European University Institute

DEPARTMENT OF ECONOMICS

Five Essays on Economics of Education

Ohto Kanninen

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 18 December 2013

DOI: 10.2870/93297

European University Institute
Department of Economics

Five Essays on Economics of Education

Ohto Kanninen

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Examining Board

Stefano Gagliarducci, University of Rome Tor Vergata
Luigi Guiso, Supervisor, Einaudi Institute for Economics and Finance
Andrea Ichino, EUI
Kjell Salvanes, Norwegian School of Economics and Business Administration

Kanninen, Ohto (2013), Five essays on economics of education
European University Institute DOI: 10.2870/93297

 DOI: 10.2870/93297

Abstract

In the first part of the thesis (Chapters 1 to 4), we analyze the near-universal gender gap reversal in secondary and tertiary education. In virtually all countries, males show a greater dispersion in ability test scores relative to females. We show that this simple fact, combined with an increase in the returns to education across cohorts, is sufficient to reproduce the gender gap reversal observed internationally. We bould a model that generates a hump-shaped relationship between the enrollment rate in education and the female-to-male ratio among the enrolled that is consistent with the data. From time-series data on enrollment rates in education by sex, we generate country estimates for gender differences in ability distribution using our model. Our estimates highly correlate with cross-country gender differences in test score distributions found in PISA. We also assess the validity of our theory against two alternative explanations for the gender gap dynamics: changes in social norms, and improvements in females' relative performance at school over time. The data does not support the predictions of the alternative hypotheses, while bringing further support to our theory.

In the second part of the thesis (Chapter 5), using Finnish high school data, I examine the relationship between peer composition and the causal effect of school choice on high school exit examination outcomes. To discern the causal effect of school choice, I exploit over 300 regression discontinuity designs that result naturally from the Finnish educational system that allocates pupils to high schools according to their ninth grade grade point average and announced preferences. I find strong evidence that high school choice matters in Finland and that it is related to peer composition. The class composition effect, however, is associated with peer homogeneity rather than average peer quality. I find that a standard deviation change in the homogeneity of peers is positively associated with a 0.02 to 0.13 standard deviation change in the exam results. I also find that the average effect of being marginally above the entrance threshold reduces slightly but significantly the performance of the pupil. This unexpected finding might be a sign of overconfidence on the part of the pupils in making their school choice.

Kanninen, Ohto (2013), Five essays on economics of education
European University Institute
DOI: 10.2870/93297

Acknowledgements

I would like to thank my supervisors Luigi Guiso and Andrea Ichino and the rest of the EUI staff for superb guidance in becoming a researcher. I would also like to thank Jerôme Adda, Christian Dustmann, Peter Hansen, Tuomas Pekkarinen, Roope Uusitalo and all of the other numerous scholars I've encountered along the way. Ilpo Suoniemi and Yrjö Vartia were early influences in pushing me in to an academic path, of which I'm grateful. My co-author Laurent Bossavie, who is on his way to becoming a relevant thinker of his own right, deserves a special mentioning. I would also like to thank Seija Ilmakunnas and Reija Lilja and all the rest of the staff at Helsinki Labour Institute for Economic Research for facilities and support. Lastly, I want to voice my gratitude to my friends (you know who you are and each one equally important), my family and Dhamma.

DOI: 10.2870/93297

Preface

Questions pertaining to education are relevant to the field of economics for a multitude of reasons. The most cited reason is probably economists' interest in the accumulation of human capital. We want to know how inputs in to education affect the outputs. Education is also a relevant factor in questions of gender equality. The two sexes' opportunities to obtain education and the realized differences in educational attainment are important factors in determining many gender related outcomes. In this doctoral thesis, I explore these two dimensions of the economics of education.

The thesis is divided in to two parts. The first part of the thesis (Chapters 1 to 4) is co-authored with Laurent Bossavie and it aims to tackle questions related to the observed near-universal gender gap reversal in secondary and tertiary education and in it we present a new framework for studying the educational gender gap. The framework allows us to present the different theories in a formal and exact manner and compare their prediction with empirical facts. We also present and assess a new theory for the educational gender gap dynamics over the last 40 years, which is based on the larger variability of males' test score distribution relative to females. We show that the theory is consistent with empirical data, and cannot be dismissed. We also showed that alternative theories proposed by the literature are unable, alone, to explain some of the patterns observed in the data. Meanwhile, our theory appears to be able to replicate all the main patterns observed in the data.

In Chapter 1 (Gender Inequality in Educational Outcomes: The Facts), we present four international facts about the gender gap in educational attainment. We show that these facts are near-universal and observed in the very large majority of countries, including developing countries. We report that most countries have nowadays a female majority in tertiary education, and that in all these countries the gender gap has reversed over time from male majority. Further, we show that there exists a similar gender gap in secondary school non-completion rates in most countries, and that it has reversed from female majority to male majority over the last decades. Although the college gender gap reversal has been noted in the literature for the US and some other western countries, we are the first to show that it applies internationally. To our knowledge, we are also the first contribution to observe the analogous gender gap reversal in high school non-completion rates.

In Chapter 2 (Explaining Gender Differences in Educational Outcomes: A Statistical Theory), we present a new theoretical framework for studying the relationship between enrollment rates and gender ratios in enrollment. The framework combines an optimal choice model of education, in which the agents' choice of educational level is a function of their test-taking ability, with different underlying distributions of test-taking ability for males and females. In particular, males exhibit higher variability in abilities, including test-taking ability. The female-to-male gender ratio in enrollment is presented as a function of the enrollment rate, some characteristics of the test-taking abilities of females and males and the possibly gender-specific changes in the entrance requirements. The fundamentally different aspect of the framework is that the enrollment rate is given a prominent role.

Our third contribution (in Chapter 3, Explaining Gender Differences in Educational Outcomes: Fitting and Testing the Theory) is to show that we are able to capture the essential dynamics of gender differences in educational outcomes. In particular, our novel framework allows us to predict the gender gap reversal at both the university level and high school level internationally. This is the first time there has been an effort to understand the college gender gap reversals in educational fortunes of boys and girls at the international level. It is also the first theory to explain the reversal in high school non-completion rates. Chapter 3 shows that our theory for the internationally observed facts about gender inequality in education is consistent with the data, and cannot be dismissed as an potential explanation for the educational gender gap dynamics.

Finally, Chapter 4 (Explaining Gender Differences in Educational Outcomes: Evaluating Alternative Hypotheses) shows that alternative theories proposed by the literature are inconsistent with some of the patterns observed in the data. Using the framework we develop in Chapter 2, we formalize the two main alternative explanations to the gender gap dynamics: changes in social norms and an increase in females' mean performance in tests. By doing so, we are able to compare the predictions of these alternative hypotheses with the ones given by our theory in a common framework, and to confront them with empirical data. We propose and perform several tests for these two alternative hypotheses using empirical data. While our theory is not rejected by these tests, all of them dismissed the change in social norms hypothesis as the main driving force between the gender gap dynamics.

The policy implications of this body of work (Chapters 1 to 4) are not obvious. Rather, the insights gained in this analysis would suggest that one should tread carefully when drawing conclusions from results showing that social norms are a major factor behind most observed gender differences in educational outcomes. We believe that the framework presented here could be useful in future analyses of educational outcomes. Also, our work suggests that one should be careful when drawing policy conclusions about the differences in male and female high school non-completion rates. Those differences might be due to tail effects or differences in the means of the distributions. Policy responses should be sensitive to this fact.

In the second part of the thesis (Chapter 5), using Finnish high school data, I study the relationship between peer composition and the causal effect of school choice on high school exit examination outcomes, a field that has been studied before. However, I use novel and more accurate data than most previous authors, since I know the exact preferences of the high school applicants. To discern the causal effect of school choice, I exploit over 300 regression discontinuity designs that result naturally from the Finnish educational system that allocates pupils to high schools according to their ninth grade grade point average and announced preferences. Although Finland offers a relatively egalitarian setting for high school students at least in terms of expenditures per student and student quality before high school, the result from this paper is that school choice does matter and that it appears to stem partly from class composition. The class composition effect, however, is associated with peer homogeneity rather than average peer quality. I find that a standard deviation change in the homogeneity of peers is positively associated with a 0.02 to 0.13 standard deviation change in the exam results. The evidence suggests that on average the students who are at the threshold are worse off getting in to their favored school than the control group that just missed their first preference, although the former group gets better-achieving and more homogenous peers on average. I propose the working hypothesis that the high-aiming applicants might actually be overconfident in the application process. They apply for a school that will be ultimately harmful for them, since the teaching is aimed at students that are on average better-achieving.

Contents

List of Figures

DOI: 10.2870/93297

List of Tables

Kanninen, Ohto (2013), Five essays on economics of education
European University Institute

DOI: 10.2870/93297

1

Gender Inequality in Educational Outcomes: The Facts

with Laurent Bossavie

1.1 Introduction

There has been surprisingly little focus on presenting stylized facts about gender inequality in education[1]. This is an important gap in the literature as gender differences in education might help to understand gender inequality in other areas, in particular on the labour market. In addition, the gender gap in educational outcomes has changed significantly over the past decades, as we will show. Some of these changes are partly unknown to economists, and may raise important questions regarding the underlying forces behind these dynamics. The first two stylized facts presented in this chapter deal with the upper tail of the educational achievement distribution, while the third and fourth facts refer to the lower tail of the distribution. In this chapter, we focus on the flow as opposed to stock measurement of human capital to capture most recent trends in gender inequality in education. We are interested in the contribution of incoming cohorts to the stock of human capital. On the other hand, stock measures of human capital are less sensitive to recent changes in gender representation.

[1] One recent exception is Pekkarinen (2012).

Kanninen, Ohto (2013), Five essays on economics of education
European University Institute

DOI: 10.2870/93297

1.2 The Gender Gap in Participation to Tertiary Education

1.2.1 Existence

There exists evidence showing that women outnumber men among university students in some countries [1]. In this chapter, we exploit new data to extend and establish this fact over a wider array of countries. Our data is from the Barro-Lee database (2010), which reports participation rates in tertiary education by gender and cohort of birth for more than 140 countries. It is in that sense a flow measure of human capital as it measures the contribution of incoming cohort by gender to the existing stock or human capital, and therefore captures latest gender-specific changes educational participation decisions. The female-to-male ratios among tertiary educated students reported in Figure 1.3 are from 2010, for cohorts born between 1976 and 1981[2]. Reported figures are computed as the ratio between the percentage of females having attended tertiary education by age 30 in 2010, over the percentage of males having attended tertiary education by age 30. It shows that women outnumber men among participants to tertiary education in virtually all OECD countries: in 26 out of 32 OECD countries, the female-to-male ratio is higher than 1, meaning that participation rates to tertiary education are higher for females than for males. The gender gap in tertiary education participation is particularly large in Northern European countries such as Finland, Iceland, Estonia or Latvia. Among non-OECD countries, the gap is the largest in Slovenia with a female-to-male ratio close to 1.9.

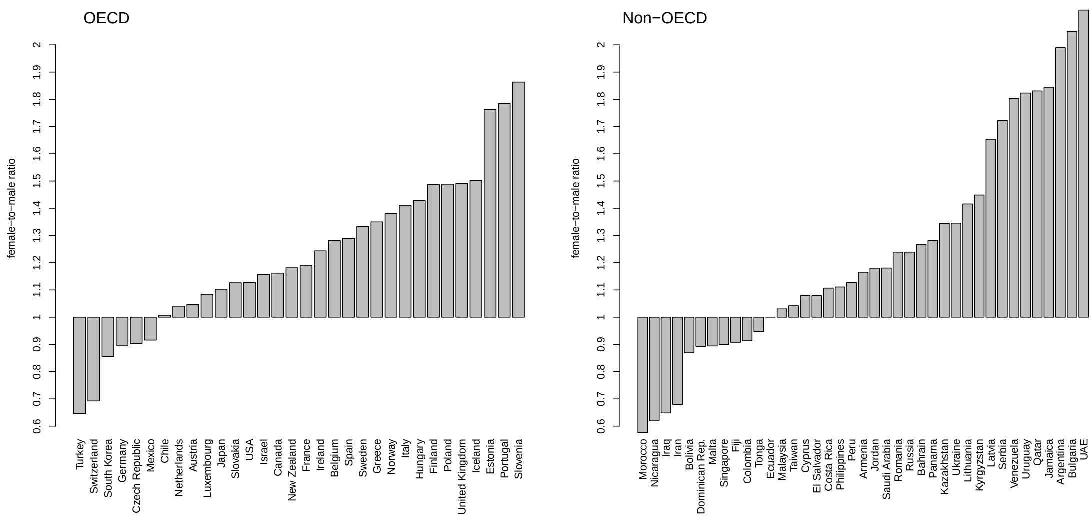
Importantly, the figure for non-OECD countries shows that this phenomenon is not restricted to most advanced economies. Females also outnumber males among participants to tertiary education in the large majority of developing countries: 29 out of 40 non-OECD countries. In Latin American countries such as Uruguay, Venezuela or Argentina, women also outnumber males by a large margin in tertiary education, with a female-to-male ratio larger than 1.8 in these 3 countries. Women are also in large majority among participants to tertiary education in ex-communist countries such as Poland, Hungary, Bulgaria or Lithuania. There exists a few notable exceptions to female dominance in participation to tertiary education among advanced economies. In South Korea, Switzerland and Germany, males are still the majority among participants to tertiary education, although the ratio is very close to one in Germany. Among OECD

[1] See, for example, Becker et al.(2010) or Pekkarinen (2012).

[2] The Barro-Lee dataset does not report participation rates at university for year-by-year cohorts of birth but for 5-year band birth cohorts

countries, Turkey has the lowest female-to-male ratio among participants to tertiary education with a ratio of approximately 0.65. Apart from these few exceptions, the larger participation of females to tertiary education can be qualified as a quasi-universal phenomenon among industrialized economies.

Figure 1.1: Female-to-Male Ratio among Individuals who Attended Tertiary Education by Age 30 - 2010



Source: Barro-Lee database 2010.

1.2.2 Reversal

One can gain further insights on the gender gap in higher education by looking at its dynamics over time. There exists solid evidence showing that, while men used to outnumber women among participants to higher education, females' participation rates progressively converged towards male levels before surpassing them during the last decades. For the US, several papers have reported a convergence, followed by a reversal in the percentage of women relative to men attending university education over the period 1960-2010 [1]. While most work in this respect has been confined to the US, we complement this evidence by showing that the reversal of the gender gap in participation to tertiary education is a phenomenon shared by the very large majority of developed countries[2]. In addition, we show the gender gap reversal in tertiary education participation is not confined to advanced economies, and is also observed in a large

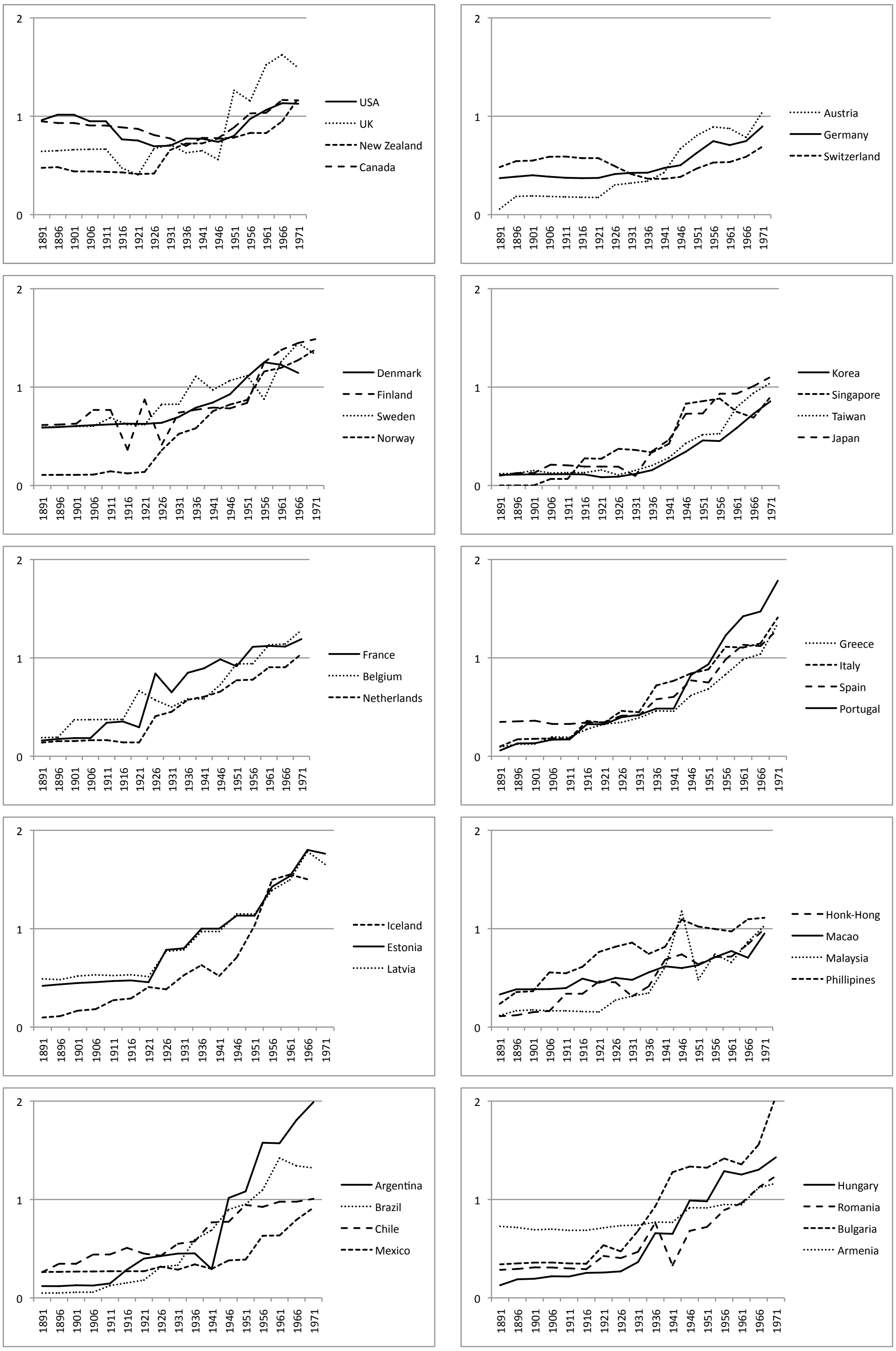
[1] See, for example, Goldin et al. (2006) or Becker et al. (2010)

[2] Pekkarinen (2012) shows that this phenomenon is also observed in Northern European countries

DOI: 10.2870/93297

number of developing countries. In order to reconstruct the evolution of participation rates to tertiary education by gender and cohort of birth, we used data from the Barro-Lee (2010) dataset. The dataset allows constructing the fraction of individuals of a given 5-year band cohorts having attended university by the year of 35, from individuals born in 1891 to 1971.

Figure 1.2: Female-to-Male Ratio among Individuals Having Attended University - by Birth Cohort



It can be seen from Figure 1.3 than the reversal occurred for virtually all industrialized countries: while males used to outnumber females in participation rates to university, females gradually became the majority over time. The timing of the reversal however varies across countries: while the percentage of women attending higher education was already higher than men in Poland and Bulgaria in the early 1970s, the reversal occurred at the beginning of the 1990s in the UK. It is also a very recent phenomenon in countries like Austria, Japan, and the Netherlands. Following the reversal, the gender gap in university attendance appear to have increased in all countries in the favor of women, before stabilizing in recent years. Interestingly, the reversal also occurred in Tunisia in the early 2000s, in spite of relatively low rates of university attendance. In Turkey, on the other hand, the reversal did not occur, in spite of a slow convergence of women's participation rate towards men's participation. More surprisingly, South Korea, Switzerland and Germany are important exceptions to the gender gap reversal phenomenon among advanced economies: the reversal did not occur in spite of a sharp increase in university attendance rates over the last decades. Strikingly, the reversal occurred in less advanced economies such as Saudi Arabia.

1.3 The Gender Gap in Secondary School Non-Completion

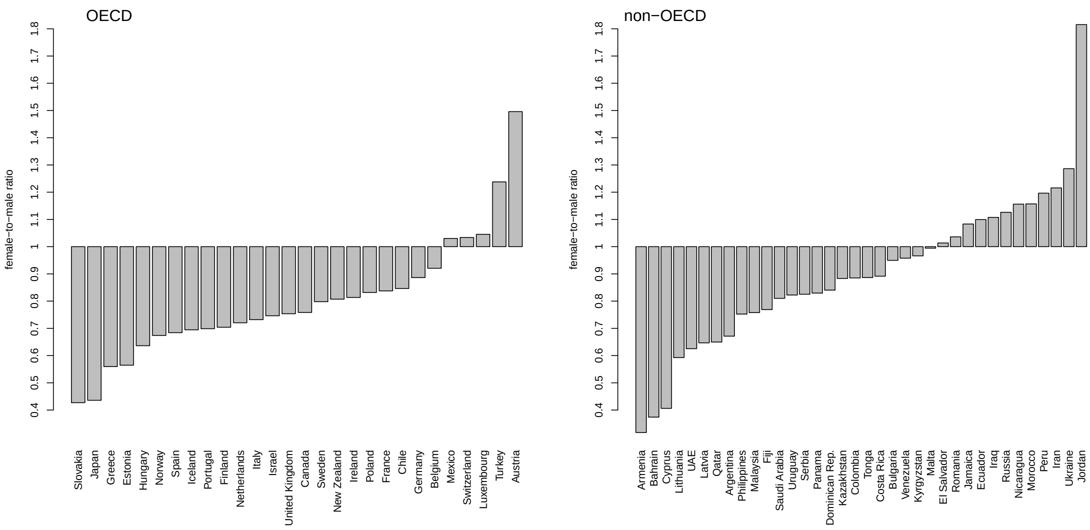
1.3.1 Existence

In this section, we establish the fact that males outnumber females among low educational achievers internationally. As compulsory education ends at the end of lower secondary school in most developed countries, one particular way to identify individuals belonging to the lower end of the educational achievement distribution is to look at individuals who did not complete upper-secondary education, whether they started it or not. Those individuals are typically referred to in the literature as secondary school non-completers, or high school non-completers. Figure 1.3 reports the upper secondary school non-completion rates by gender in 2010, from the Barro-Lee database. It shows that the over-representation of males among high school non-completers appears to be a phenomenon shared by virtually all OECD countries. Among the 28 countries for which secondary school non-completion rates are reported [1], 23 countries have a female-to-male ratio of non-completers below one, and only two have a ratio significantly larger than one. Males are highly represented among secondary school non-completers in Northern European countries such as Iceland, Finland or Norway.

[1] We only include countries with non-completion rates between 0.1 and 0.9. The countries in the extremes are likely to have more estimation error due to small numbers of non-completers or completers, respectively.

The female-to-male ratio of secondary school non-completers is also particularly low in Slovakia and Japan, where females represent less than half of the total number of secondary school non-completers. At the other extreme, Austria is a notable exception to higher non-completion rates of males among OECD countries. It is the only OECD country, with Turkey, in which females are in large majority among secondary school non-completers, with a female-to-male ratio of approximately 1.5. Importantly, the predominance of males among secondary school non-completers is not a phenomenon restricted to advanced economies, and is also observed in the majority of developing countries. Among non-OECD countries for which the data was available, the female-to-male ratio among non-completers is lower than 1 for 23 out of 36 countries. Even in some countries where social norms would a priori not favor women's education, such as Saudi Arabia, males outnumber females among secondary school non-completers.

Figure 1.3: Female-to-Male Ratio among Secondary School Non-Completers



Source: Barro-Lee database 2010.

1.3.2 Reversal

It is often assumed that the over-representation of males among dropouts is also an historical phenomenon. Interestingly, the cohort analysis reported in Figure 1.4 reveal that males have not always outnumbered females among upper-secondary school dropouts. We combined data from various data sources in order to reconstruct the evolution of the fraction of secondary school dropouts by gender and birth cohort for the main industrialized countries. We were able to compute upper secondary school dropout rates for birth cohorts over 90 years, ranging from individuals from in 1891 to individuals born in

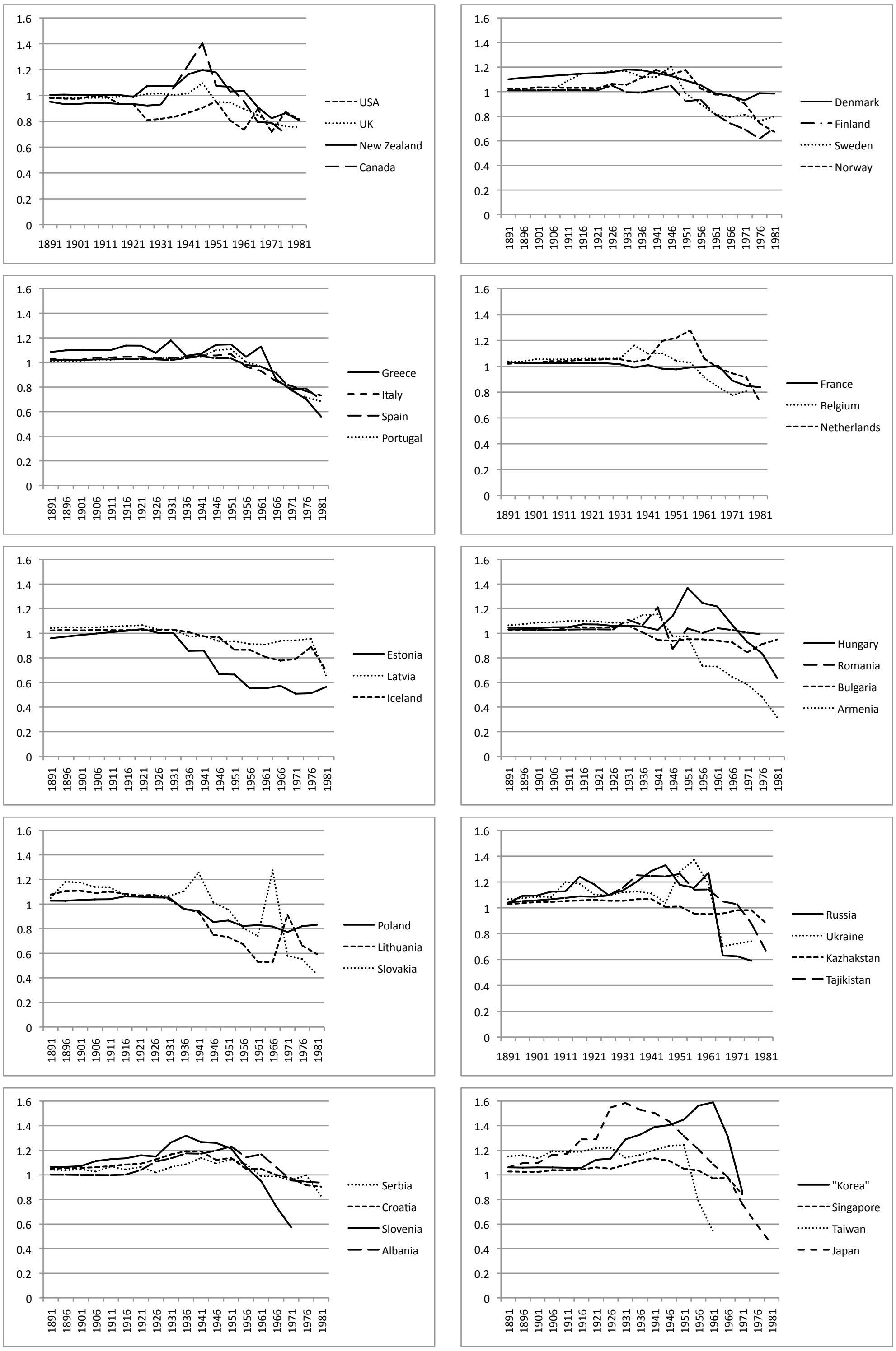
Kanninen, Ohto (2013), Five essays on economics of education
European University Institute

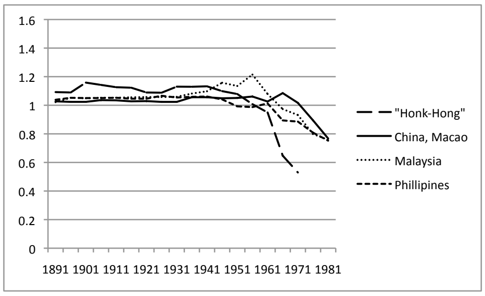
DOI: 10.2870/93297

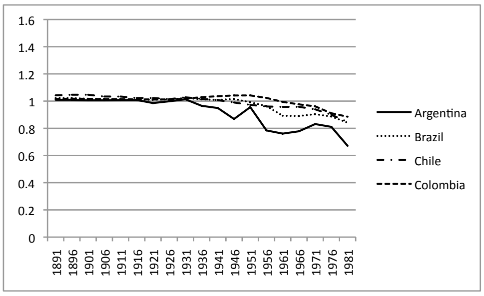
1981. The main source for the data is the Barro-Lee database (2010), which allows us to compute secondary school non-completion rates for 5-year band cohorts from 1891, and separately for males and females. The European Union Labour Force Survey was also used as a complementary source. Figure 1.4 shows that virtually all countries share similar dynamics: while females were slightly more represented among upper-secondary school non-completers in the oldest cohorts, the female-to-male ratio first increased before decreasing sharply among younger cohorts. Virtually all the graphs reported for individual countries shows a point at which the female-to-male ratio among non-completers crosses the $y = 1$ line, indicating the reversal of the gender gap in secondary school non-completion from female majority to male majority. The progressive overtake of males among secondary school non-completers over time is analogous to the overtake observed for females in tertiary education enrollment rates presented in Section 1.

 DOI: 10.2870/93297

Figure 1.4: Female-to-Male Ratio among Secondary School Non-Completers - By Birth Cohort







1.4 Conclusion

In this chapter, we presented four facts about the educational achievements of the two genders. We showed that all these facts are near-universal. Also, it appears that there is something congruent between the gender gap reversal in university attendance, and the reversal in secondary school non-completion rates. Two main theories have been proposed to account for the gender gap reversal in participation to tertiary education. The two facts we present about high school non-completion rates have not, however, been accounted for by previous literature. Focusing on the US, Chiappori et al. (2009) invoked the progressive removal of social barriers to women's education, combined with higher returns to tertiary education for women, as an explanation for the gender reversal in college attendance. Cho (2007) proposed another explanation, again using US data. He argued that the improvement of women's preparation to university proxied by high school test scores are an important driving force behind the gender gap dynamics in education.

The framework that we will establish in the following chapters allows us to present our new theory and the existing theories in a formal setting. Using this framework, we show that our theory is consistent with the four near-universal facts we presented. It is also compatible with the two other theories proposed by the literature, and the truth is likely to contain parts from each one. The new theory we present is based on the larger variability of males' test score distribution relative to females. This fact is strongly supported by empirical data, and like the phenomena we intend to explain, has been shown to be near-universal. We explore this explanation in the following chapters where we present out theoretical framework and test our theory against empirical data. We will argue that it provides a credible explanation to the gender gap dynamics in education that cannot be dismissed by the data. We will also show that some predictions of previous theories can be falsified through empirical testing, casting doubt on their ability to explain, alone, the gender gap reversal in education.

2

Explaining Gender Differences in Educational Outcomes: A Statistical Theory

with Laurent Bossavie

2.1 Introduction

This chapter proposes an explanation for the simultaneous reversal of the gender gap in university attendance and high school non-completion rates observed internationally. As shown in Chapter 1, while there was a majority of men enrolling into tertiary education at the beginning of the 1970s, women's participation to university gradually converged and overtook men's participation over the last decades. This stylized fact was first established for the US by previous literature, and we show in Chapter 1 that it is actually a near universal phenomenon observed in virtually all advanced economies, and in the large majority of developing countries. Symmetrically, we report a similar reversal for secondary school non-completers: while women used to outnumber males among secondary school non-completers, males progressively became majoritary as non-completion rates decreased over time. To our knowledge, this reversal in the gender gap among secondary school non-completers has not been put forward so far in the literature. Explaining the gender gap reversal in education is interesting in its own right, but also for efficiency purposes. First, understanding the origins of the gender gap in education might help understanding gender inequality in other areas, in particular on the labour market. In addition, it is important to identify whether observed gender differences in educational outcomes result from inefficiencies, or optimal behaviors. In particular, one would like to know whether those outcomes originate from discrimination between genders, or from optimizing behaviors based on fundamental gender differences in preferences, behaviors, or ability distributions.

To account for these two quasi-universal facts, we start from a simple Card model of optimal investment in human capital, in which the optimal length of education chosen by individuals is increasing in test-taking ability, measured by test scores. In this framework, we show that a higher variance of test scores for males relative to females, combined with an increase over time in the net benefits of university attendance for both genders, are sufficient to reproduce the empirical dynamics of the educational gender gap. In particular, our model is able to generate a relationship between the total enrollment rate and the female-to-male ratio among the enrolled that provides a very accurate fit for the data. In a similar way, it also allows us to reproduce the relationship between secondary school non-completion rates and gender ratio among non-completers observed empirically.

Several contributions have previously attempted to explain the gender gap reversal in university attendance, focusing on the US context. Goldin et al. (2006) invoke the removal of past barriers to womens' education and careers, combined with a higher college wage premium for women, as an explanation for this stylized fact over the period 1970–2010. In a similar way, Chiappori et al. (2009) combine an exogenous fall in the time required for housework with higher labor-market returns to schooling for women to explain the relative rise in womens' university education. Both contributions rely on higher returns to higher education for women to generate the gender gap reversal in university attendance observed empirically. The existence of a higher college wage premium for women is, however, a highly debatable assumption, that received little support from the empirical literature. Dougherty (2005) is one of the few contributions showing a higher college wage premium for women in the US, and the methodology underlying his estimations have been strongly criticized in Hubbard (2011)[1]. Becker et al. (2010) also report that the estimated benefits from college are still lower for women in most dimensions, although some of them increased faster from women over the past decade in the US. In the absence of a higher college premium for women, changes in social norms are unable to explain the gender gap reversal in university attendance. Also, all the above-mentioned explanations were only applied to the US.

To explain the quasi-universal reversal in the educational gender gap, we build on a quasi-universal fact: the greater dispersion in men's test score distribution relative to women's test score distribution. In addition, we show that our model is able to reproduce the reversal in the gender gap in participation to tertiary education, as well as

[1]After correcting for a bias in estimates of college wage premium, Hubbard (2011) finds that there has been essentially no gender difference in the college wage premium

the dynamics of the gender ratio among secondary school non-completers. The chapter is organized as follows: Section 2 reports the evidence about the greater dispersion of males' performance in cognitive and non-cognitive tests relative to females. Section 3 describes our modified Becker model of optimal investment in human capital. Building on our distributional assumption, Section 4 presents the predictions of the model for the joint evolution of aggregate educational enrollment, and the gender gap in educational outcomes. It also evaluates the predictions of the model against cross-sectional on dropout rates and university enrollment rates by gender.

2.2 The Greater Variability of Males' Test Score Distribution: Empirical Evidence

An important body of literature shows that males are more variable than females in a wide range of cognitive and non-cognitive tests. While this stylized fact was very recently established in the economics literature, long-standing evidence is available in the psychology literature. Ellis (1894) is typically referred to as the contribution that sparked the literature on gender differences in variability. Reviewing data from psychological, medical and anthropometric studies, he comes to the conclusion than males tend to exhibit more variability in both physical and psychological traits, including general intelligence. Frasier (1919) is the first study to compile a large dataset of more than 60,000 observations to provide support to Ellis' original claim. Using grade-level achievement tests for 13 year-olds in the US, he shows that the coefficient of variation - the ratio of the standard deviation to the mean - is larger for males, and that the gender difference is highly statistically significant.

More recently, Feingold (1992) reports the male-to-female variance ratio of the PSAT and SAT of the College Entrance Examination Board in 1960, 1966, 1974 and 1983 consisting of a verbal and quantitative test. In both mathematics and verbal tests, the male-to-female variance ratio was found to be larger than 1 (1.05 in Verbal and 1.20 in Mathematics for the SAT averaged over the 4 waves; 1.05 in Verbal and 1.24 in Mathematics for the PSAT averaged over the three waves), with little variation over the different waves. The results of Feingold (1992) have however been criticized on the ground that they are drawn from a sample of individuals taking SAT, and not representative of the entire population.

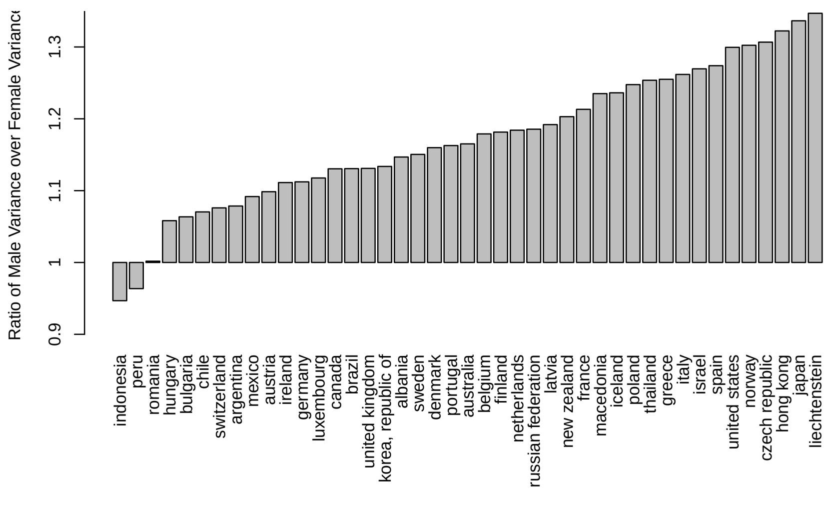
Hedges and Nowell (1995) address this issue by extending the analysis of Feingold (1992) to 6 nationally representative surveys conducted in the US between 1960 and 1992. They compute the male-to-female variance ratio of 43 ability tests extracted from the 6 surveys, and report that the variance of males is larger in 41 out of the 43 cognitive tests. The estimated variance ratio typically ranges between 1.05 and 1.30. Their results also suggest that gender representation among the top-end and bottom-end of the distributions differs depending the type of ability tested. While males substantially outnumber females among the top 10 % scoring individuals in mathematics, science, and social studies, females tend to be in the majority in the upper tail of reading comprehension perceptual speed and associative memory. In disciplines in which males dominate the upper tail, however, the extent of overrepresentation is typically much larger than the overrepresentation of females in disciplines where females dominate the upper tail. More recently, Johnson et al. (2008) also addressed the potential bias associated with non-representativeness of previous studies by using population-wide data on general intelligence. Using two population-wide surveys of 11-year-olds in Scotland, they also find greater variability among males than females in the low end and high ends of the distribution of general intelligence. They further report the greater variability for males is larger in the lower end of the distribution, than in the upper end.

Evidence also suggest that the gender ratio in the extreme scores appear to be fairly constant over-time. Hedges and Nowell (1995) find little evidence that sex differences in the variance of test scores have changed over time, although they compare surveys targeting populations with different ages. Nowell and Hedges (1998) use data from seven representative surveys of the United States twelve grade population, and from the National Assessment of Educational Progress (NEAP) long-term trend data. They find no significant change in the male-to-female ratio among extreme scores from 1960 to the beginning of the 1990s. In particular, they find no evidence for a decrease in the proportion of males among top performers over time.

Jacob (2002) provides further insight on the larger variability of males' ability distribution. Whereas most studies focus on the distribution of cognitive skills, his analysis investigates gender differences in the distribution of both cognitive and non-cognitive skills. He uses data from a nationally representative cohort of eight graders in 1988 from the National Educational Longitudinal Study. Interestingly, while he reports that while men's and women's IQ distributions are very similar in the sample, estimated distributions of non-cognitive skills show a significantly higher variance for males. He uses 4 different measures of non-cognitive skills, including a composite measure of disciplinary incidents computed from the NELS. The behavior composite score computed from the data shows a standard deviation of 1.8 for males against only 1.2 for females. More recently, Kenney-Bensen et al. (2006) also find that the standard deviation of an index of disciplinary skills is larger for boys. They use two measures of disruptive behavior in 5th and 7th grade. In 5th grade, the standard deviation of boys is 0.99 against 0.56 for girls, and 1.20 against 0.68 in 7th grade. If test scores are conceived as the observable outcome of both cognitive and non-cognitive skills, such evidence suggests that the higher dispersion of non-cognitive skills among males could be an important driving force behind the larger variability of males' performance in tests.

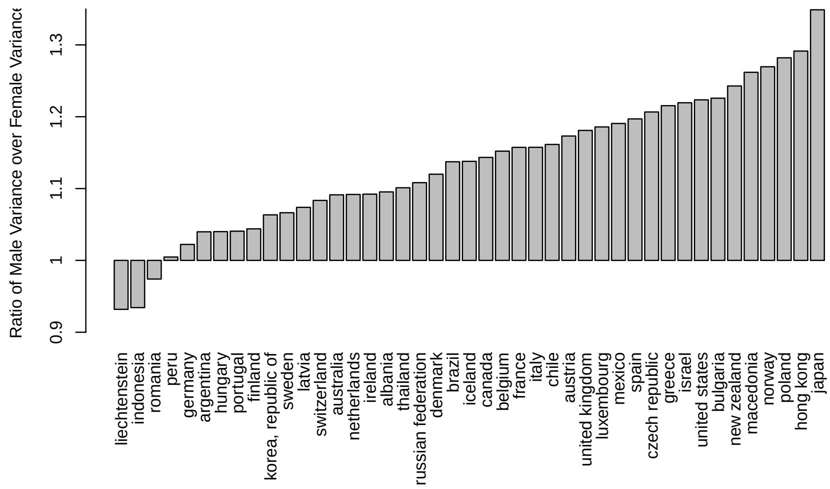
Potential gender differences in the distribution of abilities started to attract the attention of economists only recently. While evidence from the psychological literature was mostly confined to the US, Pekkarinen and Machin (2008) show that the greater variability in test scores among males is an internationally robust phenomenon. They use test score data for 15 year olds from the Program for International Student Assessment (PISA) in 2003. The PISA study tests mathematical and reading skills for a representative sample of the 15-year-old population of students, in more than 40 countries. Pekkarinen and Machin (2008) report that males' reading test score variance is strictly larger than females test score variance in 39 countries out of 40. Mathematics test score distribution exhibits a fairly similar pattern, with the male-to-female variance ratio being strictly greater than 1 in 38 countries out of 40. The gender gap in variance is statistically highly significant: in all but 5 five countries, the null hypothesis that the test score variance is equal across genders is rejected at the 5% level. The estimated variance ratio is rather large in magnitude, with an average of 1.21 for reading and 1.20 for Mathematics. Although the male-to-female variance ratio is larger than 1 in virtually all countries, it varies across countries: it ranges from a minimum of 1.00 in Indonesia to 1.45 in Honk-Hong for reading test scores, and from 0.95 in Indonesia to 1.36 in Honk-Hong for Mathematics test scores. Figures 2.1 and 2.2 summarize the main findings of Pekkarinen and Machin (2008), using data from PISA 2000.

Figure 2.1: PISA Male-to-Female Variance Ratios – Reading Scores



Source: Pisa 2000.

Figure 2.2: PISA Male-to-Female Variance Ratios – Math Scores



Source: Pisa 2000.

The greater dispersion of males' ability relative to females is already in place at very early ages. In Table 2.1, we report the variance ratio of ability test scores at 9 months, 3 years, 5 years and 7 years of age from the UK Millennium Cohort Study. At birth, we use birth weight as a proxy for ability, as it has been shown that birth weight is an important predictor of future cognitive performance, school achievement, and labour market outcomes. As shown in the first row of the table, boys already exhibit a higher variance in birth weights relative to girls, and the gender difference in variance is statistically significant at the 1 % level. At 9 months, the variance of boys in various measures of early development is larger for 9 out of 10 indicators, and the gender difference in variance is again highly statistically significant. A similar pattern is observed for test scores at ages 5 and 7, at which all male-to-female variance ratios are significantly larger than one at the 1 % level. This suggests that the larger variability of males observed at age 15 originates from very early ages, rather than being the result of an endogenous response to gender differences in educational returns at later ages.

Some contributions propose an explanation for the larger dispersion of males' ability distribution. Johnson et al (2009) with a commentary by Craig et al,(2009) discuss the possible role of the X chromosome in explaining the differences between males and females in variability of cognitive ability. They note that a large number of genes in the X chromosome are related to general intelligence. Also, the fact that females have two X chromosomes whereas males have one X chromosome and one Y chromosome seems to play an important role in producing a higher variability for males. Since "Y chromosome is very small and carries little beyond the genetic instructions for maleness", the X

DOI: 10.2870/93297

Table 2.1: Male-to-Female Variance Ratio in Ability Tests at Various Ages

	S.d. Boys	S.d. Girls	Variance Ratio
Birth			
Birth weight	0.606	0.565	1.15***
9 months			
Sits up	0.147	0.131	1.26***
Hands together	0.254	0.214	1.41***
Holds small objects	0.214	0.176	1.48***
Passes a toy	0.121	0.106	1.14***
Walks a few steps	0.344	0.331	1.07***
Gives toy	0.361	0.313	1.33***
Waves bye-bye	0.486	0.427	1.30***
Extends arms	0.212	0.194	1.19***
Nods for yes	0.377	0.405	0.86***
Age 3			
Colors Score	4.13	3.87	1.14***
Letters Score	2.61	2.68	0.95**
Numbers Score	3.79	3.56	1.13***
Size Score	2.82	2.78	1.03*
Comparisons Score	3.82	3.57	1.14***
Shapes Score	4.05	4.03	1.01
Vocabulary Score	4.81	4.75	1.03
Age 5			
Pictures Score	3.58	3.47	1.06***
Construction Score	3.53	3.34	1.12***
Vocabulary Score	3.53	3.36	1.10***
Age 7			
Reading Score	20.39	18.13	1.26***
Math Score	2.96	2.72	1.18***
Construction Score	7.29	6.91	1.11***

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Significance levels indicate that the null hypothesis of equality of the variance for males and females in the population is rejected at the given significance-level.
Source. UK Millenium Cohort Study (MCS)

chromosome functions mostly alone, as Johnson et al note. This would allow recessive genes to be expressed more frequently among males than among females, which would increase the variance of males' characteristics. They also describe a mechanism based on evolution theory that could explain why this is the case. The empirical estimates of the male-female variance ratio in general intelligence they report in their paper range between 1.06 and 1.19.

2.3 Model

2.3.1 Static Framework of Investment in Education

The economy is assumed to be populated by a continuum of agents that differ in their test-taking ability z_j. Test-taking ability z_j is continuous and perfectly observed by individuals. It can be interpreted as a combination of cognitive and non-cognitive skills relevant for educational achievement. For the sake of simplicity, we assume a single-period model in which individuals perceive the benefits of their investment in schooling in the same period as they invest.

Individuals choose their years of schooling s so that they maximize their expected discounted utility U. Building on Becker (1967), we define the utility function of individuals in the economy as:

$$U = B(s) - C(s) \tag{2.1}$$

where $B(s)$ denotes the benefit function of schooling, with $B'(s) > 0$ and $B''(s) < 0$. $C(s)$ is the cost function of schooling and is increasing and convex in s such that $C'(s) > 0$ and $C''(s) > 0$. The first-order conditions for the individual maximization problem can be expressed as:

$$B'(s) = C'(s) \tag{2.2}$$

Where $B'(s)$ is interpreted as the marginal benefit to schooling, and $C'(s)$ is the marginal cost of schooling. Following Card (1994), we linearize the model by assuming that $B'(s)$ and $C'(s)$ are linear functions with individual-specific intercepts and homogeneous slopes:

$$B'(s) = z_j - k_1 s \tag{2.3}$$

$$C'(s) = k_2 s \tag{2.4}$$

where $k_1 > 0$ and $k_2 > 0$. Intuitively, individuals with higher test-taking ability z_j perceive greater marginal benefits (or equivalently, lower costs) from attending school. In this framework, the optimal level of schooling s chosen by individual j can be expressed as:

$$s_j^* = z_j \cdot b \tag{2.5}$$

where $b \equiv \frac{1}{k_1+k_2}$, and can be interpreted as the marginal net benefit of education.
The optimal value of s_j is therefore strictly increasing in test-taking ability z_j. In this framework, let H_j denote the indicator variable taking the value 1 if individual j decides

to attend higher education, 0 otherwise. H is defined as a function of s^* such that:

$$H(s^*) = \begin{cases} 1 & \text{if } s^* \geq \bar{s} \\ 0 & \text{if } s^* < \bar{s} \end{cases}$$

where $\bar{s}$ denotes the minimum number of years of schooling to obtain a university degree.

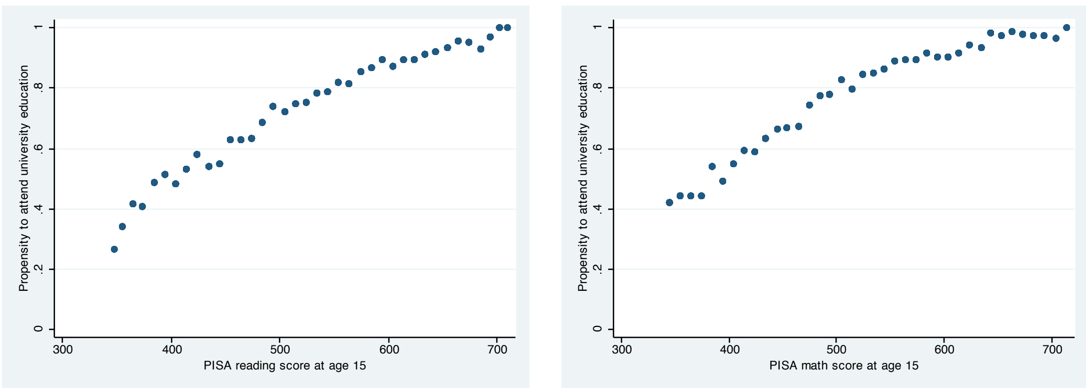
Therefore,

$$H_j = 1 \text{ if } z_j > \frac{\bar{s}}{b} \equiv \bar{z} \tag{2.6}$$

where $\bar{z}$ is the lower bound of test-taking ability such that the individual chooses to enroll at university.

Figure 2.3 provides some supporting evidence for the relationship between test-taking ability z and university attendance H. It clearly shows that propensity to attend university is an increasing function of test scores obtained at age 15.

Figure 2.3: The Empirical relationship between test score z and enrollment at university H



Source. US Educational Longitudinal Study 2002

2.3.2 Time Dynamics

We now assume that the economy is populated by successive cohorts t, with $t \in \{1, 2, .., T\}$. Each cohort comprises a continuum of agents that differ in their level of test-taking ability z. We denote $f_z(z)$ the probability density function of test-taking ability z.

Individuals belonging to the same cohort t are exposed to the same value of the exogenous parameter $b_t \equiv \frac{1}{k_{1,t}+k_{2,t}}$, regardless of their test-taking ability or gender. In this context, the enrollment rate in higher education at time t for each gender g can be expressed as:

$$E_{g,t} = 1 - F_z(\frac{\bar{s}}{b_t})$$

or, equivalently

$$E_{g,t} = G_z(\frac{\bar{s}}{b_t}) = G_z(\bar{z}_t) \tag{2.7}$$

where $G_z(\bar{z}_t)$ denotes the complementary cumulative distribution function (CCDF) of test-taking ability z, defined as: $\int_{\bar{z}}^{+\infty} f_z(z)\, dz$.

b_t is allowed to vary across cohorts and this is interpreted as a change in the net benefits of education, exogenous to the model. In this framework, an increase in the net benefits of education b_t translates mechanically into higher enrollment rates at university E_t, and individuals with lower test-taking ability z choosing to enroll at university. b_t includes monetary benefits of education as well as non-monetary benefits such as life expectancy, the propensity to marry and stay married, or household production.

There exists a important body of literature showing that returns to education, and in particular returns to university education, have increased over the last decades. First, monetary returns to tertiary education – the college wage premium – have been shown to increase sharply over the last decades. Goldin and Katz (2009) or Acemoglu and Autor (2011) among others provide consistent evidence showing a sharp increase of the college wage premium in the US since the beginning of the 1970s[1]. Card and Lemieux (2000) also report an important increase in the wage premium of university graduates relative to high school graduates in the UK and Canada over the same period. Acemoglu (2000) and Goldin and Katz (2009) invoke skill-biased technological change as the main driving force behind the increase in the university wage premium, through an increased demand for skilled workers. In particular, Goldin and Katz report a strong positive relationship between the utilization of more capital-intensive technologies and the demand for university-educated workers. Acemoglu argues that the extent of skill-biased technological was such that it allowed to absorb an increasing supply of university-educated

[1] The college wage premium is defined as the wage of college-educated workers relative to the wage of high-school educated workers.

DOI: 10.2870/93297

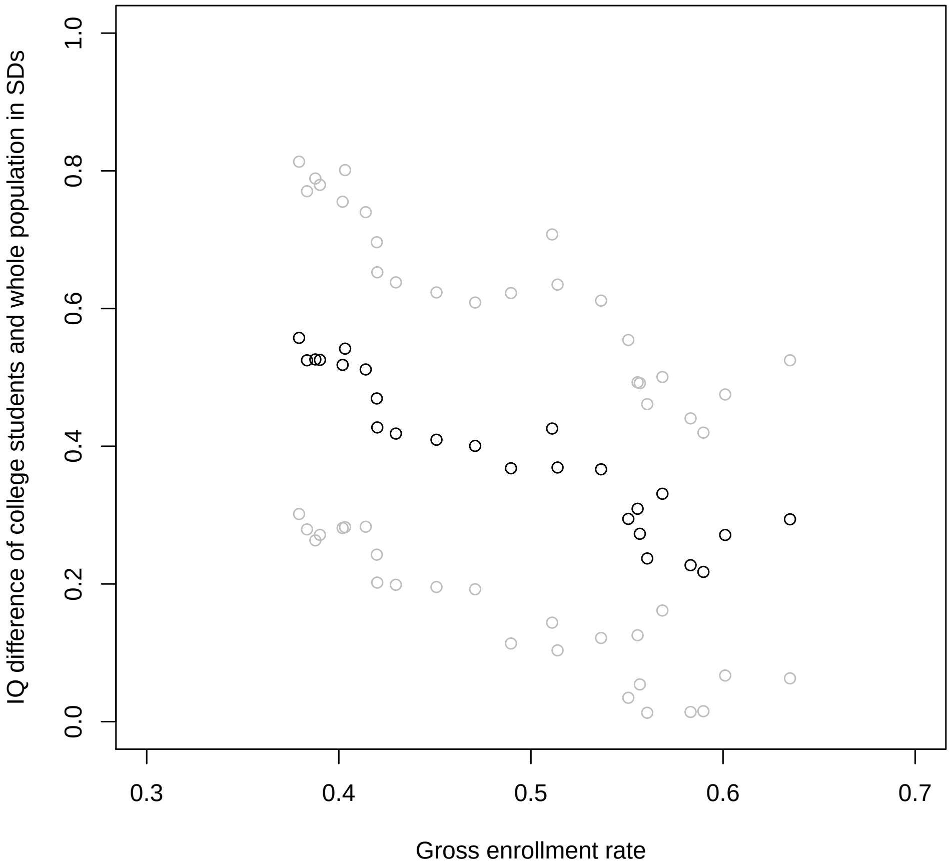
workers, without a decrease in the college wage premium over the past decades. Acemoglu (1998) suggests that the increase in the supply of university-educated workers may itself have induced further skilled-biased technological change, and therefore further increased the college wage premium in the long run although it reduced it in the short run.

The dramatic increase in life expectancy over the last century is another driving force behind increased returns to education. Life-expectancy increased dramatically for both men and women over the last half-century[1]. This increases the returns to education b through two main channels: first, by increasing the expected time-frame over which the monetary returns to education investments are received. Second, Meara et al. (2008) show that at least in the US life expectancy increased disproportionately for university graduates relative to high school-graduates, thereby reinforcing incentives to invest in tertiary education.

Our model implies that an increase in the enrollment rate into tertiary education E_t goes together with a decrease in the test-taking ability threshold $\bar{z}$ for choosing to attend university. Although the lower bound of test-taking ability for university attendance $\bar{z}$ can hardly be observed in the data, we can observe the average test-taking ability of individuals attending university, which mechanically decreases with $\bar{z}$. To check whether increased enrollment at university over time was actually accompanied by a decrease in $\bar{z}$, we computed the average IQ score of individuals attending university in the US over the period 1974–2010, using data from the General Social Survey (GSS). Each wave surveys a random sample of 1,000 to 5,000 individuals. From 1974 onwards, the GSS includes a simplified IQ test consisting of 10 questions assessing the cognitive skills of the respondents. A measure of educational attainment (in years) is also reported, and we classify individuals with more than 12 years of education as having attended college. Figure 2.4 shows the evolution of the average IQ of university students relative to the all population in number of standard deviations, over the period 1975-2010. It shows a clear downward trend: while in 1974 the average IQ of students attending tertiary education was close to 0.60 standard deviation higher than the average IQ of the all population, this relative difference decreased by half until 2005 to reach approximately 0.30. The data therefore seems to support our claim that greater access to university education was accompanied by a decrease in the average test-taking ability of students attending higher education.

[1] Between 1960 and 2010, the mean life expectancy of males increased globally by 15.2 years from 52 to 67.2. For females the increase was 16 years from 55.7 to 71.7. The numbers are calculated as an unweighted mean of all countries for which there is data available for years 1960 and 2010 in the data provided by World Bank.

Figure 2.4: The Empirical Relationship between the Enrollment Rate E and the Average Test Score of the Enrolled - All Population



Source. US General Social Survey (1975-2010)
Notes. Dashed lines represent confidence intervals at the 5%-level.

2.3.3 Implied Relationship between Total Enrollment Rate and Female-to-Male Ratio

Building on evidence about gender differences in test score variability, we now allow $f_z(z)$ to differ between genders. Let $f_{z_m}(z)$ and $f_{z_f}(z)$ denote the probability density functions of test-taking ability for males and females respectively, with $Var[z_m] > Var[z_f]$. In words, males and females in a given cohort are assumed to draw their test-taking ability from two different distributions[1]. Each cohort is assumed to be split equally between males and females, and the distribution of test-taking ability for each

[1]We assume the the distribution function types to be the same, allowing the parameters to vary. For all empirical applications, we also assume normality of the two distributions.

gender $g \in \{m, f\}$ is assumed to be invariant over time:

$$f_{z_g,t}(z) = f_{z_g}(z) \tag{2.8}$$

Panel A of Figure 2.5 illustrates the two test-taking ability distributions, when $\sigma_m^2 > \sigma_f^2$ and $\mu_m^2 < \mu_f^2$. It also depicts the two CCDFs of z_m and z_f, denoted $G_{z_m}(\bar{z})$ and $G_{z_f}(\bar{z})$, respectively and defined as:

$$G_z(\bar{z}) = \int_{\bar{z}}^{+\infty} f_z(z)\,dz$$

. In the illustrational graph, test-taking ability z is assumed to be normally distributed in the population for both genders, with $z_m \sim N(\mu_m, \sigma_m^2)$ and $z_f \sim N(\mu_f, \sigma_f^2)$. By combining the two complementary cumulative distributions $G_{z_m}(\bar{z})$ and $G_{z_f}(\bar{z})$, it is possible to compute the total enrollment rate in tertiary education in the economy as:

$$x = E(\bar{z}) \equiv \frac{G_{z_f}(\bar{z}) + G_{z_m}(\bar{z})}{2} \tag{2.9}$$

which is represented by the thin dotted line in panel B of Figure 2.5, and obtained by averaging the two complementary cumulative distributions, assuming that males and females are equally split in the population. In this framework, the female-to-male ratio among the enrolled, denoted $R(\bar{z})$, can be expressed as:

$$y = R(\bar{z}) \equiv \frac{G_{z_f}(\bar{z})}{G_{z_m}(\bar{z})} \tag{2.10}$$

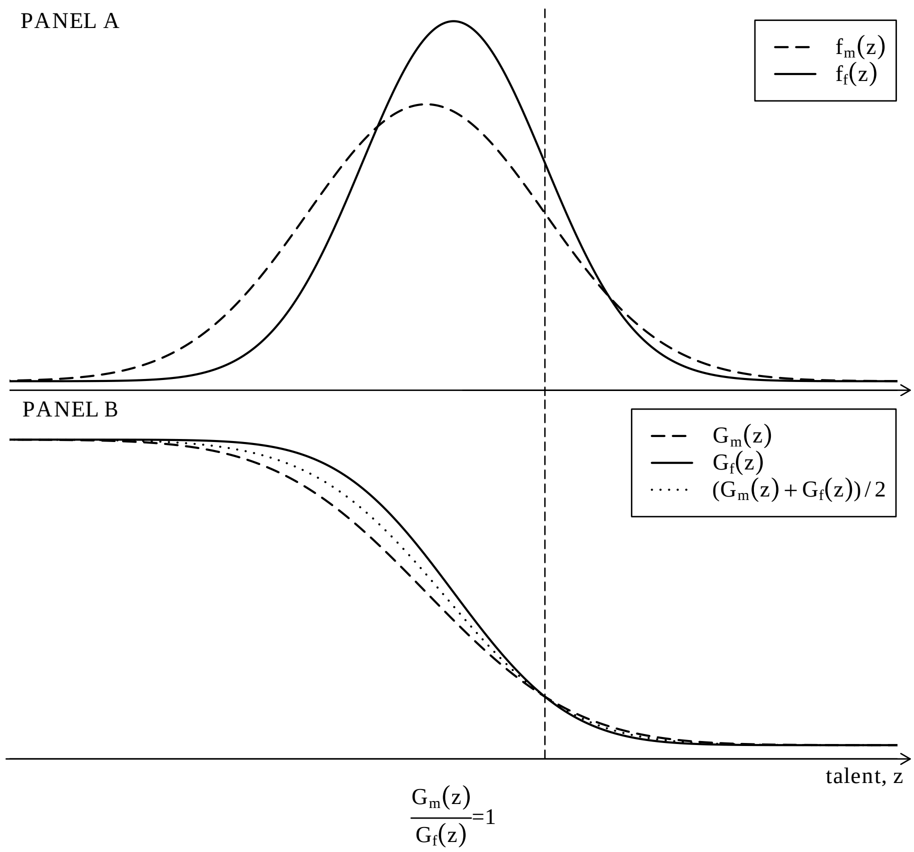
From Panel B of Figure 2.5, we can derive the relationship between total enrollment rate and the female-to-male ratio among the enrolled. Figure 2.6 illustrates the expected relationship between the total enrollment rate and the female-to-male ratio among the enrolled, when $\sigma_m^2 > \sigma_f^2$, as depicted in Panel A and B of Figure 2.5.

$R(\bar{z})$ and $E(\bar{z})$ are both functions of the lower bound of test-taking ability for enrolling $\bar{z}$, which varies with the exogenous parameter b. Under the assumption that $Var[z_m] > Var[z_f]$, it can be shown analytically that the relationship between $R(\bar{z})$ and $E(\bar{z})$ has 3 notable properties:

Proposition 1 *The female-to-male ratio $R(\bar{z})$ tends to zero when the total enrollment rate $E(\bar{z})$ tends to zero.*

Proposition 2 *The female-to-male ratio $R(\bar{z})$ tends to one when the total enrollment rate $E(\bar{z})$ tends to one.*

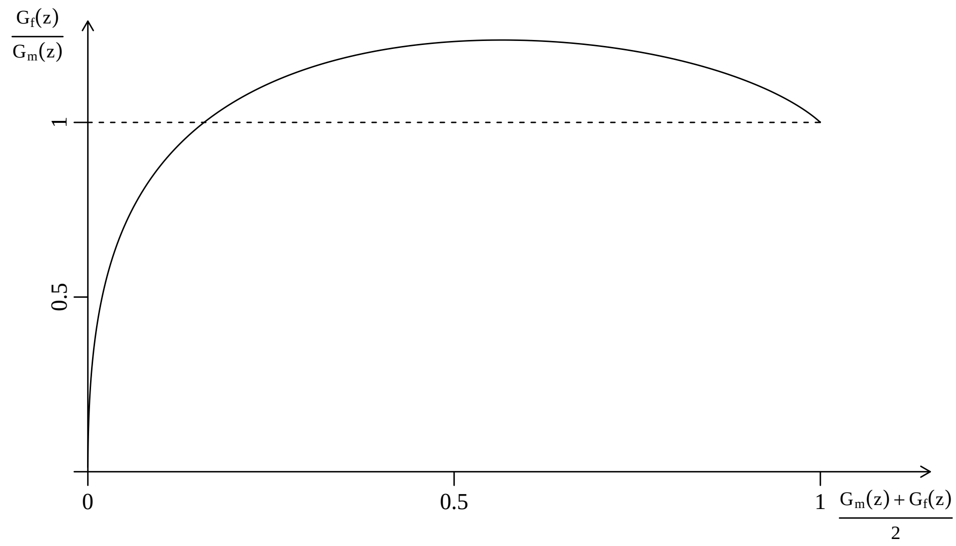
Figure 2.5: Distribution Functions of Test Scores by Gender when $\sigma_m^2 > \sigma_f^2$ - Illustration



Notes. Panel A shows the probability distribution functions of test-taking ability z among males (full line) and females (dashed line), when test-taking ability z is normally distributed and $\sigma_m > \sigma_f$. Panel B shows the complementary cumulative distributions, resulting from the integration from $+\infty$ to z of $f_{z_f}(z)$ and $f_{z_m}(z)$.

Proposition 3 *There exists a value of $E(\bar{z}) \in [0, 1[$ such that $R(\bar{z}) = 1$. This value is unique and always exists.*

Figure 2.6: Expected Relationship between Total Enrollment Rate and Female-to-Male Ratio of the Enrolled – Illustration



Proof. See the Appendix. Importantly, the normality assumption is not required for Proposition 1 to 3 to hold. As shown in the Appendix, Proposition 1 to 3 also hold when z follows alternative two-parameter probability distribution functions.

An analogous reasoning can be applied to extract the relationship between secondary school non-completion rates and gender ratio among non-completers. The only difference is that we work with the cumulative distribution functions (CDF) instead of the complementary cumulative distribution functions (CCDF) since we are now dealing with the lower tail of the probability density functions. In our setting, the secondary school non-completion rate for each gender in a given cohort t is simply:

$$N_t = F_z(\frac{\bar{s}}{b_t}) = F_z(\underline{z_t})$$

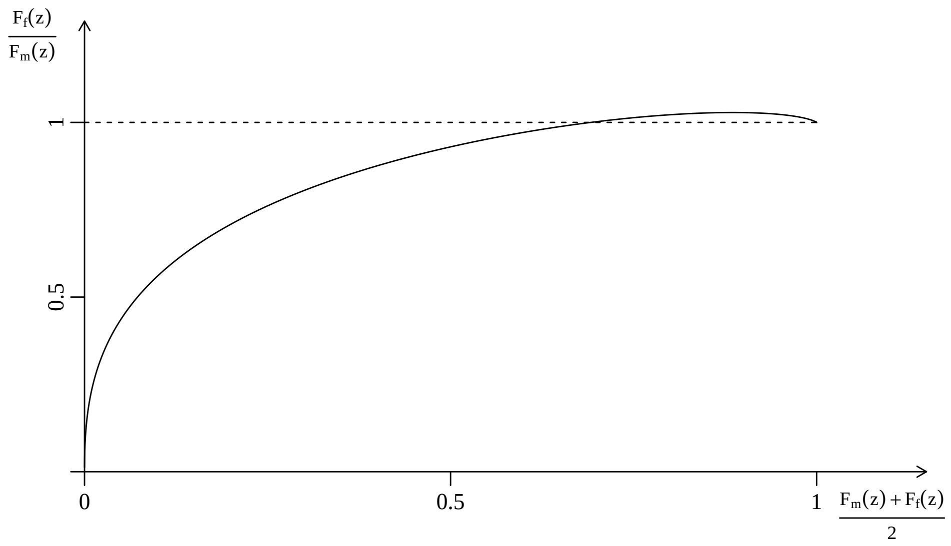
where

$$F_z(\underline{z}) = \int_{-\infty}^{\underline{z}} f_z(z)\, dz$$

and $\underline{z}$ denotes the lower bound of test-taking ability such that individuals complete secondary school. The total non-completion rate for both genders in a given cohort is:

$$N(\underline{z}) \equiv \frac{F_{z_f}(\underline{z}) + F_{z_m}(\underline{z})}{2} \tag{2.11}$$

Figure 2.7: Expected Relationship between Total Non-Completion Rate and Female-to-Male Ratio among Non-Completers – Illustration



Note. The parameters are the same as in Figure 2.6. Since the process goes from the lower tail up, the curve takes a different form.

2.4 Conclusion

In this chapter, we have introduced a theory which allow us to study the coevolution of university enrollment rates and gender ratios, as well as high school non-completion rates and gender ratios. We have also shown three properties that the evolution should exhibit in the case that male variance in test-taking ability is higher than female variance.

2.5 Technical Appendix

2.5.1 Proof of Proposition 1 to 3 - Normal Distributions

Let $f_{zf}(z)$ and $f_{zm}(z)$ denote the probability distribution functions of talent z for females and males, respectively. We assume for the sake of the argument that:

$$z_f \sim N(\mu_f, \sigma_f^2)$$

and

$$z_m \sim N(\mu_m, \sigma_m^2).$$

where $\sigma_m^2 > \sigma_f^2$.

Proof of Proposition 1. *The female-to-male ratio $R(\bar{z})$ tends to zero when the* *total enrollment rate $E(\bar{z})$ tends to zero.*

First, it is immediate to see that $\lim_{\bar{z}\to\infty} E(\bar{z}) = \frac{G_{z_f}(\bar{z})+G_{z_m}(\bar{z})}{2} = \frac{0+0}{2} = 0$. where $G_z(\bar{z})$ denotes the complementary cumulative distribution function (or tail distribution function) of talent z, defined as $\int_{\bar{z}}^{+\infty} f_z(z)\,dz$.

Let us now study $\lim_{\bar{z}\to\infty} R(\bar{z})$. Using the analytical expression of the probability distribution function of the normal distribution, the ratio $R(\bar{z})$ can be expressed as:

$$R(\bar{z}) = \frac{\int_{\bar{z}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_f^2}} e^{\frac{(\bar{z}-\mu_f)^2}{2\sigma_f^2}}\,dz}{\int_{\bar{z}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{\frac{(\bar{z}-\mu_m)^2}{2\sigma_m^2}}\,dz}$$

Taking the integral, one can express the ratio as:

$$R(\bar{z}) = \frac{\frac{1}{2}\left(1 - erf\left[\frac{\bar{z}-\mu_f}{\sqrt{2\sigma_f^2}}\right]\right)}{\frac{1}{2}\left(1 - erf\left[\frac{\bar{z}-\mu_m}{\sqrt{2\sigma_m^2}}\right]\right)},$$

where $erf(\cdot)$ denotes the Gauss error function, expressed as $erf(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt$.

Using the analytical expression of $R(\bar{z})$, we get:

$$\lim_{\bar{z}\to\infty} \frac{\frac{1}{2}\left(1 - erf\left[\frac{\bar{z}-\mu_f}{\sqrt{2\sigma_f^2}}\right]\right)}{\frac{1}{2}\left(1 - erf\left[\frac{\bar{z}-\mu_m}{\sqrt{2\sigma_m^2}}\right]\right)} = \lim_{z\to\infty} \frac{1-(erf\,[\bar{z}])}{1-(erf\,[\bar{z}])} = \frac{1-1}{1-1} = \frac{0}{0},$$

where the second to last step follows from the fact that $\lim_{\bar{z}\to\infty} erf(\bar{z}) = 1$.

Thus, we need to use the l'HÃ´pital rule. We take the derivative for the denominator and the numerator to get the following expression:

$$\lim_{\bar{z}\to\infty} \frac{\sigma_m}{\sigma_f} exp\left\{\frac{(\bar{z}-\mu_m)^2}{\sigma_m^2} - \frac{(\bar{z}-\mu_f)^2}{\sigma_f^2}\right\} = \lim_{\bar{z}\to\infty} \frac{\sigma_m}{\sigma_f} exp\left\{\frac{(\bar{z}-\mu_m)^2\sigma_f^2}{\sigma_m^2\sigma_f^2} - \frac{(\bar{z}-\mu_f)^2\sigma_m^2}{\sigma_f^2\sigma_m^2}\right\} =$$

$$= \lim_{\bar{z}\to\infty} \frac{\sigma_m}{\sigma_f} exp\left\{\frac{\bar{z}^2\sigma_f^2 - 2\bar{z}\mu_m\sigma_f^2 + \mu_m^2\sigma_f^2 - \bar{z}^2\sigma_m^2 + 2\bar{z}\mu_f\sigma_m^2 - \mu_f^2\sigma_m^2}{\sigma_m^2\sigma_f^2}\right\} =$$

$$= \lim_{\bar{z}\to\infty} \frac{\sigma_m}{\sigma_f} exp\left\{ \frac{\bar{z}}{\sigma_m^2\sigma_f^2} \left[\bar{z}\{\sigma_f^2 - \sigma_m^2\} - 2\mu_m\sigma_f^2 + 2\mu_f\sigma_m^2 + \frac{\mu_m^2\sigma_f^2}{\bar{z}} - \frac{\mu_f^2\sigma_m^2}{\bar{z}} \right] \right\} =$$

$$= 0,$$

since by assumption $\sigma_m^2 > \sigma_f^2$, and both are positive by definition.

Proof of Proposition 2. *The female-to-male ratio $R(\bar{z})$ tends to one when the total enrollment rate $E(\bar{z})$ tends to one.*

First, it is immediate to see that $\lim_{\bar{z}\to infty} E(\bar{z})\frac{G_{z_f}(\bar{z})+G_{z_m}(\bar{z})}{2} = \frac{1+1}{2} = 1$.

Let us now study the behavior of $R(\bar{z})$ when $\bar{z}$ tends to $-\infty$.

$$\lim_{\bar{z}\to-\infty} \frac{\frac{1}{2}\left(1 - erf\left[\frac{\bar{z}-\mu_f}{\sqrt{2\sigma_f^2}}\right]\right)}{\frac{1}{2}\left(1 - erf\left[\frac{\bar{z}-\mu_m}{\sqrt{2\sigma_m^2}}\right]\right)} = \lim_{\bar{z}\to-\infty} \frac{1-(erf\,[\bar{z}])}{1-(erf\,[\bar{z}])} = \frac{1+1}{1+1} = 1.$$

where we use the fact that $\lim_{\bar{z}\to-\infty} erf(\bar{z}) = -1$.

Proof of Proposition 3. *There exists a value of $E(\bar{z})$ such that $R(\bar{z}) = 1$. This value is unique and always exists.*

Let us now show that given our distributional assumptions, there exists a value of z denoted z^*, such that the numerator and denominator are of equal value, thus the ratio is one. Again, we invoke the ratio

$$R(\bar{z}) = \frac{\frac{1}{2}\left(1 - erf\left[\frac{\bar{z}-\mu_f}{\sqrt{2\sigma_f^2}}\right]\right)}{\frac{1}{2}\left(1 - erf\left[\frac{\bar{z}-\mu_m}{\sqrt{2\sigma_m^2}}\right]\right)}.$$

Since we know that the error function is monotonously increasing on the whole domain, $R(\bar{z}) = 1$ when

$$\frac{\frac{\bar{z}-\mu_f}{\sqrt{2\sigma_f^2}}}{\frac{\bar{z}-\mu_m}{\sqrt{2\sigma_m^2}}} = 1 \Leftrightarrow \frac{\bar{z}-\mu_f}{\sigma_f^2} = \frac{\bar{z}-\mu_m}{\sigma_m^2} \Leftrightarrow \bar{z} = \frac{\mu_m\sigma_f - \mu_f\sigma_m}{\sigma_f - \sigma_m}.$$

This equation has a unique solution given $\sigma_m > \sigma_f$. Since the support of $E(\bar{z})$ is the whole real line, there always exists a value of $\bar{z}$ denoted $\bar{z}^*$ such that

$$\bar{z}^* = \frac{\mu_m\sigma_f - \mu_f\sigma_m}{\sigma_f - \sigma_m}.$$

In addition, $\bar{z}^*$ is unique given the vector of exogenous parameters $\{\mu_f, \mu_m, \sigma_f, \sigma_m\}$.

2.5.2 Proof of Proposition 1 to 3 - Log-normal Distributions

Proof of Proposition 1. *The female-to-male ratio $R(\bar{z})$ tends to zero when the total enrollment rate $E(\bar{z})$ tends to zero.*

First, it is immediate to see that $\lim_{\bar{z}\to\infty} C(\bar{z}) = \frac{G_{z_f}(\bar{z})+G_{z_m}(\bar{z})}{2} = \frac{0+0}{2} = 0.$

The ratio $R(\bar{z})$ of two log-normal complementary CDFs can be expressed as:

$$R(\bar{z}) = \frac{\frac{1}{2}\left(1 - erf\left[\frac{log\,\bar{z}-\mu_f}{\sqrt{2}\sigma_f}\right]\right)}{\frac{1}{2}\left(1 - erf\left[\frac{log\,\bar{z}-\mu_m}{\sqrt{2}\sigma_m}\right]\right)}$$

Now,

$$\lim_{\bar{z}\to\infty} \frac{\frac{1}{2}\left(1 - erf\left[\frac{log\,\bar{z}-\mu_f}{\sqrt{2}\sigma_f}\right]\right)}{\frac{1}{2}\left(1 - erf\left[\frac{log\,\bar{z}-\mu_m}{\sqrt{2}\sigma_m}\right]\right)} = \frac{1 - erf\,[\infty]}{1 - erf\,[\infty]} = \frac{1-1}{1-1} = \frac{0}{0},$$

since $\lim_{\bar{z}\to\infty} erf(\bar{z}) = 1$.

We then use the l'HÃ´pital rule:

$$\lim_{\bar{z}\to\infty} \frac{\sigma_m}{\sigma_f} exp\left\{\frac{(log\,\bar{z}-\mu_m)^2}{\sigma_m^2} - \frac{(log\,\bar{z}-\mu_f)^2}{\sigma_f^2}\right\} = \lim_{log\,\bar{z}\to\infty} \frac{\sigma_m}{\sigma_f} exp\left\{\frac{(log\,\bar{z}-\mu_m)^2\sigma_f^2}{\sigma_m^2\sigma_f^2} - \frac{(log\,\bar{z}-\mu_f)^2\sigma_m^2}{\sigma_f^2\sigma_m^2}\right\} =$$

$$= \lim_{\bar{z}\to\infty} \frac{\sigma_m}{\sigma_f} exp\left\{\frac{(log\,\bar{z})^2\sigma_f^2 - 2\,log\,\bar{z}\,\mu_m\sigma_f^2 + \mu_m^2\sigma_f^2 - (log\,\bar{z})^2\sigma_m^2 + 2\bar{z}\mu_f\sigma_m^2 - \mu_f^2\sigma_m^2}{\sigma_m^2\sigma_f^2}\right\} =$$

$$= \lim_{\bar{z}\to\infty} \frac{\sigma_m}{\sigma_f} exp\left\{\frac{log\,\bar{z}}{\sigma_m^2\sigma_f^2}\left[log\,\bar{z}(\sigma_f^2 - \sigma_m^2) - 2\mu_m\sigma_f^2 + 2\mu_f\sigma_m^2 + \frac{\mu_m^2\sigma_f^2}{log\,\bar{z}} - \frac{\mu_f^2\sigma_m^2}{log\,\bar{z}}\right]\right\} =$$

$$= 0,$$

since by assumption $\sigma_m^2 > \sigma_f^2$, and both are positive by definition.

Proof of Proposition 2. *The female-to-male ratio $R(\bar{z})$ tends to one when the total enrollment rate $E(\bar{z})$ tends to one.*

First, it is immediate that $\lim_{\bar{z}\to 0} E(\bar{z})\frac{G_{z_f}(\bar{z})+G_{z_m}(\bar{z})}{2} = \frac{1+1}{2} = 1$. Since the support of the log-normal distribution is $(0, +\infty)$.

Let us now study $\lim_{\bar{z}\to 0} R(\bar{z})$:

$$\lim_{\bar{z}\to 0} R(\bar{z}) = \lim_{\bar{z}\to 0} \frac{\frac{1}{2}\left(1 - erf\left[\frac{log\,\bar{z}-\mu_f}{\sqrt{2}\sigma_f}\right]\right)}{\frac{1}{2}\left(1 - erf\left[\frac{log\,\bar{z}-\mu_m}{\sqrt{2}\sigma_m}\right]\right)} = \frac{1 - erf\,[-\infty]}{1 - erf\,[-\infty]} = \frac{1+1}{1+1} = 1.$$

and

$$\lim_{\bar{z}\to 0}(R\bar{z}) = \frac{\frac{1}{2}\left(1 - erf\left[\frac{log\,\bar{z}-\mu_f}{\sqrt{2}\sigma_f}\right]\right) + \frac{1}{2}\left(1 - erf\left[\frac{log\,\bar{z}-\mu_m}{\sqrt{2}\sigma_m}\right]\right)}{2} =$$

$$= \frac{2 - erf\,[-\infty] - erf\,[-\infty]}{4} = 1$$

Proof of Proposition 3. *There exists a value of $E(\bar{z}) \in [0, 1)$ such that $R(\bar{z}) = 1$. This value is unique and always exists.*

As with the normal distribution, we use the fact that the error function is monotonously increasing. Thus, $R(\bar{z}) = 1$ when:

$$\frac{\frac{log\,\bar{z}-\mu_f}{\sqrt{2}\sigma_f^2}}{\frac{log\,\bar{z}-\mu_m}{\sqrt{2}\sigma_m^2}} = 1 \Leftrightarrow \frac{log\,\bar{z}-\mu_f}{\sigma_f^2} = \frac{log\,\bar{z}-\mu_m}{\sigma_m^2} \Leftrightarrow log\,\bar{z} = \frac{\mu_m\sigma_f - \mu_f\sigma_m}{\sigma_f - \sigma_m}.$$

if $\mu_f = \mu_m = \mu$, $R(\bar{z}) = 1$ when $log\,\bar{z} = \mu$.

Since the support of $E(\bar{z})$ is the positive the real line, there always exists a $C(log\,\bar{z})$ such that

$$log\,\bar{z} = \frac{\mu_m\sigma_f - \mu_f\sigma_m}{\sigma_f - \sigma_m} \Leftrightarrow \bar{z}^* = exp\left\{\frac{\mu_m\sigma_f - \mu_f\sigma_m}{\sigma_f - \sigma_m}\right\}.$$

In addition, $\bar{z}^*$ is unique given the vector of exogenous parameters $\{\mu_f, \mu_m, \sigma_f, \sigma_m\}$.

2.5.3 Proof of Proposition 1 to 3 - Uniform Distributions

Proof of Proposition 1. *The female-to-male ratio $R(\bar{z})$ tends to zero when the total enrollment rate $E(\bar{z})$ tends to zero.*

We study the behavior of the ratio $R(\bar{z})$ when $\bar{z}$ tends to b_f, since before that point it

DOI: 10.2870/93297

is evident that no female attends college, and thus the ratio has to be 0. Thus,

$$lim_{\bar{z}\to b_f}\frac{(b_m-a_m)(b_f-\bar{z})}{(b_f-a_f)(b_m-\bar{z})}=\frac{(b_m-a_m)(b_f-b_f)}{(b_f-a_f)(b_m-b_f)}=\frac{(b_m-a_m)(0)}{(b_f-a_f)(b_m-b_f)}=0.$$

From the definition of the CDF of a uniform distribution (from the top): $F_z(\bar{z}) = \begin{cases} 0 & \text{for } \bar{z}\geq b \\ \frac{b-\bar{z}}{b-a} & \text{for } \bar{z}\in(a,b)\cdot \\ 1 & \bar{z}\leq a. \end{cases}$

Proof of Proposition 2. *The female-to-male ratio $R(\bar{z})$ tends to one when the total enrollment rate $E(\bar{z})$ tends to one.*

To study the other extreme we set $\bar{z}$ to tend to $a_m < a_f$. By definition, $lim_{\bar{z}\to a_m}R(\bar{z}) = lim_{\bar{z}\to a_m}\frac{CDF_f(\bar{z})}{CDF_m(\bar{z})}=\frac{1}{1}=1$.

Proof of Proposition 3. *There exists a value of $E(\bar{z})\in[0,1[$ such that $R(\bar{z})=1$. This value is unique and always exists.*

$R(\bar{z})=1$ when $\frac{b_f-\bar{z}}{b_f-a_f}=\frac{b_m-\bar{z}}{b_m-a_m}\Leftrightarrow \bar{z}^*=\frac{a_f b_m-a_m b_f}{(a_f-a_m)-(b_f-b_m)}$, which is unique and exists when $b_f\neq b_m$ or $a_f\neq a_m$ and $(a_f-a_m)\neq(b_f-b_m)$. One of the first two inequalities will hold as long as the two distributions are not identical, as we assume. The last inequality holds since, by assumption, $Var[z_m]>Var[z_f]$.

2.5.4 Proof of Proposition 1 to 3 - Logistic Distributions

We assume for the sake of the argument that:

$$z_f\sim Logi(\mu_f,s_f)$$

and

$$z_m\sim Logi(\mu_m,s_m).$$

Proof of Proposition 1. *The female-to-male ratio $R(\bar{z})$ tends to zero when the total enrollment rate $E(\bar{z})$ tends to zero.*

Using the analytical expression of the logistic probability density function, the ratio $R(\bar{z})$ can be expressed as:

$$R(\bar{z}) = \frac{1 - \frac{1}{1+exp\left\{-\frac{z-\mu_f}{\sigma_f}\right\}}}{1 - \frac{1}{1+exp\left\{-\frac{z-\mu_m}{\sigma_m}\right\}}}$$

Thus,

$$\lim_{\bar{z}\to\infty} R(\bar{z}) = \frac{1 - \frac{1}{1+exp\{-\infty\}}}{1 - \frac{1}{1+exp\{-\infty\}}} = \frac{0}{0}$$

We thus need to us L'Hopital rule:

Proof of Proposition 2. *The female-to-male ratio $R(\bar{z})$ tends to one when the total enrollment rate $E(\bar{z})$ tends to one.*

First, it is immediate that: $\lim_{\bar{z}\to-\infty} C(\bar{z}) = \frac{1+1}{2} = 1$.

In addition,

$$\lim_{\bar{z}\to-\infty} R(\bar{z}) = \frac{1 - \frac{1}{1+exp\{\infty\}}}{1 - \frac{1}{1+exp\{\infty\}}} = \frac{1}{1} = 1$$

Proof of Proposition 3. *There exists a value of $C(\bar{z}) \in [0, 1[$ such that $R(\bar{z}) = 1$. This value is unique and always exists.*

Therefore,

$$R(\bar{z}) = 1$$

is true if and only if

$$1 - \frac{1}{1 + \exp\frac{-(\bar{z}-\mu_f)}{s_f}} = 1 - \frac{1}{1 + \exp\frac{-(\bar{z}-\mu_m)}{s_m}}$$

$$\frac{-(\bar{z} - \mu_f)}{s_f} = \frac{-(\bar{z} - \mu_m)}{s_m}$$

Reorganizing yields:

$$\bar{z}^* = \frac{\mu_m\sigma_f - \mu_f\sigma_m}{\sigma_f - \sigma_m}.$$

which always exists since $\bar{z}$ is defined on the entire real line. In addition, this $\bar{z}^*$ is unique given the vector of exogenous parameters $\{\mu_f, \mu_m, \sigma_f, \sigma_m\}$.

 DOI: 10.2870/93297

3

Explaining Gender Differences in Educational Outcomes: Fitting and Testing the Theory

with Laurent Bossavie

3.1 Introduction

We show in this chapter that the theory proposed in Chapter 2 can explain the four stylized facts proposed in Chapter 1. The first two of those facts were the gender gap in participation to university, and its reversal from male majority to female majority over time. The other two were the gender gap in secondary school non-completion rates, and its reversal from female majority to male majority. To our knowledge this is the first attempt to account for all four facts with a single theory using international data.

The theory, which is based on the higher dispersion of male test-taking ability, is able to capture the empirically observed dynamics of the gender ratio in both the university enrollment rate and high school non-completion rate. Simulated predictions for gender ratios using parameters from our model fits and PISA assessment estimates are significantly correlated. This gives further support for our theory. In addition, it shows that the model resonates country-specific differences in test-taking ability distributions between genders.

Previous attempts in the literature have focused on the college gap reversal in the US. The theory proposed in Chiappori at al. (2009) is able to account for the reversal, although Hubbard (2011) criticizes the assumption that women have a higher college wage premium in the US. Goldin et al. (2006) invoke the removal of past barriers to womens' education and careers, combined with a higher college wage premium for women, as an explanation for this stylized fact over the period 1970–2010.

The structure of this chapter is as follows. In section 2, we describe the data used in the analysis and arguments as to why the particular dataset was used. We explain the method we use for the numerical estimation of the model in section 3. In section 4, we describe the fit with the data and show that it captures some characteristic of the country level differences in gender distributions of test-taking ability.

3.2 Data

Our data on total enrollment rates and gender ratios are from two main sources. Data on tertiary education enrollment rates is from the UNESCO Institute of Statistics, which records information on enrollment rates by gender for almost 200 countries over the period 1970–2010. The available measure for tertiary enrollment rates is the Gross Enrollment Ratio (ger), defined as the total number of students registered in tertiary education *regardless of their age*, expressed as a percentage of total mid-year population in the 5 year age group after the official secondary school leaving age (typically between 18 and 23). Formally, it can be expressed as:

$$ger_t = \frac{E^t}{P^t} \cdot 100$$

where E^t is the total number of individuals enrolled in tertiary education at time t. It includes all students officially enrolled in ISCED 5 and 6 levels of tertiary education[1]. P^t is the number of individuals belonging to the five-year age group following on the secondary school leaving age in year t. ger_t is therefore not bounded to be lower than 100%. It is a noisy measure of its theoretical counterpart in our model $x_t \equiv C_i(\bar{z}_t) \equiv \frac{G_{z_f}(\bar{z}_t)+G_{z_m}(\bar{z}_t)}{2}$, which is the fraction of individuals belonging to a synthetic age-cohort enrolling into tertiary education. We are, however, mainly interested in the comparative evolution of this ratio by gender, rather than in its absolute value. In addition, the ger is the only measure of tertiary enrollment available by gender on a yearly basis for a period of 40 years, in a large sample of countries.

Our data for upper-secondary school non-completion rates is from the Barro-Lee database 2010. The dataset records aggregate information on upper-secondary school

[1]ISCED 5 refers to the first stage of tertiary education, and includes both practically-oriented/occupationally specific programs and theory-based programs, respectively referred as 5B and 5A in the International classification of the United Nations. ISCED 6 refers to the second stage of tertiary education leading to the award of an advanced research degree

completion by gender, for more than 140 countries. Contrary to the UNESO data which measures the stock of tertiary educated individuals in a given country in a given year, the Barro-Lee dataset contains information on upper-secondary school completion rate by cohort of birth. It is therefore a flow measure of human capital, and is in this respect more sensitive to cohort-by-cohort changes in educational choices. The Barro-Lee dataset allows us to observe secondary school non-completion rates by gender for 5-year band birth cohorts born from 1891–1895 to 1981–1985 [1]. In total, we can extract the total secondary school non-completion rate and the female-to-male ratio among non-completers for 16 five-year-band cohorts in more than 150 countries. The drawback of the Barro-Lee dataset is that the measurement of educational attainment may vary for some countries. In particular, the data appears to be less reliable for developing countries.

3.3 Estimation

3.3.1 Maximum Likelihood Estimation

Our dataset allows to observe the following $2 \times T$ matrix for each country i in our sample:

$$\begin{pmatrix} x_{i1} & y_{i1} \\ x_{i2} & y_{i2} \\ \dots & \dots \\ x_{iT} & y_{iT} \end{pmatrix}$$

where x_{it} denotes the total enrollment rate in country i and year t, and y_{it} denotes the female-to-male ratio of country i in year t. In the context of our model, the total enrollment rate x_{it} is defined as $x_{it} = E_i(\bar{z}_{it}) \equiv \frac{G_{z_f}(\bar{z}_{it})+G_{z_m}(\bar{z}_{it})}{2}$ where $x = C(.) : \bar{z} \rightarrow [0, 1]$, given the two underlying distributions G_{z_f} and G_{z_m}. Contrary to $\bar{z}_{it}$, x_{it} presents the advantage of being observable in the data. Assuming normality, the two distributions are fully characterized by the two-parameter vectors (μ_m, σ_m^2) and (μ_f, σ_f^2), respectively.

The parameters of our model can easily be reduced to two, by normalizing one of the two probability density functions. We choose to standardize the female probability density function such that $f_f(\bar{z}_{it}) \sim N(0, 1)$, and denote ($\mu_i$, σ_i^2) the first two moments

[1] The aggregate Barro-Lee database was mostly constructed from nationally-representative surveys in which the exact year of birth of the respondent is typically not available for anonymity reasons. Instead, a five-year window of the individual's age is usually given.

of males' test taking ability distribution relative to females in country i. Formally,

$$\mu_i = \frac{\mu_{i,m} - \mu_{i,f}}{\mu_{i,f}} = \mu_{i,m} \tag{3.1}$$

$$\sigma_i = \frac{\sigma_{i,m}}{\sigma_{i,f}} = \sigma_{i,m} \tag{3.2}$$

In this setting, our model of investment in human capital predicts a unique value $\hat{y}_{it}$ of y_{it}, conditional on the triplet $\{x_{it}, \mu_i, \sigma_i\}$. Given the $2 \times T$ matrix, it is possible to estimate the vector of parameters $\{\mu_i; \sigma_i\}$ for country i by maximum-likelihood estimation, such that the distance between the actual data points and the ones predicted by our model is minimized.

Let $\bar{z}_{it}$ denote the test-taking ability cutoff in year t in country $i \in \{1, 2, ..., n\}$ above which individuals attend tertiary education. Test-taking ability for males and females are random variables, denoted z_m and z_f, respectively. Both are assumed to be normally distributed and their mean and standard deviation are allowed to differ across countries. We are interested in estimating the following model:

$$y_{it} = \frac{G_{z_f}(\bar{z}_{it})}{G_{z_m}(\bar{z}_{it})} \cdot \exp(\epsilon_{it}), \tag{3.3}$$

where $\exp(\epsilon_{it}) \sim \ln N(\mu_\epsilon, \sigma_\epsilon^2)$ Taking the logs of equation (13) yields:

$$\log y_{it} = \log G_{z_f}(\bar{z}_{it}) - \log G_{z_m}(\bar{z}_{it}) + \epsilon_{it}, \tag{3.4}$$

where $\epsilon_i \sim N(0, \sigma_\epsilon^2)$.

In addition, the analytical expressions for $G_{z_f}(\bar{z}_{it})$ and $G_{z_m}(\bar{z}_{it})$ are given by:

$$G_{z_f}(\bar{z}_{it}) = \int_{\bar{z}_{it}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_f^2}} e^{\frac{(z-\mu_f)^2}{2\sigma_f^2}} \, dz = \frac{1}{2}\left(1 - erf\left[\frac{\bar{z}_{it} - \mu_f}{\sqrt{2}\sigma_f^2}\right]\right) \tag{3.5}$$

$$G_{z_m}(\bar{z}_{it}) = \int_{\bar{z}_{it}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{\frac{(z-\mu_m)^2}{2\sigma_m^2}} \, dz = \frac{1}{2}\left(1 - erf\left[\frac{\bar{z}_{it} - \mu_m}{\sqrt{2}\sigma_m^2}\right]\right) \tag{3.6}$$

where $erf(\cdot)$ denotes the Gauss error function, expressed as $erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} \, dt$. The x-axis in the model is however not the test-taking ability variable, but the average of the two CCDFs:

$$x_{it} = E_{it}(\bar{z}_{it}) \equiv \frac{G_{z_f}(\bar{z}_{it}) + G_{z_m}(\bar{z}_{it})}{2} = \frac{1}{4}\left(2 - erf\left[\frac{\bar{z}_{it} - \mu_f}{\sqrt{2\sigma_f^2}}\right] - erf\left[\frac{\bar{z}_{it} - \mu_m}{\sqrt{2\sigma_m^2}}\right]\right) \tag{3.7}$$

We are interested in the inverse of this function. For practical purposes, we calculate numerically the inverse value $\bar{z}_{it} = C_{it}^{-1}(x_{it})$. The likelihood function can be expressed as:

$$L(\theta_i|\bar{z}_{it}) = L(\theta_i|E_i^{-1}(x_i)) = \log G_{z_m}(E_i^{-1}(x_{it})) - \log G_{z_m}(E_i^{-1}(x_{it})) + \epsilon_{it}, \tag{3.8}$$

where $\theta_i = \{\mu_i, \sigma_i\}$. Since the error term is normal, the model can be fitted by minimizing the sum of squared errors of the likelihood function, given different values of the parameters. Without any loss of generality, the number of parameters can be decreased to two by normalizing one of the two test-taking ability distributions, namely the one denoting females, to a standard normal i.e $\forall i = \{1, 2, ..., n\} : z_f \sim N(0, 1)$. Thus, the model is a non-linear mapping from x_i to y_i, whose form is defined by the parameters of the male distribution, when the female distribution in normalized to a standard normal distribution.

The maximum-likelihood estimator of $\theta_i = \{\mu_i, \sigma_i, \alpha_i\}$ can be expressed as:

$$L(\{\theta_i|z\}) = \frac{1}{(2\pi\sigma_\epsilon^2)^{\frac{n}{2}}} \cdot \exp\left\{-\frac{1}{2\sigma_\epsilon^2} \cdot \sum_{i=1}^{n} \left(\log y_{it} - \log G_{z_f}(E_i^{-1}(x_{it})) + \log G_{z_m}(E_i^{-1}(x_{it}))\right)^2\right\} \tag{3.9}$$

To obtain the values of $\theta_i = \{\mu_i, \sigma_i\}$ that maximize this likelihood, we take the Least-Squares fit given by:

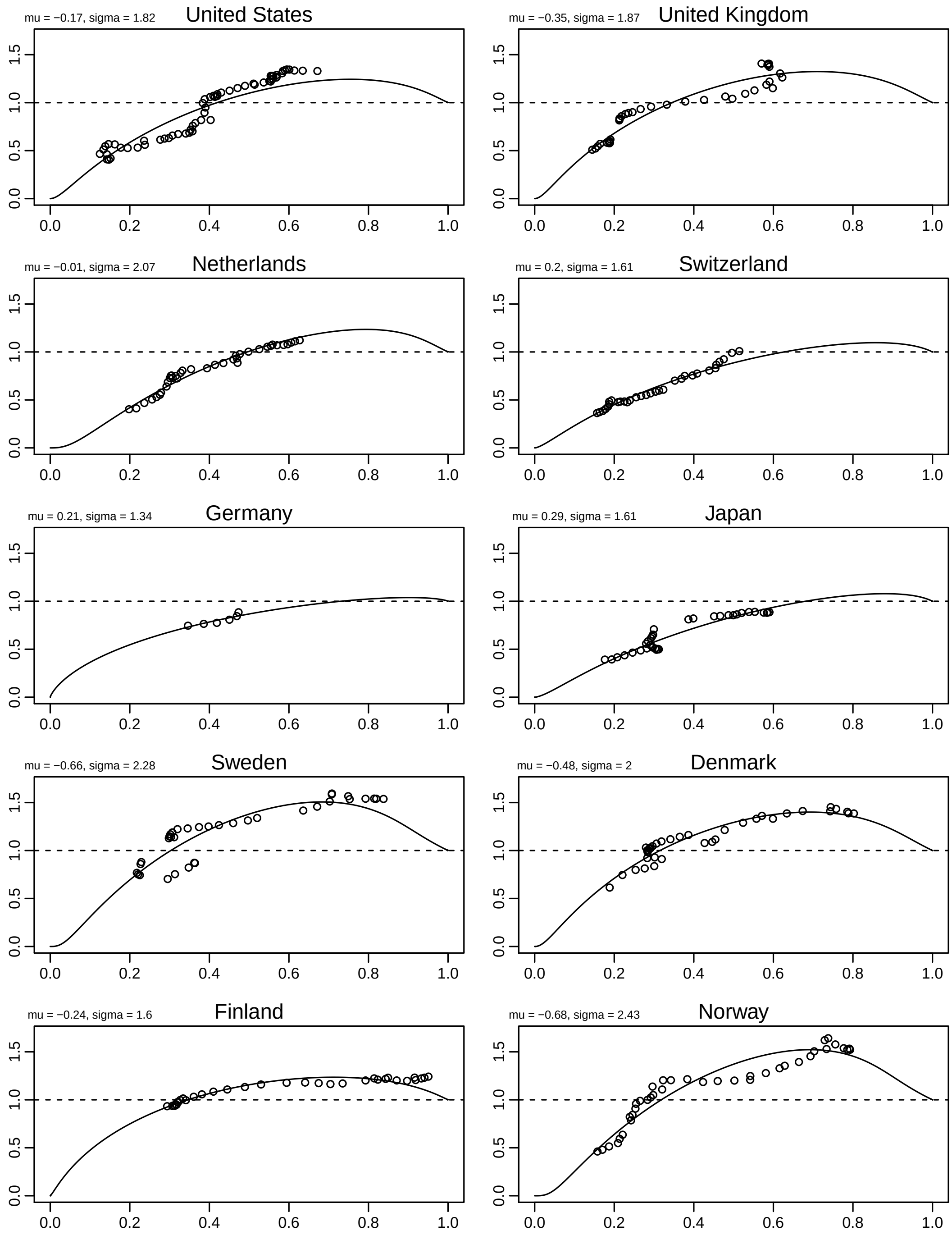
$$\hat{\theta_{MLE}} = \min_{\theta_i \in \Theta} \sum_{i=1}^{n} \left\{\log y_{it} - \log G_{z_f}(E_i^{-1}(x_{it}|\theta_i)|\theta_i) + \log G_{z_f}(E_i^{-1}(x_{it}|\theta_i)|\theta_i)\right\}^2 \tag{3.10}$$

3.4 Results

3.4.1 Model Fit

Figure 3.1 depicts the estimated relationship between the total enrollment rate in tertiary education x and the female-to-male ratio y when $\{\mu_i; \sigma_i\}$ are estimated from the $2 \times T$ matrix, as described in the previous section. As shown in the figure, our model generates an accurate fit for the relationship between x and y observed in each individual country. In a similar way, Figure 3.2 shows our model fit for the relationship between the secondary school non-completion rate, and the gender ratio among non-completers. As depicted in the figure, our model also generates a very satisfactory fit for the gender ratio in the lower tail of educational distribution. Contrary to university enrollment rates data, the advantage of the secondary school non completion data is that it ranges from virtually 100% for cohorts born at the end of the 19th century until less than 10% nowadays in some countries. It therefore allows us to reconstruct the almost entire path of the gender ration among non-completers, as a function of secondary school non completion rates. In particular, it allows to confirm that when the secondary non-completion rate increases at high levels of non-completion, the gender-ratio gradually decreases and converges to 1 for a 100% non-completion rate, as predicted by our model. To our knowledge, this is the first contribution to account for the gender gap reversal among secondary school non- completers, in addition to university students.

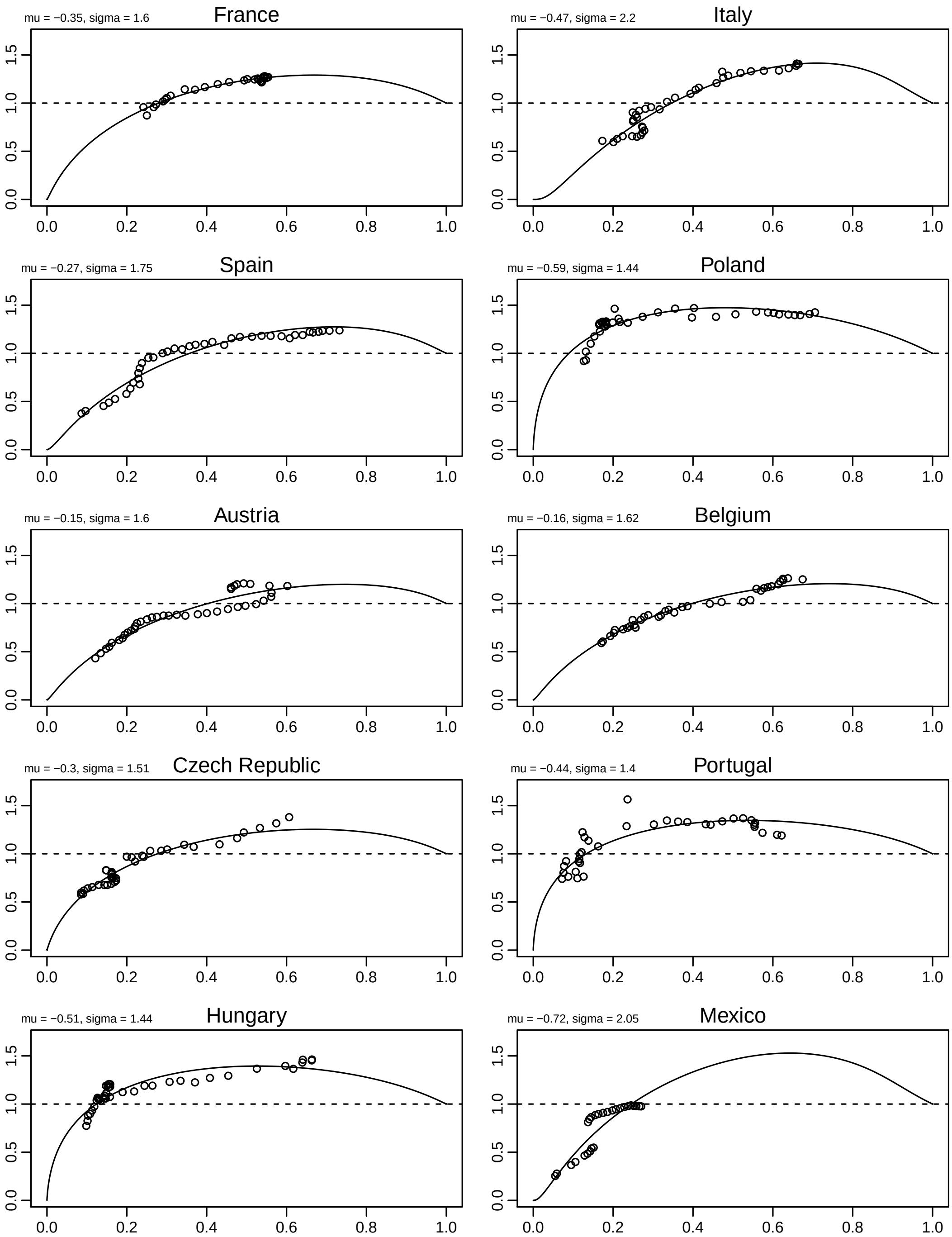
Figure 3.1: Model Fit - Gender Ratio in Participation to Tertiary Education



Notes. The x-axis measures the total gross enrollment ratio in tertiary education for country i. The y-axis measures the females-to-males ratio in tertiary education for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ for country i from the UNESCO Institute of Statistics. The full line depicts the estimated relationship between y and x from our model when $\{\mu_i; \sigma_i\}$ are estimated by maximum likelihood to minimize the error sum of squares.

3.4.2 Matching our Estimates with PISA Distributions

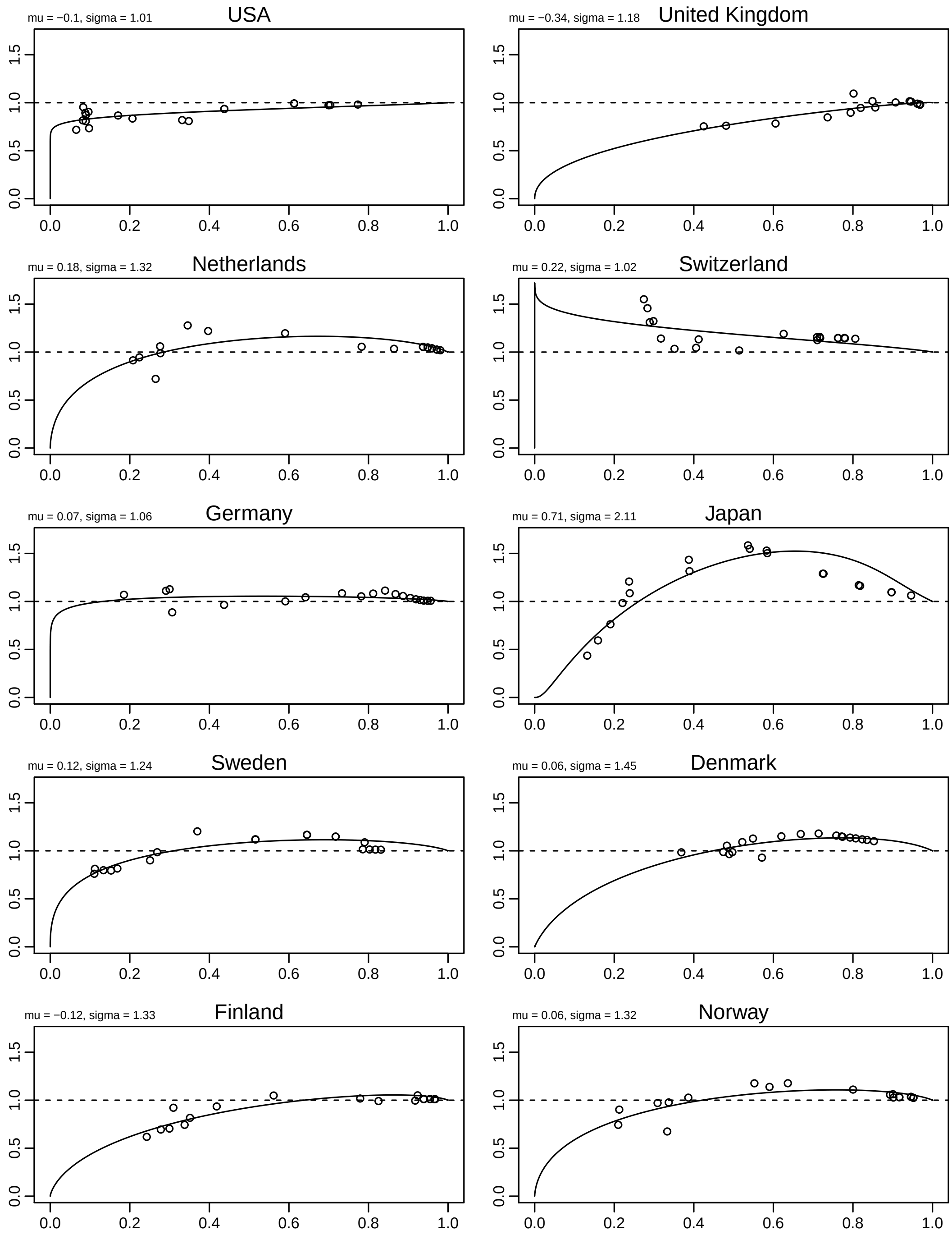
Our model is: $\log y_t = \log R(C^{-1}(x_t, \mu, \sigma), \mu, \sigma) + \epsilon_t$. Given our estimates for μ and σ, the model predicts the gender ratio $\hat{y}$ among the enrolled, for a given value of the



Notes. The x-axis measures the total gross enrollment ratio in tertiary education for country i. The y-axis measures the females-to-males ratio in tertiary education for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ for country i from the UNESCO Institute of Statistics. The full line depicts the estimated relationship between y and x from our model when $\{\mu_i; \sigma_i\}$ are estimated by maximum likelihood to minimize the error sum of squares.

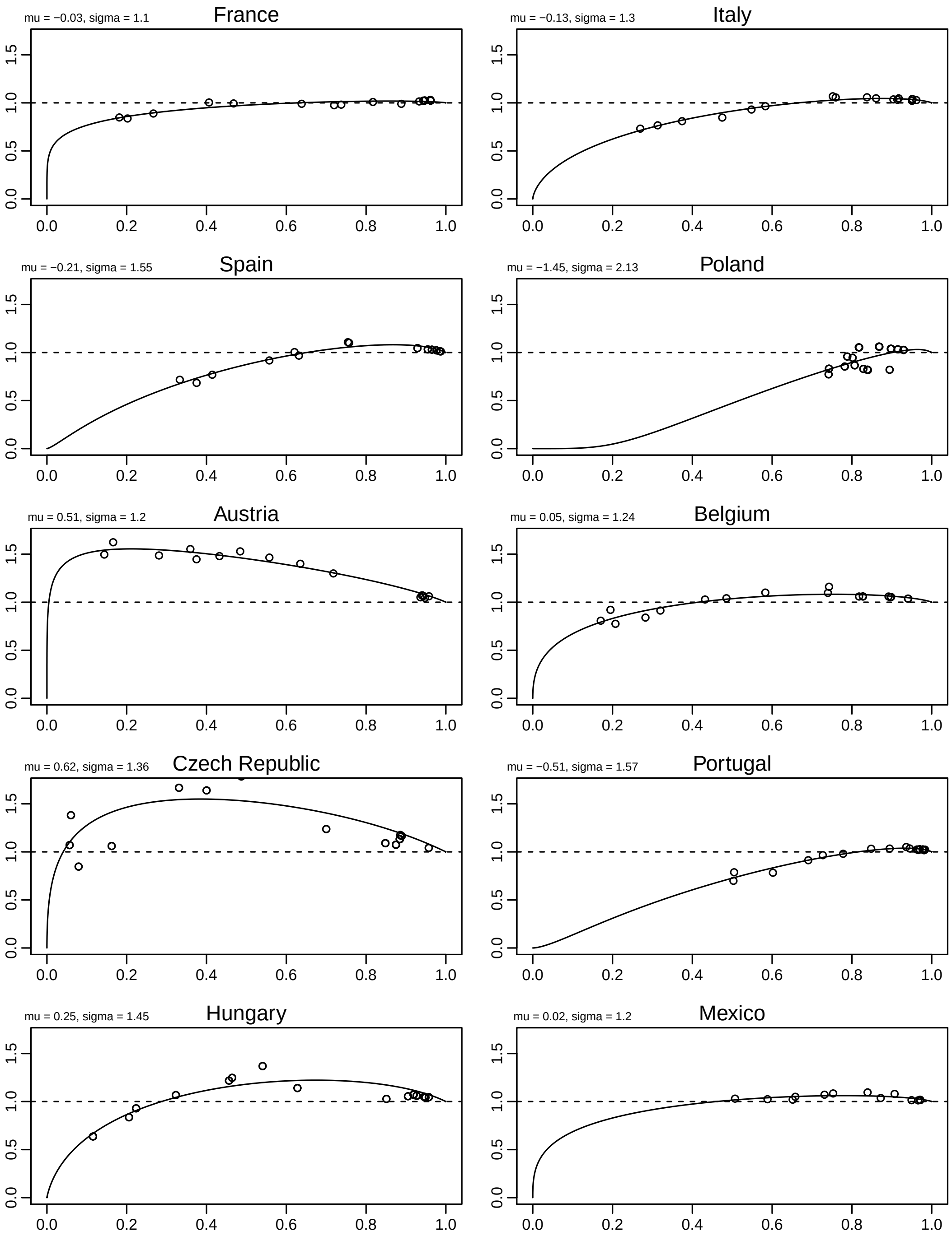
enrollment rate x. To assess its validity, we simulate the model with $\{\hat{\mu}_i\ \hat{\sigma}_i\}$ obtained from our fit with the UNESCO enrollment data, and a fixed value of of x. We then repeat the same procedure by inputing $\{\hat{\mu}_i\ \hat{\sigma}_i\}$ extracted from PISA test score distri-

Figure 3.2: Model Fit - Gender ratio among Secondary School Non-Completers



Notes. The x-axis measures the rate of secondary school non-completion for country i. The y-axis measures the females-to-males ratio among secondary school non-completers for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ *for country i from Barro-Lee (2010). The full line depicts the estimated relationship between y and x from our model when* $\{\mu_i; \sigma_i\}$ *are estimated by maximum likelihood to minimize the error sum of squares.*

butions, and the same fixed value of of x. We obtain two vectors of y_i's and measure to which extend they correlate for countries that both available in the UNESCO and PISA



Notes. The x-axis measures the total gross enrollment ratio in tertiary education for country i. The y-axis measures the females-to-males ratio in tertiary education for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ for country i from the UNESCO Institute of Statistics. The full line depicts the estimated relationship between y and x from our model when $\{\mu_i; \sigma_i\}$ are estimated by maximum likelihood to minimize the error sum of squares. Column (1) depicts our result using Estimator 1, column (2) shows our result using Estimator 2, as described in the previous section.

datasets. In total, 40 countries are common to both sources. In that sense, the PISA parameters are used as a benchmark to which we compare the predictions of our model.

The motivation behind this test is to check whether our model can capture underlying differences in gender distributions on a national level that shows in PISA exams at 15 and our model fit at around 18.

PISA assessments provide a suitable benchmark for underlying test-taking ability distributions by gender. First, the PISA sample was designed to be representative of the entire population of 15 year olds in a given country, since it surveys individuals in schools before the end of compulsory education. Second, it has been designed to be comparable across countries. Finally, it contains information on the gender of each individual, therefore allowing to construct estimates of ability test score distribution by gender in each country.

Table 3.2 shows the correlation between the $\hat{y}_i$'s simulated from the two sets of $\{\hat{\mu}_i, \hat{\sigma}_i\}$ obtained from our model fit, and the PISA distributions. We use the PISA dataset from year 2000 to get as close to the median year of the UNESCO data to minimize any attenuation bias that might emerge if there is a systematic shift in the relative distributions of females and males. We report correlations for 3 different values of x, $x = 0.20$, $x - 0.50$ and $x0.70$. Correlations between our predictions and PISA estimates are large, and do not vary much depending on the value of x we consider. The correlation of the gender ratio in a given quantile is approximately 0.4 with PISA reading ability, and significant at the 5% level. The magnitude of the correlation is slightly lower with PISA mathematics ability but remain larger than 0.3, and statistically significant at the 10% level.

Table 3.1: Correlations between Predicted Gender Ratio by Quantile from our Estimates and PISA Estimates

	Predicted Gender Ratio - Model Fit			
	2nd decile	5th decile	7th decile	9th decile
Predicted Gender Ratio - PISA Math	0.242	0.309	0.366*	0.339*
	(0.214)	(0.109)	(0.056)	(0.078)
Predicted Gender Ratio - PISA Reading	0.060	0.245	0.427**	0.530***
	(0.761)	(0.208)	(0.023)	(0.004)

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Sample size is 28 countries. The sample is all available countries for which there is are at least 20 observations in time. p-values are reported in parentheses.

3.5 Conclusion

Building on a simple framework of optimal investment in human capital, our model is able to reconcile three internationally robust stylized facts: the greater dispersion of men'a ability distribution relative to women, he progressive convergence and reversal in the university attendance gender gap and a reversal in the high school non-completion

Table 3.2: Correlations between Predicted Gender Ratio by Quantile from our Estimates and PISA Estimates

	Predicted Gender Ratio - Model Fit			
	2nd decile	5th decile	7th decile	9th decile
Predicted Gender Ratio - PISA Math	0.358*	0.394**	0.396**	0.370**
	(0.052)	(0.031)	(0.031)	(0.044)
Predicted Gender Ratio - PISA Reading	0.145	0.052	0.157	0.176
	(0.446)	(0.784)	(0.406)	(0.351)

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Sample size is 30 countries. The sample is all available OECD countries. p-values are reported in parentheses.

gender gap from female majority to male majority. Given data limitations and the inherent difficulty in measuring test-taking ability, our model provides a very satisfactory fit for the empirical relationship between total enrollment rate in tertiary education and the female-to male ratio. It is also the first theory to account for the convergence and reversal of the gender gap in secondary school non-completion rates. Beyond the particular question addressed in this paper, our findings suggest that gender differences in ability distribution might be relevant to account for other gender-related stylized facts in labour economics. Further research using this empirical fact to analyze other aspects of the gender gap in economic outcomes would therefore be of particular interest. Importantly, the larger variability of men's ability distribution observed empirically remains mostly unexplained, and stands as another promising area for future research, beyond the field of Economics.

4

Explaining Gender Differences in Educational Outcomes: Evaluating Alternative Hypotheses

with Laurent Bossavie

4.1 Introduction

In previous chapters, we have shown that our theory fits the empirical data quite accurately. Some alternative explanations, however, are also consistent with the dynamics observed in the data and can generate a similar reversal of the gender ratio. Theories proposed in the literature have been focusing on explaining the gender gap reversal in university attendance. Two main hypotheses have been formulated. First, as argued in Chiappori et al. (2009), changes in social norms combined with higher returns to education for females can produce a reversal from male majority to female among university students. Second, a relative increase in females' mean test-taking ability over time can also generate a reversal in the college gender gap, as suggested by Cho (2007).

In this chapter, we propose several tests to assess the validity of these two alternative theories against empirical data. We first formulate them in the framework we developed in the previous chapters, and then assess their predictions against the ones of our competing theory. All three hypotheses can be stated within the framework we have established in the previous chapters:

Hypothesis 1, *higher male variability hypothesis*: The function $G_z(\cdot)$ is gender-specific, but static over time, i.e. the higher dispersion in male test-taking ability explains the college gender gap reversal. The lower bound of test-taking ability $\bar{z}$ for attending university is the same for both genders.

Hypothesis 2, *change in social norms hypothesis*: The lower bound of test-taking ability required for college varies between men and women, i.e. there exist gender-specific $\bar{z}$'s, $\bar{z}_f$ and $\bar{z}_m$ for females and males respectively, although the test-taking ability distribution $G_z(\cdot)$ is the same for both genders. In the past, $\bar{z}_f > \bar{z}_m$ and at some point in time the situation reversed. This theory can also explain the college gender gap reversal.

Hypothesis 3, *increase in females' mean performance hypothesis*: Over the last decades, the female mean test-taking ability has increased relative to males. The variability of the distributions is the same and the $\bar{z}$ is also the same for both genders. Again, this hypothesis can explain the reversal.

These three hypotheses could theoretically be combined into joint hypotheses. However, we analyze them separately to retain maximum simplicity and assess their respective explanatory powers. In this chapter, we focus on Hypotheses 2 and 3 since we have already closely scrutinized Hypothesis 1, our main hypothesis, in the preceding chapters. We will only evoke it for comparison purposes.

To evaluate the change in social norms hypothesis (Hypothesis 2), we perform three different tests against our theory. All three tests reject the hypothesis that changes in social norms are the main driving force behind the gender gap dynamics, and bring further support to our theory.

Regarding the increase in females' mean performance over time (Hypothesis 3), we argue that the results of previous studies might suffer from sample restriction biases. Since evidence suggests that males' and females' variances of test-taking ability differ, sample restriction to one side of the distributions will bias estimates of gender differences in mean ability. To address this issue, we use repeated cross sections of representative samples of a given age population, for a wide range of countries. While females' mean performance appears to have increased relative to males' in reading, we find that females' tend to do relatively worse in mathematics compared to what they were used to. Evidence in this respect is therefore ambiguous, and offers little support for Hypothesis 3.

4.2 Change in Social Norms (Hypothesis 2)

4.2.1 The Change in Social Norm Hypothesis in our Framework

In our framework, the enrollment rate at university for each gender is:

$$E = 1 - F_z(\frac{\bar{s}}{b}) = G_z(\frac{\bar{s}}{b}) = G_z(\bar{z}) \tag{4.1}$$

where $G_z(.)$ denotes the CCDF of test-taking ability z. $\bar{z} \equiv \frac{\bar{s}}{b}$ is the lower bound of ability z such that individuals choose to attend university, where $\bar{s}$ is the number of schooling years required to obtain a university degree. b is an exogenous parameter, which defines the net benefits to education. We formulate the change in social norms hypothesis by allowing b to differ and change over time between the two genders, with $G_z(.)$ being identical for males and females. In this setting, it is possible to generate the gender gap dynamics observed in the data if $b_f < b_m$ originally, before gradually converging and overtaking b_m over time. Optimal levels of investment in schooling are expressed separately for males and females as:

$$s_i^* = \begin{cases} z_i \cdot b^m & \text{if male} \\ z_i \cdot b^f & \text{if female} \end{cases}$$

And the enrollment rate in higher education for each gender is:

$$E = \begin{cases} G_z(\frac{\bar{s}}{b_m}) = G_z(\bar{z}_m) & \text{for males} \\ G_z(\frac{\bar{s}}{b_f}) = G_z(\bar{z}_f) & \text{for females} \end{cases}$$

where $G_z(.)$ is identical for males and females.

In this context, $b_f > b_m$ in most recent years is a necessary condition for the gender gap reversal in university participation. This means that net benefits for females have to be higher than for males at the margin. Empirical evidence on larger returns to education for females is ambiguous, however. An important component of monetary returns to higher education is the college wage premium. Chiappori at al. (2009) Card and DiNardo (2002), or Charles and Luoh (2003) for example find a higher college wage premium for women, but their estimations are restricted to the US. In addition, the methodology behind the findings of these studies has been strongly criticized by Hubbart (2011). In particular, he finds no gender difference in the US college wage premium after correcting for a bias associated with income topcoding in the dataset used by US studies. Cho (2007) further points out that the trends in the college wage premium has been very similar for men and women over the last decades, making it an unlikely explanation for the college gender gap reversal. Even if the college wage premium is higher for females and some other other benefits increased faster for females over the last decades, Becker et al. (2010) argue that benefits to higher education are still lower for women in most dimensions. In the absence of $b_f > b_m$ in recent years, the change in social norms hypothesis alone would be unable to account for the reversal of the gender ratio.

Let us now fit and assess Hypothesis 2, by allowing b_f and b_m to take any value. Hypothesis 2 is very flexible for fitting the data. Since the relative changes of $\bar{z}_f$ and $\bar{z}_m$ have not been constrained, we can even make an exact fit with the data. The fit is depicted in Figure 4.1. The method for fitting is straightforward. For each time period, there is a system of two equations:

$$y_t = \frac{G_z(\bar{z}_{f,t})}{G_z(\bar{z}_{m,t})}$$

and

$$x_t = \frac{G_z(\bar{z}_{f,t}) + G_z(\bar{z}_{m,t})}{2},$$

where x_t and y_t are known. By substituting we get:

$$y_t = \frac{G_z(\bar{z}_{f,t})}{2x_t - G_z(\bar{z}_{f,t})}$$

We can easily solve numerically for the unknown and unique values of $\bar{z}_{f,t}$ and $\bar{z}_{m,t}$. To extrapolate outside the actual data range, we assume that $\bar{z}_m$ and $\bar{z}_f$ continue to change at the average estimated pace of change between the years 1946 and 2009. The estimated values of $\bar{z}_{f,t}$ and $\bar{z}_{m,t}$ and their change over time is depicted in Figure 4.2.

4.2.2 Hypothesis 2 Versus Hypothesis 1: Test 1

Hypothesis 2 appears to do an excellent job at explaining the reversal. However, it can be shown that Hypothesis 1 and Hypothesis 2 have opposite implications regarding the relationship between the total enrollment rate defined as:

$$C(\bar{z}_f, \bar{z}_m) = \frac{G_{z_f}(\bar{z}_f) + G_{z_m}(\bar{z}_m)}{2} \tag{4.2}$$

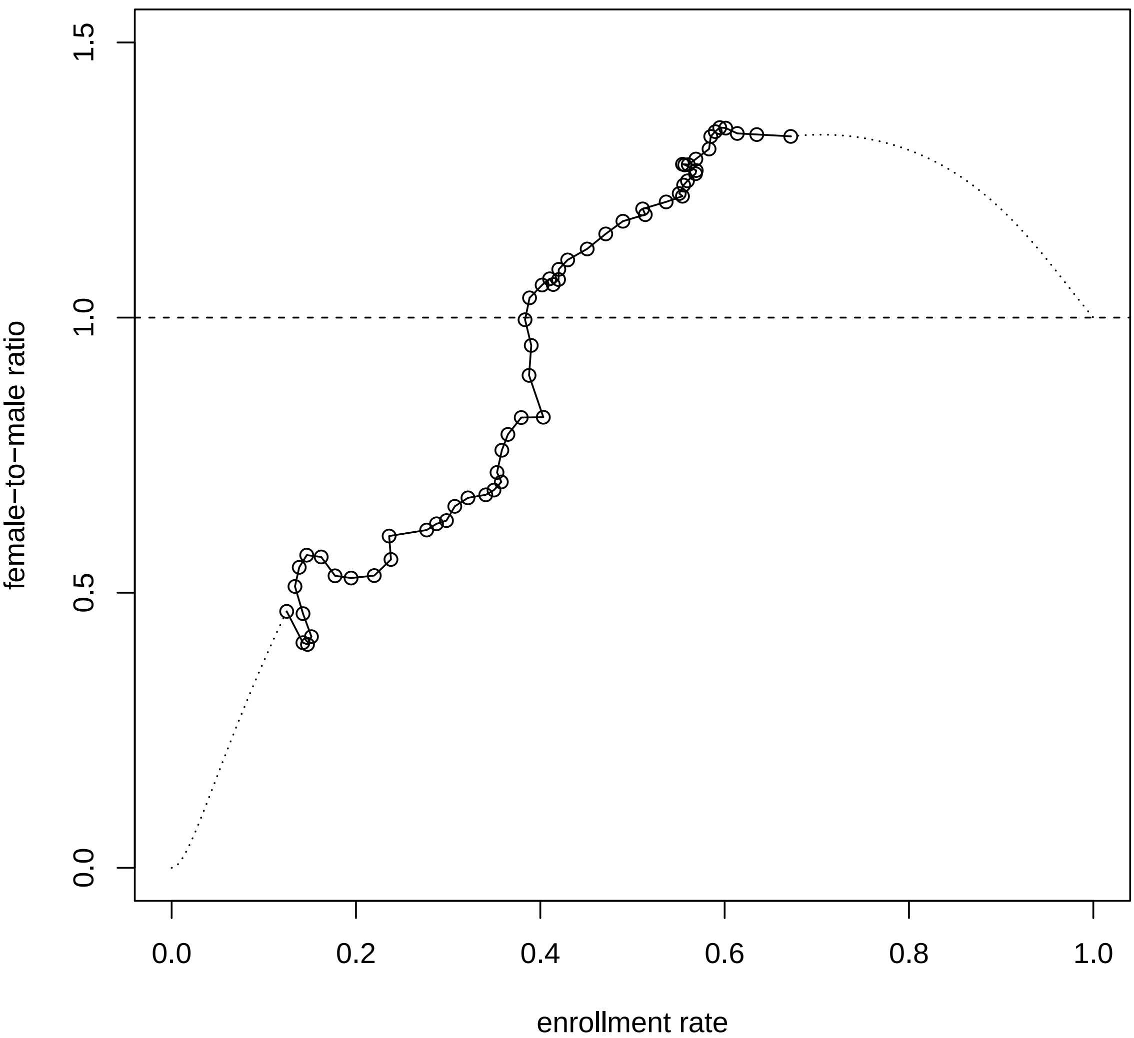
and

$$E[z_f|z_f > \bar{z}_f] - E[z_m|z_m > \bar{z}_m], \tag{4.3}$$

which is the difference between the mean test-taking ability of females and males selected into university education. This comes from the fact that Hypothesis 1 implies that z_f and z_m have different distributions and $\bar{z}_f = \bar{z}_m$, where as Hypothesis 2 would imply that z_f and z_m are distributed according to the same distribution and that $\bar{z}_f$ and $\bar{z}_m$ may differ.

The change in social norms hypothesis (Hypothesis 2) implies that the average test-taking ability is initially higher for enrolled females than for enrolled males, and progressively converges towards it before taking lower values. On the other hand, Hypothesis 1 implies that the females' average gets higher relative to males as the the

Figure 4.1: The change in social norms hypothesis fitted, with projections.

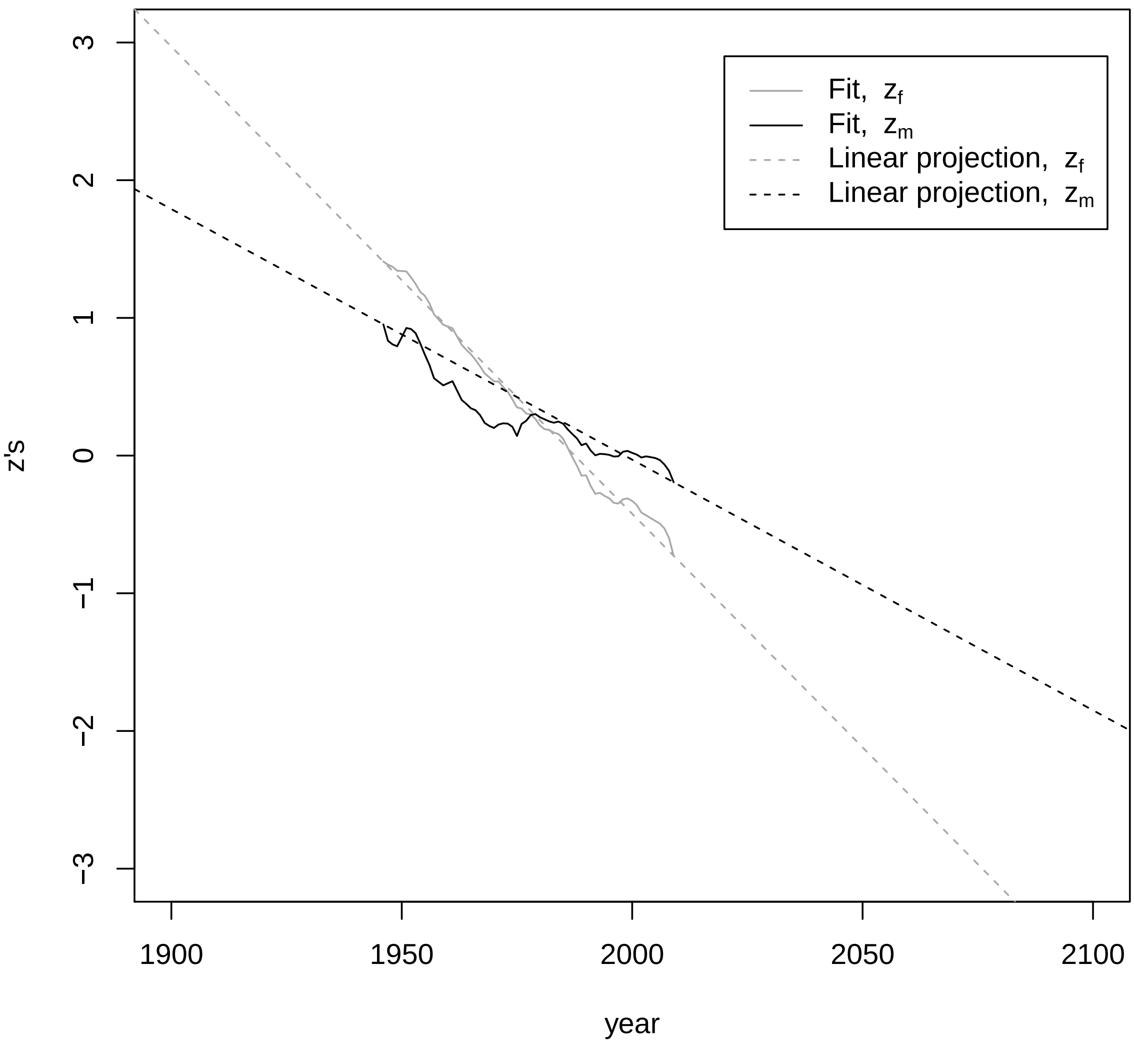


Notes. The graph shows the exact fit with the data obtained by varying $\bar{z}_f$ and $\bar{z}_m$, as explained in the text. The projections are made assuming an evolution of $\bar{z}_f$ and $\bar{z}_m$ that follows the average of the fitted years before and after the data range used.

fraction of population taking the test increases. When the test takers are a representative sample of the whole population, the observed mean difference becomes an estimate of the mean difference of the whole population. To test these two opposite predictions against the data, we analyze data on both the average performance in cognitive test by gender, and the proportion of a given cohort taking the test.

Our data is from various sources. SAT mean test scores by gender are from the

Figure 4.2: Change in social norms hypothesis fitted values of $\bar{z}_f$ and $\bar{z}_m$, with projections.



Notes. The graph shows the values of fitted z's from the exact fit with the data obtained by varying $\bar{z}_f$ and $\bar{z}_m$ as explained in the text. The projections are made assuming an evolution of $\bar{z}_f$ and $\bar{z}_m$ that follows the average of the fitted years before and after the data range used.

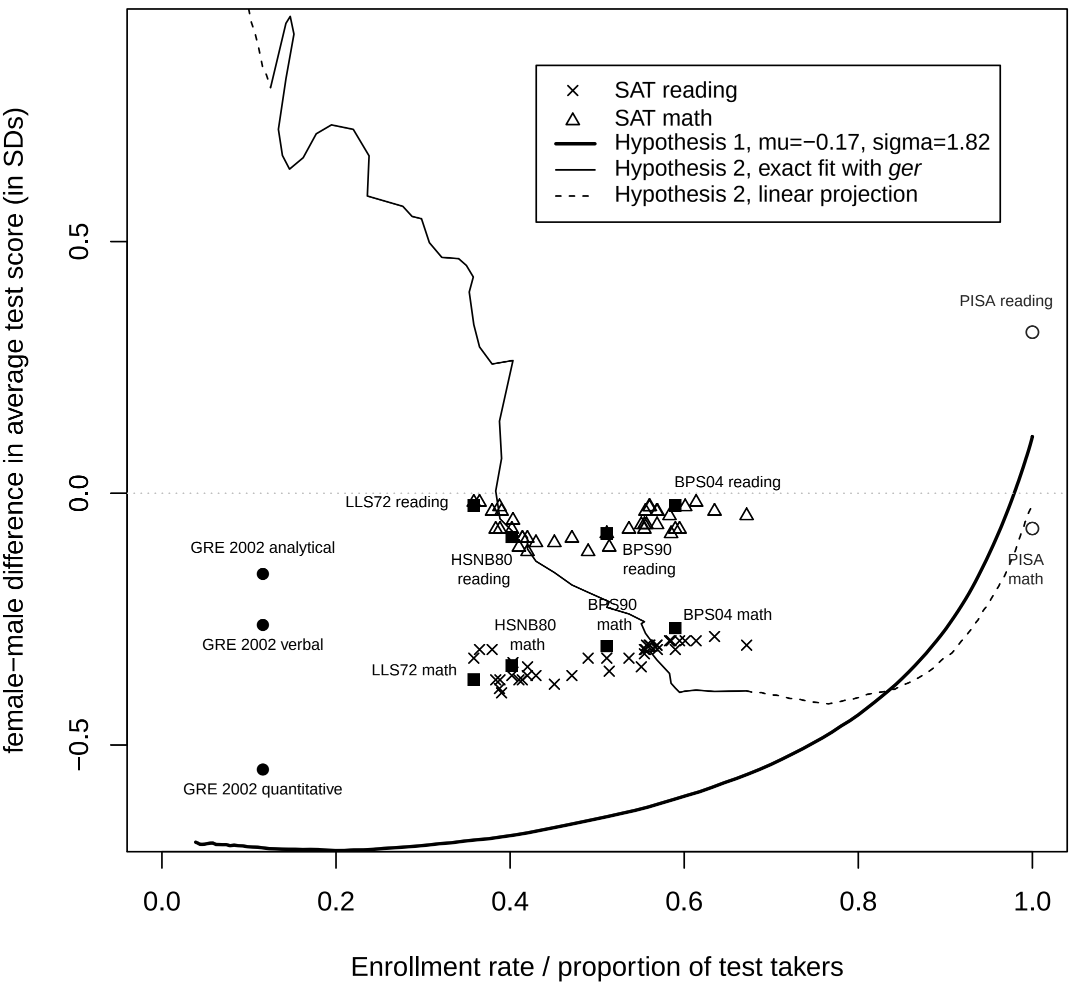
College Board which provides average mean scores for mathematics and reading by gender from 1970 to 2010, as well as the total number of females and males taking the test in a given cohort. We complemented this data with 4 US longitudinal surveys: The National Longitudinal Study of the high school class of 1972 (NLS72), High School and Beyond 1980 (HS&B), the National Educational Study of 1988 (NELS 88) and the Educational Longitudinal Study of 2002 (ELS 2002). From these data sets, we obtain the average test score of college students by gender over time, and can follow with total enrollment rate through these 4 data points. We also use test score data from PISA for both mathematics and reading, which are taken by a representative sample of the entire population of 15 year olds in a given country. Finally, we also add Graduate Record Examinations data from Graduate Record Examinations Board for the year 2000 for quantitative, analytical and verbal skills. This is necessary to study results in tests that are taken by only a small fraction of the population.

As depicted in Figure 4.3, while Hypothesis 1 generates an increase in the average level of test-taking ability of females attending university relative to males over time, the change in social norms hypothesis implies an evolution of the opposite sign, until the sample restriction reaches a proportion of around 0.6 to 0.8 of the cohort taking the test.

Figure 4.3 shows that Hypothesis 1 performs well at predicting the relationship between the average gender in cognitive scores and the fraction of the population taking the test, implied by $\bar{z}$. It provides a good fit for both the shape of the relationship observed in the data, and the sign of the gender gap in average cognitive tests. The data shows that males do better relative to females in a restricted sample selected from the top of the distribution, than in the entire population. In PISA, which is a sample of the entire population of 15 year-olds, females obtain higher average test scores relative to males in reading. On the other hand, males perform on average better in the same discipline with the SAT test. The higher male variability hypothesis (Hypothesis 1), although not a perfect fit, provides a simple explanation for the main facets of this puzzle, relying on the fact that test-takers are drawn from different ranges of the ability distribution in these tests. This matters for the observed average performance by gender, given different underlying distributions of test-taking ability by gender assumed in our model.

The alternative hypothesis of change in social norms (Hypothesis 2) neither predicts the sign of the gender gap for the whole range, nor the sign of the relationship between the gender gap in average test performance and the proportion of population taking the test. According to this theory, the college gender gap reversal would be expected to be accompanied by an analogous reversal in the test-taking ability of college students. This does't appear to happen, although earlier SAT data would improve inference based on test scores. The GRE exam scores especially follow the prediction of Hypothesis 1 rather than Hypothesis 2.

Figure 4.3: Change in Social Norms Hypothesis Vs Model Fit for the Relationship between Enrollment Rate and Gender Gap in Cognitive Score of the Enrolled



Notes. The x-axis represents the proportion of the cohort taking a given cognitive test. The y-axis represents the female-to-male difference in means of the given cognitive test, expressed in standard deviations units. The thick line depicts the expected relationship between the gender difference in average cognitive score of university students as a function of total enrollment, as predicted by the higher male variability hypothesis (Hypothesis 1). The thin solid line represents the evolution of the same variable according to the fit of the change in social norms hypothesis (Hypothesis 2), when males and females cognitive score distributions are assumed to be the same, but women are initially facing a higher $\bar{z}$ relative to males that progressively converges to the males' level when university enrollment increases. The thin dashed line represents the predicted evolution according to Hypothesis 2, assuming a linear evolution of z_m and z_f that follows the average estimated evolution between 1946 and 2009. The crosses and triangle dots represent the actual value of the gender difference in SAT test scores for the entire population of SAT test-takers, as a function of the fraction of SAT test-takers in the population. The filled squares represent similar values computed from US post-secondary longitudinal surveys, for college students only. The filled dots represent GRE exam results for the year 2002.

DOI: 10.2870/93297

4.2.3 Hypothesis 2 Versus Hypothesis 1: Test 2

We also test the explanatory power of our theory against the change in social norms hypothesis in a setting where we fit our model against a proxy variable for female emancipation at a country group level. We use data from the International Labour Organization on female labour force participation rate (FLFPR) from 1980 to 2010 as a proxy for women's emancipation. Table 4.2 shows the results of a numerical fit of the model when FLFPR is added as an explanatory variable as a proxy for females' emancipation. Also, each country within the country group is controlled for with a dummy variable to produce the following model:

$$log\, y_{i,t} = log\, G_{z_f}(E^{-1}(x_{i,t}, \mu, \sigma), \mu, \sigma) - log\, G_{z_m}(E^{-1}(x_{i,t}, \mu, \sigma), \mu, \sigma) + \alpha_i + \beta\, log\, FLFPR_{i,t}.$$

The coefficient associated with σ is significant at the one percent level in all country groups with bootstrapped standard errors. On the other hand, the coefficient associated with the proxy for women's emancipation is insignificant in all the regions, with the exceptions of catholic Europe and Africa.

Table 4.1: Competing Hypothesis - Numerical Estimation 1980-2010

	Africa	Anglo-Saxon	Catholic Europe	Protestant Europe	Eastern Europe	Latin America	Islamic	Asia
	Dependent variable: females-to-males ratio in tertiary education							
$\hat{\mu}$	0.17	-0.28	-0.07	-0.58	0.18**	-0.58**	0.38	0.03
	(0.33)	(0.31)	(0.35)	(0.41)	(0.09)	(0.27)	(0.24)	(0.23)
$\hat{\sigma}$	1.16*	1.56***	1.34***	1.74***	1.18***	1.39***	1.23***	1.21***
	(0.1)	(0.09)	(0.07)	(0.15)	(0.05)	(0.07)	(0.07)	(0.06)
Women's	-0.19	0.14	-0.14	0.11	-0.46***	-0.09	-0.17**	0.07
LFPR	(0.28)	(0.16)	(0.17)	(0.27)	(0.08)	(0.17)	(0.08)	(0.14)
Country F.E.	✓	✓	✓	✓	✓	✓	✓	✓
N. Countries	30	6	12	8	22	23	21	16
N. Observations	291	298	156	427	199	271	333	243

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Standards errors are bootstrapped.

4.2.4 Hypothesis 2 Versus Hypothesis 1: Test 3

Another way to discriminate between our theory and the change in social norms hypothesis is to look at the evolution over time of the relationship between test score and enrollment at university by gender. Hypothesis 1 assumes that test-taking ability is the only relevant choice variable for choosing to attend university. This implies that the empirical relationship between z and H should be identical for males and females, in every given cohort. In particular, there should not be any gender-specific change in the

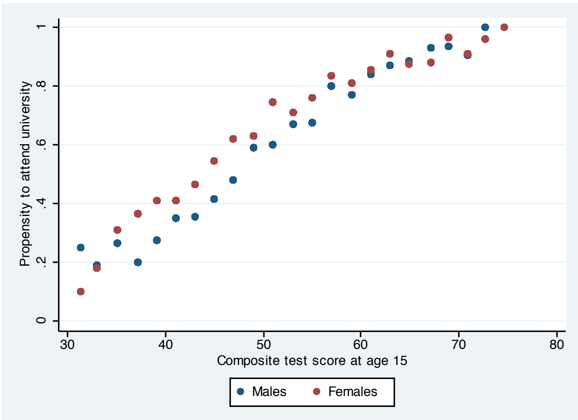
Table 4.2: Competing Hypothesis - Numerical Estimation 1980-2010

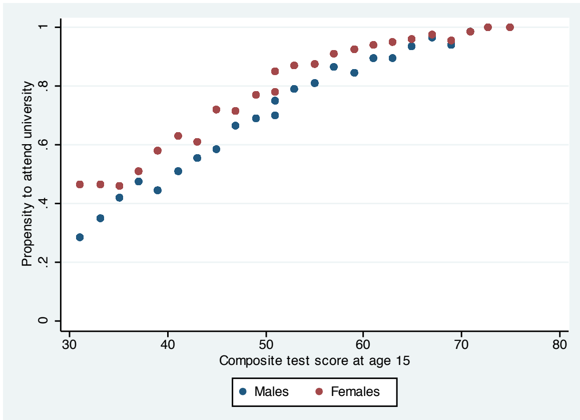
	Africa	Anglo-Saxon	Catholic Europe	Protestant Europe	Eastern Europe	Latin America	Islamic	Asia
	Dependent variable: females-to-males ratio in tertiary education							
$\hat{\mu}$	-0.58	-0.01	0.13	0.05	-0.18	-0.33	0.09	0.07
	(0.39)	(0.29)	(0.15)	(0.37)	(0.21)	(0.22)	(0.12)	(0.2)
$\hat{\sigma}$	1.4***	1.46***	1.26***	1.48***	1.28***	1.31***	1.27***	1.16***
	(0.12)	(0.13)	(0.05)	(0.16)	(0.06)	(0.07)	(0.05)	(0.06)
Women's	0.83***	0.44	0.73***	0.16	0.26*	0.43*	0.16	-0.29
LFPR	(0.31)	(0.29)	(0.22)	(0.32)	(0.15)	(0.24)	(0.18)	(0.21)
Country F.E.	✓	✓	✓	✓	✓	✓	✓	✓
N. Countries	30	6	12	8	22	23	21	16
N. Observations	291	298	156	427	199	271	333	243

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Standards errors are bootstrapped.

relationship between z and H over time. On the other hand, hypothesis 2 implies that females used to enroll less than males at university conditional on test scores ($b_f < b_m$), while they now have a higher propensity to enroll for a given test-taking ability, since $b_f > b_m$.

Figure 4.4: The Relationship between z and H by Gender: 1980 Vs 2002





Source. High School and Beyond survey of 1980 (HS&B) and Educational Longitudinal Study (ELS) of 2002

To assess these competing hypothesis, we use data from two longitudinal surveys conducted in the US in 1980 and 2002. The first of these surveys is the High School and Beyond 1980, which follows a cohort of 10th graders in 1980 until university studies. The second of these surveys is the US Educational Longitudinal Study 2002, which also follows 10th graders in 2002 until university studies. These two datasets both survey a representative sample of the same age group in the US population, and provide test score information in 10th grade on a comparable scale. In addition, they both allow us to know whether individuals attended university education, and to match this information with individual test scores in 10th grade. Using these two surveys, we constructed the empirical relationship between z measured by test scores at age 15, and the propensity to enroll at university H at two different points in time: 1980 and 2002. We believe the 22-year time span to be sufficient to detect asymetric changes by gender in the relationship between z and H, especially in a period during which enrollment at university increased dramatically in the US. The results of the analysis are depicted in Figure 4.4.

Figure 4.4 shows two main patterns. First, females seem to be more likely to enroll in college, conditional on test scores. Second, and most importantly in our context, this was already the case in 1980. In other words, we do not observe any change in the propensity of females to enroll at university conditional on test scores relative to males, which goes against the change in social norms hypothesis. We further investigate changes in the relationship between z and H quantitatively, using regression analysis. The results are presented in Table 4.3 and Table 4.4. As shown in Table 4.3, regression analysis suggests that males enroll significantly less at university than females for given test scores. Interestingly, the point estimate of this negative effect of being male is very stable between 1980 and 2002: -0.083 in 1980, against -0.078 in 2002. In other words, there is no evidence for women being penalized in university enrollment given their test scores, as the change in social norm hypothesis would suggest. In addition, the effect of gender on university enrollment conditional on test scores is fairly stable over time, which goes against the predictions of Hypothesis 2.

Table 4.4 brings further insights by running separate regressions by gender in 1980 and 2002. The intercept for both males and females increased sharply from 1980 to 2002, reflecting the fact that both genders have a higher propensity to enroll at university in 2002 than in 1980, conditional on their test scores. Therefore, although the propensity to enroll at university increased sharply over the period, there does not seem to be any asymmetrical change in the relationship between test scores and university enrollment: both genders enroll more at university conditional on their test scores, and males enroll less at university than females given their test score both in 1980 and 2002. Interestingly, the association between university enrollment and test scores have decreased for both genders over the period. This may reflect the multiplication of US postgraduate institutions in recent years implementing less strict criteria of admissions. Again, this evolution appears to be quite symmetric between genders: the association decreased from 0.25 to 0.19 for males, and from 0.21 to 0.16 for females. The interaction changes from positive and significant to negative and insignificant, but has only a distributional effect within boys, since the mean is very close to zero.

 DOI: 10.2870/93297

Table 4.3: The Relationship between z and H: 1980 Vs 2009

	Dep. variable: Propensity to Enroll at University			
	1980		2002	
Male	-0.086***	-0.083***	-0.076***	-0.078***
	(0.01)	(0.008)	(0.007)	(0.04)
Composite Test Score	0.22***	0.21***	0.17***	0.18***
	(0.007)	(0.008)	(0.004)	(0.006)
Interaction		0.027**		-0.036
		(0.01)		(0.028)
Intercept				
R-squared	0.142		0.151	
N. observations	11,641	11,641	13,240	13,240

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Standards errors are bootstrapped.

Table 4.4: The Relationship between z and H in 1980 and 2009 - Separate Regressions by Gender

	Dep. Variable: Propensity to Enroll at University			
	1980		2002	
	Males	Females	Males	Females
Composite Test Score	0.25***	0.21***	0.19***	0.16***
	(0.007)	(0.007)	(0.005)	(0.005)
Intercept	0.21***	0.43***	0.64***	0.91***
	(0.018)	(0.018)	(0.018)	(0.019)
R-squared	0.163	0.120	0.146	0.145
N. observations	5,554	6,087	6,305	6,878

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Standards errors are bootstrapped.

4.3 Increase in Females' Mean Performance (Hypothesis 3)

Our theory relies on the assumption that test score distributions for males and females are fixed over time, and in particular that the difference μ_f - μ_m is time-invariant. Now, however, we study the possibility that the average performance of girls relative to boys increased over time. In our setting, this corresponds to a shift of f_{z_f} to the left relative to f_{z_m}, leading to an increase in E_f relative to E_m, without any change in $\bar{z}$. This possibility has been investigated by Cho (2007) and Fortin (2011) for the US. Fortin, however, finds no relative increase in girls' self-reported grades over the period 1970–2010. Using data from the Monitoring the Future study, she reports an increase in high school grades for both boys and girls with parallel trends over the period, and

DOI: 10.2870/93297

girls already outperforming boys in the early 1970s. Cho, on the other hand, finds that women's performance in high school, measured by test scores, increased more rapidly than for men over the last three decades.

An important caveat of the analysis is that it applies to high school seniors, which are approximately 18 at the time of the survey, and therefore beyond the compulsory age of high school attendance. This means that if the hypothesis about higher male variance is true, the mean of a restricted sample of students is a function of the first two moments of the underlying distribution, and the point of restriction.[1]. Since the proportion of students taking the tests in Cho (2007) has changed over time, which means that the point of restriction has changed, one would expect there to be changes in the implied means even if the mean of the underlying distribution did not change.

To evaluate whether the mean performance of the female population increased over time relative to boys, one should use a representative sample of the entire country population of a given age group. This can be achieved by using school test scores taken at an age at which schooling is still compulsory. In this respect, the Project for International Student Assessment (PISA) surveys a representative sample of the 15-year-old population in more than 40 countries. In addition, test results have been designed to be comparable over time. The drawback of this data, however, is that it is only available from 2000 onwards, and therefore allows tracking relative changes in mean performance between genders over the period 2000-2010. Figure 4.5 and Figure 4.6 depict the evolution of girls' mean average performance relative to boys in reading and mathematics over the period 2000-2010 for around 40 countries included in PISA. They show that while female relative average performance in reading seem to have increased over the period 2000-2009, females appear to do worse in mathematics relative to males in 2009 compared to 2000. Therefore, international evidence is mostly inconclusive regarding the increase of female mean performance.

Given the short time-spam of the PISA study, we complement our analysis by looking at the evolution of the performance of high school students by gender in the US, over the period 1980–2002. To this purpose, we use two nationally representative longitudinal surveys of high school students in the US conducted in 1980 and 2002. These surveys both contain information on test scores in mathematics and reading when individuals were in 10th grade. The results are depicted in Table 4.5. The table shows that the average test score of female 10th graders has increased relative to boys in the US, between 1980 and 2002. In mathematics, girls' disadvantage decreased from -0.121 to -0.107, whereas the girls advantage in reading score increased from 0.057 standard

[1] In the case of a random variable x that follows a normal distribution that is truncated from below, $E(x) = \mu + \frac{\phi(\alpha)}{1-\Phi(\alpha)}\sigma$, where $\phi(\cdot)$ is a standard normal PDF and $\Phi(\cdot)$ is a standard normal CDF.

DOI: 10.2870/93297

Figure 4.5: Average PISA performance of males relatives to females in maths: 2000-2009

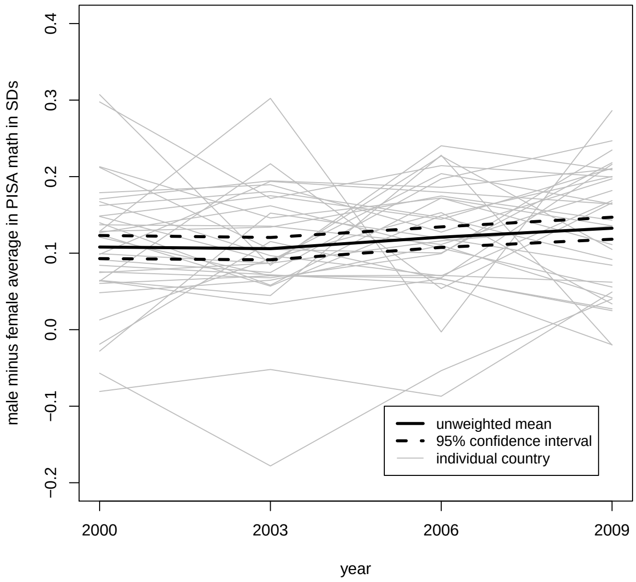
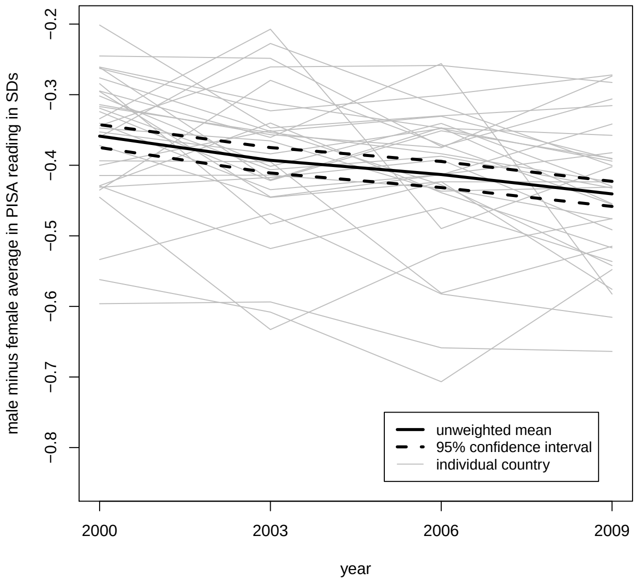


Figure 4.6: Average PISA performance of males relatives to females in reading: 2000-2009



deviations from 0.148 standard deviation. Although these figures seem to suggest that girls' average performance in high school increased relative to boys over the period,

DOI: 10.2870/93297

they should be interpreted with care. First of all, there is the usual sample restriction issue since 10th graders in the US are already beyond the end of compulsory schooling. Therefore, if high school dropout rates differ between boys and girls and have changed between 1980 and 2002, which is likely, changes in average performance of 10th graders will give a biased estimate of the entire population. Second, one should keep in mind that such evidence is restricted to the US. Data constraints do not allow to repeat a similar exercise for other countries over the same period.

Table 4.5: Gender Difference in Average Test Score at Age 15: 1980 Vs 2002

	1980			2002		
	Female Mean	Male Mean	F-M Diff. in S.d.	Female Mean	Male Mean	F-M Diff. in S.d.
Mathematics Score	49.41	50.62	-0.121***	49.49	50.56	-0.107***
Reading Score	50.29	49.72	0.057***	50.77	49.29	0.148***
Composite Score	49.52	50.50	-0.097***	50.13	49.94	0.019

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Standards errors are bootstrapped.
Sources. High School and Beyond 1980 and US Educational Longitudinal Study 2002.

4.4 Conclusion

In this chapter, we have formulated the two main competing hypotheses in the same framework as the one developed for our theory. This allowed us to assess the predictions of alternatives explanations against our theory in a common setting. We have shown that alternative theories to the gender gap reversal in education appear to be inconsistent with the data. On the other hand, our theory is not dismissed by the exact same tests. Although we cannot exclude joint hypotheses as explanations behind the gender gap reversal, it appears that a change in social norms or an increase in females' mean performance alone cannot explain the patterns observed in the data.

5

Do Peers Matter in School Performance and How? Evidence from a Finnish High School Quasi-Experiment

5.1 Introduction

The question of how students are affected by their peers is an important one for education policies, and the answer remains to be fully understood. We are both interested in how school choice affects a single student's results as well as the distributional effects of the whole student allocation on all the students' results. One of the fundamental reasons for our relative lack of understanding in this field is that educational systems and the process of education are complex systems that rarely allow experimentation. In addition, the fact that different schooling systems in different countries and regions have distinct properties makes comparisons difficult.

One of the properties of the Finnish high school system is a nation-wide application system that matches the preferences of the applicants and preset student quotas for the schools in a centralized manner. This system offers a quasi-experimental regression discontinuity (hereafter RD) design, which I exploit in this paper using data on the whole universe of Finnish high school students spanning nine years to isolate and study the causal effects of high school choice on high stakes high school exit exam performance. I also study the relationship of these causal effects with the composition of the students' peers. In Finland, expenditures per student vary only moderately between high schools. This, along with the fact that Finnish 9th grade students are known to be among the brightest on the planet in international comparisons, makes the setup exceptionally intriguing.

In the first part of the analysis, I take an ancillary step in which I study the general implications of the Finnish high school application system. I find that, in general, being just above the required selection threshold to their school of first preference has a negative causal effect on student performance, although on average one gets peers that are clearly better according to their pre-high school grades compared to the peers one would get in the counterfactual case of being just below the threshold. In the next step, I study each school by year observation as a separate quasi-experiment. This yields more than 300 valid RD designs, which allow the study of the relationship between the effects on outcomes of school choice and characteristics of peers in those schools.

I answer two questions in this paper. The first is simply whether school choice actually has an effect on exam success at the end of the high school. The negative effect at the margin found in this paper differs from the majority of the literature.

The second question I tackle in this paper is whether these effects stem from a particular characteristic of the peers. In the Finnish high school system school results are positively associated with peer homogeneity rather than average peer quality. More precisely, I find that a standard deviation change in the homogeneity of peers is positively associated with a 0.02 to 0.13 standard deviation change in the exam results. The main contribution of the paper is the finding that in Finland it is the homogeneity of the class that seems to matter for the students at the entrance threshold. This result complements findings in other comparable studies.

According to the second result, average peer quality does not improve exam outcomes. The hypothesis laid out in this paper is that the students that aim high on average behave overconfidently in their school choices. The negative effect at the margin could stem from the fact that the level of teaching is too high for those students who just make it to their preferred school versus those who are just below the threshold and thus end up with worse-performing peers on average. Teachers teach to the median student and thus those at the entrance threshold don't get the optimal teaching.

In the literature, there are some previous examples of quasi-experimental designs being used to study school performance, whether at the university level, the high school level or an earlier stage of schooling. An early example of such a design is by Sacerdote (2001), who exploits a randomized allocation of students to dorm rooms. He finds that peers matter at the college level, and that better-performing peers improve school performance. (15) examine a quasi-experimental setup, where students are allocated to high schools partly according to a lottery. They find little impact on academic outcomes, but some effect on other social outcomes.

The RD design has been used by some authors. (17) study an actual experiment, where a number of first grade pupils were allocated to classes according to exam scores in Kenya. They find no effect of being allocated with higher-achieving students compared to those who were allocated with lower-achieving students. (13) examines the effect of attending selective schools from the sixth to the ninth grade. Clark discovers a small positive effect on exam scores and a stronger effect on long-run outcomes such as university attendance. (36) study intention-to-treat effects in high schools. They find a positive effect of attending a school with better-achieving peers. (26) runs an RD analysis on a flagship university and discovers a positive income effect of being treated with attending a particular flagship university versus a control group that did not attend that university.

The rest of the article is structured in the following way. Section 2 describes the institutional setting and the data. Sections 3 and 4 discuss the methodology and establish the validity of the quasi-experiment. Section 6 presents the results, and Section 7 is the conclusion.

5.2 Institutional Setting and Data Description

In Finland, there are two types of schools that provide secondary education, upper secondary schools[1] and vocational schools. The application to both is conducted through the same centralized process. The Finnish National Board of Education, which is an agency that is subordinate to the Ministry of Education, conducts the automated allocation.

Students submit a list of five ranked choices usually at the end of their ninth grade of comprehensive schooling, which typically takes place in the year during which they turn 16 years old. The selection itself is based on the announced preferences and students' GPAs[2] according to the announced available slots in the high schools. This ensures that the entrance threshold for a particular school changes yearly and is unpredictable to the student. I define the threshold to be the level of GPA where the jump in the treatment proportion is maximized. Some schools also have another threshold, which is based on GPA and other points, which can usually be gathered via hobbies. Any applicant that surpasses either of these thresholds is eligible. These schools are also included by finding both criteria[3].

[1] I use the terms upper secondary school and high school interchangeably.

[2] The relevant GPA is the arithmetic mean of the following subjects: mother tongue and literature, the second national language, foreign languages, religion or education on ethics and moral history, history, citizenship education, mathematics, physics, chemistry, biology, health education and geography. Grades run between 4 and 10. 4 means fail.

[3] However, only the academic GPA is used for purposes of describing peer characteristics.

 DOI: 10.2870/93297

The choice is made randomly for those applicants that have the same GPAs to the 2nd decimal. If an applicant is awarded a place at any of the preferences, all the lower ranked preferences become obsolete since applicants are not allowed to refuse a place in a school in order to opt for a lower choice. In this paper, an applicant to a high school is defined to be anyone who has that particular high school as the first choice.

The schools, however, do have some discretion in taking in a part of their students according to other criteria. This allows them to choose some of the students with GPAs below the official threshold. Although this discretion is not used extensively, it slightly complicates the analysis, and I will take this into account in the later stages of this study. Furthermore, if the authority that organizes the schooling, such as the municipality, has multiple schools or tracks to offer, and a student is not admitted to one of them, the organizer is allowed to offer the student a place in another school or track.

During the years under scrutiny, the majority of Finland's more than 300 high schools were owned by the local municipalities. Under a tenth of the schools were run by private owners and 10 schools were owned by the state. Municipal schools do not charge for the schooling. Private schools may charge small fees, but high schooling is not allowed to be organized for profit. Funding is granted by the state and municipalities based on the number of students. Expenditures per student do vary somewhat between schools[1].

In upper secondary schools, students take courses according to their individual study plan and aim to graduate in two to four years[2]. Upper secondary school and some vocational school tracks conclude with a matriculation examination. The outcome of interest in this study is the matriculation examination score in the Finnish language test, since this test is taken universally by all pupils with Finnish as their mother tongue. Also, the schools are considered at the school level without separating between tracks within schools. This is necessary, since the data only give the school of graduation and not the particular track. This should not affect the results much, since most schools analyzed have only one track.

The matriculation examination is organized twice a year. The student can attend up to three examination periods. Thus, the student can spend up to 1.5 years for the whole matriculation examination. If the student fails one or more of the exams, he or she can make even further attempts to achieve a passing grade. The candidate has to complete at least four tests, which make up the matriculation examination. Until 2005, the compulsory tests were the mother tongue (usually Finnish), the second national

[1]Unfortunately there is data available on the Finnish schools' expenditures per student only at the ownership level. Since for example the city of Helsinki owns many high schools, we do not know what is the expenditure per student in each individual school is.

[2]Read (30) for a more comprehensive description of Finnish high schools.

language (usually Swedish), a foreign language (usually English) and either mathematics or general studies. Since 2005, the only compulsory test has been the candidate's mother tongue. The tests are graded nationally by the Matriculation Examination Board.

The students are given a verbal grade with seven different values. The grades are normalized nationally within each matriculation period so that ideally an equal share of students would get a particular grade[1]. For numerical equivalents of the verbal grades, I will use the same convention as the Finnish National Board of Education does for some other purposes. The verbal grades are assigned values from 7 to 0 with grade 1 missing and 7 being the best one. In most tests, with the notable exception of the mother tongue test, the student can choose whether to take the A-level, B-level or C-level test, A-level being the most advanced[2].

The main part of the data used in this study is combined from two administrative registers. The application data for the years between 1996 and 2004 consist of a dataset collected by the the Finnish National Board of Education. These data reveal among other things the applicants' ranked high school preferences, GPAs, the final school choice, gender, school of origin and home address. The number of first choice non-vocational high school applicants for these nine years is about 300,000, or about 33,000 per year. The other part of the data consists of the matriculation examination outcomes, which are collected by the Matriculation Examination Board. These data cover the years between 1990 and 2010 and have grade, time and place of all the matriculation exams taken by each student. To assign each student to a particular school, I will use the matriculation exam data. This is due to two reasons. First, if these two differ and a student has switched schools during their high school studies, it is likely that the school where the pupil graduated and thus probably also did his or her preparatory courses for the matriculation exams is more influential in defining their exam outcomes. Secondly, the data about the initial school assignment appear to be missing or corrupt for some of the schools for some years.

Only the largest schools with a minimum of 600 first applicants in the combined nine years are used in this study to ensure large enough sample size for the estimates. The data and the relevant variables are described in Table 5.1. Given the strict criteria I set for a valid year by school quasi-experiment, around 30000 pupil observations in 326 different school by year observations are used for the main analysis.

[1] The grades from top to bottom and the target shares are: laudatur (5%), eximia cum laude approbatur (15%), magna cum laude approbatur (20%), cum laude approbatur (24%), lubenter approbatur (20%), approbatur (11%) and improbatur (5%, a fail)

[2] Statistics and other information about the matriculation examination can be found in the statistics book "Finnish Matriculation Examination 2007" (1).

5.3 Empirical Methodology

The Finnish schooling system does not offer a fully randomized setup as e.g. (37) is able to exploit. However, the RD nature of the Finnish high school process gives us a close approximation of a randomized experiment at the entrance threshold.

The theory, as laid out by (24) and (6), states that an RD design arises when the treatment status depends on an observable characteristic S of the agent, where S stands for the score. In this case S is the GPA, and the preference ranking of schools for those with GPAs at the threshold. A second requirement is that the treatment status of the agents or students has to jump discontinuously at a known cutoff point.

In the case of a *sharp RD design* all the subjects above the cutoff take the treatment and below it none of them do. Let Y_0 denote the outcome in the case of no treatment and let $\bar{s}$ be the cutoff so that superscript minus and plus signs denote the values right below and above it. In this rare case the only condition (Condition 1) for the identification of the treatment effect is that in the counterfactual world, where no one takes the treatment, there would be no discontinuity in the outcomes at the cutoff, i.e.

$$E\{Y_0|\bar{s}^+\} = E\{Y_0|\bar{s}^-\}.$$

If this condition holds, we are able to identify the local average treatments effect (LATE) for the individuals immediately above the cutoff, since locally we have a situation where treatment is assigned randomly with full compliance. If the same condition holds true for Y_1, the outcome with treatment, we can identify the mean impact of extending the programme to marginally excluded individuals.

In a *fuzzy RD design* the participation rate jumps by less than one. Some of the subjects below the cutoff might participate and some of them above the cutoff might not. Two additional conditions are required for this framework to achieve a clean identification of the treatment effects. One additional condition (Condition 2) says that there is no direct effect of S on the outcome for a fixed treatment status in the neighborhood of the cutoff and that assignment around the cutoff is as randomized. Thus, if Y_1 is the outcome with treatment, and I(s) is the binary treatment status, the triple $(Y_0, Y_1, I(s))$ needs to be stochastically independent of S in a neighborhood of $\bar{s}$. The other additional condition (Condition 3) is that there are no defiers, i.e. there are no subjects who would take the treatment below the cutoff and not take it above the cutoff. If the above conditions hold, we can identify the LATE. Most years in most schools studied in this paper do not offer a sharp, but a fuzzy setup. I discuss the RD design in more detail in the next subsection.

An RD design is not a method, rather it is a description of a data generating process, as (31) note. Once the RD setup is confirmed to be valid, the method is a rather straightforward means comparison or a regression analysis. The setup becomes statistically reminiscent of a controlled experiment.

Since each school for each of the 9 years offers a potential quasi-experiment, one can be very selective ex-ante about the necessary conditions for inclusion of that particular school/year in the study. I study only schools that have at least 600 first applicants in all the 9 years combined. This leaves us with 184 schools. Also, in my main specification, I require there to be at least 20 control subjects and 30 students in the treatment group within an interval of 0.6 9th grade grades of the threshold. This is to ensure a decent signal to noise ratio in the estimates.[1]

The statistical method itself used for the estimates is the following regression :

$$y_{i,t,k} = \alpha + h_k(S_i) + x_i'\beta + x_{2,k}'\beta_2 + x_{3,i,k}'\beta_3 + \delta_{t,k} I(S_{i,k} \geq \bar{s}_{t,k}) I(Enroll_i = t) + u_i,$$

where $y_{i,t,k}$ is the dependent variable, such as the matriculation examination result, $i \in \{1, 2, ..., n\}$ is the pupil index for the population of first applicants to a particular high school, $t \in \{1996, 1997, ..., 2004\}$ is the year of enrollment and $k \in \{1, 2, ..., K\}$ is the school of graduation. $h_k(S_i)$ is the control function, a fifth degree polynomial of GPA, which is different for each school[2]. x_i' is a vector of yearly controls, $x_{2,k}'$ is a vector of school controls and $x_{3,i,k}'$ is a vector of school by year controls. $I(\cdot)$ is an indicator function, $\bar{s}_{t,k}$ is the threshold for that school for that year and $Enroll_i$ stands for the year of enrolment of the individual. The parameter of interest is $\delta_{t,k}$, which is the estimate of causal effect for the given year for the given school. For these regressions, I combine the data for all the analyzed schools and years. Then, I normalize the GPA to zero at each year by school threshold.

In a first step ancillary of the analysis, I study the whole sample for general properties of the quasi-experiment. More importantly, though, I estimate and extract school by year estimates of the effects on the matriculation examination and peer quality of eligibility for the first choice school using the above regression. In the second step, I study the statistical association between school effects and the statistical properties of the peers.

[1]The idea behind this restriction is that in the second step of the analysis, the school by year observations are considered individual quasi-experiments. Low sample size would induce measurement error and would attenuate the second step estimates. The band around the threshold and other specifications are varied later for robustness checks.

[2]The reasoning behind a separate control polynomial for each school is that the populations are different, with a band around a different threshold. Thus it is appears relevant to control for a different polynomial of the score variable.

 DOI: 10.2870/93297

5.4 Establishing the Validity of the Randomization

In a valid regression discontinuity design, the treatment proportion has to jump at the threshold. Table 5.2 and Figure 5.1 illustrate the jump in the treatment proportions below and above the threshold as estimated with the above method. The jump is about 0.84 on aggregate for the whole sample. There are some treated subjects below the threshold and above some of the subjects do not take the treatment.

The first necessary condition for identification in a fuzzy RD i.e. the condition $E\{Y_0|\bar{s}^+\} = E\{Y_0|\bar{s}^-\}$ is not testable, since we do not observe $E\{Y_0|\bar{s}^+\}$. However, we can test the continuity of other pre-treatment covariates. The remaining three diagrams in Figure 5.1 and columns in Table 5.2 show that non-academic grades, gender and a dummy variable about whether the pupil's home municipality is the same as the the municipality of the school are all continuous at the 5% significance level. In aggregate, the variables are continuous at the threshold. Although some of these variables are statistically significantly non-continuous for some individual school by year observations, this appears to be a result of randomness, not systematic.

There are a couple possible reasons as to why the first condition would not hold. One possibility is the presence of "gaming" at the cutoff, i.e. manipulating the GPA to get just above the cutoff. This is possible but unlikely, since the annual intake quota is a preset number[1] and the number of applicants varies between years. In our sample, the average threshold is 8.2 on a scale from 4 to 10. The lowest threshold in the sample is 7.2 and the highest threshold is 9.36. The average standard deviation of the threshold between different years for our sample is 0.18. Even if a pupil were able to manipulate their 9th grade GPA to ensure that they qualify to the high school of their choice, they would have to aim clearly above the previous year's threshold to ensure selection. Since we are interested in the effect at the threshold, their impact in our estimates should be minimal. The kernel density estimate of applicants according to normalized GPA is shown in Figure 5.3. There doesn't seem to be anything out of the ordinary in the distribution of applicants just above the threshold.

Another reason why Condition 1 might not hold is the presence of attrition. Attrition in this setup could arise from a selected part of the students just above or below the threshold not participating in the matriculation examination. This would bias the sample at the baseline.

An obvious feature of the setup is the fact that the schools are allowed to handpick a small portion of the students after the official application process and also some

[1] The quota itself does not vary much annually. The exact number, however, is not typically observed by the applicant.

individuals who were eligible for the treatment, actually graduate from another high school. Due to these reasons, the DGP is not a sharp RD, but instead it is a fuzzy one. As mentioned above, that increases the requirements for a valid quasi-experiment. The above-mentioned third necessary condition for a fuzzy RD that requires there to be no defiers, appears to be a valid assumption in our setup, especially since we only consider the candidates that applied to that particular school as their first choice.

The second condition imposes that assignment around the threshold take place as if randomized. There is no guarantee that this condition holds, since one might expect that the students who switch out of the school of first choice are not a random sample. Even more strongly, the students that find a way to take the treatment even when not assigned are most likely not a random selection. I deal with this problem in the following way. Since I cannot be sure about the second condition, I will call my estimates intention to treat (ITT) effects and will not correct for the fact that not all of the assigned subjects were actually treated. Thus, the causal estimate I get is simply the effect of being eligible for treatment. In other words, I estimate the effect of being eligible to attend the school of first choice, when the control is having to attend a school of one of the lower choices.

5.5 Results

In the ancillary step, I estimate the causal effects of school choice on peer composition and exam outcomes. I use the same regression discontinuity setup as when estimating the continuity of covariates. The treatment is described according to the composition of the class in terms of ninth grade GPAs. This is done by comparing the standard deviations, mean 9th grade GPAs and mean 9th grade GPAs of the top and bottom decile of the high school year. The absolute level of the threshold is of interest as well. The outcome is the Finnish language mother tongue matriculation examination.

The estimation is exactly to the estimation of pre-treatment continuities. As above, the results of these estimations are shown in Table 6.2 and Figures 6.2 and 5.4. As the table shows, the causal effect estimates for the Finnish mother tongue exam as shown in the first column are significant but negative. Being eligible for attending the school of first preference at the threshold is associated with a 0.01σ to 0.12σ change in matriculation exam performance at the 0.95 confidence level. All the other four columns show significant estimates for the composition of the class in terms of 9th grade GPAs in the expected direction. These are the pre-treatment descriptions of the pupils. Thus on average, when the pupils that are on the margin get to the school of their first choice, they get better and more homogeneous peers, but they do slightly worse than

Kanninen, Ohto (2013), Five essays on economics of education
European University Institute

DOI: 10.2870/93297

by attending the control schools. This is a surprising result and clearly in contrast with a large portion of the literature.

This result could stem from the fact that teachers aim their teaching towards the median or average students. As the students who just make it to their preferred school are at the bottom of the class according to 9th grade GPAs, they end up worse off on average than the comparable students who attend a school lower in their preferences. Assuming that the students choose particular schools in order to achieve the best possible matriculation examination results, the working hypothesis put forth here is that at least a subset of the students who make their school choices behave in an overconfident way.

At the ancillary step, I extract the estimates for the ITT causal effect of high school selection on Finnish language mother tongue exam and the four variables describing the peers for each of the 326 year by school observations that have at least 20 control subjects and 30 subjects in the treatment group within the band of 0.6 grades from the threshold. Table 6.4 shows the descriptive statistics for these estimated observations.

In the second step, the causal effect estimates are compared to the estimated class composition variables. Table 6.5 shows the correlations between these 5 variables and the absolute level of the entrance threshold for the school by year observation. The class homogeneity as measured by the standard deviation of the 9th grade GPAs of the class and the mean 9th grade GPA of the top 10% of the peers have the strongest correlation with the estimated causal effects. Mean GPA of peers is highly correlated with the other class composition variables and the entrance threshold level.

Figure 6.5 shows the scatter plots of the estimates for the causal effect on exam performance with respect to the other five variables. Evidently, the four variables describing the statistical properties of the class composition seem to have some outliers, the relationships seem quite linear and class homogeneity as proxied by peer GPA standard deviations seems to be relevant in explaining the effects. The black dots are significant estimates at the 0.1 significance level. 17.2% of the observations are significant at the 0.1 level.

Since the data is pooled from nine different years and 184 schools and there is an unbalanced number of observations from each year and school, one should control for each year effects to sift out the true estimate for the linear relationship between the variables. Table 6.6 shows the results in regression analysis framework. The table shows the results with and without the yearly control dummies. The lower half of the table shows the results from the exact same analysis with weights in the second stage for the size of the control group. This is to correct for possible inaccurate estimates of the first stage with small control groups. It appears that of the class composition variables, class homogeneity as measured as the standard deviation of 9th grade peer GPAs is the most significant variable.

The estimated significant effect of -0.53 means that a σ change in the standard deviations of the class 9th grade GPAs is associated with a 0.02σ to 0.13σ change in matriculation exam performance at 0.95 confidence level. Interestingly, peer mean GPA doesn't have a significant statistical association with the causal effects in any of the specifications.

Table 6.7 indicate regressions where multiple class composition variables are inserted as explanatory variables. These regressions show that indeed class homogeneity seems to have a robust statistical association with the causal effect even when controlled for other class composition variables, many of which are highly correlated with one another as shown in Table 6.5. The high correlations make it difficult to discern the statistical effects.

Tables 7.1, 7.2, 7.3 and 7.4 show results from robustness checks with a band of 0.2 and 0.6 around the threshold with different third, fifth and seventh degree control polynomials. The results stay unchanged with respect to class homogeneity especially once we use the fifth and seventh degree specifications in the control polynomial. However, in some specifications of the robustness check as shown in the appendix peer mean quality becomes positive and significant as an explanatory variable to exam performance.

5.6 Conclusions

The aim of this paper was to assess whether peer effects in high school are significant in defining success in high stakes school exit exams in Finland. Although Finland offers a relatively egalitarian setting for high school students at least in terms of expenditures per student and student quality before high school, the result from this paper is that school choice does matter and that it appears to stem partly from class composition. Surprisingly, the evidence suggests that on average the students who are at the threshold are worse off getting in to their favored school than the control group that just missed their first preference, although the former group gets better-achieving and more homogenous peers on average. I propose the working hypothesis that the high-aiming applicants might actually be overconfident in the application process. They apply for a school that will be ultimately harmful for them, since the teaching is aimed at students that are on average better-achieving.

The main contribution of this paper, however, is to show that the homogeneity of the class is robustly associated with the exit exam results of the students. The association between results and average peer quality so often explored is rather more vague in the

DOI: 10.2870/93297

Finnish setting. A standard deviation change in the homogeneity of peers is positively associated with a 0.02 to 0.13 standard deviation change in the exam results. The association between results and homogeneity is little explored and thus these results suggest avenues for future peer effect research.

Table 5.1: Descriptive statistics.

	Mean	S.D.	Min	Max	n
Individual level					
9th grade GPA	8.23	1	4	10	205696
Finnish mother tongue matriculation exam taken	0.87	0.33	0	1	205894
Finnish mother tongue matriculation exam grade	4.39	1.26	0	7	179966
Mean 9th grade GPA of high school peers	8.26	0.37	7.02	9.65	182257
Standard deviation of 9th grade GPA of high school peers	0.73	0.13	0.19	1.52	182255
Mean 9th grade GPA of top 10% high school peers	9.43	0.26	7.38	10	182257
Mean 9th grade GPA of bottom 10% high school peers	6.99	0.54	5.23	9.42	182257
Sex (female = 1)	0.56	0.50	0	1	205892
Municipality of residence same as municipality of school	0.44	0.50	0	1	161116
GPA of non-academic grades: music	8.54	0.88	5	10	58063
GPA of non-academic grades: household care	8.42	0.83	5	10	58069
GPA of non-academic grades: visual arts	8.48	0.84	5	10	58166
GPA of non-academic grades: handicraft	8.40	0.81	4	10	58065
GPA of non-academic grades: physical education	8.58	0.95	4	10	58504
School level					
Applicants	1118.99	434.19	600	2534	184
School by year level					
Applicants	124.71	54.19	25	418	1651
Applicants for valid quasi-experiments	181.13	56.46	85	418	326
No. of years	4.13	2.26	1	9	326

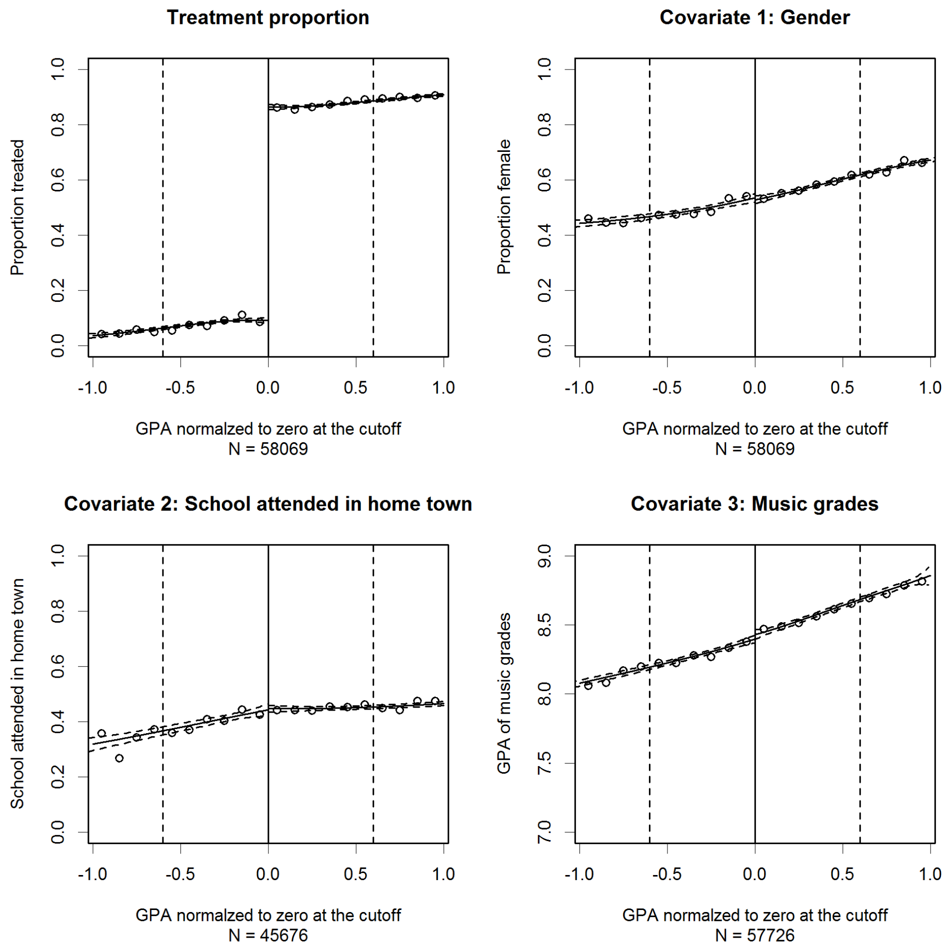
Note: This table describes those Finnish high school applicants that applied in the years 1996–2004 to the sample of Finnish speaking schools with 600 or more first applicants in aggregate for those years.

Table 5.2: Treatment proportion and continuity of pre-treatment variables.

Dependent variable	Treatment status	Gender (female = 1)	School of matriculation exam in home town	Non-academic grades: music
$I(S > Thres)$	0.843***	-0.027*	0.022	0.023
	(0.009)	(0.016)	(0.013)	(0.027)
n	30478	30478	26034	30119
Dependent variable	Non-academic grades: household chores	Non-academic grades: visual arts	Non-academic grades: handicraft	Non-academic grades: physical education
$I(S > Thres)$	0.024	-0.011	0.053**†	0.119***††
	(0.024)	(0.026)	(0.025)	(0.029)
n	30131	30179	30138	30332

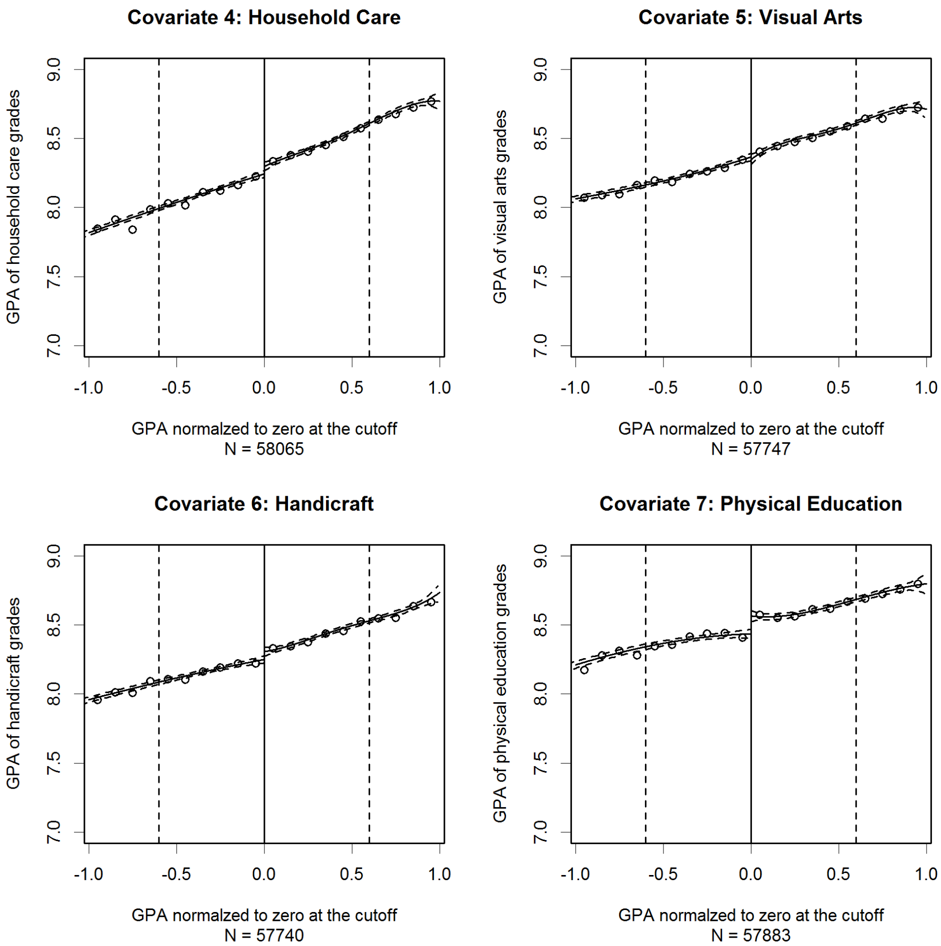
Note: Significance level: *** 0.01, ** 0.05, * 0.1. Each column is a separate regression with a different dependent variable. The control variables are school fixed effects, yearly fixed effects, school by year fixed effects and a school-specific fifth degree polynomial control polynomial of score variable. The key independent variable is a dummy for whether the student was eligible to attend the school of his or her first choice.
†An estimate of 0.053 means that being eligible for attending the school of first preference at the threshold is associated with a 0.03σ to 0.11σ increase in 9th grade handicraft score at the 0.95 confidence level.
††An estimate of 0.119 means that being eligible for attending the school of first preference at the threshold is associated with a 0.07σ to 0.19σ increase in 9th grade physical education score at the 0.95 confidence level.

Figure 5.1: Treatment status of applicants and continuity of covariates.



Note: The dashed line is the 95 % confidence band. The horizontal dashed lines represent the band used in the main regressions. The fitted line represents two separate polynomial functions of the score variable for each side of the threshold. The treatment status jumps by an estimated 84% at the threshold. The pre-treatment variables are continuous at the 95% confidence level.

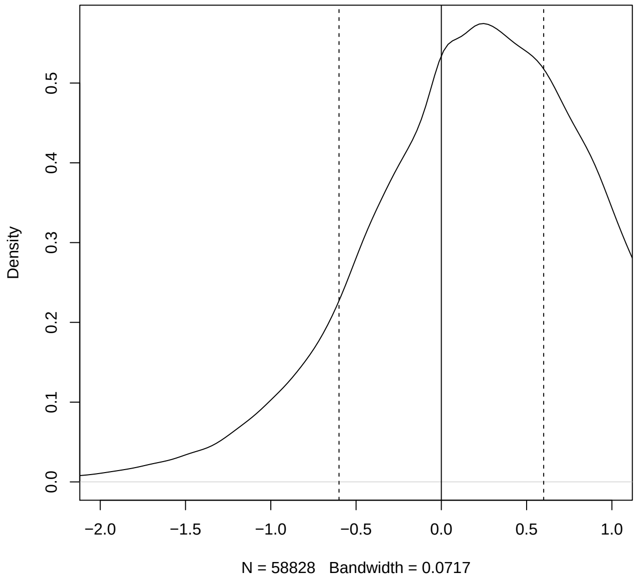
Figure 5.2: Treatment status of applicants and continuity of covariates.



Note: The dashed line is the 95 % confidence band. The horizontal dashed lines represent the band used in the main regressions. The fitted line represents two separate polynomial functions of the score variable for each side of the threshold. The pre-treatment variables are continuous at the 95% confidence level.

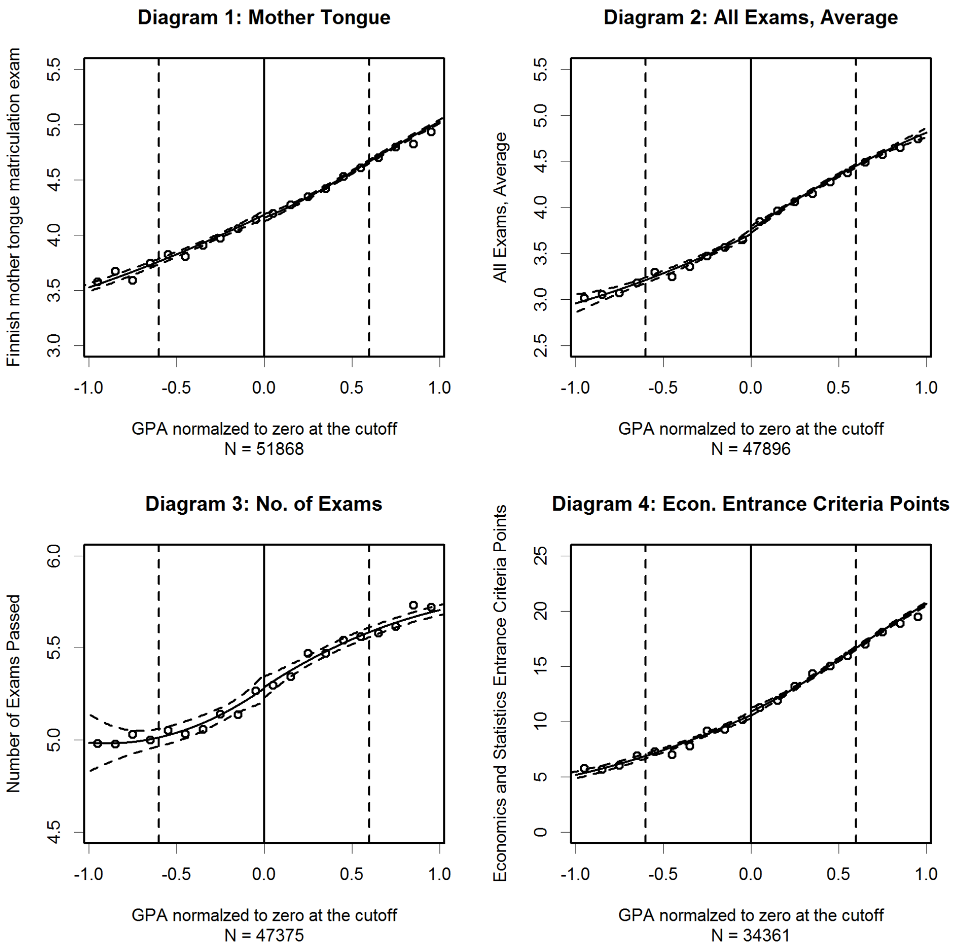
DOI: 10.2870/93297

Figure 5.3: Kernel estimate of the density function of first applicants and the threshold.



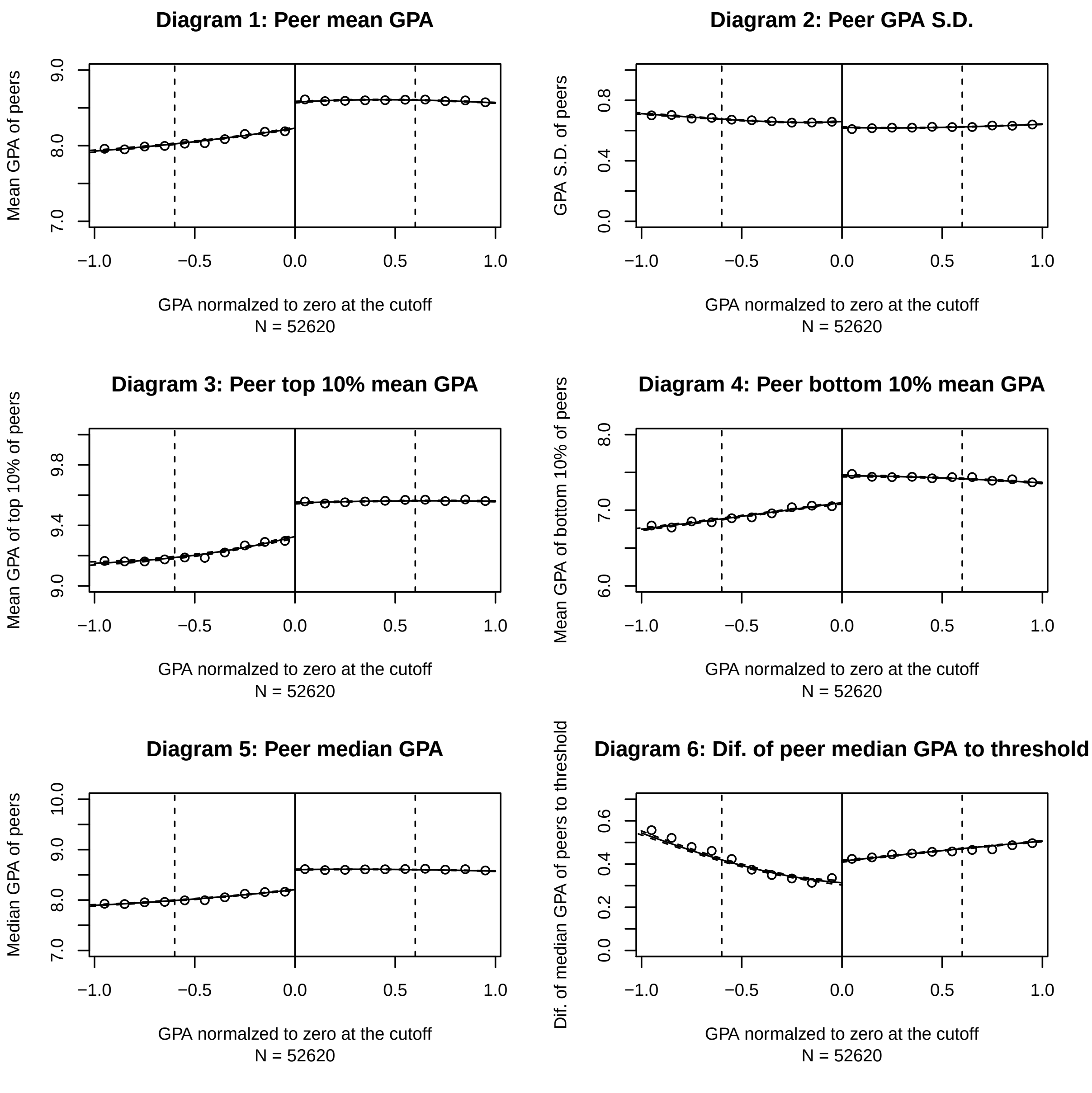
Note: The function used to estimate the density function is the Gaussian kernel function. The bandwidth used is the default one in R statistical package and suggested by Silverman's "rule of thumb"(38). The x-axis depicts the GPAs normalized to the annual threshold level.

Figure 5.4: Matriculation examination performance.



Note: The y-axis depicts the exam performance of the students in the Finnish language matriculation examination. The x-axis depicts the GPAs normalized to the annual threshold level. The dashed line is the 95 % confidence band. The horizontal dashed lines represent the band used in the main regressions. The fitted line represents two separate polynomial functions of the score variable for each side of the threshold. With the main specification, the matriculation examination results are statistically significantly worse above the threshold than below.

Figure 5.5: Description of treatment versus control in terms of class composition variables.



Note: The y-axes depict the statistical properties of the GPA of the year by school peers of each student. The x-axis depicts the GPAs normalized to the annual threshold level. The dashed line is the 95 % confidence band. The horizontal dashed lines represent the band used in the main regressions. The fitted line represents two separate polynomial functions of the score variable for each side of the threshold. All these class composition variables are non-continuous at the threshold for the whole sample.

DOI: 10.2870/93297

Table 5.3: Estimated effects from being eligible for treatment.

Panel A: Peer Description

	Peer 9th grade mean GPA	Peer 9th grade GPA standard deviation	Peer 9th grade top 10% mean GPA	Peer 9th grade bottom 10%
$I(S > Thres)$††	0.396***	-0.044***	0.245***	0.386***
	0.008	0.003	0.007	0.011
n	30349	30349	30349	30349
	Peer Median GPA	Distance between Peer Median GPA and Threshold		
$I(S > Thres)$†††	0.426***	0.100***		
	0.009	0.006		
n	30349	30349		

Panel B: Outcome Variables

	Mother tongue, Finnish	All exams, average	Number of exams passed	Economics and statistics entrance criteria points
$I(S > Thres)$	-0.078**†	0.040	0.019	-0.155
	(0.034)	(0.029)	(0.044)	(0.272)
n	30478	30478	30450	23937

Note: Standard errors in parentheses. Significance level: *** 0.01, ** 0.05, * 0.1. Each column is a separate regression with a different dependent variable. The control variables are school fixed effects, yearly fixed effects, school by year fixed effects and a school-specific fifth degree polynomial control polynomial of score variable. The key independent variable is a dummy for whether the student was eligible to attend the school of his or her first choice.
†An estimate of -0.078 means that being eligible for attending the school of first preference at the threshold is associated with a 0.01σ to 0.12σ change in matriculation exam performance at the 0.95 confidence level.
††The estimates in this column mean that being eligible for attending the school of first preference at the threshold is associated with a 0.81σ to 0.87σ, 0.28σ to 0.37σ, 0.71σ to 0.80σ and 0.55σ to 0.62σ change in the respective class composition variables at the 0.95 confidence level in order from first to last column.
†††The estimates in this column mean that being eligible for attending the school of first preference at the threshold is associated with a 0.82σ to 0.89σ and 0.38σ to 0.48σ change in the respective class composition variables at the 0.95 confidence level in order from first to last column.

Table 5.4: Descriptive statistics, second step.

	Mean	S.D.	Min	Max	n
Estimated causal effect	-0.06	0.43	-1.13	1.51	326
Entrance threshold level	8.14	0.40	7.20	9.36	326
Difference in peer mean 9th grade GPA	0.38	0.21	-0.25	0.94	326
Difference in peer top 10% mean 9th grade GPA	0.23	0.19	-0.53	0.74	326
Difference in peer bottom 10% mean 9th grade GPA	0.38	0.37	-0.72	1.45	326
Difference in peer 9th grade GPA standard deviation	0.38	0.37	-0.72	1.45	326
Difference in peer 9th grade median GPA	0.41	0.24	-0.44	1.16	326
Distance between peer median GPA and Threshold	0.11	0.20	-0.39	0.71	326

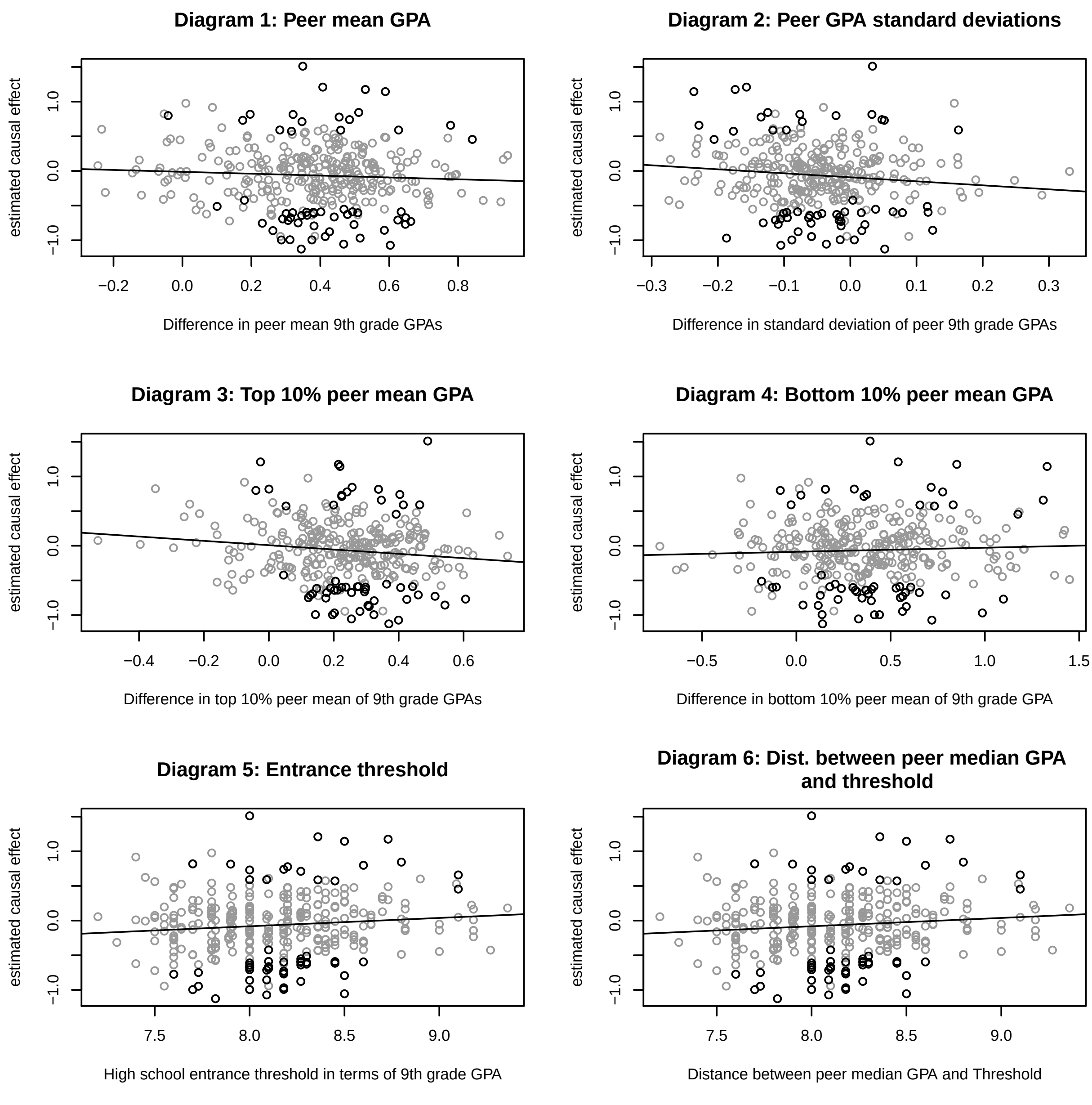
Note: This table gives the descriptive statistics for the estimates of the year by school causal effects at the margin of school eligibility on Finnish language mother tongue matriculation examination and four class composition variables. In addition it describes the entrance threshold levels. All the class composition variables range from negative to far in the positive values.

Table 5.5: Correlations of class composition variables, entrance thresholds and the estimated causal effects. Band of 0.6, at least 20 controls.

	Dif. in peer mean 9th grade GPA	Dif. in peer 9th grade GPA standard deviation	Dif. in peer 9th grade top 10% mean GPA	Dif. in peer 9th grade bottom 10% mean GPA	Dif. in peer 9th grade median GPA	Distance between Peer median GPA and Threshold	Entrance threshold	Estimated causal effect on test performance
Dif. in peer mean 9th grade GPA	1.00	-0.37	0.76	0.75	0.87	0.15	0.47	-0.07
Dif. in peer 9th grade GPA standard deviation	-0.37	1.00	0.18	-0.81	-0.21	-0.09	-0.16	-0.12
Dif. in peer 9th grade top 10% mean GPA	0.76	0.18	1.00	0.36	0.70	0.14	0.30	-0.13
Dif. in peer 9th grade bottom 10% mean GPA	0.75	-0.81	0.36	1.00	0.57	0.11	0.38	0.05
Dif. in peer 9th grade median GPA	0.87	-0.21	0.70	0.57	1.00	0.24	0.39	-0.10
Distance between Peer median GPA and Threshold	0.15	-0.09	0.14	0.11	0.24	1.00	-0.41	-0.10
Entrance threshold	0.47	-0.16	0.30	0.38	0.39	-0.41	1.00	0.11
Estimated causal effect on test performance	-0.07	-0.12	-0.13	0.05	-0.10	-0.10	0.11	1.00

Note: The school by year sample size is 326. Each cell represents the correlations between the column and the row variable. These correlations arise from the second step where class level variables are studied so that each school by year causal effect of school selection on matriculation examination and class composition variables is estimated separately. Thus each individual data point here means the effect of attending the school of first choice versus attending a mix of other schools for the students who are near the threshold. The table gives the results of a simple correlation between these variables and also the entrance threshold level of that particular school by year observation.

Figure 5.6: Peer description and causal effect estimates.



Note: 326 data points in each figure. These estimates are extracted from the main specification with a fifth degree polynomial and with at least 20 school by year control subjects and a treatment group of at least 30 treated subjects at a 0.6 band from the threshold. The black dots are estimates of the causal effect of the treatment school that are significant at the 0.1 significance level. 17.2% of the estimates are significant at the 0.1 significance level, 10.4% at the 0.05 significance level and 3.7% at the 0.01 significance level.

Table 5.6: Regression analysis, second step.

Panel A: Non-weighted

	Entrance threshold		Peer 9th grade mean GPA		Peer 9th grade top 10%		Peer 9th grade bottom 10%		Peer 9th grade GPA standard deviation	
	0.12**	0.1*†	-0.14	-0.12	-0.31**	-0.29**††	0.06	0.06	-0.58**	-0.53**†††
	(0.06)	(0.06)	(0.11)	(0.12)	(0.13)	(0.13)	(0.07)	(0.07)	(0.28)	(0.27)
Year fixed effects		Yes		Yes		Yes		Yes		Yes
Fifth degree control polynomial in the ancillary step	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Peer 9th grade median GPA		Distance between peer median GPA and entrance threshold							
	-0.19	-0.16	-0.22	-0.21						
	(0.10)	(0.10)	(0.12)	(0.13)						
Year fixed effects		Yes								
Fifth degree control polynomial in the ancillary step	Yes	Yes	Yes	Yes						

Panel B: Weighted

	Entrance threshold		Peer 9th grade mean GPA		Peer 9th grade top 10%		Peer 9th grade bottom 10%		Peer 9th grade GPA standard deviation	
	0.11**	0.09	-0.03	-0.01	-0.2	-0.18	0.08	0.08	-0.47**	-0.48**
	(0.06)	(0.06)	(0.11)	(0.11)	(0.13)	(0.13)	(0.06)	(0.06)	(0.23)	(0.24)
Weighted by the control group size in the ancillary step	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects		Yes		Yes		Yes		Yes		Yes
Fifth degree control polynomial in the ancillary step	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Peer 9th grade median GPA		Distance between peer median GPA and entrance threshold							
	-0.08	-0.05	-0.18	-0.18						
	(0.10)	(0.10)	(0.11)	(0.12)						
Weighted by the control group size in the ancillary step	Yes	Yes	Yes	Yes						
Year fixed effects		Yes		Yes		Yes		Yes		Yes
Fifth degree control polynomial in the ancillary step	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. The school by year sample size is 326. Significance level: *** 0.01, ** 0.05, * 0.1. Each column represents a different regression. Rows show the explanatory variable. The dependent variable is the estimated causal effect of that school by year observation for the students at the margin. The band is 0.6. The sample contains only schools with more than 20 control subjects and 30 subject in the treatment group within the band. The number of student observations in valid school by year combinations is 30349. The different degrees of polynomials indicate the degree of school individual control polynomial of the score variable used in the ancillary step estimate the effects.

†An estimate of 0.1 for the threshold level means that a standard deviation change in the threshold level is associated with a 0σ to 0.08σ change in matriculation exam performance at 0.95 confidence level.

††An estimate of -0.29 for the threshold level means that a standard deviation change in the threshold level is associated with a 0.01σ to 0.15σ change in matriculation exam performance at 0.95 confidence level.

†††The estimated significant effect in standard deviation with yearly fixed effects and a fifth degree polynomial of -0.53 means that a σ change in the standard deviations of the class is associated with a 0.02σ to 0.13σ change in matriculation exam performance at 0.95 confidence level.

Table 5.7: Regressions, second step, band of 0.6, at least 20 controls, multiple explanatory class composition variables, weighted by control group size.

	1	2	3	4	5	6	7	8	9
Peer mean 9th grade GPA	-0.15 (0.12)	-0.15 (0.12)	-0.29* (0.15)						
Top 10% peer mean 9th grade GPA				-0.17 (0.13)	-0.15 (0.13)	-0.24 (0.16)	-0.3** (0.14)	-0.28** (0.14)	-0.38** (0.16)
Bottom 10% peer mean 9th grade GPA							0.13** (0.06)	0.13** (0.06)	0.14* (0.08)
Peer 9th grade GPA standard deviation	-0.62** (0.26)	-0.63** (0.27)	-0.8** (0.31)	-0.44* (0.24)	-0.45* (0.24)	-0.46 (0.29)			
Weighted by the control group size in the ancillary step	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
School fixed effects			Yes			Yes			Yes
Fifth degree polynomial	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses. Significance level: *** 0.01, ** 0.05, * 0.1. The dependent variable is the estimated causal effect of that school by year observation for the students at the margin. The school by year sample size is 326. Each column represents a different regression. Rows show the explanatory variable. The band is 0.6. The sample contains only schools with more than 20 control subjects and 30 subject in the treatment group within the band. The number of student observations in valid school by year combinations is 30349. Class homogeneity retains its explanatory power even when other class composition variables are used in the same regression. The specifications were chosen in the manner that would not use two explanatory variables with more than 0.5 correlation between each other in the same regression. The degree of polynomial indicates the degree of polynomial used in the first stage to estimate the effects.

DOI: 10.2870/93297

5.7 Appendix

Table 5.8: Regressions, second step, band of 0.6, at least 20 controls

	1	2	3	4	5	6	7	8	9
Entrance threshold	0.13**	0.11*	0.07	0.12**	0.1*†	0.08	0.07	0.06	0.05
	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.08)	(0.07)	(0.07)	(0.09)
Year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
School fixed effects			Yes			Yes			Yes
Third degree polynomial	Yes	Yes	Yes						
Fifth degree polynomial				Yes	Yes	Yes			
Seventh degree polynomial							Yes	Yes	Yes
Peer mean 9th grade GPA	-0.06	-0.05	-0.14	-0.14	-0.12	-0.21	-0.17	-0.15	-0.2
	(0.12)	(0.12)	(0.14)	(0.11)	(0.12)	(0.14)	(0.12)	(0.13)	(0.16)
Year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
School fixed effects			Yes			Yes			Yes
Third degree polynomial	Yes	Yes	Yes						
Fifth degree polynomial				Yes	Yes	Yes			
Seventh degree polynomial							Yes	Yes	Yes
Top 10% peer mean 9th grade GPA	-0.19	-0.17	-0.24	-0.31**	-0.29**††	-0.37**	-0.34**	-0.32**	-0.39**
	(0.13)	(0.13)	(0.15)	(0.13)	(0.13)	(0.15)	(0.14)	(0.14)	(0.17)
Year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
School fixed effects			Yes			Yes			Yes
Third degree polynomial	Yes	Yes	Yes						
Fifth degree polynomial				Yes	Yes	Yes			
Seventh degree polynomial							Yes	Yes	Yes
Bottom 10% peer mean 9th grade GPA	0.06	0.06	0.04	0.06	0.06	0.06	0.06	0.07	0.09
	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.09)
Year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
School fixed effects			Yes			Yes			Yes
Third degree polynomial	Yes	Yes	Yes						
Fifth degree polynomial				Yes	Yes	Yes			
Seventh degree polynomial							Yes	Yes	Yes
Peer 9th grade GPA standard deviation	-0.45*	-0.45*	-0.42	-0.58**	-0.53**†††	-0.6*	-0.64**	-0.65**	-0.76**
	(0.26)	(0.27)	(0.31)	(0.28)	(0.27)	(0.31)	(0.29)	(0.3)	(0.35)
Year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
School fixed effects			Yes			Yes			Yes
Third degree polynomial	Yes	Yes	Yes						
Fifth degree polynomial				Yes	Yes	Yes			
Seventh degree polynomial							Yes	Yes	Yes

Note: Standard errors in parentheses. The school by year sample size is 326. Significance level: *** 0.01, ** 0.05, * 0.1. Each column represents a different regression. Rows show the explanatory variable. The dependent variable is the estimated causal effect of that school by year observation for the students at the margin. The band is 0.6. The sample contains only schools with more than 20 control subjects and 30 subject in the treatment group within the band. The number of student observations in valid school by year combinations is 30349. The different degrees of polynomials indicate the degree of school individual control polynomial of the score variable used in the ancillary step estimate the effects.
†An estimate of 0.1 for the threshold level means that a standard deviation change in the threshold level is associated with a 0σ to 0.08σ change in matriculation exam performance at 0.95 confidence level.
††An estimate of -0.29 for the top 10% peer mean 9th grade GPA means that a standard deviation change in the ttop 10% peer mean GPA is associated with a 0.01σ to 0.15σ change in matriculation exam performance at 0.95 confidence level.
†††The estimated significant effect in standard deviation with yearly fixed effects and a fifth degree polynomial of -0.53 means that a σ change in the standard deviations of the class is associated with a 0.02σ to 0.13σ change in matriculation exam performance at 0.95 confidence level.

Table 5.9: Regressions, second step, band of 0.6, at least 20 controls, weighted by the number of control subjects

	1	2	3	4	5	6	7	8	9
Entrance threshold	0.12**	0.1*	0.05	0.11**	0.09	0.03	0.06	0.05	0
	(0.05)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.08)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Peer mean	0.01	0.03	-0.09	-0.03	-0.01	-0.13	-0.02	-0.01	-0.11
9th grade GPA	(0.11)	(0.11)	(0.13)	(0.11)	(0.11)	(0.14)	(0.12)	(0.12)	(0.15)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Top 10% peer mean	-0.09	-0.07	-0.16	-0.2	-0.18	-0.28*	-0.18	-0.17	-0.27
9th grade GPA	(0.13)	(0.13)	(0.15)	(0.13)	(0.13)	(0.15)	(0.14)	(0.14)	(0.17)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Bottom 10% peer mean	0.06	0.07	0.06	0.08	0.08	0.08	0.11*	0.11*	0.12
9th grade GPA	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.08)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Peer 9th	-0.35	-0.36	-0.37	-0.47**	-0.48**	-0.54*	-0.62**	-0.62**	-0.71**
grade GPA standard deviation	(0.24)	(0.25)	(0.28)	(0.23)	(0.24)	(0.28)	(0.26)	(0.26)	(0.32)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes

Note: Standard errors in parentheses. The school by year sample size is 326. Significance level: *** 0.01, ** 0.05, * 0.1. Each column represents a different regression. Rows show the explanatory variable. The dependent variable is the estimated causal effect of that school by year observation for the students at the margin. The band is 0.6. The sample contains only schools with more than 20 control subjects and 30 subject in the treatment group within the band. The second step regression is weighted by the number of control subjects to correct for possible problems with small sample size. The number of student observations in valid school by year combinations is 30349. The different degrees of polynomials indicate the degree of school individual control polynomial of the score variable used in the ancillary step estimate the effects.

†An estimate of -0.29 for the threshold level means that a standard deviation change in the threshold level is associated with a 0.01σ to 0.15σ change in matriculation exam performance at 0.95 confidence level.

††The estimated significant effect in standard deviation with yearly fixed effects and a fifth degree polynomial of -0.53 means that a σ change in the standard deviations of the class is associated with a 0.02σ to 0.13σ change in matriculation exam performance at 0.95 confidence level.

DOI: 10.2870/93297

Table 5.10: Regressions, second step, band of 0.2, at least 10 controls

	1	2	3	4	5	6	7	8	9
Entrance threshold	0.104	0.096	0.062	0.087	0.087	0.011	0.098	0.095	0.013
	(0.088)	(0.091)	(0.111)	(0.091)	(0.094)	(0.116)	(0.098)	(0.101)	(0.128)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Peer mean	0.283*	0.277*	0.284	0.331**	0.336**†	0.325*	0.332**	0.347**	0.303
9th grade GPA	(0.147)	(0.152)	(0.182)	(0.14)	(0.145)	(0.183)	(0.149)	(0.153)	(0.198)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Top 10% peer mean	0.111	0.095	0.109	0.16	0.154	0.14	0.16	0.164	0.086
9th grade GPA	(0.172)	(0.177)	(0.208)	(0.165)	(0.17)	(0.209)	(0.177)	(0.181)	(0.228)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Bottom 10% peer mean	0.231***	0.24***	0.294***	0.279***	0.294***	0.324***	0.323***	0.336***	0.385***
9th grade GPA	(0.088)	(0.09)	(0.103)	(0.085)	(0.087)	(0.104)	(0.089)	(0.091)	(0.11)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Peer 9th	-0.949***	-1.013***	-1.122***	-0.878**	-1.163***††	-1.201***	-1.282***	-1.34***	-1.51***
grade GPA standard deviation	(0.35)	(0.36)	(0.399)	(0.369)	(0.347)	(0.396)	(0.346)	(0.357)	(0.41)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes

Note: Standard errors in parentheses. The school by year sample size is 257. Significance level: *** 0.01, ** 0.05, * 0.1. Each column represents a different regression. Rows show the explanatory variable. The dependent variable is the estimated causal effect of that school by year observation for the students at the margin. The band is 0.2. The sample contains only schools with more than 10 control subjects and 15 treated subjects. The number of student observations in valid school by year combinations is 10452. The different degrees of polynomials indicate the degree of polynomial used in the first stage to estimate the effects.
†The estimated significant effect in peer mean 9th grade GPA with yearly fixed effects and a fifth degree polynomial of 0.336 means that a σ change in the standard deviations of the class is associated with a 0.02σ to 0.26σ change in matriculation exam performance at 0.95 confidence level.
††An estimate of -1.163 for the peer 9th grade GPA standard deviation means that a standard deviation change in the homogeneity of the class is associated with a 0.06σ to 0.22σ change in matriculation exam performance at 0.95 confidence level.

 DOI: 10.2870/93297

Table 5.11: Regressions, second step, band of 0.2, at least 10 controls, weighted by the number of control subjects

	1	2	3	4	5	6	7	8	9
Entrance threshold	0.11	0.1	0.05	0.1	0.1	0.01	0.11	0.11	0.01
	(0.08)	(0.08)	(0.11)	(0.09)	(0.09)	(0.11)	(0.09)	(0.09)	(0.12)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Peer mean 9th grade GPA3	0.39***	0.38***	0.32*	0.44***	0.44***†	0.38**	0.45***	0.45***	0.34*
	(0.14)	(0.14)	(0.18)	(0.13)	(0.14)	(0.18)	(0.14)	(0.14)	(0.19)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Top 10% peer mean 9th grade GPA2	0.27	0.24	0.12	0.31*	0.3*	0.17	0.33*	0.32*	0.1
	(0.18)	(0.18)	(0.21)	(0.17)	(0.17)	(0.21)	(0.18)	(0.18)	(0.23)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Bottom 10% peer mean 9th grade GPA	0.28***	0.28***	0.32***	0.33***	0.34***	0.36***	0.37***	0.38***	0.41***
	(0.08)	(0.08)	(0.1)	(0.08)	(0.08)	(0.1)	(0.08)	(0.08)	(0.1)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes
Peer 9th grade GPA standard deviation	-1.07***	-1.12***	-1.27***	-1.22***	-1.28***††	-1.36***	-1.41***	-1.47***	-1.66***
	(0.33)	(0.34)	(0.38)	(0.32)	(0.33)	(0.38)	(0.33)	(0.34)	(0.39)
year fixed effects		Yes	Yes		Yes	Yes		Yes	Yes
school fixed effects			Yes			Yes			Yes
third degree polynomial	Yes	Yes	Yes						
fifth degree polynomial				Yes	Yes	Yes			
seventh degree polynomial							Yes	Yes	Yes

Note: Standard errors in parentheses. The school by year sample size is 257. Significance level: *** 0.01, ** 0.05, * 0.1. Each column represents a different regression. Rows show the explanatory variable. The dependent variable is the estimated causal effect of that school by year observation for the students at the margin. The band is 0.2. The sample contains only schools with more than 10 control subjects and 15 treated subjects. The second step regression is weighted by the number of control subjects to correct for possible problems with small sample size. The number of student observations in valid school by year combinations is 10452. The different degrees of polynomials indicate the degree of polynomial used in the first stage to estimate the effects.

†The estimated significant effect in peer mean 9th grade GPA with yearly fixed effects and a fifth degree polynomial of 0.44 means that a σ change in the standard deviations of the class is associated with a 0.07σ to 0.30σ change in matriculation exam performance at 0.95 confidence level.

††An estimate of -1.28 for the peer 9th grade GPA standard deviation means that a standard deviation change in the homogeneity of the class is associated with a 0.08σ to 0.24σ change in matriculation exam performance at 0.95 confidence level.

DOI: 10.2870/93297

References

[1] (2009), *Finnish Matriculation Examination 2007 - Statistics from the matriculation examination board*. The Matriculation Examination Board. 65

[2] Acemoglu, D. (1998), "Why do new technologies complement skills? directed technical change and wage inequality." *The Quarterly Journal of Economics*, 113, 1055–1089. 21

[3] Acemoglu, D. (2000), "Technical change, inequality, and the labor market." Technical report, National Bureau of Economic Research. 20

[4] Acemoglu, D. et al. (2011), "Skills, tasks and technologies: Implications for employment and earnings." *Handbook of Labor Economics*, 4, 1043–1171. 20

[5] Barro, R.J. and J.W. Lee (2010), "A new data set of educational attainment in the world, 1950–2010." Technical report, National Bureau of Economic Research. 2, 4, 8

[6] Battistin, E. and E. Rettore (2008), "Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs." *Journal of Econometrics*, 142, 715–730. 66

[7] Becker, G.S., W.H.J. Hubbard, and K.M. Murphy (2010), "Explaining the worldwide boom in higher education of women." *Journal of Human Capital*, 4, 203–241. 2, 3, 12, 47

[8] Card, D. and J.E. DiNardo (2002), "Skill-biased technological change and rising wage inequality: Some problems and puzzles." *Journal of Labor Economics*, 20, 4. 47

[9] Card, D. and T. Lemieux (2000), "Can falling supply explain the rising return to college for younger men? a cohort-based analysis." Technical report, National Bureau of Economic Research. 20

[10] Charles, K.K. and M.C. Luoh (2003), "Gender differences in completed schooling." *Review of Economics and Statistics*, 85, 559–577. 47

[11] Chiappori, P.A., M. Iyigun, and Y. Weiss (2009), "Investment in schooling and the marriage market." *The American Economic Review*, 99, 1689–1713. 10, 12, 33, 45, 47

[12] Cho, D. (2007), "The role of high school performance in explaining women's rising college enrollment." *Economics of Education Review*, 26, 450–462. 10, 45, 47, 56, 57

[13] Clark, D. (2010), "Selective schools and academic achievement." *The B.E. Journal of Economic Analysis and Policy*, 10:1. 63

[14] Craig, I.W., C.M.A. Haworth, and R. Plomin (2009), "Commentary on "a role for the x chromosome in sex differences in variability in general intelligence?"(johnson et al., 2009)." *Perspectives on Psychological Science*, 4, 615–621. 16

[15] Cullen, J.B., B.A. Jacob, and S. Levitt (2006), "The effect of school choice on participants: Evidence from randomized lotteries." *Econometrica*, 74, 1191–1230. 62

[16] Dougherty, C. (2005), "Why are the returns to schooling higher for women than for men?" *Journal of Human Resources*, 40, 969–988. 12

[17] Duflo, E., P. Dupas, M. Kremer, L. Center, and T. Floor (2008), "Peer effects and the impact of tracking: Evidence from a randomized evaluation in Kenya." *NBER Working Paper*. 63

[18] Ellis, H. (1894), *Man and woman*. Scott. 13

[19] Feingold, A. (1992), "Sex differences in variability in intellectual abilities: A new look at an old controversy." *Review of Educational Research*, 62, 61–84. 13

[20] Fortin, N.M., P. Oreopoulos, and S. Phipps (2011), "Leaving boys behind: Gender disparities in high academic achievement." Technical report, Working Paper, Department of Economics, UBC. 56

[21] Frasier, G.W. (1919), "A comparative study of the variability of boys and girls." *Journal of Applied Psychology*, 3, 151. 13

[22] Goldin, C. and L.F. Katz (2009), "The race between education and technology: The evolution of us educational wage differentials, 1890 to 2005." Technical report, National Bureau of Economic Research. 20

[23] Goldin, C., L.F. Katz, and I. Kuziemko (2006), "The homecoming of american college women: The reversal of the college gender gap." *Journal of Economic Perspectives*, 20, 133–156. 3, 12, 33

[24] Hahn, J., P. Todd, and W. Van der Klaauw (2001), "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica*, 69, 201–209. 66

[25] Hedges, L.V. and A. Nowell (1995), "Sex differences in mental test scores, variability, and numbers of high-scoring individuals." *Science*, 269, 41–45. 13, 14

[26] Hoekstra, M. (2009), "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *The Review of Economics and Statistics*, 91, 717–724. 63

[27] Hubbard, W.H.J. (2011), "The phantom gender difference in the college wage premium." *Journal of Human Resources*, 46, 568–586. 12, 33, 47

[28] Johnson, W., A. Carothers, and I.J. Deary (2008), "Sex differences in variability in general intelligence: A new look at the old question." *Perspectives on Psychological Science*, 3, 518. 14

[29] Johnson, W., A. Carothers, and I.J. Deary (2009), "A role for the x chromosome in sex differences in variability in general intelligence?" *Perspectives on Psychological Science*, 4, 598–611. 16

[30] Kirjavainen, Tanja (2009), *Essays on the Efficiency of Schools and Student Achievement*, volume 53 of *VATT Publications*. Government Institute for Economic Research, VATT. 64

[31] Lee, D. and T. Lemieux (2009), "Regression discontinuity designs in economics." *NBER working paper.* 67

[32] Machin, S. and T. Pekkarinen (2008), "Global sex differences in test score variability." *Science.* 15

[33] Meara, E.R., S. Richards, and D.M. Cutler (2008), "The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000." *Health Affairs*, 27, 350–360. 21

[34] Nowell, A. and L.V. Hedges (1998), "Trends in gender differences in academic achievement from 1960 to 1994: An analysis of differences in mean, variance, and extreme scores." *Sex Roles*, 39, 21–43. 14

[35] Pekkarinen, T. (2012), "Gender differences in education." *forthcoming in: Nordic Economic Policy Review.* 1, 2, 3

[36] Pop-Eleches, C. and M. Urquiola (2008), "The consequences of going to a better school." *Department of Economics. Columbia University. Mimeo.* 63

[37] Sacerdote, B. (2001), "Peer effects with random assignment: Results for dartmouth roommates*." *Quarterly Journal of Economics*, 116, 681–704. 62, 66

[38] Silverman, B.W. (1986), *Density estimation for statistics and data analysis*, volume 26. Chapman & Hall/CRC. 77