The Gender Gap Reversal in Education: the Higher Male Dispersion Theory *

Laurent Bossavie & Ohto Kanninen †

August 25, 2014

Abstract

In virtually all countries, males show a larger dispersion in test scores relative to females. Using this fact, this paper proposes a comprehensive theory that can explain the gender gap reversal in educational attainment. We show that a larger dispersion of abilities among males, together with an increase in the returns to education over time, are sufficient to reproduce the gender gap dynamics observed in all countries. From time-series on enrollment rates, our model generates country estimates for gender differences in ability distributions. Our estimates strongly correlate with gender differences in test score distributions found in international student assessments. Using our framework, we also test our theory against previous explanations proposed by the literature. The data do not support the predictions of alternative hypotheses, while being more consistent with the higher male dispersion theory.

Keywords: Educational attainment, gender gap, test scores.

JEL-Codes: I21, J16

*We are grateful to Andrea Ichino and Luigo Guiso for constant advice and discussions. We also would like to thank Jerôme Adda, Christian Dustmann, Peter Hansen, Steve Machin, Tommaso Naniccini,Tuomas Pekkarinnen and Kjell Salvanes as well as all participants to the micro-econometrics working group at the European University Institute in Florence for helpful comments and suggestions. Finally, we also thank the participants to the EEA conference 2012, Malaga, the EALE conference 2013, Turin and the 2nd Conference on Higher Education and Innovation Research 2012, Helsinki for their suggestions. All remaining errors are ours.

†Department of Economics, European University Institute, Villa San Paolo, Via della Piazzuola 43, 50133 Florence, Italy. E-mail: laurent.bossavie@eui.eu.

1 Introduction

The gender gap in educational attainment has reversed over the last decades, and this reversal has been quasi-universal. The first contribution of the paper is to establish these two facts internationally, and to show that the reversal occurred in the two tails of the educational attainment distribution. In the upper tail, females progressively surpassed males among participants to tertiary education. In the lower tail, males are now the majority among secondary school non-completers, while females used to be the majority. We show that these two facts are observed in virtually all developed countries, but also in the majority of developing countries.

The central contribution of the paper is to propose a simple and comprehensive framework to account for the gender gap reversal in education. To explain this quasi-universal phenomenon, we rely on a quasi-universal fact evidenced by Pekkarinen and Machin (2008): the larger variance of males' test score distribution relative to females'. The model has two other building blocks: the optimal choice of schooling based on individual ability, and the increase in the returns to tertiary education over time.

Our model builds on the micro-foundations presented in Card (1994). In this framework, the optimal length of schooling chosen by individuals is an increasing function of individual ability, measured by test scores. Building on the international evidence presented in Machin and Pekkarinen (2008), we assume a larger variance of test scores among the male population. In this setting, we show that an increase in the returns to education over time generates both an increase in enrollment, and a change in the gender ratio among the enrolled. The joint evolution of the enrollment rate and the gender ratio produced by our model closely mimics the dynamics observed in each individual country in our sample. In particular, it is able to generate the gender gap reversal observed among participants to tertiary education, and among secondary school non-completers.

The model can replicate the within-country dynamics of the gender ratio, but is also able to account for differences in the evolution of the gender gap across countries. Using our model, we are able to extract estimates for country-specific gender differences in test score distributions. To assess the validity of our theory, we confront our estimates to gender differences in test score distributions found in the Project for International Student Assessment (PISA). We find a strong correlation between gender differences in test score distribution parameters estimated from our model, and those found in PISA. This suggests that our model resonates actual gender differences in test scores distributions observed across countries, and gives further credit to our theory.

Finally, a third contribution of the paper is to formulate and evaluate alternative explanations to the reversal within the theoretical framework we developed. To the best of our knowledge, this is the first attempt to evaluate alternative explanations to this phenomenon in a common setting. In particular, we show that some predictions of these alternative hypotheses conflict with the predictions of our theory, and can be tested using empirical data. We develop and propose several tests for the validity of these conflicting hypotheses. These tests appear to reject the predictions of alternative explanations, while bringing further support to our theory.

Understanding the gender gap reversal in education is important in its own right, but also for efficiency purposes. 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 crucial to identify whether differences in observable outcomes originate from discrimination or other distortions between genders, or from optimizing behaviors based on gender differences in preferences, behaviors, or traits.

Previous literature has established the gender gap reversal in participation to tertiary education for the US, and more recently for Scandinavian countries[1]. It has also proposed potential explanations to the gender gap reversal in the US context. Goldin et al. (2006) in particular invoke the removal of female carrier barriers combined with greater returns to college for females as a potential driving force behind the gender gap reversal in the US. In a similar spirit, Chiappori et al. (2009) argue that a change in the social roles driven by the improvement of home technology combined with greater returns education for women might have driven the reversal. Cho (2007) suggests that the improvement of high school performance of girls could explain a large part of the gender gap reversal observed in the US.

Although informative, they do not offer a framework to disentangle between competing explanations to the reversal. In addition, they are restricted to the US context. This paper proposes a new explanation to the reversal that fits the dynamics of the gender gap in a large sample of countries. It is also the first contribution to offer a comprehensive framework to evaluate competing explanations for the reversal in a common setting using empirical data.

The paper is organized as follows. Section 2 presents the empirical facts motivating the paper. It summarizes the existing literature on the larger variance of males' test score distribution,

[1]For evidence for the US, see among others Goldin at al. (2006) or Chiappori at al.(2009). For evidence about Scandinavian countries, see Pekkarinen (2012).

and establishes the reversal of the gender gap in both the upper and lower tail of educational attainment. Section 3 introduces our theoretical framework and derives its implications for the relationship between the enrollment rate and the gender ratio among the enrolled. Section 4 describes the data and our estimation strategy. Section 5 presents our main results. Section 6 formulates alternative theories in our framework and test their implications against our theory using empirical data. Section 7 concludes.

2 Empirical Motivation

2.1 The Larger Dispersion of Males' Test Score Distribution

There exists long-standing evidence showing that males exhibit a higher variance in a wide range of ability tests compared to females. 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 notes than males 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 further support for Ellis' original observation. 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 statistically significant.

More recently, Feingold (1992) reports the gender variance ratio of the PSAT and SAT of the College Entrance Examination Board in 1960, 1966, 1974 and 1983. In both mathematics and verbal tests the male-to-female variance ratio was found to be larger than 1, with little variation over successive waves. The results of Feingold (1992) have however been criticized as that they are drawn from a selected sample of individuals taking SAT, which are 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 these surveys, and report that the variance of males is larger in 41 out of the 43 cognitive tests. The estimated variance ratio ranges from 1.05 to 1.30. Deary et al. (2003) also use population-wide data on general intelligence of 11 year-olds in Scotland. While they report no significant mean differences in cognitive test scores between boys and girls, they find that the gender difference in standard deviations is highly statistically significant. Using a nationally-representative sample of 320,000 11-year-old pupils in the UK, Strand et al. (2006) also find that the test score variance is significantly larger among boys. Johnson et al. (2008) also find greater variability among boys' test scores using two population-representative samples of 11-years-olds in Scotland.

While early evidence is mostly confined to the US and anglo-saxon countries, Pekkarinen and Machin (2008) show that the greater variability in test scores among males is a universal phenomenon. They use test score data for 15-year-olds from the Program for International Student Assessment (PISA) conducted in 2003. The PISA study tests mathematical and reading skills of a representative sample of the population of 15-year-olds in 40 countries. The authors report that males' tests score variance is strictly larger than females' in 38 countries for mathematics and 39 countries for reading, out of a total of 40 countries. The gender difference in variance is statistically significant in all but 5 five countries with an average male-to-female variance ratio of 1.21 for reading and 1.20 for mathematics.

Johnson et al (2009) with a commentary by Craig et al (2009) discuss the possible role for the X chromosome in explaining the differences between males and females in ability dispersion. They note that a large number of genes in the X chromosome are related to general intelligence. This, combined with the fact that females have two X chromosomes while males have one X chromosome and one Y chromosome seems to play an important role in producing a higher variability. Since "Y chromosome is very small and carries little beyond the genetic instructions for maleness", the X chromosome functions mostly alone, as Johnson et al note. This allows recessive genes to function more often among males than among females, which increases 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.

2.2 The Gender Gap Reversal in Educational Attainment

Fact 1. The gender gap in participation to tertiary education has reversed from male majority to female majority

For the US, several papers have reported a convergence, followed by a reversal in the percentage of women relative to men attending tertiary education over the last decades [2]. While most evidence is confined to the US, we show in Figure **??** that the gender gap reversal in participation to tertiary education is a quasi-universal phenomenon. To reconstruct the evolution of participation rates to university by gender and cohort of birth, we used data from the Barro-Lee (2010) database. It allows to compute the fraction of individuals of a given 5-year band cohorts that attended university by age 35, for individuals born from 1891 to 1971. Figure 1 reports that the gender composition of participants to tertiary education education has already reversed from male majority to female majority in virtually all developed countries. In addition, it shows that this reversal also occurred in many developing countries.

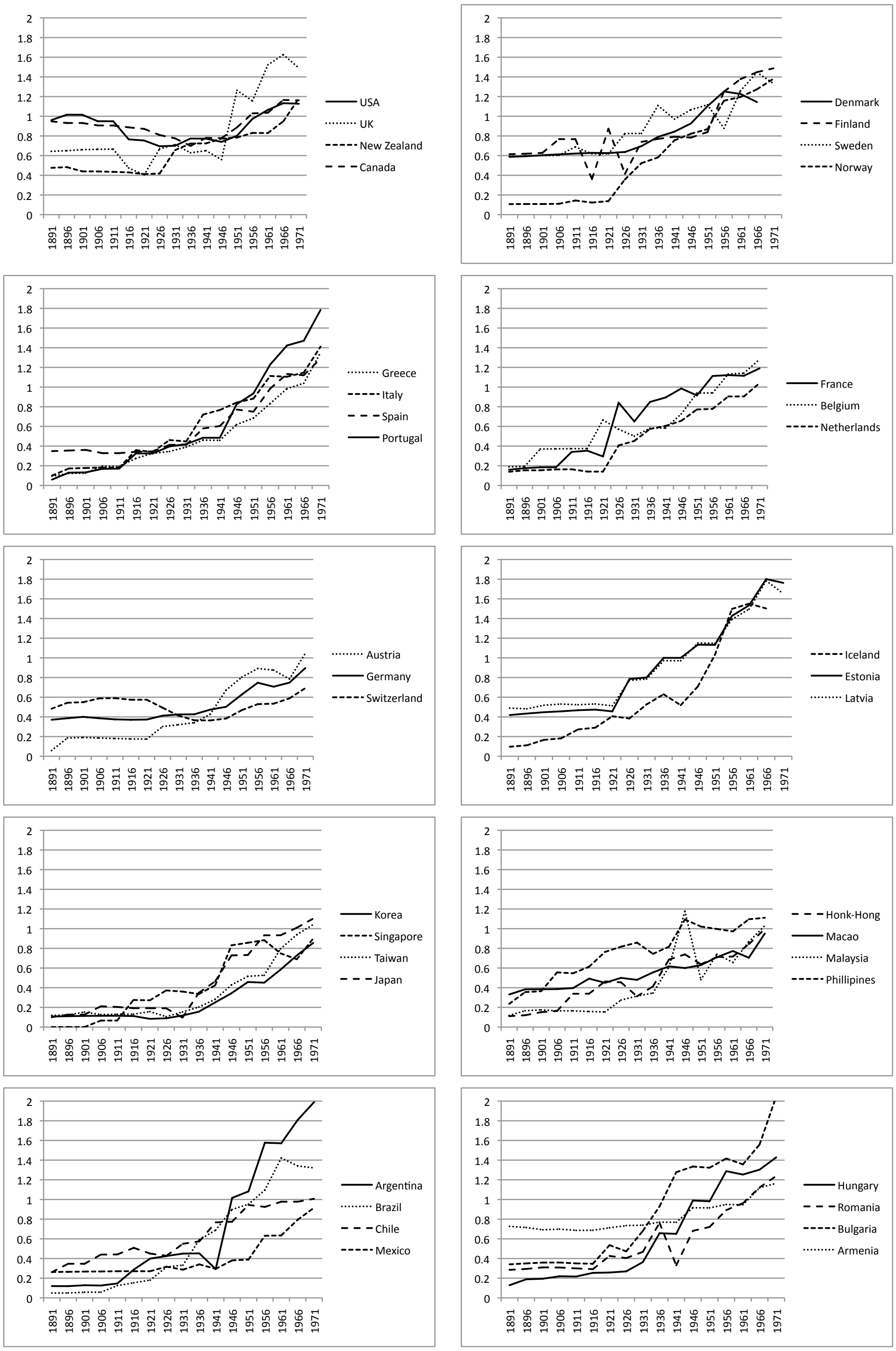
The timing of the reversal varies across countries. While the fraction of women attending higher education was already higher than for men in some Eastern European countries in the early 1970s, the reversal occurred only at the beginning of the 1990s in the UK. It is also a recent phenomenon in countries like Austria, Japan, and the Netherlands. South Korea, Switzerland and Germany are the only exceptions to the gender gap reversal among advanced economies. A sharp increase in the female-to-male ratio among participants to tertiary education is however observed in these countries, and one may expect the reversal to occur shortly.

Fact 2. The gender gap in secondary school non-completion has reversed from female majority to male majority

In Figure 2, we show that an analogous reversal occurred in the lower tail of educational achievement. As compulsory education ends at the end of lower secondary school in most developed countries, one 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. Using the Barro-Lee database (2010), we were able to reconstruct the evolution of the fraction of secondary school dropouts by gender and birth cohort for the main industrialized countries. The database allows to compute the secondary school non-completion rates for 5-year band cohorts from 1891 to 1981, for males and females separatly. The cohort analysis reported in Figure 2 reports that the gender composition of non-completers has reversed from female majority to male majority. As for the upper tail of educational achievement, it shows the reversal in the gender composition of low educational achievers is quasi-universal.

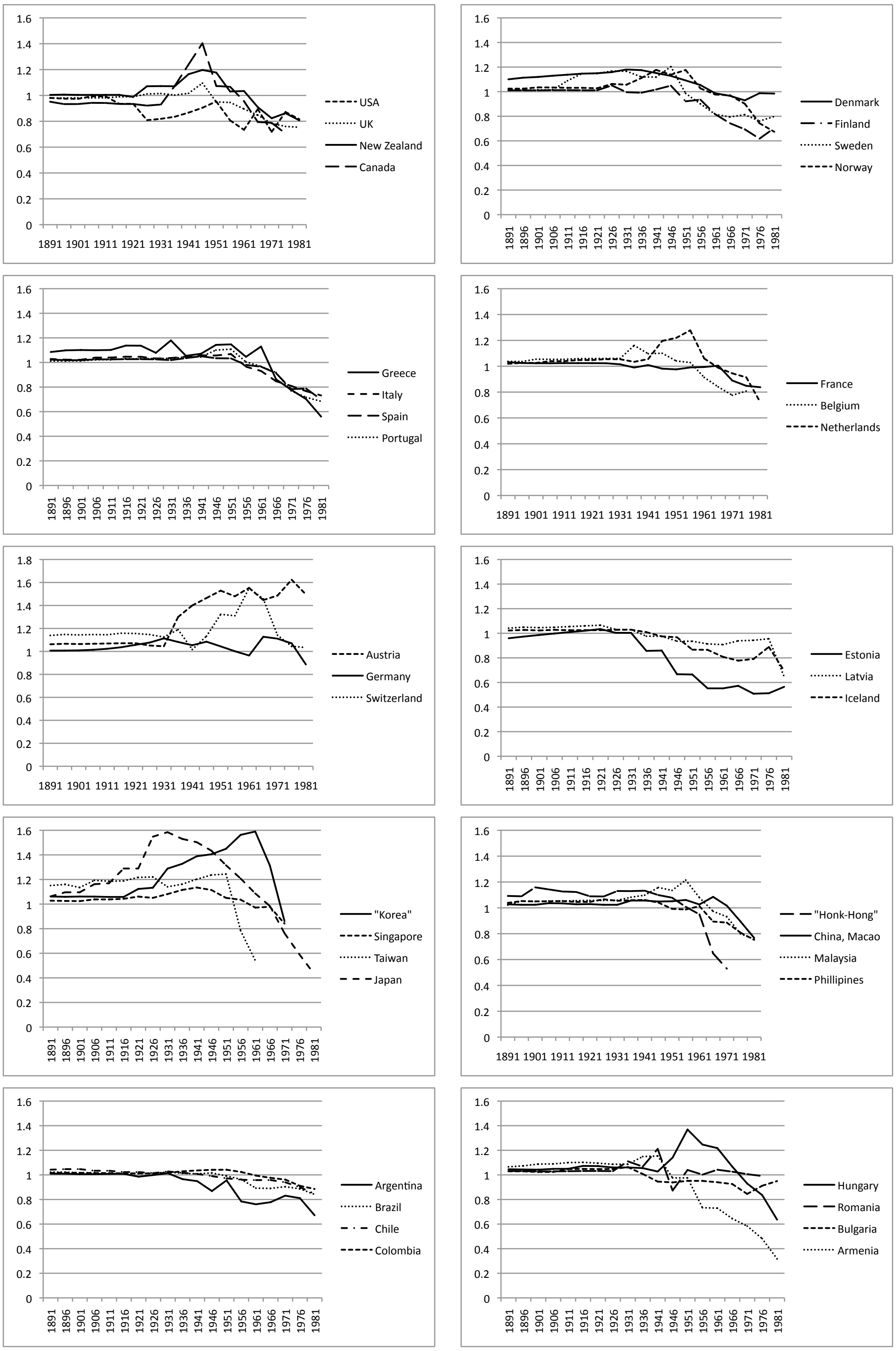
[2] See, for example, Goldin et al. (2006), Chiappori et al.(2009) or Becker et al. (2010)

Figure 1: Female-to-male ratio among participants to tertiary education - by cohort of birth



Notes. The y-axis reports the ratio of females to males having attended tertiary education by age 35. The x-axis indicates the cohort of birth.
Source. Barro-Lee database (2010)

Figure 2: Female-to-male ratio among secondary school non-completers - by cohort of birth



Notes. The y-axis reports the ratio of females to males among individuals that did not complete secondary education by age 25. The x-axis indicates the cohort of birth.
Source. Barro-Lee database (2010)

3 Theoretical Framework

Individuals are assumed to differ in their test-taking ability z_j and gender g_j, where $g_j \in \{m; f\}$. 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 affecting test scores and educational outcomes [3]

For the sake of simplicity, we assume a single-period model in which individuals perceive the benefits of their investment in education in the same period as they invest. Individual j chooses his years of schooling s so that he maximizes his utility U. Building on Becker (1967) and Card (1994), we define the utility function of individuals in the economy as:

$$U = B(s) - C(s) \tag{1}$$

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 with $C'(s) > 0$ and $C''(s) > 0$. The first order conditions for the individual maximization problem can simply be expressed as:

$$B'(s) = C'(s) \tag{2}$$

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 of s with $B'(s)$ having an individual-specific slope:

$$B'(s) = z_j - k_1 s \tag{3}$$

$$C'(s) = k_2 s \tag{4}$$

where $k_1 > 0$ and $k_2 > 0$. Intuitively, equation (2.3) states individuals with higher test-taking ability z_j perceive greater benefits (or equivalently, lower costs) from attending education. Importantly, k_1 and k_2 and are assumed to be identical for all individuals, irrespective of gender.

In this framework, the optimal level of schooling s chosen by individual j can be expressed

[3] As Heckman and Kautz (2012) note, test scores are the observable product of a complex combination of cognitive and non-cognitive skills, as well as effort and motivation. Understanding the mapping from ability to tests score is a tremendous task beyond the scope of this paper, and we will therefore refer to the combination of cognitive and non-cognitive abilities captured by test scores as test-taking ability.

as:

$$s_j^* = z_j \cdot b \tag{5}$$

where $b \equiv \frac{1}{k_1+k_2}$, and can be interpreted as the marginal return to education.
In the simple case in which b is not gender-specific, the optimal value of s_j is strictly increasing in individual 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:

$$H(s^*) = \begin{cases} 1 & \text{if } s^* \geq \bar{s} \\ 0 & \text{if } s^* < \bar{s} \end{cases} \tag{6}$$

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

Therefore,

$$H_j = 1 \text{ if } z_j > \frac{\bar{s}}{b} \equiv \bar{z} \tag{7}$$

Equation 7 simply states that individuals whose test-taking ability is below a given threshold $\bar{z}$ will choose not to enroll into tertiary education, while individuals whose ability is equal or greater than $\bar{z}$ with enroll. It also shows that the value of the ability threshold for enrolled is determined by the net benefits of education, that are common to all individuals in a given cohort.

Since the positive relationship between test-taking ability and enrollment in tertiary education is at the heart of our model, we provide empirical support for this assumption. Figure **??** and Table 1 report the relationship between test scores at age 15 and tertiary education attendance in the US. They show that test-scores are indeed a major determinant of enrollment in tertiary education. The coefficient associated with test scores is a highly significant predictor of university enrollment, and test scores alone explain virtually 80 percent of the variability in tertiary education attendance a few years later. Empirically, however, the propensity to attend tertiary education seem to increase monotonically with test scores rather than shifting upward discontinuously beyond a given threshold. Our model is indeed a very simplified version of educational enrollment decisions. Empirically, agents certainly enroll into education on the basis of not only individual test-taking ability z_j, but also a set of individual circumstances, such as parental income, personal network or taste for schooling that may or may not be correlated with test-taking ability[4]. This can justify why the empiri-

[4]The original Card (1994) model incorporates another variable r_j called circumstances as an intercept in

cal relationship between test-taking ability end enrollment decisions is fuzzy, instead of being discontinuous as in our model.

Table 1: The positive relationship between test score z and enrollment in tertiary education H - Probit regressions

	Dep. Variable: Dummy for Tertiary Education Attendance		
Composite test score	4.49***		
	(0.03)		
Reading test score		4.49***	
		(0.03)	
Mathematics test score			4.50***
			(0.03)
N. of observations	12,585	12,585	12,585

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Reported coefficients are marginal effects. Regressions do no include an intercept.
Source. US Educational Longitudinal Survey (ELS) - base year (2002) and 2nd follow-up (2006).

the expression of $C'(s)$. This variable captures differences in marginal costs of education across individuals that are not abilities, such as parental network or taste for schooling. For the sake of simplicity, we abstract from this variable in our model.

3.1 Aggregate Level and 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 and their gender g. Each cohort is assumed to be split equally by gender and the distribution of z. We denote $f_z(z)$ the probability density function of test-taking ability z.

All 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 talent or gender. In this context, the enrollment rate in higher education at time t for each gender can be expressed as:

$$E_t = 1 - F_z(\frac{\bar{s}}{b_t}) \tag{8}$$

or, equivalently

$$E_t = G_z(\frac{\bar{s}}{b_t}) = G_z(\bar{z}_t) \tag{9}$$

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

b_t is however allowed to vary across cohorts and those variations are interpreted as changes in the net returns to education, exogenous to the model. Therefore,

$$\frac{\partial E_t}{\partial b_t} > 0 \tag{10}$$

and

$$\frac{\partial \bar{z}}{\partial b_t} < 0 \tag{11}$$

An increase in b_t over time is the third building block of our model. To provide empirical evidence on the increase in b_t over time, we estimate equation 5 of our model using US data from 1980 and 2002. This data is from two US longitudinal surveys of 10-graders that are followed until their post-secondary education. Both surveys include test scores in 10th grades and allow to link individual test scores z to post-secondary education attendance H. In addition, they have have been designed to be comparable. This allows us to estimate the value of b using individual data on test scores at age 15 and tertiary education enrollment decisions.

Table 4 reports our estimate for b in the US in 1980 and 2002 using equation 5 in our model. It shows that the estimated net benefits to education $\hat{\beta}$ increased significantly from 1980 to

2002, which is consistent with the third fundamental assumption underlying our model. The null hypothesis of equality between β_{1980} and β_{2002} is strongly rejected by statistical tests. The data therefore seems to support the assumption of increased net benefits to education over time we make in our model.

Table 2: Estimates of the net benefits to education b for the US in 1980 and 2002 - Probit regressions

	1980	2002
$\hat{\beta}$	2.78***	4.50***
	(0.08)	(0.03)
Observations	11,687	12,585

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Reported coefficients are marginal effects. Regressions do not include an intercept.
Sources. High School and Beyond -base year (1980) and 2nd follow-up (1984), US Educational Longitudinal Study - base year (2002) and 2nd follow-up (2006)

There also exists a important body of literature showing that returns to education, and in particular returns to university education, have increased over the last decades. Monetary returns to tertiary education – the college wage premium – have been shown to increase sharply over the period 1970–2010. 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[5]. 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.

In our setting, an increase in b goes hand in hand with a decrease in $\bar{z}$, as individuals are assumed to enroll according to ablity. We do not observe $\bar{z}$ empirically, but one implication of a decrease in $\bar{z}$ is a decrease in the average ability of the enrolled, which is observable in the data. To check for this phenomenon, we computed the average IQ score of individuals attending teriary in the US over the period 1975-2010, using data form the General Social Survey (GSS)[6]. 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

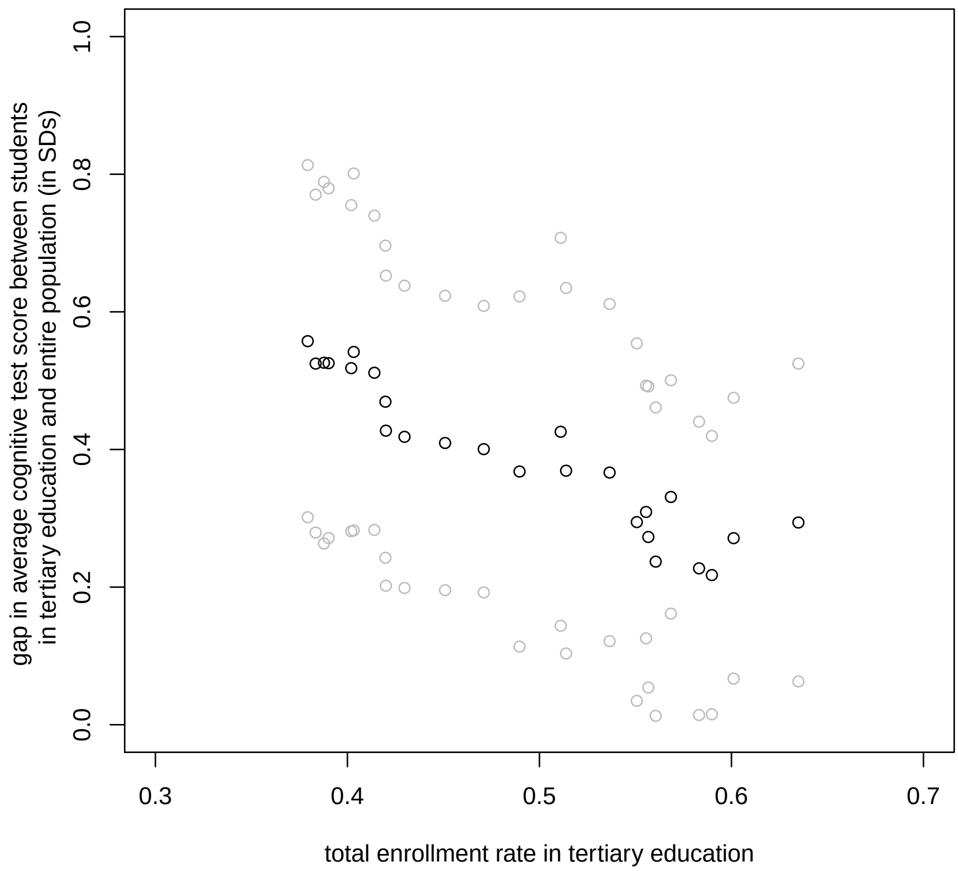
[5]The college wage premium is defined as the wage of college-educated workers relative to the wage of high-school educated workers.

[6]Although the lower bound of talent for university attendance can hardly be observed in the data, we can observe the average level of pre-university skills of students attending university, which mechanically decreases with $\bar{z}$

skills of the respondents. A measure of educational attainment (in years) is also reported, and we consider individuals with more than 12 years of education as having attended tertiary education. Figure 3 depicts the evolution of the average IQ of university students relative to the all population in number of standard deviations, from 1974 to 2010. It shows a clear downward trend. In 1974 the average IQ of students attending tertiary education was 0.60 standard deviation higher than the average IQ of the all population, but this relative difference decreased progressively until 2005 to reach approximately 0.30. The fact greater access to tertiary education was accompanied by a decrease in the average ability of the enrolled is consistent with less and less able individuals enrolling into education as enrollment expands. This corresponds to a decrease in $\bar{z}$ over time z in our framework.

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 attending university education. b_t can be thought as including monetary benefits of education as well as non-monetary benefits such as life expectancy, the propensity to marry and stay married, or household production.

Figure 3: The negative relationship between the enrollment rate in tertiary education E and the average ability of the enrolled - macro-level



Source. US General Social Survey (1975-2010)
Notes. Darks dots represents our estimates. Light dots represent confidence intervals at the 5%-level.

3.2 The Relationship between Total Enrollment Rate and Female-to-Male Ratio

Building on the evidence on gender differences in test score variability, we allow $f_z(z)$ to differ between genders. We denote $f_{z_m}(z)$ and $f_{z_f}(z)$ the probability density functions (pdf) 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[7]. Each cohort is assumed to be split equally between males and females, and

[7] We assume the distribution function types to be the identical for both genders, but allow the distributional parameters to differ. For all empirical applications, we also assume normality of the two distributions.

the distribution of test-taking ability for each gender $g \in \{m, f\}$ is assumed to be invariant over time:

$$f_{z_g,t}(z) = f_{z_g}(z) \tag{12}$$

Panel A of Figure 4 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 complementary cumulative distribution functions (ccdfs) of z_m and z_f, denoted $G_{z_m}(\bar{z})$ and $G_{z_f}(\bar{z})$, respectively and defined as:

$$G_z(\bar{z}) = \int_{\bar{z}}^{+\infty} f_z(z)\, dz \tag{13}$$

In the illustration, 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 CCDFs $G_{z_m}(\bar{z})$ and $G_{z_f}(\bar{z})$, it is possible to compute the total enrollment rate in tertiary education in the economy as:

$$E(\bar{z}) \equiv \frac{G_{z_f}(\bar{z}) + G_{z_m}(\bar{z})}{2} \tag{14}$$

which is represented by the thin dotted line in panel B of Figure 4, 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:

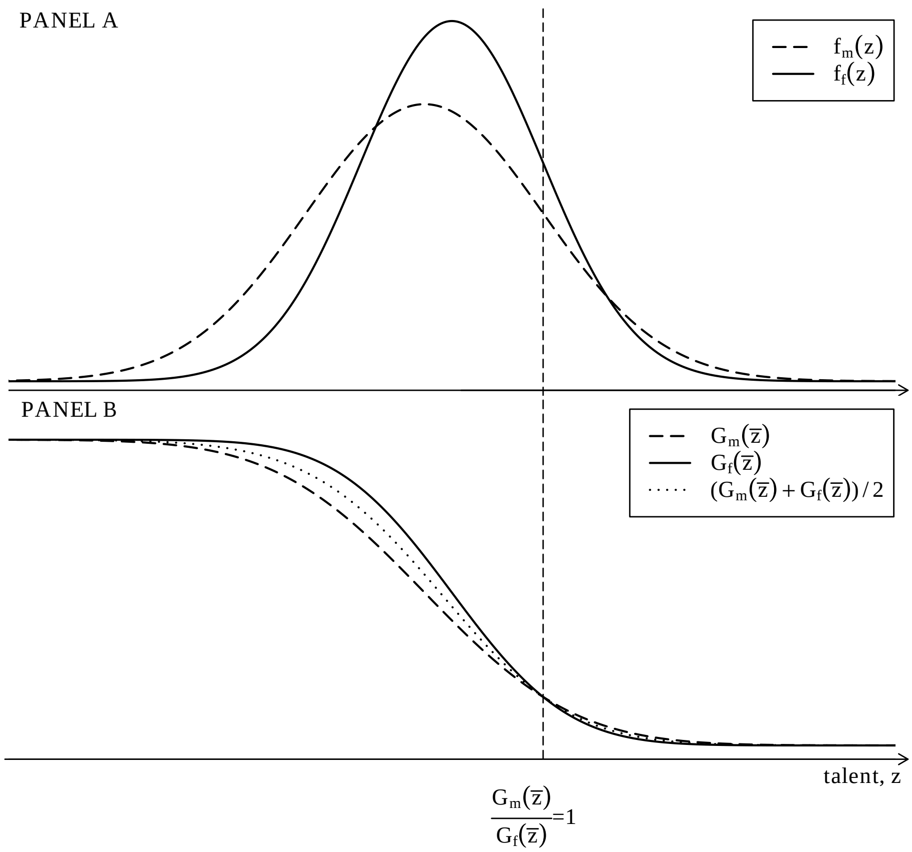
$$R(\bar{z}) \equiv \frac{G_{z_f}(\bar{z})}{G_{z_m}(\bar{z})} \tag{15}$$

From Panel B of Figure 4, we can derive the relationship between total enrollment rate and the female-to-male ratio among the enrolled. Figure 5 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 4.

$R(\bar{z})$ and $E(\bar{z})$ are both functions of the lower bound of test-taking ability to enroll $\bar{z}$, which varies with the exogenous parameter b. Under the assumption that $Var[z_m] > Var[z_f]$, it can be shown that the relationship between the female-to-male ratio $R(\bar{z})$ 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*

Figure 4: 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 with $\sigma_m > \sigma_f$ and $\mu_f > \mu_m$. Panel B shows the complementary cumulative distributions, resulting from the integration from $+\infty$ to z of $f_{z_f}(z)$ and $f_{z_m}(z)$.

rate $E(\bar{z})$ tends to zero.

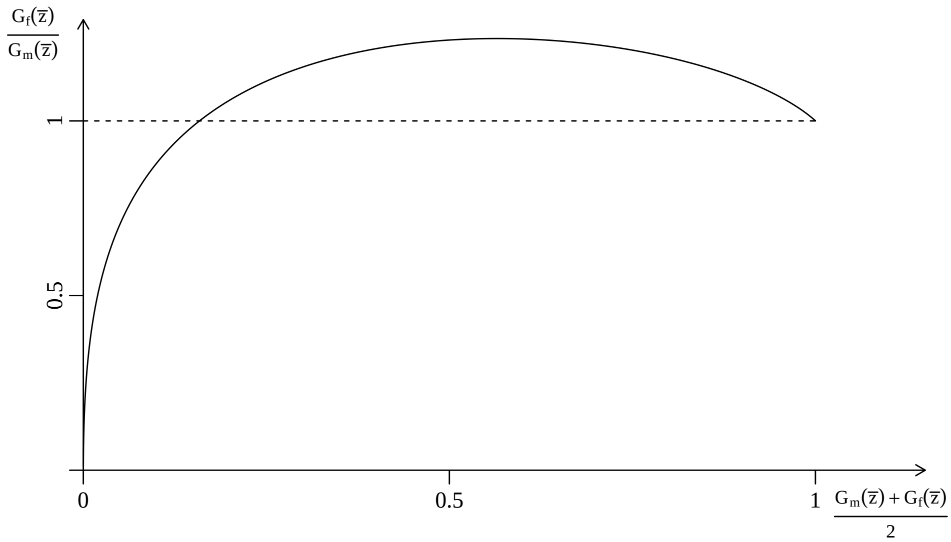
Proposition 2 *The female-to-male ratio $R(\bar{z})$ tends to one when the total enrollment rate $E(\bar{z})$ tends to one.*

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.*

Proof. See the Appendix.

Importantly, the assumption $Var[z_m] > Var[z_f]$ is necessary and sufficient for Proposition

Figure 5: Expected relationship between enrollment rate in tertiary education and female-to-male ratio among the enrolled – illustration



1 to 3 to hold, without any restriction required on μ_f and μ_m. In addition, the normality assumption on $f_{z_f}(z)$ and $f_{z_m}(z)$ is not necessary for Proposition 1 to 3 to be true. As shown in the Appendix, Proposition 1 to 3 also hold when z follows alternative two-parameter probability distribution functions commonly used in the literature. Our theory therefore appears to be quite robust as the only necessary condition on the distribution of z to generate the reversal is the larger dispersion of males' distribution $Var[z_m] > Var[z_f]$.

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

$$N_t = F_z(\frac{s}{b_t}) = F_z(\underline{z_t}) \tag{16}$$

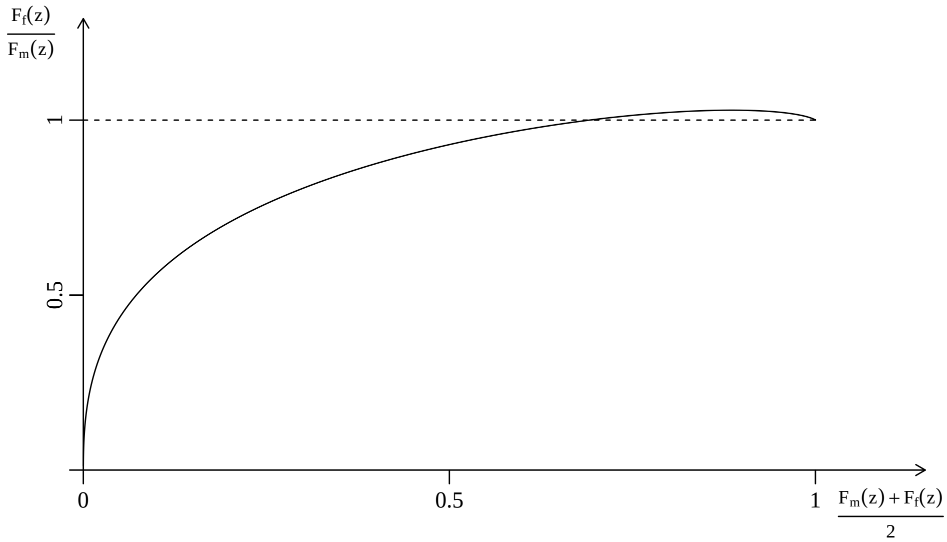
where

$$F_z(\underline{z}) = \int_{-\infty}^{\underline{z}} f_z(z)\, dz \tag{17}$$

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:

$$N(\underline{z}) \equiv \frac{F_{z_f}(\underline{z}) + F_{z_m}(\underline{z})}{2} \tag{18}$$

Figure 6: Expected relationship between secondary school non-completion and female-to-male ratio among non-completers – illustration



Note. The distribution parameters of F_{z_f} and F_{z_m} are the same as in Figure 5.

Figure 6 shows that our simple theoretical framework can also generate a reversal in the gender composition of secondary school non-completers. In an analogous way as for tertiary education enrollment, it can be shown that the relationship between the female-to-male ratio among non-completers $D(\underline{z})$ and the secondary school non completion rates $N(\underline{z})$ has 3 notable properties:

Proposition 4 *The female-to-male ratio among non-completers $D(\underline{z})$ tends to zero when the total non-completion rate $N(\underline{z})$ tends to zero.*

Proposition 5 *The female-to-male ratio among non-completers $D(\underline{z})$ tends to one when the total non-completion rate $N(\underline{z})$ tends to one.*

Proposition 6 *There exists a value of $N(\underline{z}) \in [0, 1[$ such that $D(\underline{z}) = 1$. This value is unique and always exists.*

The proof is analogous to the proof for Proposition 1 to 3.

4 Empirical Estimation

4.1 Data

Our data on enrollment rates and gender ratios is from two sources. Data on tertiary education enrollment is from the UNESCO database, which records information on enrollment rates by gender for about 200 countries over the period 1970–2010. it was compiled for national censuses. 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 is 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[8]. 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 therefore 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, fortunately, mostly interested in the comparative evolution of this enrollment rate between genders, rather than in its absolute value. In addition, it is, to the best of our knowledge, the only enrollment measure 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 educational attainment dataset 2010. It allows to compute upper-secondary school completion rates disaggregated by gender for 146 countries. The dataset was constructed from census/survey data as compiled by UNESCO and Eurostat. Contrary to publicly available UNESCO data which measures the stock of individuals attending education in a given year, the Barro-Lee dataset allows to compute educational attainment by cohort of birth. It is therefore a more accurate measure of its theoretical counterpart $N(\underline{z}) \equiv \frac{F_{z_f}(\underline{z})+F_{z_m}(\underline{z})}{2}$, which is the fraction of individuals belonging to a synthetic age-cohort enrolling into tertiary education. This flow measure of human capital, is more sensitive to cohort-by-cohort changes in educational

[8]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

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[9]. In total, we can compute the total secondary school non-completion rate and the female-to-male ratio among non-completers fo sixteen 5-year-band cohorts born from 1891 to 1985 in 146 countries. One drawback of the Barro-Lee dataset is that the measurement of educational attainment varies for some countries [10]. This is however not a major concern in our context, as we are mostly interested in variations of the enrollment rate by gender within countries, rather than across countries.

4.2 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 among the enrolled 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 $x = E(.) : \bar{z} \to [0,1]$, given the two underlying distributions G_{z_f} and G_{z_m}. Contrary to $\bar{z}_{it}$, x_{it} is observed in the data. Assuming normality, the two distributions are fully characterized by the two-parameter vectors $(\mu_m,\, \sigma_m^2)$ and $(\mu_f,\, \sigma_f^2)$, respectively.

Without loss of generality, the parameters of our model can 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 $(\mu_i,\, \sigma_i^2)$ the first two moments of the males' test taking ability distribution relative to females in country i. Formally,

$$\mu_i = \frac{\mu_{i,m} - \mu_{i,f}}{\mu_{i,f}} = \mu_{i,m} \tag{19}$$

[9]The aggregate Barro-Lee database is 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.

[10]For more details about the construction of the Barro-Lee dataset, see Barro and Lee (2010)

$$\sigma_i = \frac{\sigma_{i,m}}{\sigma_{i,f}} = \sigma_{i,m} \tag{20}$$

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, and assumed to be normally distributed. Their mean and standard deviation are allowed to differ across countries. We are interested in estimating the following model:

$$y_{it} = \frac{G_{z_f}(\bar{z}_{it})}{G_{z_m}(\bar{z}_{it}, \mu_{it}, \sigma_{it})} \cdot \exp(\epsilon_{it}), \tag{21}$$

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

$$\log y_{it} = \log G_{z_f}(\bar{z}_{it}) - \log G_{z_m}(\bar{z}_{it}, \mu_{it}, \sigma_{it}) + \epsilon_{it}, \tag{22}$$

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

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

$$G_{z_f}(\bar{z}_{it}) = \int_{\bar{z}_{it}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}}\, dz = \frac{1}{2}\left(1 - erf\left[\frac{\bar{z}_{it}}{\sqrt{2}}\right]\right) \tag{23}$$

$$G_{z_m}(\bar{z}_{it}) = \int_{\bar{z}_{it}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{\frac{(z-\mu_m)^2}{2\sigma_m^2}}\, dz = \frac{1}{2}\left(1 - erf\left[\frac{\bar{z}_{it} - \mu_m}{\sqrt{2}\sigma_m^2}\right]\right) \tag{24}$$

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

$$x_{it} = E(\bar{z}_{it}) \equiv \frac{G_{z_f}(\bar{z}_{it}) + G_{z_m}(\bar{z}_{it})}{2} = \frac{1}{4}\left(2 - erf\left[\frac{\bar{z}_{it}}{\sqrt{2}}\right] - erf\left[\frac{\bar{z}_{it} - \mu_m}{\sqrt{2\sigma_m^2}}\right]\right) \tag{25}$$

Therefore, $\bar{z}_{it} = E^{-1}(x_{it}, \mu_{it}, \sigma_{it})$. Since the error term ϵ_{it} is normally distributed, the model can be fitted numerically by finding the values that minimize the sum of squared errors of the likelihood function. 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 standardized.

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

$$L(\{\theta_i|z\}) = \frac{1}{(2\pi\sigma_\epsilon^2)^{\frac{n}{2}}} \cdot \exp\left\{-\frac{1}{2\sigma_\epsilon^2} \cdot \sum_{i=1}^{n} \left(\log y_{it} - \log G_{z_f}(E_i^{-1}(x_{it})) + \log G_{z_m}(E_i^{-1}(x_{it}))\right)^2\right\} \tag{26}$$

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

$$\hat{\theta_{MLE}} = \min_{\theta_i \in \Theta} \sum_{i=1}^{n} \left\{\log y_{it} - \log G_{z_f}(E_i^{-1}(x_{it}|\theta_i)|\theta_i) + \log G_{z_f}(E_i^{-1}(x_{it}|\theta_i)|\theta_i)\right\}^2 \tag{27}$$

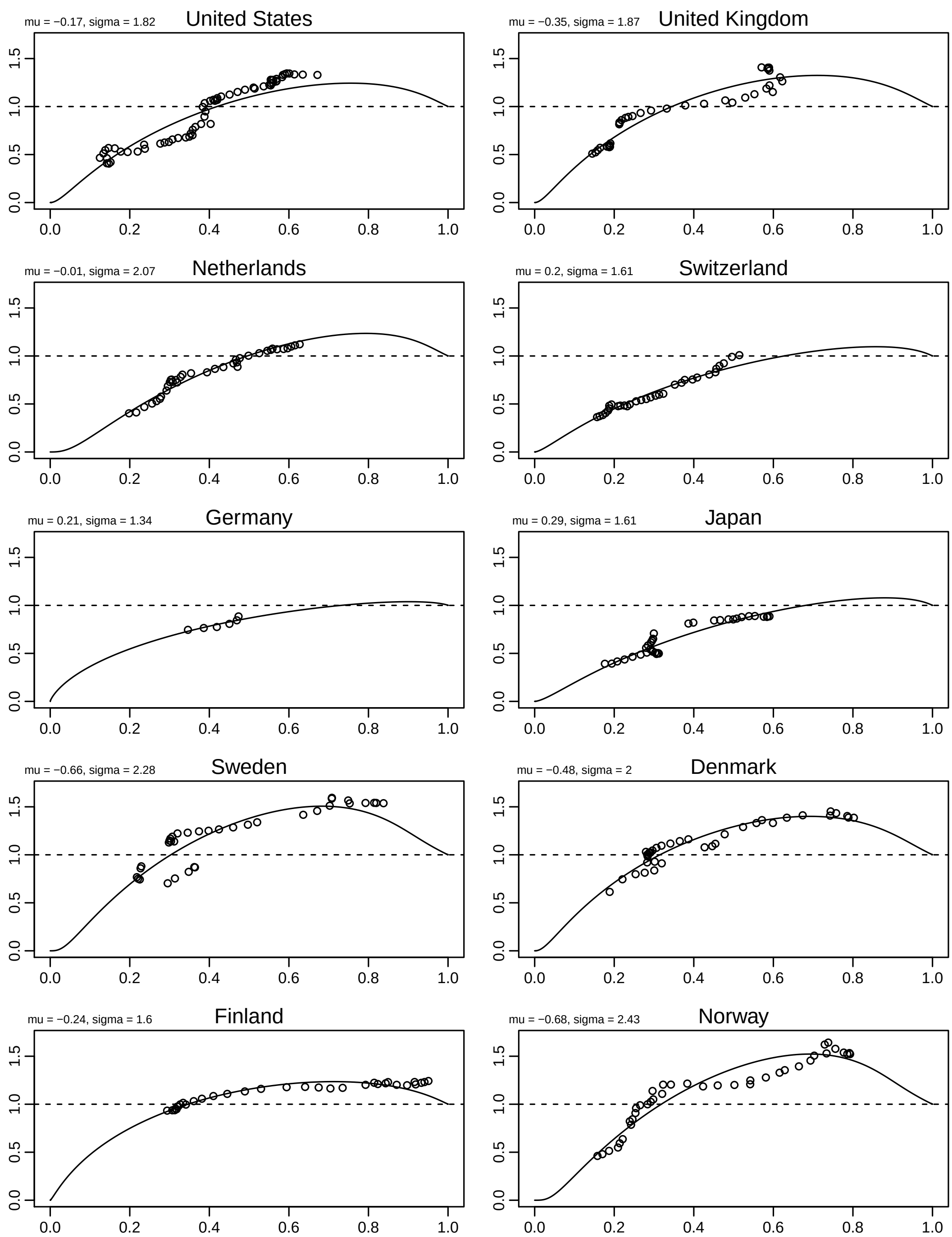
5 Results

5.1 Model Fit

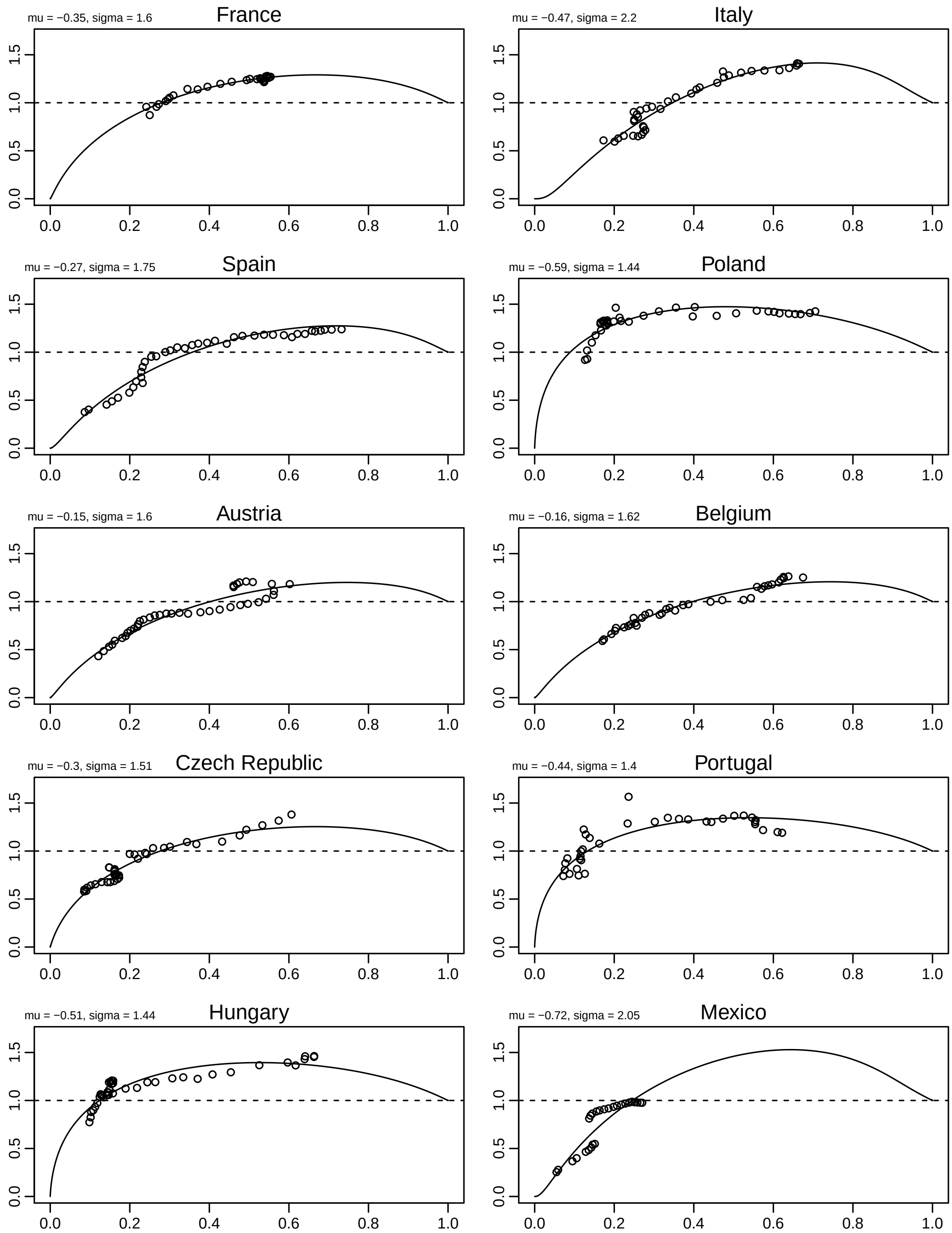
Figure 7 reports the estimated relationship between the total enrollment rate in tertiary education x and the female-to-male ratio y when $\{\mu_i; \sigma_i\}$ are estimated from the $2 \times T$ matrix, as described in the previous section. As shown in the figure, our model generates an accurate fit for the relationship between x and y observed in each individual country. Figure 8 and Figure 9 show our model fit for the relationship between the secondary school non-completion rate, and the gender ratio among non-completers. Our model also appears to generate a very satisfactory fit for the gender ratio in the lower tail of educational attainment.

Contrary to data on tertiary enrollment, the secondary school non-completion rate ranges from virtually 100% of non-completion for cohorts born at the end of the 19th century until less than 10% nowadays in some countries. It therefore allows to reconstruct almost the entire path of the gender ratio among non-completers as a function of non-completion rates. Figure 9 reports our model fit for two countries, France and Spain for which we were able to retrieve time-series of cohort born from the end of the 19th century onwards. It shows a very peculiar non-linear relationship between the non-completion rate and the gender ration among non-completers. At low levels of secondary school non-completion, the female-to-male ratio among non-completers gradually increases with the non-completion rate before reaching a point at which the female-to-male ratio among non-completers is larger than 1. However, once the non-completion rate has reached a certain threshold, the female-to-male ratio starts decreasing before converging back to 1 for a non-completion rate of 100%. To the best of our knowledge, our theory is the only contribution being able to account for this particular non-linear relationship between the secondary school non-completion rate and the gender ratio among non-completers.

Figure 7: Model Fit - Gender Ratio among the Enrolled to Tertiary Education

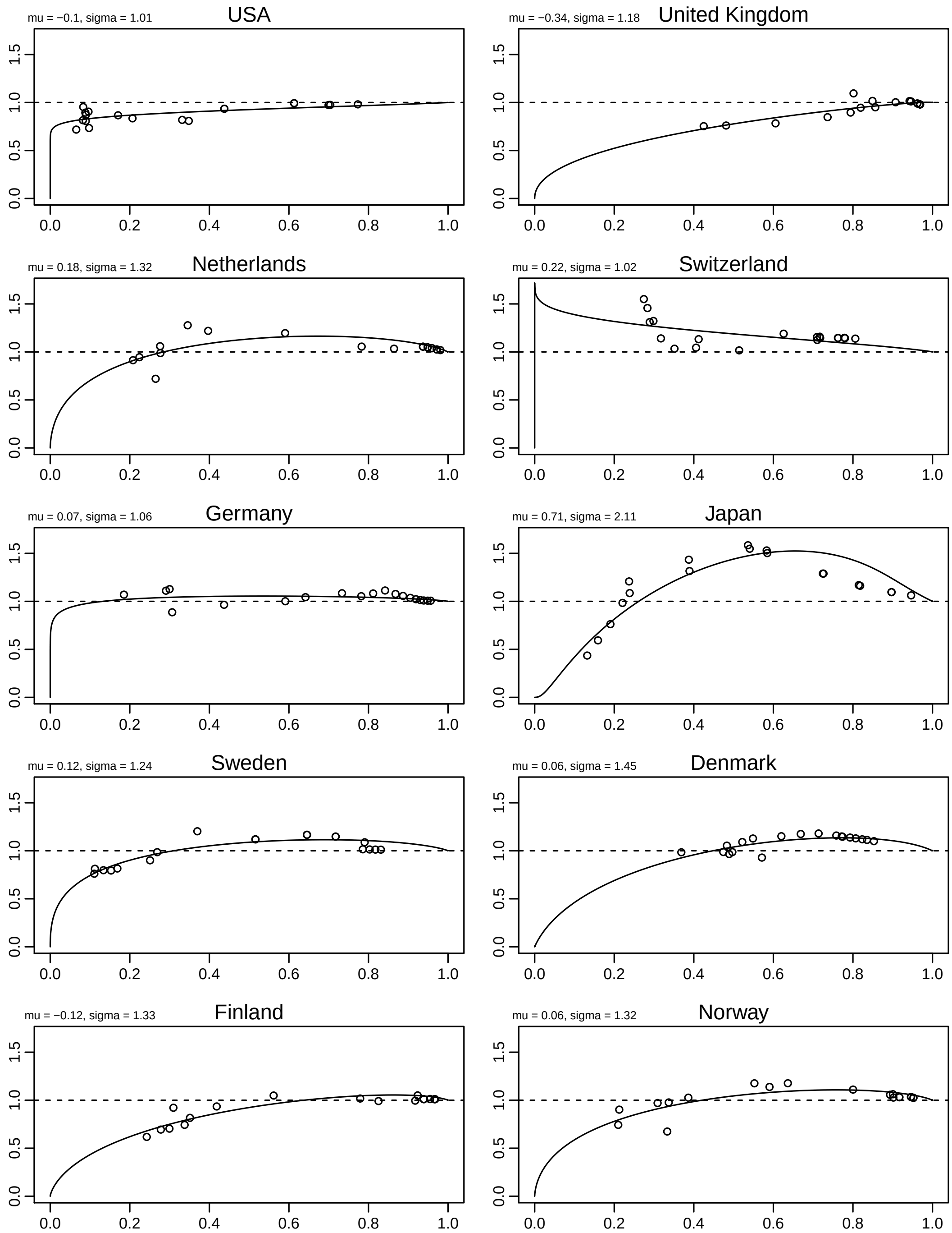


Notes. The x-axis measures enrollment in tertiary education for country i. The y-axis measures the females-to-males ratio in tertiary education for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ *for country i from the UNESCO Institute of Statistics. The full line depicts the 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.*

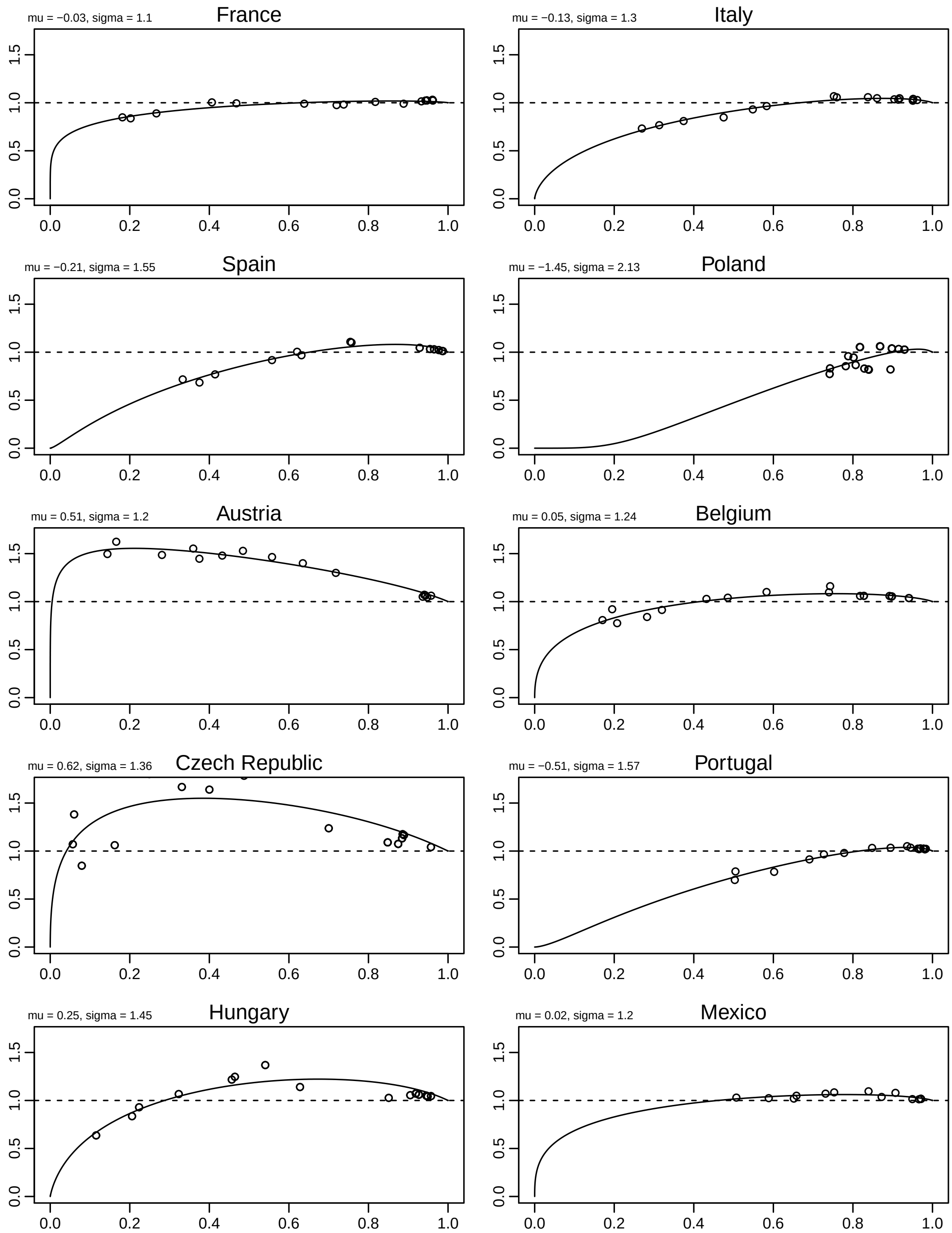


Notes. The x-axis measures the enrollment rate in tertiary education for country i. The y-axis measures the females-to-males ratio in tertiary education for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ *for country i from the UNESCO Institute of Statistics. The full line 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.*

Figure 8: Model Fit - Gender ratio among Secondary School Non-Completers

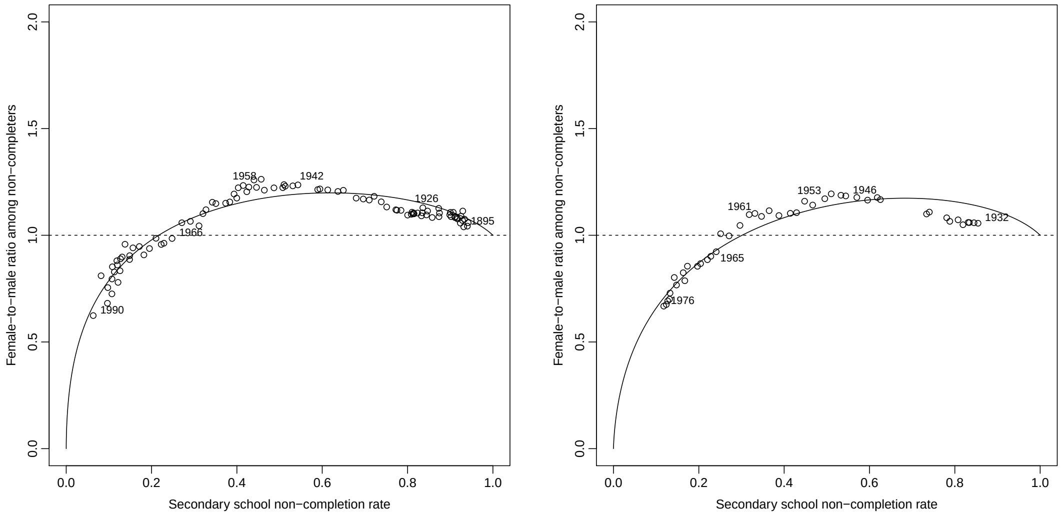


Notes. The x-axis measures the secondary school non-completion rate for country i. The y-axis measures the female-to-male ratio among secondary school non-completers for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ for country i from Barro-Lee (2010). The full line depicts 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.



Notes. The x-axis measures the enrollment in tertiary education for country i. The y-axis measures the female-to-male 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.*

Figure 9: Model Fit - Gender ratio among Secondary School Non-Completers with Full time-series



5.2 Evaluating our Estimates against Gender Differences in PISA Score Distributions

Our model is: $\log y_t = \log R(E^{-1}(x_t, \mu, \sigma), \mu, \sigma) + \epsilon_t$. Given our estimates for the parameters μ and σ, the model predicts the gender ratio $\hat{y}$ among the enrolled, for a given value of the 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 distributions, 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 are both available in the UNESCO and PISA datasets. In total, 40 countries are common to both sources. The intuition behind this exercise is to test whether our model captures true country-specific gender differences in test score distributions. If we find no or little association between $\{\hat{\mu}_i\ \hat{\sigma}_i\}$ estimated from our model and PISA distributions, this could indicate that our estimates mostly capture noise, and that the gender gap reversal has little to do with gender differences in test score distributions. On the other hand, a strong correlation would reveal that time-series on enrollment rates by gender carries substantial information on gender differences in test scores for a given country, given our framework. This would suggest that gender differences in test score distributions are an important driver of the empirical relationship between total enrollment rate and gender ratio among the enrolled.

PISA assessments provide a suitable benchmark to evaluate estimates from our model. 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. It is therefore reasonable to think of PISA as capturing true gender differences in test-taking ability at age 15 for the entire population of a given country.

Table 3 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. We report correlations for 3 different values of x, $x = 0.20$, $x = 0.50$ and $x = 0.70$. 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. This suggests that our estimates from time-series on enrollment rates using our model capture actual gender differences in test score distributions observed across countries. Importantly, it should be noted that a country like Korea, where the reversal has not been observed yet despite a high total enrollment, exhibits one of the lowest male-to-female variance ratio of PISA. This is consistent with our model, which predicts that countries with low male-to-female variance ratio should expect a later reversal, for a given enrollment in tertiary education.

The fact that our model estimates does not perfectly correlates with gender differences found in PISA can be explained by several factors. One important source of measurement error is that we are comparing 18-year-olds with 15-year-olds. PISA assessment at indeed taken at age 15, while we estimate our model for individuals who are enrolled in tertiary education, and are therefore 18 and older. Therefore, any country-specific change in the relative distributions during that age gap will induce measurement errors. In this respect, Kanazawa and Lynn (2011) report evidence suggesting that male cognitive abilities mature later than females in the UK, especially after age 16. As cognitive abilities are certainly captured by test scores, this would imply that gender differences in mean test score are not the same at age 15 as they are age 18 leading to discrepancies between our model estimates and PISA estimates.

Second, our estimates use the gross enrollment ratio as a proxy for its theoretical counterpart $E(\bar{z})$in our model. As we already emphasized, the gross enrollment rate is an imperfect proxy, which is also likely to generate measurement error in our estimates.

Third, test scores are an imperfect measure of its test-taking ability, its theoretical counterpart, that is unobservable by the econometrician. As emphasized by Heckman et al. (2012), measured test scores are the observable outcome of a complex combination of cognitive and non-cognitve abilities, as well as unobservable effort. Understanding the mapping from test-taking ability to test score is a challenging task that is beyond the scope of this paper. The role played by effort in test score production is likely to introduce noise in measured ability, especially if effort is correlated with gender.

Finally, as emphasized in Hoxby (2012), test-taking ability is certainly not the only relevant variable for enrollment in education decisions. Variables such as parental envrironements or networks are also likely to play a role. As a result, observable educational enrollment variables are likely to capture these additional factors, thereby generating measurement error in our estimates for country specific-gender differences in ability distributions.

Table 3: Correlations between predicted gender ratio $\hat{y}$ from our model estimates and actual PISA distributions, for alternative values of the enrollment rate x

	Predicted $\hat{y}$ - Model estimates		
	x=0.2	x=0.5	x=0.7
Predicted $\hat{y}$ - PISA Math	0.332*	0.335*	0.317*
	(0.059)	(0.057)	(0.072)
Predicted $\hat{y}$ - PISA Reading	0.488***	0.429**	0.362**
	(0.004)	(0.013)	(0.039)

Source. PISA 2000 *Notes.* ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Sample size is 40 countries. p-values are reported into parenthesis.

6 Test Against Alternative Hypothesis

In previous sections, we have shown that our theory is consistent with empirical data and can generate the gender gap reversal observed empirically. The second step is to show that alternative explanations proposed by the literature are inconsistent with some patterns of the data that our theory can explain. Previous theories have been focusing on accounting for the gender gap reversal in participation to tertiary education in the US context. To the best of our knowledge, 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 majority in tertiary education. Second, a relative increase in females' mean test-taking ability over time can also generate a reversal in the educational gender gap, as suggested by Cho (2007).

In this section, we propose and perform several tests to assess the validity of these two alternative theories against empirical data. We first formulate them in the common framework we developed in the previous chapters, and then assess their implications against our alternative theory. All three hypotheses can be stated within the theoretical framework we established in previous sections:

Hypothesis 1: *Higher male dispersion'*: The function $G_z(\cdot)$ is gender-specific, with $Var[z_m] > Var[z_f]$. The progressive increase in the net benefits to education b (or, equivalently, a decrease in $\bar{z}$), combined with the higher dispersion in test-taking ability for males, explains the reversal.

Hypothesis 2: *Change in social norms*: The test-taking ability distribution $G_z(\cdot)$ is the same for both genders. The net benefits of education differ between gender, i.e. there exist gender-specific b_f and b_m (or, equivalently, $\bar{z}_f$ and $\bar{z}_m$) for females and males, respectively. In the past, $b_f < b_m$ (or, equivalently $\bar{z}_f > \bar{z}_m$) before progressively converging and surpassing b_m over time, generating the reversal.

Hypothesis 3: *Increase in females' mean performance*: The variance of the ability distributions and the net benefits to education b (or, equivalently, the test-taking ability threshold $\bar{z}$) are the same for both genders. The mean of test-taking ability for females has increased over time and progressively surpassed the male mean, generating the gender gap reversal.

These three hypotheses could be combined into joint hypotheses. We however analyze them separately to compare their respective explanatory power and consistency with empirical data.

6.1 The Change in Social Norms Hypothesis (Hypothesis 2)

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

$$E = 1 - F_z(\frac{\bar{s}}{b}) = G_z(\frac{\bar{s}}{b}) = G_z(\bar{z}) \tag{28}$$

We can formulate the change in social norms hypothesis in this setting by allowing b to differ between gender and have different evolutions over time, 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 surpassing b_m over time. Under hypothesis 2, 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:

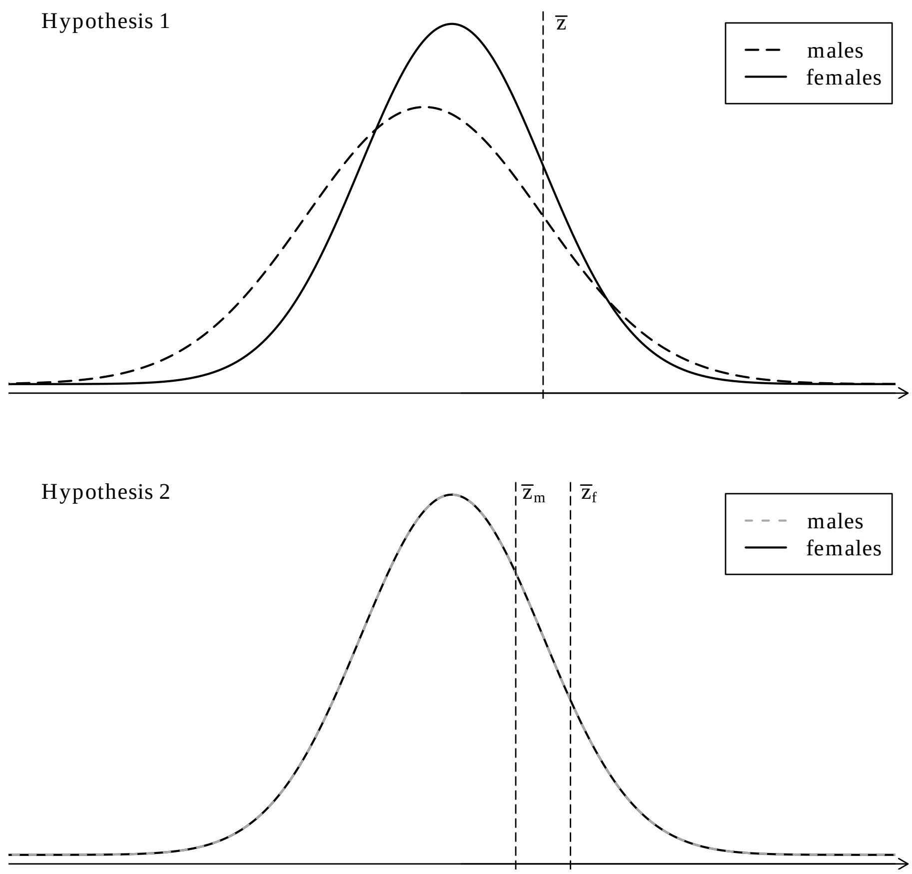
$$E = \begin{cases} G_z(\frac{\bar{s}}{b^m}) = G_z(\bar{z}_m) & \text{if male} \\ G_z(\frac{\bar{s}}{b^f}) = G_z(\bar{z}_f) & \text{if female} \end{cases}$$

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 participation to education. In other words, the net benefits to education 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 the monetary returns to tertiary education is what is typically referred as the college wage premium, which is the wage difference between individuals that attended tertiary education and those who did not. Chiappori at al. (2009) Card and DiNardo (2002), or Charles and Luoh (2003) ind 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 challenged by Hubbart (2011). In particular, the gender difference in the college premium vanished once a bias associated with income top-coding in the dataset used by US studies is corrected for. Cho (2007) further points out that trends in the college 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, Becker et al. (2011) argue that most non-monetary 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 is unable to account for the reversal of the gender ratio among the enrolled.

Our framework allows to formally test for the validity of this alternative hypothesis proposed by previous literature. We can first fit Hypothesis 2 using time-series data on enrollment

Figure 10: Hypothesis 1 and 2 - Illustration



rates, by allowing b^f and b^m to take different values. Hypothesis 2 is by nature very flexible for fitting the data. Since the relative changes of $\bar{z}_f$ and $\bar{z}_m$ have not been constrained, we can calculate the values algebraically. The calculated values are depicted in Figure 11. For each time period, we solve a system of two equations:

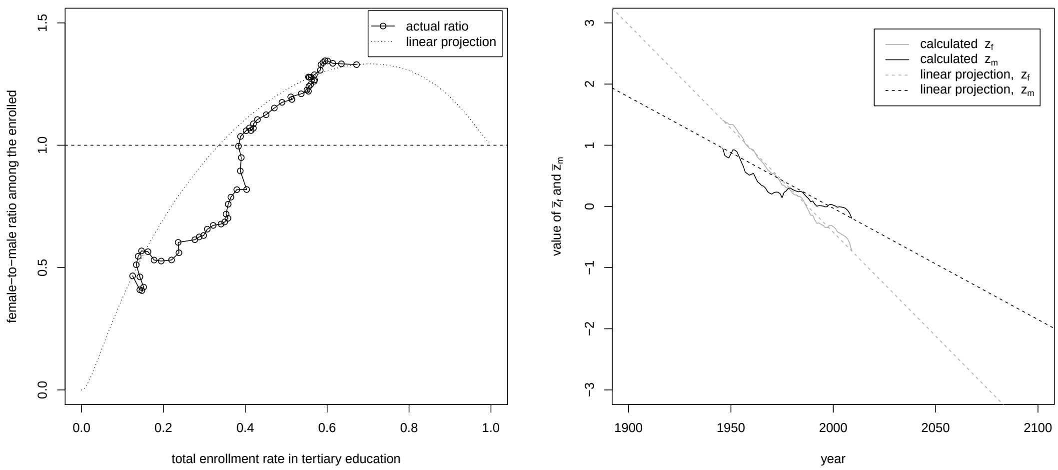
$$\begin{cases} y_t = \frac{G_z(\bar{z}_{f,t})}{G_z(\bar{z}_{m,t})} \\ x_t = \frac{G_z(\bar{z}_{f,t})+G_z(\bar{z}_{m,t})}{2} \end{cases}$$

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

$$y_t = \frac{G_z(\bar{z}_{f,t})}{2x_t - G_z(\bar{z}_{f,t})} \tag{29}$$

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 11.

Figure 11: The change in social norms hypothesis fitted, with projections.



Notes. The graphs show the values of calculated $\bar{z}_f$ and $\bar{z}_m$, given the gender ratio and gross enrollment rate in 1946 to 2009 for the US. The projections are made assuming an evolution of $\bar{z}_f$ and $\bar{z}_m$ that follows the average of the calculated years before and after the data range used.

6.2 Hypothesis 2 Versus Hypothesis 1: Test 1

Hypothesis 2 appears to do an excellent work at explaining the reversal, which is natural as it allows for an exact fit. However, it can be shown that Hypothesis 1 and Hypothesis 2 have opposite implications regarding the relationship between the total enrollment rate defined as:

$$E(\bar{z}_f, \bar{z}_m) = \frac{G_{z_f}(\bar{z}_f) + G_{z_m}(\bar{z}_m)}{2} \tag{30}$$

and the difference in mean test-taking ability of females and males enrolled in tertiary education:

$$W(\bar{z}_f, \bar{z}_m) = E[z_f | z_f > \bar{z}_f] - E[z_m | z_m > \bar{z}_m], \tag{31}$$

This comes from the fact that Hypothesis 1 claims that z_f and z_m have different distributions with $\bar{z}_f = \bar{z}_m$, and *vice versa* for Hypothesis 2. 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 fraction of population enrolling in tertiary education 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.

Formally, Hypothesis 1 implies:

$$\frac{\partial W(\bar{z}_f, \bar{z}_m)}{\partial E(\bar{z}_f, \bar{z}_m)} > 0 \tag{32}$$

While Hypothesis 2 predicts:

$$\frac{\partial W(\bar{z}_f, \bar{z}_m)}{\partial E(\bar{z}_f, \bar{z}_m)} < 0 \tag{33}$$

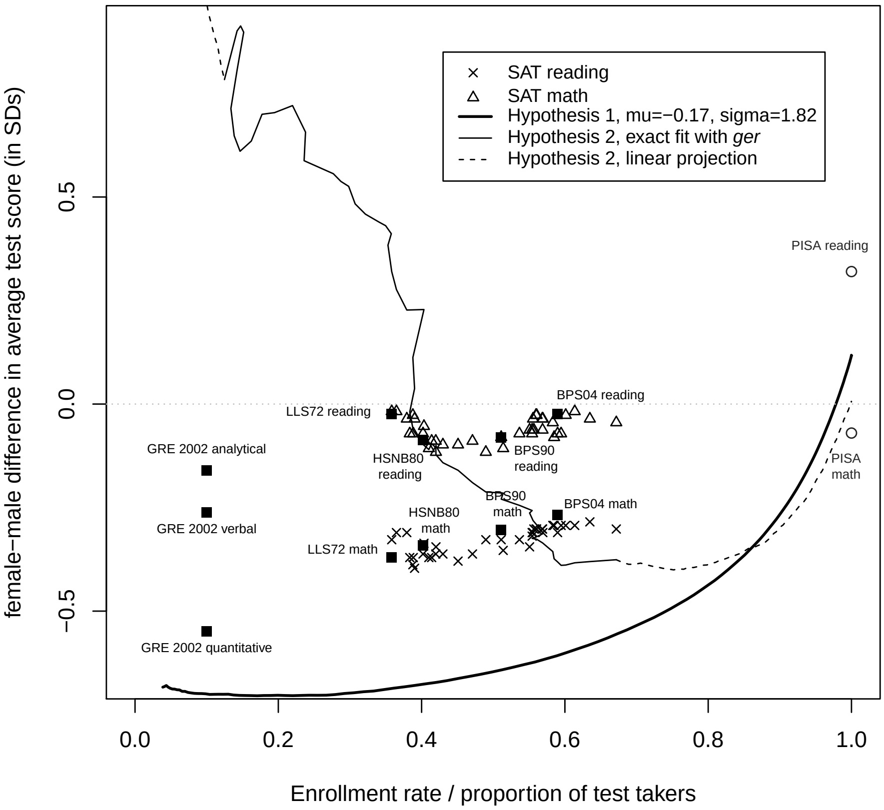
Testing for these two opposite predictions requires data on the average test score by gender among individuals enrolled in tertiary education. To the best of our knowledge, there does not exist a standard achievement test taken by post-secondary students that is comparable nationwide. Instead, we use data from the Scholastic Achievement Test (SAT) by gender that provides average mean scores in mathematics and reading by gender from 1970 to 2010. We are also able to link this data to the total number of females and males taking the test in a given cohort. We also use test score data from PISA in both mathematics and reading, which are taken by a representative sample of the entire population of 15-year-olds in a given country.

Figure 12 shows that Hypothesis 1 generates an increase in the average level of test-taking ability of females attending university relative to males over time. On the other hand, it reports that the change in social norms hypothesis (Hypothesis 2) 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 12 reveals that Hypothesis 1 performs well at predicting the relationship between the mean gender gap in test scores and the fraction of the population taking the test or enrolling into tertiary education. First, it is consistent with a negative female-to-male difference in mean test score among the enrolled at low level of enrollment observed in the data.. Second, it also consistent with a progressive decrease in the negative female-to-male difference in mean test score among the enrolled.On the other hand, the 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.

In addition, our theory is also consistent with males performing better in SAT while males do relatively better in PISA in the same discipline. The data shows that in a sample of the entire population as in PISA, females obtain higher average test scores relative to males in reading. On the other hand, males perform on average better in the same discipline in the SAT test. The higher male variability hypothesis (Hypothesis 1) 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 sample truncation, combined given the larger of test-taking ability by gender assumed in our model can explain the differences in the mean gender gap observed in PISA and SAT scores. On the other hand, hypothesis 2 (change in social norms hypothesis) is unable to explain why females would perform better in PISA, while doing worse in PISA in a same given year.

Figure 12: Fit for the empirical relationship between enrollment rate and gender gap in the mean score of the enrolled: Our model Vs the social norm hypothesis



Notes. The x-axis reports the fraction of individuals in the cohort taking the achievement test. The y-axis measures the female-to-male difference in means of the given achievement test, expressed in standard deviations. The thick line depicts the expected relationship between y and x predicted by our model (hypothesis 1). The thin line depicts the relationship between x and y predicted by the change in social norms hypothesis (hypothesis 2). The cross and triangle dots represent the actual value of the gender difference in SAT test scores for the entire population of SAT takers, as a function of the fraction of SAT takers in the population. The squared-dots represent the gender difference in mean test scores at age 15 for individuals who enrolled into tertiary education, from US longitudinal surveys.

6.3 Hypothesis 2 Versus Hypothesis 1: Test 2

Another way to assess the higher male dispersion hypothesis against the change in social norms hypothesis is to look at the evolution over time of the relationship between test scores and enrollment in tertiary education by gender. Under Hypothesis 1, the relationship between enrollment and test scores at the individual level at time t is given by:

$$s = z \cdot b_t \tag{34}$$

where the relationship between s and z is identical across genders, and given by the parameter b. This implies that the empirical relationship between z and s should be identical for both genders and there should not be any gender-specific change in the relationship between test-taking ability z and tertiary enrollment s over time.

Under Hypothesis 2, however, the relationship between individual enrollment decision and test scores differs between genders and is given by:

$$s = \begin{cases} z \cdot b_m, t & \text{for males} \\ z \cdot b_f, t & \text{for females} \end{cases}$$

Hypothesis 2 implies that the relationship between enrollment decisions and test scores is gender-specific and changes over time. In particular, a testable implication is 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 (b_f>b_m).

Since G_z is identical for both genders under hypothesis 2, a necessary condition for the gender gap reversal is:

$$\begin{cases} b_{f,t_1} - b_{m,t_1} > 0 \\ b_{f,t_1} - b_{m,t_1} < 0 \end{cases}$$

Where t_1 and t_2 denote a time prior and posterior to the gender gap reversal, respectively. On the other hand, our theory (Hypothesis 1) predicts:

$$\begin{cases} b_{f,t_1} - b_{m,t_1} = 0 \\ b_{f,t_1} - b_{m,t_1} = 0 \end{cases}$$

Testing these alternative hypotheses empirically is not a straightforward task, as this requires data on test scores prior and posterior to the gender gap reversal that can be linked to tertiary education attendance. According to the National Center of Educational Statistics (NCES) figures, the reversal in participation to tertiary education occurred in the mid-1980s for the US. We therefore need to observe the relationship between test scores and tertiary education enrollment both prior to 1985 and posterior to 1985, with a sufficient time gap to capture potential trends over time. Two longitudinal surveys conducted in the US in 1980 and 2002 allow to do so. The first of these surveys is the High School and Beyond 1980, which follows a cohort of 10th graders in 1980 until their post-secondary studies. The second survey is the US Educational Longitudinal Study 2002, which also follows 10th graders in 2002 until their tertiary education studies. In addition, tests scores are on the same scale in the two surveys and have been designed to be comparable. These tests are also taken at an age at which school is still compulsory, which ensures having a consistent estimate of average test scores by gender for the entire age population. Finally, since the data is longitudinal, it allow linking test-scores at age 15 with tertiary education attendance a few years later.

By using these two surveys, we are able to estimate the empirical relationship between test-taking ability z, and enrollment in tertiary education H both prior to the reversal in 1980 and after the reversal in 2002. In particular, we can obtain an estimate β of the net benefits to education b in our simplified Card model:

$$H_j = z_j \beta + \epsilon_j \tag{35}$$

where β is allowed to be gender-specific and to take different values in 1980 and 2002. H_j is a dummy for tertiary education attendance and z_j is the test score at age 15. Table 4 reports the results of our estimation of $(\beta_{m,y=1980}, \beta_{f,y=1980})$ and $(\beta_{m,y=2002}, \beta_{f,y=2002})$

Table 4 reveals several important patterns regarding the net benefits to education by gender in 1980 and 2002. The estimated $\hat{\beta_f} - \hat{\beta_m}$ is positive in both 1980 and 2002, meaning that girls receive higher net benefits to tertiary education before and after the reversal. In other words, girls enroll significantly more at university than boys conditional on test scores. Importantly in our context, this was already true in 1980 prior to the reversal and the estimated $\hat{\beta_f} - \hat{\beta_m} -$ is very stable between 1980 and 2002. If anything, there was a slight decrease in the females' advantage in the benefits benefits to education from 0.74 to 0.65. This is inconsistent with the predictions of the change in social norm hypothesis, which implies a relative increase in the net benefits to education for females, with $\hat{\beta_f} - \hat{\beta_m} > 0$ prior to the reversal and $\hat{\beta_f} - \hat{\beta_m} < 0$ after the reversal. On the other hand, the stability of $\hat{\beta_m} - \hat{\beta_f}$ over time seems to give further credit to our theory.

Table 4: Estimated net benefits to education β by gender in 1980 and 2002 - Probit regressions

	Dep. Variable: Dummy for Tertiary Education Attendance					
	1980			2002		
	Males	Females	F-M Diff.	Males	Females	F-M Diff.
$\hat{\beta_m}$	2.55***			4.14***		
	(0.11)			(0.06)		
$\hat{\beta_f}$		3.29***			4.79***	
		(0.10)			(0.02)	
$\hat{\beta_f} - \hat{\beta_m}$			0.74***			0.65***
N. observations	5,083	5,859		6,023	6,515	

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Reported coefficients are marginal effects. Regressions do not include an intercept.
Sources. High School and Beyond -base year (1980) and 2nd follow-up (1984), US Educational Longitudinal Study - base year (2002) and 2nd follow-up (2006)

6.4 Increase in Females' Mean Performance (Hypothesis 3)

Our higher male dispersion hypothesis (Hypothesis 1) relies on the assumption that test score distributions for males and females are fixed over time and, in particular, that the gender difference in mean test score μ_f - μ_m is time-invariant. We now consider the possibility that the average performance of girls relative to boys increased over time, which can also lead to the observed gender gap reversal in participation to education. In our framework, this corresponds to a shift of f_{z_f} to the right relative to f_{z_m}, leading to an increase in $E(z_f)$ relative to $E(z_m)$. This possibility has been investigated by Cho (2007) and Fortin et al. (2013) for the US. Fortin et al. (2013) find no relative increase in girls' self-reported grades over the period 1970–2010. Using data from the Monitoring the Future study, they report an increase in high school grades for both boys and girls with parallel trends over the period, with girls already outperforming boys in the early 1970s. Cho (2007), 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.

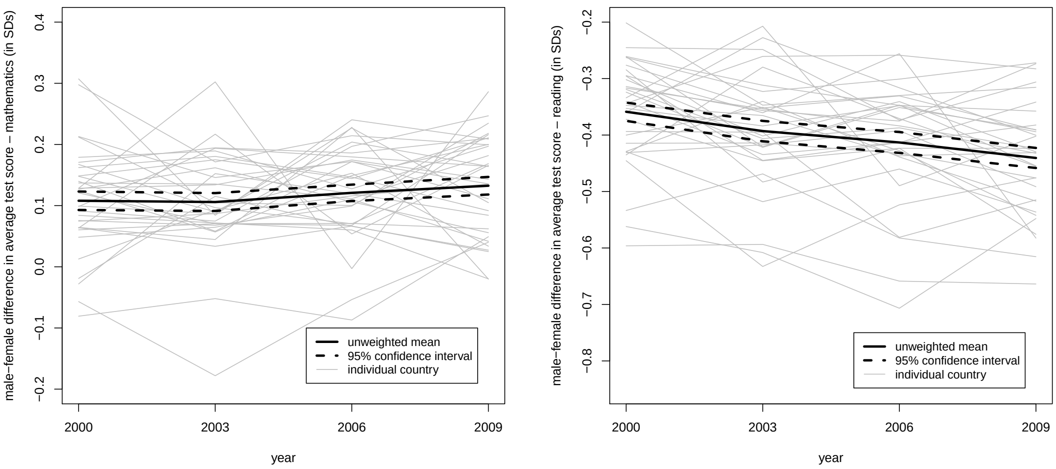
An important limitation of the analysis in our context is that it uses a sample of high-school seniors, which are beyond the compulsory age of high school attendance. Early high-school dropouts are therefore excluded from the sample, which may generate biases in the estimated gender differences in mean test scores for the entire population, and its change over time. This phenomenon is referred as attrition bias or sample restriction in the literature, and has strong implications in our context where the variance of test scores can differ between genders[11]. In the presence of sample restriction, the mean of the truncated distribution of student test scores is a function of the first two moments of the underlying distribution, and of the ability threshold for truncation. With a higher variance of test scores for men in the entire population, sample restriction will generate changes in the observed mean performance between genders over time, even if the mean of the distribution for the entire population did not change.

To evaluate whether the mean performance of the female population increased over time relative to boys for the entire population, one should use a representative sample of the 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. 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 13 depicts the evolution of girls' mean average performance relative to boys in reading and mathematics over the period 2000-2010, for 41 countries sampled in all waves PISA. It shows 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-span of the PISA study, we complement our analysis by looking at the evolution of the performance of secondary school students by gender in the US, over the period 1980–2002. To this purpose, we use two nationally representative longitudinal surveys of secondary 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 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 deviations from 0.148 standard deviation. Although these figures seem to suggest that girls' average performance in high

[11] Using data from the 1970 British Cohort Study, Deary et al. (2007) have shown that observed differences in mean IQ scores can be partly created by the combination of sample restriction, and a larger variance of cognitive abilities for males.

Figure 13: Average PISA performance of males relative to females in mathematics and reading: 2000-2009



Note. The sample includes 41 countries that are included in all waves of PISA in 2000, 2003, 2006 and 2009

school increased relative to boys over the period, they should be interpreted with care. One should keep in mind that such evidence is restricted to the US. Data availability unfortunately does not allow to repeat a similar exercise for other countries over the same period.

Table 5: Gender difference in mean test score at 15 in the US: 1980 Vs 2002

	1980				2002			
	Female	Male	M-F mean diff. (in s.d.)	M-to-F Var. ratio	Female	Male	M-F mean diff (in s.d.)	M-to-F Var. ratio
Mean test score	-0.075	0.084	0.159***		-0.053	0.053	70.106***	
Test score sd	0.952	1.044		1.20***	1.037	0.959		1.17***
Observations	5,083	5,859			6,023	6,515		

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level.
Sources. High School and Beyond 1980 (base year) and US Educational Longitudinal Study 2002 (base year).

7 Conclusion

We developed a comprehensive framework that can explain gender differences in educational enrollment. Our model allows to reconcile three stylized facts that are robust internationally: the higher variance of test score distribution among males, the gender gap reversal in participation to tertiary education, and the gender gap reversal in secondary school non-completion. Despite data limitations and the inherent difficulty in measuring test-taking ability, the model

Table 6: Gender difference in mean test score at 11 in the UK: 1969 Vs 2010

	1969				2010			
	Female	Male	M-F mean diff. (in s.d.)	M-to-F Var. ratio	Female	Male	M-F mean diff (in s.d.)	M-to-F Var. ratio
Mean test score	-0.034	0.034	0.068***		-0.034	0.035	0.069***	
Test score sd	0.960	1.038		1.17***	0.973	1.025		1.12***
Observations	868	868			6,557	6,611		

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level.
Sources. High School and Beyond 1980 (base year) and US Educational Longitudinal Study 2002 (base year).

provides a very satisfactory fit for the empirical relationship between the total enrollment rate in tertiary education and the gender composition of the enrolled. It is also the first theory to account for the reversal of the gender gap in secondary school non-completion. Importantly, the model can not only account for the gender gap dynamics within countries, but can also explain cross-country differences in enrollment by gender. Last but not least, this is the first contribution to formulate and test competing explanations to the gender gap reversal in a common setting. It turns out that our model can explain some patterns of the data that previous theories are inconsistent with.

This paper suggests that when looking at gender differences in observable outcomes, it is important to go beyond the analysis of means by looking at entire distributions. We showed that gender differences in the variance of abilities or traits is a strong candidate in explaining gender differences in educational attainment. Likewise, it is legitimate to expect that the larger variance of traits among males may be relevant to explain observed gender differences in other areas. Further research building on this fact in labor economics and other fields of economics has a strong potential. In addition, the larger variability of men's test score distribution observed empirically remains mostly unexplained, and stands as another promising area for future research.

8 References

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

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

Acemoglu, D. and D. Autor (2011), "Skills, tasks and technologies: Implications for employment and earnings." Handbook of Labor Economics, 4, 1043–1171.

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.

Becker, G.S., Hubbard, W.H. and K.M. Murphy. 2010 "The Market for College Graduates and the Worldwide Boom in Higher Education of Women" American Economic Review, 100 (2): 229–233

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.

Card, D. 1994. "Earnings, Schooling, and Ability Revisited," NBER Working Papers 4832, National Bureau of Economic Research, Inc.

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.

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.

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

Chiappori, P.A., Iyigun, M. and Y. Weiss. 2009. "Investment in Schooling and the Marriage Market." American Economic Review, 99(5): 1689–1713.

Cho, D. (2007), "The role of high school performance in explaining womenÕs rising college enrollment." Economics of Education Review, 26, 450–462.

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.

Deary, I.J., D. Dykiert, D., and C.R. Gale (2009), "Are apparent sex differences in mean IQ scores created in part by sample restriction and increased male variance?" Intelligence, 37, 42–47.

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

Ellis, H. (1894), Man and woman. Scott.

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

Fortin, N.M., Oreopoulos P. and S. Phipps (2013), "Leaving Boys Behind: Gender Disparities in High Academic Achievement." Technical report, National Bureau of Economic Research.

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

Goldin, C., Katz, L.F. and I. Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." Journal of Economic Perspectives 20(4): 133–56.

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 26.

Hoxby, C.M. and C. Avery (2012), "The Missing "One-Offs": The Hidden Supply of Hich-Achieving, Low income Students," Technical report, National Bureau of Economic Research.

Heckman, J.J. and T. Kautz (2012), "Hard evidence on soft skills", Labour Economics, 19 (4), 451–464 **Hedges, Larry V. and A. Nowell. (1995)** "Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals". Science 269 (5220): 41–45.

Hubbard, W.H.J. 2011. "The Phantom Gender Difference in the College Wage Premium. "Journal of Human Resources 46(3): 568–586.

Jacob, Brian A., (2002), "Where the boys are not: non-cognitive skills, returns to school and the gender gap in higher education." Economics of Education Review, 21, 589–598.

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.

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.

Kanazawa, S. and R. Lynn (2011), "A longitudinal study of sex differences in intelligence at ages 7, 11 and 16 years." Personality and Individual differences, 51, 321–324.

Machin, S. and T. Pekkarinen, 2008. "Global Sex Differences in Test Score Variability". Science 322 (5906): 1331–2.

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.

Pekkarinen, T. (2012), "Gender differences in education." forthcoming in: Nordic Economic Policy Review.

9 Appendix

10 Technical Appendix

10.1 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'Hôpital 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 $\{\mu_f, \mu_m, \sigma_f, \sigma_m\}$.

10.2 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'Hôpital 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 $\{\mu_f, \mu_m, \sigma_f, \sigma_m\}$.

10.3 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 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, $Var[z_m] > Var[z_f]$.

10.4 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 $\{\mu_f, \mu_m, \sigma_f, \sigma_m\}$.