time to graduation in the main results. This could be due to several reasons. For example, if more women influence gender norms, a higher share of women in a cohort could lead to that male students take more paternity leave. This mechanism might arise because (1) men are more likely to have children during the doctoral education when they have more female peers and/or (2) men with children are more inclined to take paternal leave for those children when they have more female peers. This would not affect net time to graduation, but only gross time to graduation.

In the analysis in which the effect of a larger share of females is studied separately for each of the three largest fields within my sample, I find no gender differences for social sciences and science. However, a highly significant, positive impact on female performance is found for all three performance measures within engineering science. Specifically, I estimate that if the share of women increases by 10 percentage points, then on average, a female engineering student is 1.2 percentage points more likely to obtain a doctoral

degree within 7 years. In line with this, a 10 percentage point increase in the share of women is estimated to decrease the female time to graduation by about one-sixth and one-fourth of a semester in net and gross time, respectively. Hence, I find evidence in support of Hypothesis 4 for engineering sciences in particular. This is not too surprising given that, as discussed in Section 1, the theory presented by Oosterbeek and Van Ewijk (2014) concerning the importance of stereotypes for performance might be more prominent in STEM programs. The reason for this would be that they are math intensive, which might intensify the importance of stereotypes of female mathematical capabilities.

As is common in studies using large registry data set I have been forced to make assumptions that do not perfectly correspond to the reality. Each one of the assumptions made naturally constitutes a threat to identifying the actual effect of gender composition of ones cohort on the academic performance. When defining cohorts, I did so by the use of research topic codes that in some cases may not lead to that the individual who makes out a peer group are defined as such. Moreover, I needed to assume that only individuals defined as beginner doctoral students by Statistics Sweden[21] within a program constitutes the cohort some could have been missed. If an individual either begin at different doctoral education or retake the first semester in the same one, she will only be counted as belonging to the peer group in one of the two cases. Together, these reasons could lead to that the cohort size and calculated gender composition in some cohorts do not correspond to the actual one, even if this is likely to be rare.

I also addressed possible limitations in section 7. In the main analysis, I assumed that it is irrelevant if a cohort begins during the fall or spring semester. When I tested how robust my results were if I instead assume the opposite - that cohorts beginning during the fall in one doctoral program are not comparable to cohorts beginning during the spring - I find that it especially affected the statistical significance of my results. This could be due both to that the original assumption is false or the fact that the degrees of freedom in the robustness checks model were much lower. Moreover, the statistical significance of my main results was not robust to neither changes in the included years in the analysis nor changes in the required number of cohort-observations per program. Due to these inconsistent results in the robustness tests, the main results must be interpreted with some caution.

A final limitation of my study could be that the three performance measures used are bad proxies for actual academic performance. I would argue that it is probable that high-

---

[21]Statistics Sweden definition of a beginner doctoral student as an individual who for the first time is registered at a doctoral education with at least 10 percent degree of activity.

performing students, in general, will complete their education quicker due to financial incentives and career ambitions. Nonetheless, it is possible that time to graduation and obtaining a degree does not perfectly correspond to either knowledge or performance in all cases.

In this master thesis, the knowledge on gender peer effects within education has been strengthened by extending the research to education at the doctoral level. This topic is of importance since knowledge about such effects can allow the administration of educational programs to more optimally organise the education, which can result in greater welfare (Hoxby, 2000). As this master's thesis is the first to study the gender peer effect at the doctoral level, more research is needed at this level of education in other contexts. An interesting extension on the research on peer gender effects in education would be to conduct more research on such effects in Scandinavian countries at all levels of education to explore whether there are geographical differences. *hej & hej*

9   References

Anelli, M., & Peri, G. (2016). Peers' Composition Effects in the Short and in the Long Run: College Major, College Performance and Income. unpublished manuscript.

Antonovics, K., Arcidiacono, P., & Walsh, R. (2009). The effects of gender interactions in the lab and in the field. *The Review of Economics and Statistics, 91*(1), 152-162.

Black, S., Devereux, P., & Salvanes, K. (2013). Under Pressure? The Effect of Peers on Outcomes of Young Adults. *Journal of Labor Economics, 31*(1), 119-153.

Booth, A., Cardona-Sosa, L., & Nolen, P. (2014). Gender Differences in Risk Aversion: Do Single-Sex Environments Affect Their Development? *Journal of Economic Behavior and Organization, 99*, 126–154.

Booth, A., & Nolen, P. (2012a). Gender Differences in Risk Behaviour: Does Nurture Matter? *Economic Journal, 122* (558).

Booth, A., & Nolen, P. (2012b). Choosing to Compete: How Different Are Girls and Boys? *Journal of Economic Behavior and Organization, 81* (2), 542–555.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics, 145*, 27-41.

Burow, N., Beblo M., Beninger, D., & Schröder M. (2017). Why Do Women Favor Same-Gender Competition? Evidence from a Choice Experiment. *DIW Berlin Discussion Paper* No. 1662.

Brenoe, A. (2017). High School Gender Composition and STEM Choice in Tertiary Education. Unpublished manuscript.

Cárdenas, J.C., Dreber, A., Von Essen, E., & Ranehill, E. (2012). Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *Journal of Economic Behavior and Organization, 83*(1), 11-23.

Carrell, Scott E., Richard L. Fullerton, & James E. West. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics, 27*, 439–464.

Choi, E. J., Moon, H. R. & Ridder, G. (2014). Estimation of an Education Production Function under Random Assignment with Selection. *American Economic Review: Papers and Proceedings, 104*(5), 206-211.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *The coleman report. Equality of Educational Opportunity.*

Dreber, A., Essen, E., & Ranehill, E. (2014). Gender and competition in adolescence:

Task matters. *Experimental Economics, 17*(1), 154-172.

Dreber, A., Essen, E., & Ranehill, E. (2011). Outrunning the gender gap - boys and girls compete equally. *Experimental Economics, 14* (4), 567-582.

Duflo, E., Dupas p., & Kremer M. (2011). Peer effects and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review 101*, 1739–1774.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics, 14*(4), 583-610.

Feld, J., & Zölitz, U. (2017). The effect of peer gender on major choice (No. 270). Department of Economics-University of Zurich.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics, 118*(3), 1049-1074.

Grover, S. S., Ito, T. A., & Park, B. (2017). The effects of gender composition on women's experience in math work groups. *Journal of personality and social psychology, 112*(6), 877.

Grunspan, D. Z., Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, A. J., & Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PloS one, 11*(2), e0148405.

Haraldsson, J. (2010). "Det ska ju vara lite äventyr" - styrning av svensk forskarutbildning utifrån reformen 1998.["There has to be an element of adventure" governing Swedish graduate studies according to the 1998 reform] (Doctoral dissertation, Faculty of Education, University of Gothenburg).

Hill, A. J. (2017). The positive influence of female college students on their male peers. *Labour Economics, 44*, 151-160.

Hogarth, R. M., Karelaia, N., & Trujillo, C. A. (2012). When should I quit? Gender differences in exiting competitions. *Journal of economic behavior and organization, 83*(1), 136-150.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No. 7867.

Jackson, C. K. (2012). Single-Sex Schools, Student Achievement, and Course Selection: Evidence from Rule-Based Student Assignments in Trinidad and Tobago. *Journal of Public Economics, 96* (1-2), 173-187

Kuhnen, C. M., & Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science, 58*(1), 94-113.

Lavy, V., & Schlosser, A. (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics, 3*(2), 1-33.

Lee, S., Turner, L.J., Woo, S., & Kim, K., (2014). All or Nothing? The Impact of School and Classroom Gender Composition on Effort and Academic Achievement. NBER Working Paper. No. w20722.

Mengel, F., Sauermann, J., & Zölitz, U. (2017). Gender bias in teaching evaluations. *Journal of the European Economic Association.*

Oosterbeek, H., & Van Ewijk, R. (2014). Gender peer effects in university: Evidence from a randomized experiment. *Economics of Education Review, 38*, 51-63.

Park, H., Behrman, J. R. & Choi, J. (2013). Causal Effects of Single-Sex Schools on College Entrance Exams and College Attendance: Random Assignment in Seoul High Schools. *Demography, 50*(2), 447-469.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics, 116*, 681–704.

Sacerdote, B. (2011). Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?, Editor: Hanushek, E., Machin S., Woessmann L., In: *Handbook of the Economics of Education*, Elsevier, Volume 3, 2011, Pages 249-277.

Schneeweis, N., & Zweimüller, M. (2012). Girls, girls, girls: Gender Composition and Female School Choice. *Economics of Education Review, 31*(4), 482-500.

Sjögren Lindquist, G., & Säve-Söderbergh, J. (2011). 'Girls Will Be Girls', Especially among Boys: Risk-Taking in the 'Daily Double' on Jeopardy. *Economics Letters, 112*(2), 158–60

Sousa, J. & Hollard, G. (2015). Gender differences: evidence from field tournaments. CEPREMAP Working paper No. 1506.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.

Whitmore, D. (2005). Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. *American Economic Review, 95*(2), 199-203.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*:9–23.