The Gender Gap Reversal in Education: An Explanation Based on Gender Differences in Test Score Dispersion [*]

Laurent Bossavie & Ohto Kanninen [†]

October 14, 2013

Abstract

In virtually all countries, males show a greater dispersion in test scores relative to females. We show that this fact, together with an increase in the returns to education over the last decades, can explain the gender gap reversal in educational attainment observed internationally. Our model generates an accurate fit for the relationship between the enrollment rate and the gender composition of the enrolled in each individual country. From time-series on enrollment rates, we generate country estimates for gender differences in ability distribution parameters using our model. Our estimates highly correlate with country-specific gender differences in test score distributions found in international assessments. We also assess the validity of our theory against previous explanations proposed by the literature. The data does not support the predictions of these alternative hypotheses, while bringing further support to our theory.

Keywords: Educational outcomes, gender inequality, test scores.

JEL-Codes: I21, J16

[*]We are grateful to Andrea Ichino and Luigo Guiso for constant advice and encouragement. We also would like to thank Jerôme Adda, Christian Dustmann, Peter Hansen, Gábor Kézdi and Tuomas Pekkarinnen, as well as all participants to the EUI micro-econometrics working group for helpful comments and 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. As we show, this reversal has been quasi-universal, and 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 over-represented in secondary school non-completion.

This paper establishes empirically these two quasi-universal facts and proposes a comprehensive theory to account for them. The gender gap reversal in participation to tertiary education has been established for the US, and for the Nordic countries by Pekkarinen (2012)[1]. We show in this paper that this reversal is actually a quasi-universal phenomenon observed in virtually all developed countries, and in the majority of developing countries. In addition, we report an analogous reversal in the lower tail of the educational achievement distribution, among secondary school non-completers. We establish that the gender composition of secondary non-completers has reversed from female majority to male majority over the last decades, and that this reversal is also quasi-universal. To the best of our knowledge, this second fact has not been touched upon by the literature.

Understanding the origins of the gender gap reversal in education is important in its own right, but also for efficiency purposes. In particular, it is crucial to know whether these differences in outcomes observed between genders originate from discrimination and other forms of distortions, or from optimizing behavior based on gender differences in preferences, attitudes, or ability distributions. In addition, explaining the dynamics of the gender gap in education may help understanding gender inequality of outcomes in other areas, in particular on the labour market. This is crucial from both a theoretical, and a policy point of view.

Our explanation for these reversals builds on the insights of Becker et al. (2012). Using the fact that males' test score distribution is more spread out than females', the proposed higher male dispersion theory generates a relationship between the enrollment rate into education and the gender composition of the enrolled. It relies on two main building blocks. The first building block is the larger variance of males' test score distribution, which has received strong empirical support. The second building block is the optimal choice of education made by individuals based on ability, and on the exogenous benefits to education that vary across cohorts.

[1] For evidence for the US, see among others...

Our theoretical framework builds on the behavioral model described in Card (1994), in which the optimal length of schooling chosen by individuals is increasing in individual ability, and in the net benefits to education common to all individuals in the cohort. At the aggregate level, our basic mechanism relies on two ability distributions - one for each gender - whose common ability threshold for truncation determines which individuals enroll into education. The ability threshold decreases with the net benefits to education, as it becomes optimal for less and less able individuals to enroll when the returns to education education increase. In this framework, the net benefits to education and the test score distribution of males and females in the population determine the total enrollment rate and the gender ratio among the enrolled. Since the ability distribution for males and females are assumed to differ, a change in the ability threshold for enrollment not only affects the fraction of individuals enrolling, but also the gender composition of the enrolled.

The relationship between the total enrollment rate and the gender composition generated by our model provides an accurate fit for empirical data. In particular, we are able to reproduce the gender gap reversal observed in both the upper tail and the lower tail of educational attainment. Importantly, we show that the larger dispersion of males' test scores is the only necessary condition required about ability distributions to generate the reversal observed in the data. We demonstrate that no further restriction is needed on the gender difference in mean test score, as long as this condition about the variance holds. Using time-series data for more than 40 countries, we are able to capture the dynamics of the gender ratio of both the upper tail and the lower tail of educational attainment. For the upper tail, our model provides an accurate fit for the joint evolution of the total enrollment rate into tertiary education and the gender composition of the enrolled. In the lower tail, it is also able to replicate the joint evolution of the secondary school non-completion rate and the gender composition of non-completers for all countries in the sample.

Fitting enrollment rate data with our model also allows to extract estimates for country-specific gender differences in test score distribution parameters. To assess the validity of our theory, we confront model estimates to gender differences in test score distributions found in PISA. We find a strong correlation between gender differences in test score distribution parameters estimated from our model, and those found in PISA. To further assess the validity of our theory, we confront it to alternative explanations proposed by the literature in a common framework using empirical data. The higher male dispersion theory appears to perform strongly to alternative theories proposed by the literature

This paper contributes to the literature on the topic in a number of respects. First, it establishes the quasi-universal reversal of the gender gap in both the upper and in the lower tail of educational achievement. Second, it proposes a theory to account or these reversals, based on a stylized fact that received strong empirical support. Finally, it provides a comprehensive and common framework to assess competing explanations that have been proposed by the literature. 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, 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 maximum-likelihood estimation strategy. Section 5 presents our main results. Section 6 tests the validity of our theory against alternative explanations proposed by the literature. Section 7 concludes.

2 Empirical Motivation

2.1 The Larger Dispersion of Males' Test Score Distribution

An important body of literature shows that males' test score distribution exhibits a higher variance than females' for a wide range of cognitive and non-cognitive assessments. While this stylized fact was recently out forward in economics, long-standing evidence is available from psychology. 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 tend to exhibit more variability in both physical and psychological traits, including general intelligence. Frasier (1919) is the first study to compile a large dataset of more than 60,000 observations to provide 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 used 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 for boys is significantly higher than for girls. More recently, Johnson et al. (2008) also find greater variability among boys using two population-representative sample 11-years-olds in Scotland.

While evidence from the psychological literature is mostly confined to the US, 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

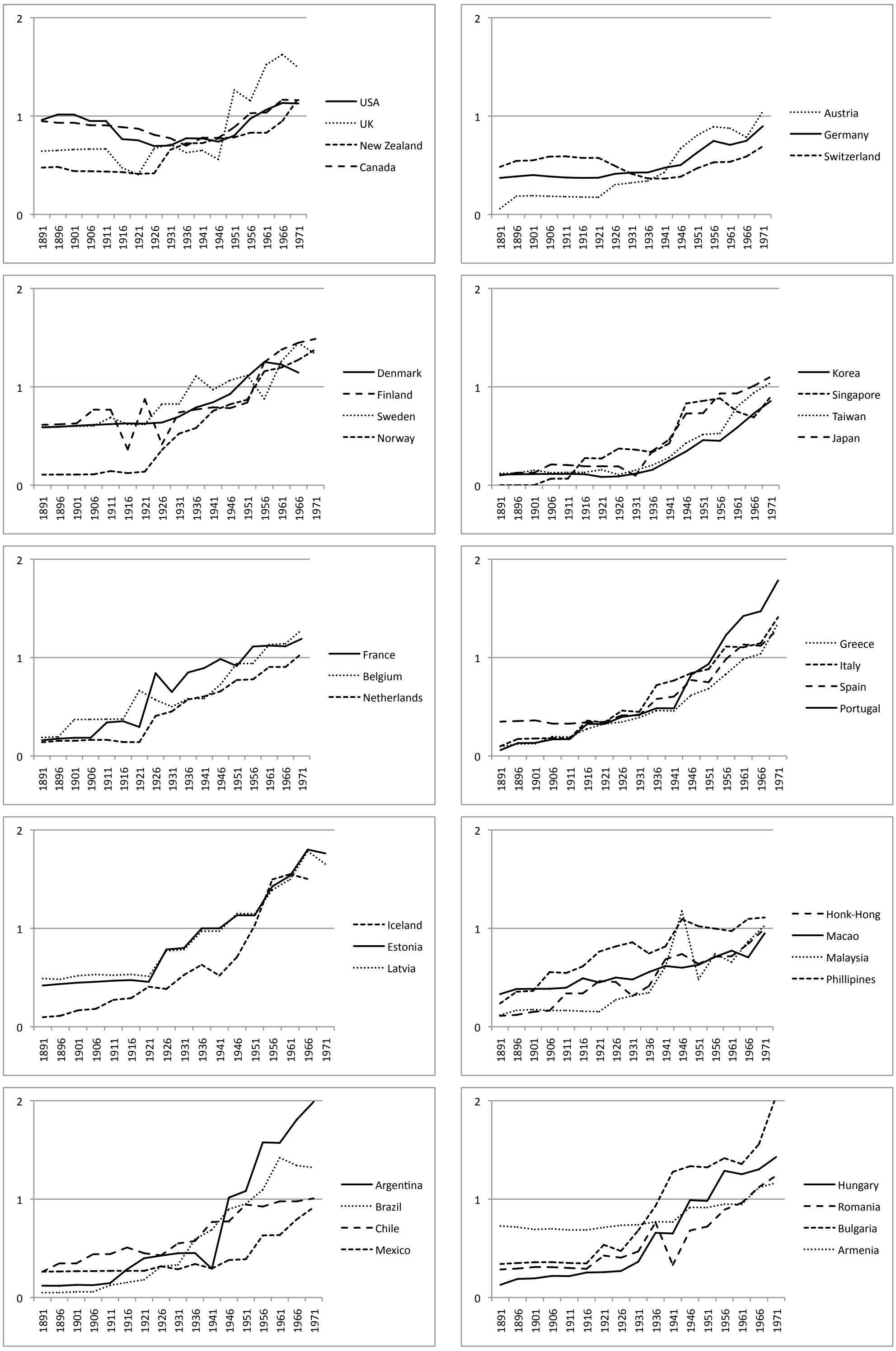
For the US, several papers have reported a convergence, followed by a reversal in the percentage of women relative to men attending tertiary education over the last decades [2]. While most evidence is confined to the US, we show in Figure 1 that the gender gap reversal in participation to tertiary education 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 shows 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 reports 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. Following the reversal, the gender gap in university attendance appear to have increased in the favor of women in all countries. 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. Strikingly, the reversal already occurred in less advanced economies such as Saudi Arabia.

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

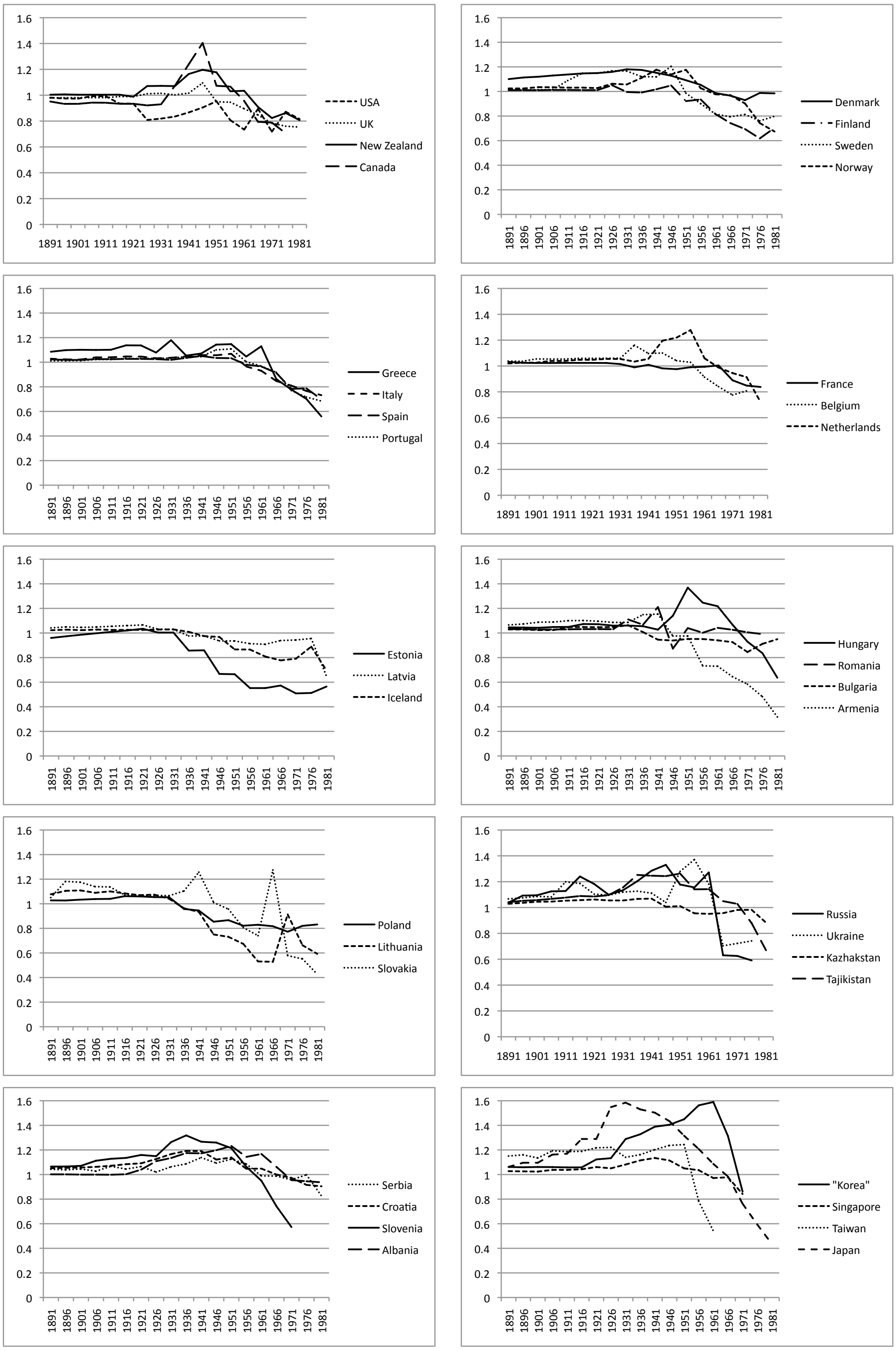
[2] See, for example, Goldin et al. (2006), Chiappori et al.(2009) or Becker et al. (2010)

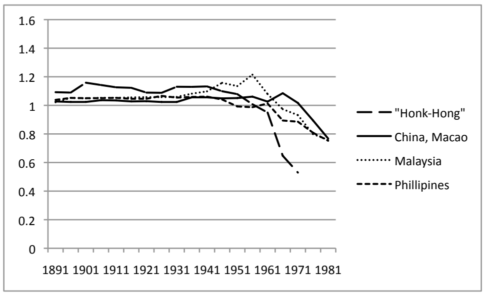
Figure 1: Female-to-male ratio among participants to tertiary education - by cohort of birth

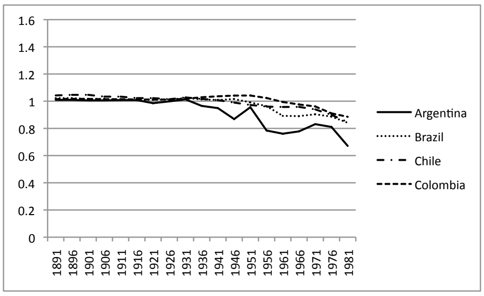


In Figure 2 we show that an analogous reversal occurred in the lower tail of educational achievement. As compulsory education ends at the end of lower secondary school in most developed countries, one way to identify individuals belonging to the lower end of the educational achievement distribution is to look at individuals who did not complete upper secondary education. We combined data from various data sources in order to reconstruct the evolution of the fraction of secondary school dropouts by gender and birth cohort for the main industrialized countries. We were able to compute upper secondary school dropout rates for birth cohorts over 90 years, ranging from individuals from in 1891 to individuals born in 1981. The main source for the data is the Barro-Lee database (2010), which allows to compute secondary school non-completion rates for 5-year band cohorts from 1891, and separately for males and females. 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.

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







3 Theoretical Framework

3.1 Optimal Investment in Education

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 effecting test scores and educational outcomes. As Heckman et al. (2012) note, test scores are the observable product of a complex combination of cognitive and non-cognitive skills, as well as effort. 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. 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 expected discounted utility U. Building on Becker (1967), we define the utility function of individuals in the economy as:

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

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 $k_2 > 0$. Importantly, both k_1 and k_2 and are assumed to be the same for females and males. Intuitively, individuals with higher test-taking ability z_j perceive greater benefits (or equivalently, lower costs) from attending education. In this framework, the optimal level of schooling s chosen by individual j can be expressed 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:

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

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

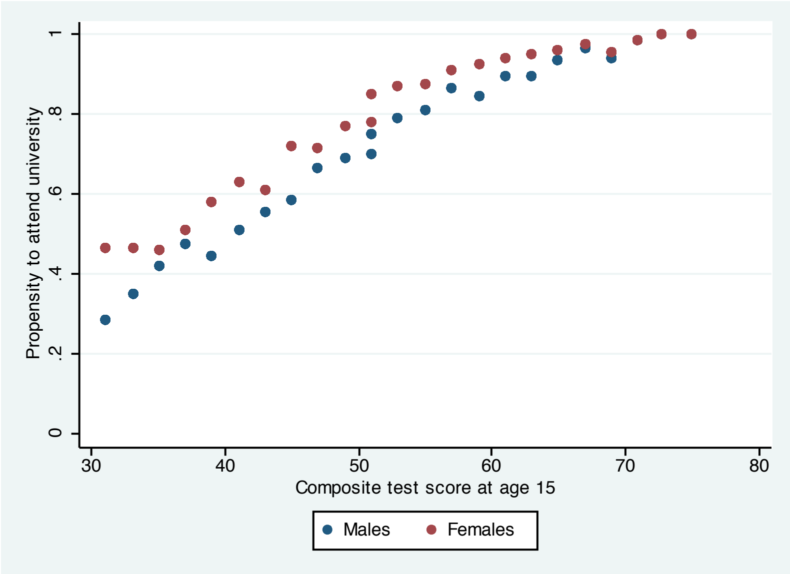
Therefore,

$$H_j = 1 \text{ if } z_j > \frac{\bar{s}}{b} \equiv \bar{z} \tag{6}$$

One may argue that individuals enroll in education on the basis of not only individual test-taking ability z_j, but also a set of individual circumstances, such as parental income or personal network[3].

[3]See Hoxby (2012) for recent evidence in the US

Figure 3: The positive relationship between test score z and university enrollment H - empirical support



Source. US General Social Survey (1975-2010)
Notes.

3.2 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:

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

or, equivalently

$$E_t = G_z(\frac{\bar{s}}{b_t}) = G_z(\bar{z}_t) \tag{7}$$

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,

$$\frac{\partial E_t}{\partial b_t} > 0 \tag{8}$$

and

$$\frac{\partial \bar{z}}{\partial b_t} < 0 \tag{9}$$

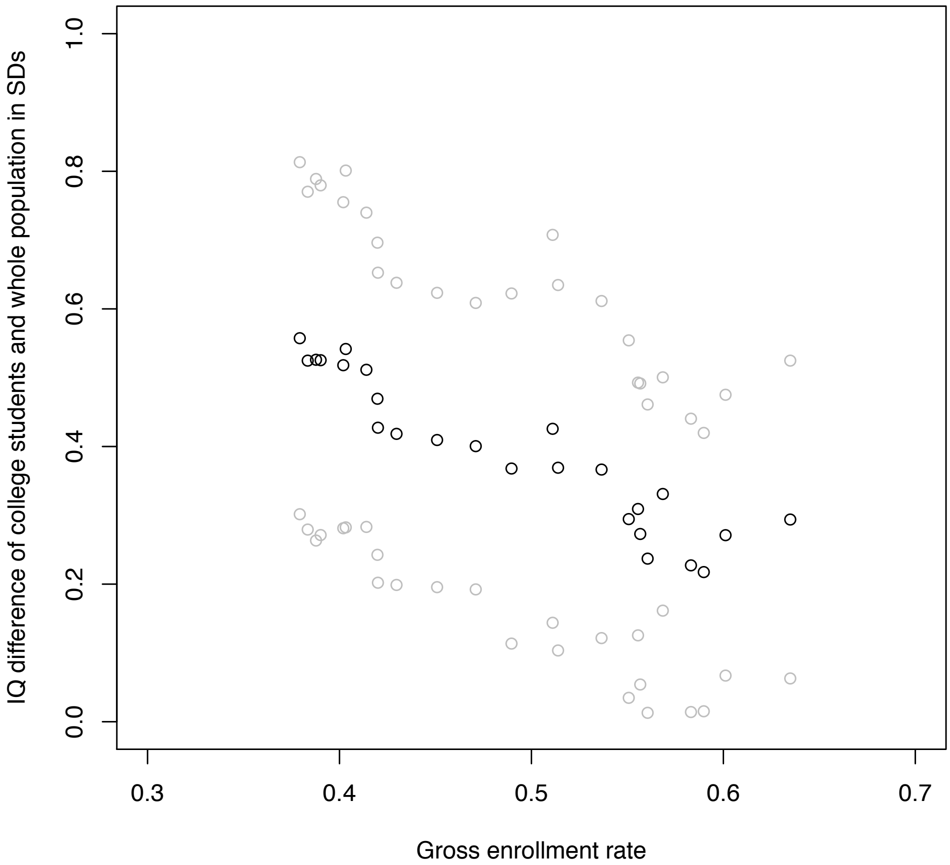
To check whether increased enrollment at university over time was accompanied by a decrease in $\bar{z}$, we computed the average IQ score of individuals attending university in the US over the period 1975-2010, using data form the General Social Survey (GSS)[4]. Each wave surveys a random sample of 1,000 to 5,000 individuals. From 1974 onwards, the GSS includes a simplified IQ test consisting of 10 questions assessing the cognitive skills of the respondents. A measure of educational attainment (in years) is also reported, and we consider individuals with more than 12 years of education as having attended tertiary education. Figure 4 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.

There exists a important body of literature showing that returns to education, and in particular returns to university education, have increased over the last decades. First, monetary returns to tertiary education – the college wage premium – have been shown to increase sharply over the 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

[4] 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 $\bar{z}$

Figure 4: The negative relationship between total enrollment rate in tertiary education and the average ability of the enrolled - empirical evidence



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

the US since the beginning of the 1970s[5]. Card and Lemieux (**?**) also report an important increase in the wage premium of university graduates relative to high school graduates in the UK and Canada over the same period. Acemoglu (2000) and Goldin and Katz (2009) invoke skill-biased technological change as the main driving force behind the increase in the university wage premium, through an increased demand for skilled workers. Goldin and Katz (2009) report a strong positive relationship between the utilization of more capital-intensive technologies and the demand for university-educated workers. Acemoglu (2000) argues the the extent of skill-biased technological was such that it allowed to absorb an increasing sup-

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

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

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

Building on the evidence about 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 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[6]. Each cohort is assumed to be split equally between males and females, and the distribution of test-taking ability for each gender $g \in \{m, f\}$ is assumed to be invariant over time:

$$f_{z_g,t}(z) = f_{z_g}(z) \tag{10}$$

Panel A of Figure 5 illustrates the two test-taking ability distributions, when $\sigma_m^2 > \sigma_f^2$ and $\mu_m^2 < \mu_f^2$. It also depicts the two 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:

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

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:

$$E(\bar{z}) \equiv \frac{G_{z_f}(\bar{z}) + G_{z_m}(\bar{z})}{2} \tag{11}$$

which is represented by the thin dotted line in panel B of Figure 5, and obtained by averaging the two complementary cumulative distributions, assuming that males and females are equally split in the population. In this framework, the female-to-male ratio among the enrolled, denoted $R(\bar{z})$, can be expressed as:

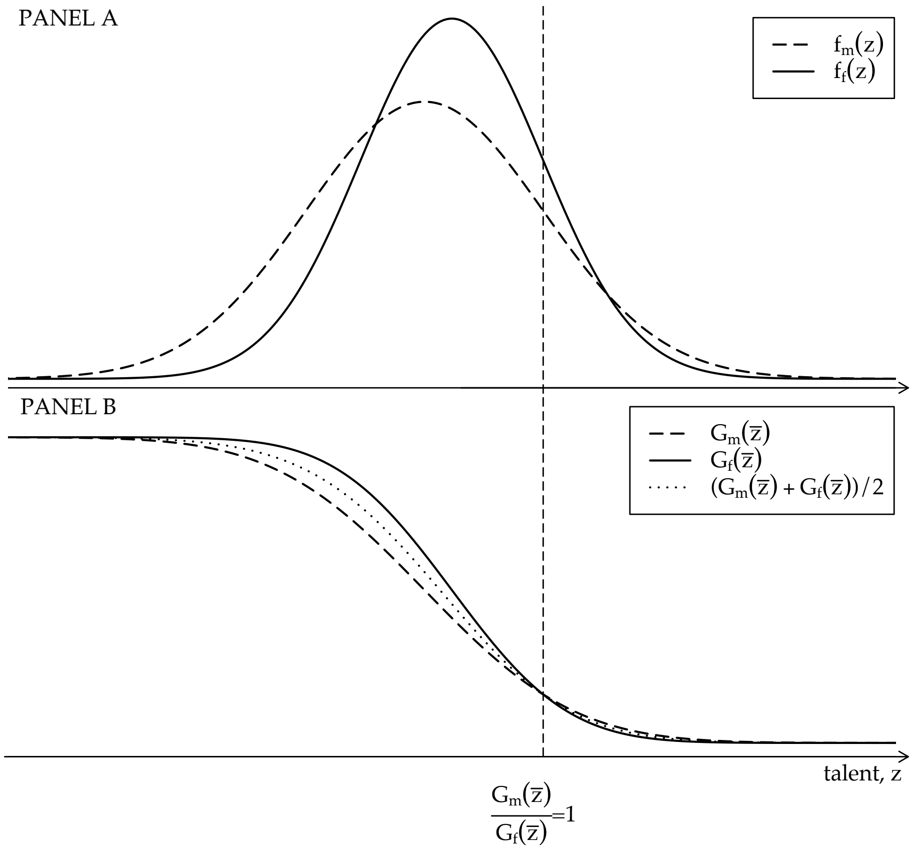
$$R(\bar{z}) \equiv \frac{G_{z_f}(\bar{z})}{G_{z_m}(\bar{z})} \tag{12}$$

From Panel B of Figure 5, we can derive the relationship between total enrollment rate and

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

the female-to-male ratio among the enrolled. Figure 6 illustrates the expected relationship between the total enrollment rate and the female-to-male ratio among the enrolled, when $\sigma_m^2 > \sigma_f^2$, as depicted in Panel A and B of Figure 5.

Figure 5: Distribution functions of test scores by gender when $\sigma_m^2 > \sigma_f^2$ - illustration

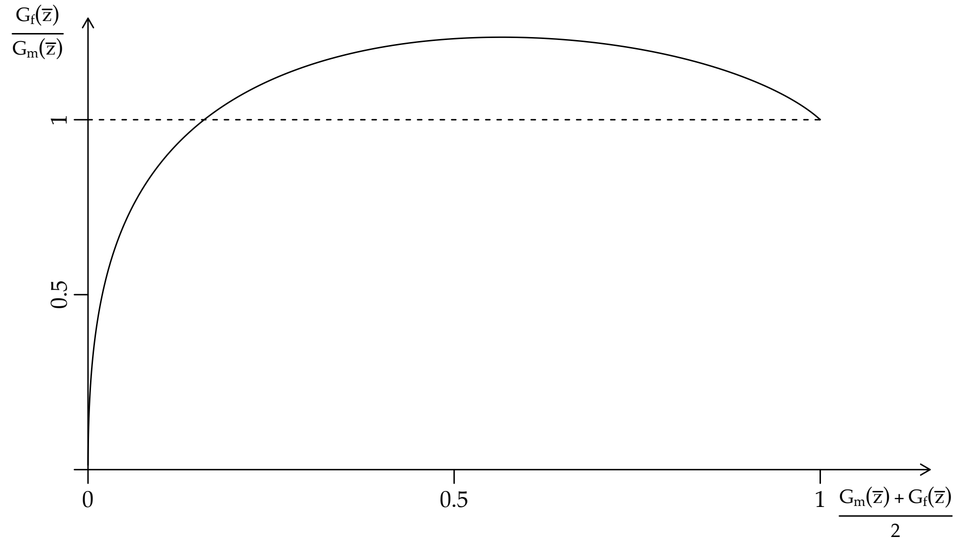


Notes. Panel A shows the probability distribution functions of test-taking ability z among males (full line) and females (dashed line), when test-taking ability z is normally distributed with $\sigma_m > \sigma_f$ and $\mu_f > \mu_m$. Panel B shows the complementary cumulative distributions, resulting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

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

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

Figure 6: Expected relationship between enrollment rate in tertiary education and female-to-male ratio among the enrolled – illustration



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 (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 assumed to follow alternative two-parameter probability distribution functions the are commonly used in the literature. Our theory therefore appears to be quite robust as the only necessary and sufficient condition on the distribution of z to generate the reversal is the larger dispersion of males' test-taking ability 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 only difference is that we work with the cumulative distribution functions (CDF) instead of the complementary cumulative distribution functions (CCDF) since we are now dealing with the lower tail of the probability density functions. In our setting, the secondary school non-completion rate for each gender in a given cohort t is simply:

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

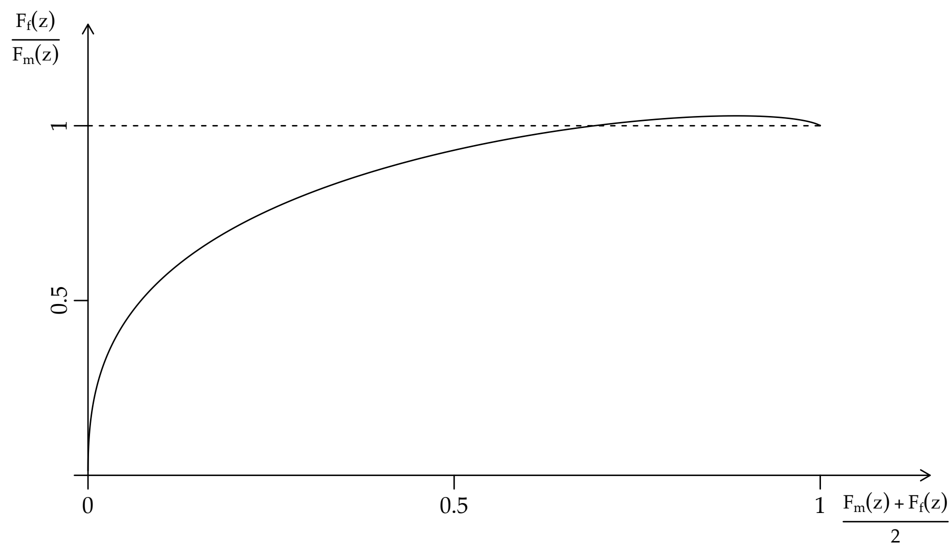
where

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

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

$$N(\underline{z}) \equiv \frac{F_{z_f}(\underline{z}) + F_{z_m}(\underline{z})}{2} \tag{13}$$

Figure 7: Expected relationship between non-completion rate in secondary education and female-to-male ratio among non-completers – illustration



Note. The parameters are the same as in Figure 6.

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[7]. 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

[7]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[8]. 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 [9]. 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 Maximum Likelihood Estimation

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

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

where x_{it} denotes the total enrollment rate in country i and year t, and y_{it} denotes the female-to-male ratio 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 (μ_m, σ_m^2) and (μ_f, σ_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 (μ_i, σ_i^2) the first two moments of the males' test taking ability distribution relative to females in country i. Formally,

$$\mu_i = \frac{\mu_{i,m} - \mu_{i,f}}{\mu_{i,f}} = \mu_{i,m} \tag{14}$$

[8]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.

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

$$\sigma_i = \frac{\sigma_{i,m}}{\sigma_{i,f}} = \sigma_{i,m} \tag{15}$$

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:

$$y_{it} = \frac{G_{z_f}(\bar{z}_{it})}{G_{z_m}(\bar{z}_{it}, \mu_{it}, \sigma_{it})} \cdot \exp(\epsilon_{it}), \tag{16}$$

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

$$\log y_{it} = \log G_{z_f}(\bar{z}_{it}) - \log G_{z_m}(\bar{z}_{it}, \mu_{it}, \sigma_{it}) + \epsilon_{it}, \tag{17}$$

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:

$$G_{z_f}(\bar{z}_{it}) = \int_{\bar{z}_{it}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}} \, dz = \frac{1}{2}\left(1 - erf\left[\frac{\bar{z}_{it}}{\sqrt{2}}\right]\right) \tag{18}$$

$$G_{z_m}(\bar{z}_{it}) = \int_{\bar{z}_{it}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{\frac{(z-\mu_m)^2}{2\sigma_m^2}} \, dz = \frac{1}{2}\left(1 - erf\left[\frac{\bar{z}_{it} - \mu_m}{\sqrt{2}\sigma_m^2}\right]\right) \tag{19}$$

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

$$x_{it} = E_{it}(\bar{z}_{it}) \equiv \frac{G_{z_f}(\bar{z}_{it}) + G_{z_m}(\bar{z}_{it})}{2} = \frac{1}{4}\left(2 - erf\left[\frac{\bar{z}_{it}}{\sqrt{2}}\right] - erf\left[\frac{\bar{z}_{it} - \mu_m}{\sqrt{2\sigma_m^2}}\right]\right) \tag{20}$$

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

$$L(\theta_i|y_{it}) = \log G_{z_m}(E_i^{-1}(x_{it})) - \log G_{z_m}(E_i^{-1}(x_{it})) + \epsilon_{it}, \tag{21}$$

where $\theta_i = \{\mu_i, \sigma_i\}$. Since the error term is normal, the model can be fitted by finding numerically 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 normalized to a standard normal distribution.

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

$$L(\{\theta_i|z\}) = \frac{1}{(2\pi\sigma_\epsilon^2)^{\frac{n}{2}}} \cdot \exp\left\{-\frac{1}{2\sigma_\epsilon^2} \cdot \sum_{i=1}^{n} \left(\log y_{it} - \log G_{z_f}(E_i^{-1}(x_{it})) + \log G_{z_m}(E_i^{-1}(x_{it}))\right)^2\right\} \tag{22}$$

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

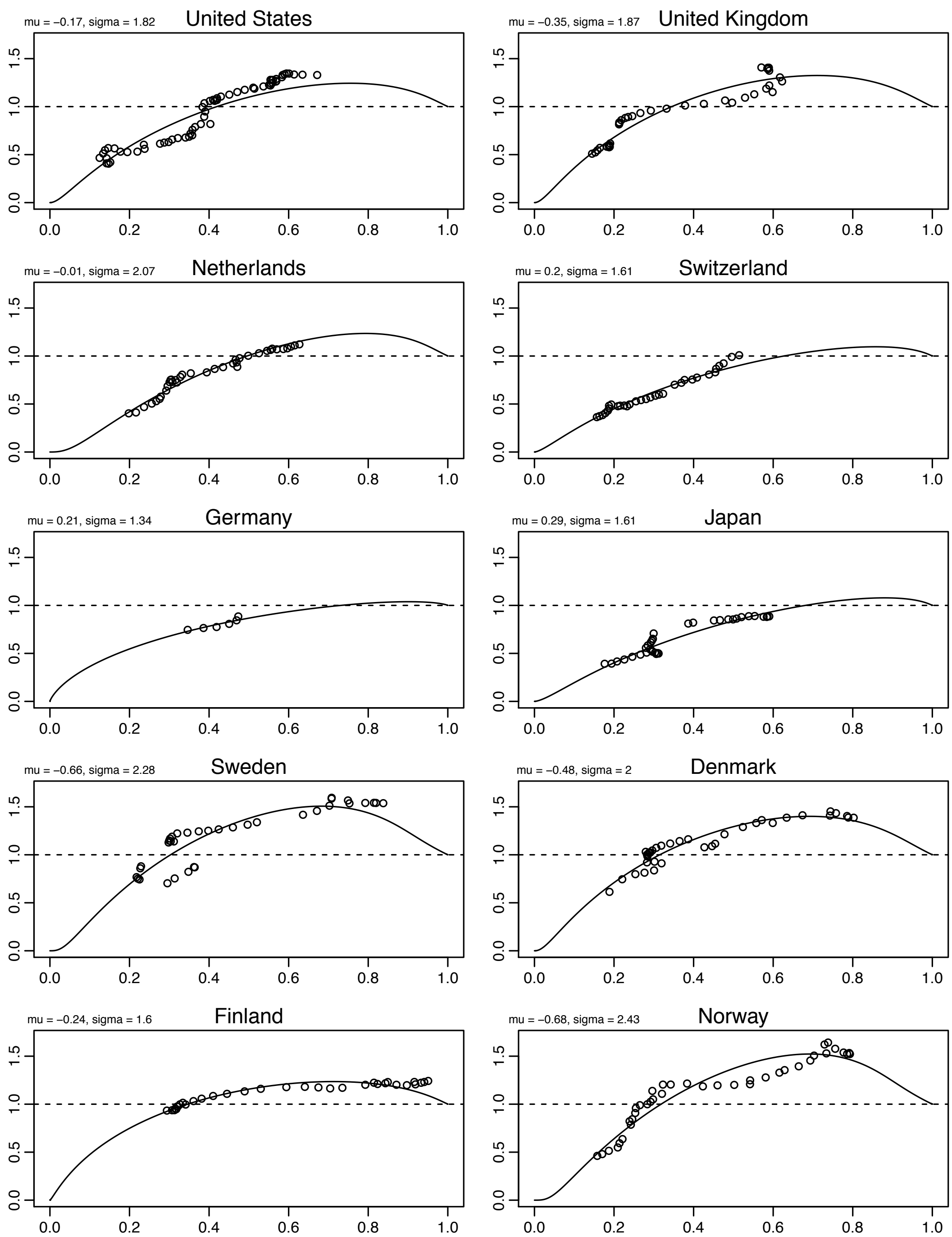
$$\hat{\theta_{MLE}} = \min_{\theta_i \in \Theta} \sum_{i=1}^{n} \left\{\log y_{it} - \log G_{z_f}(E_i^{-1}(x_{it}|\theta_i)|\theta_i) + \log G_{z_f}(E_i^{-1}(x_{it}|\theta_i)|\theta_i)\right\}^2 \tag{23}$$

5 Results

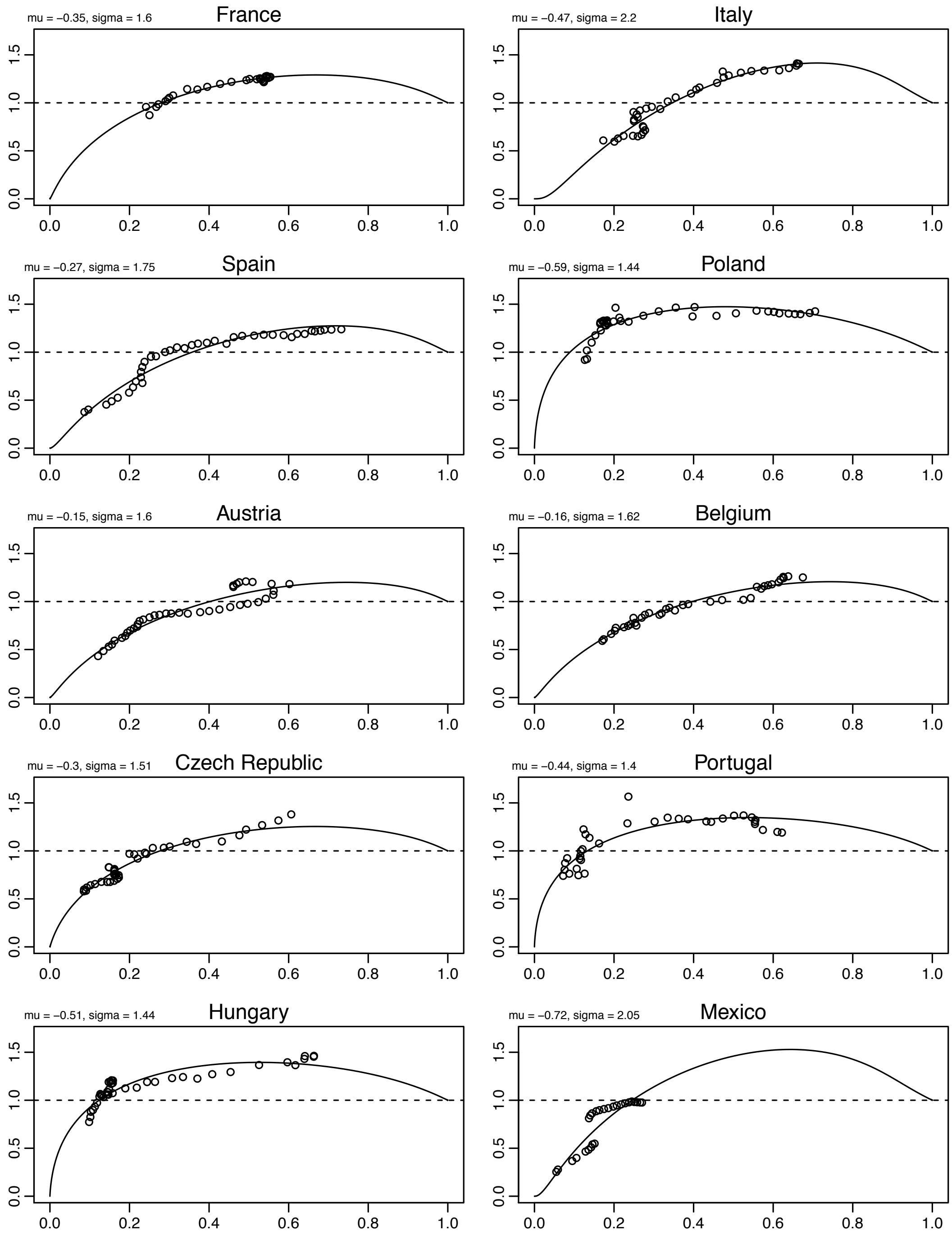
5.1 Model Fit

Figure 8 depicts the estimated relationship between the total enrollment rate in tertiary education x and the female-to-male ratio y when $\{\mu_i; \sigma_i\}$ are estimated from the $2 \times T$ matrix, as described in the previous section. As shown in the figure, our model generates an accurate fit for the relationship between x and y observed in each individual country. Figure 9 shows our model fit for the relationship between the secondary school non-completion rate, and the gender ratio among non-completers. As depicted in the figure, our model also generates a very satisfactory fit for the gender ratio in the lower tail of educational distribution. Contrary to the 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 ration among non-completers as a function of non-completion rates. It shows a 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 gradualy 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 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 first contribution to account for this very particular non-linear pattern in the relation between the secondary school non-completion rate and the gender ratio among non-completers.

Figure 8: Model Fit - Gender Ratio in Participation to Tertiary Education

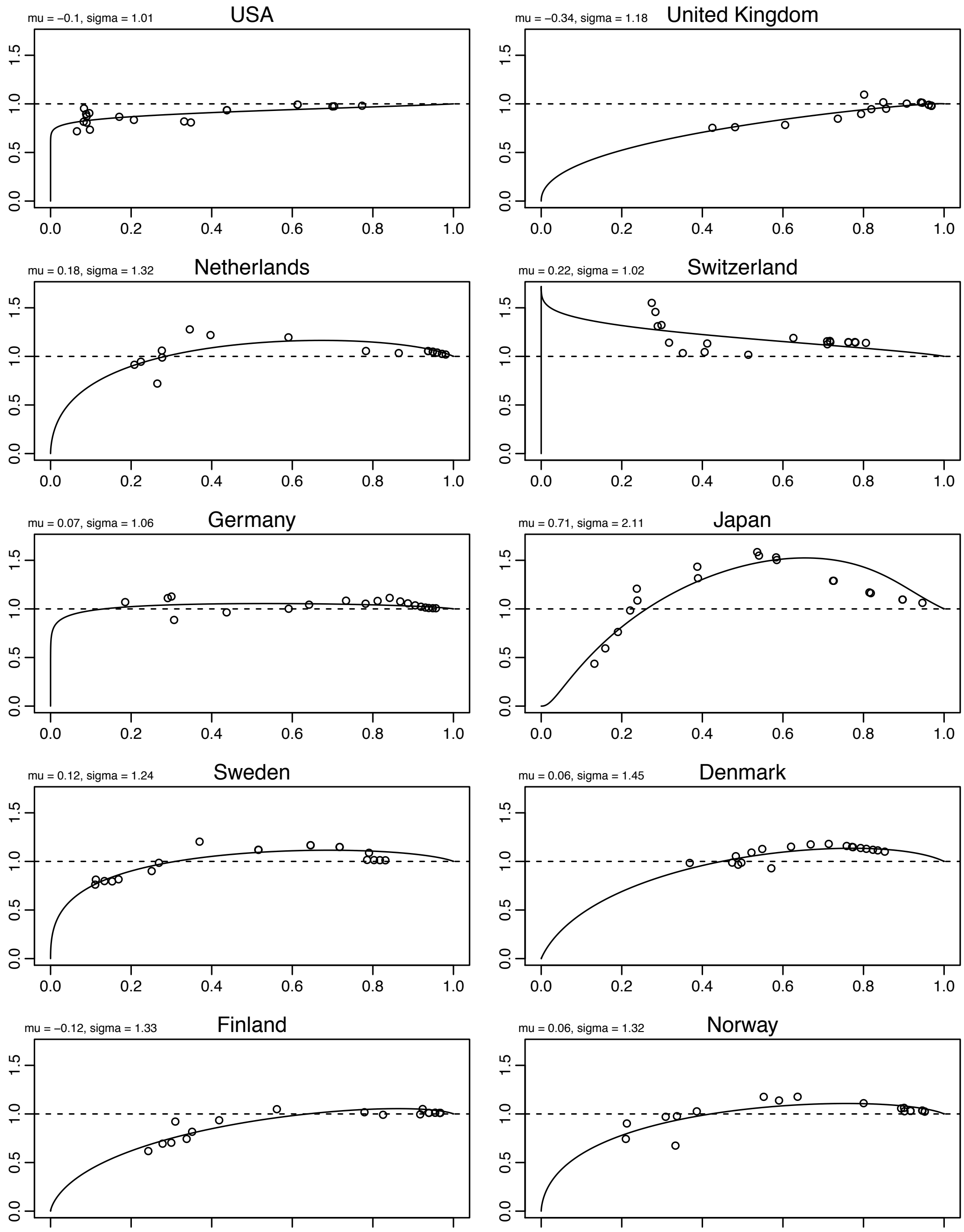


Notes. The x-axis measures the total gross enrollment ratio in tertiary education for country i. The y-axis measures the females-to-males ratio in tertiary education for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ for country i from the UNESCO Institute of Statistics. The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

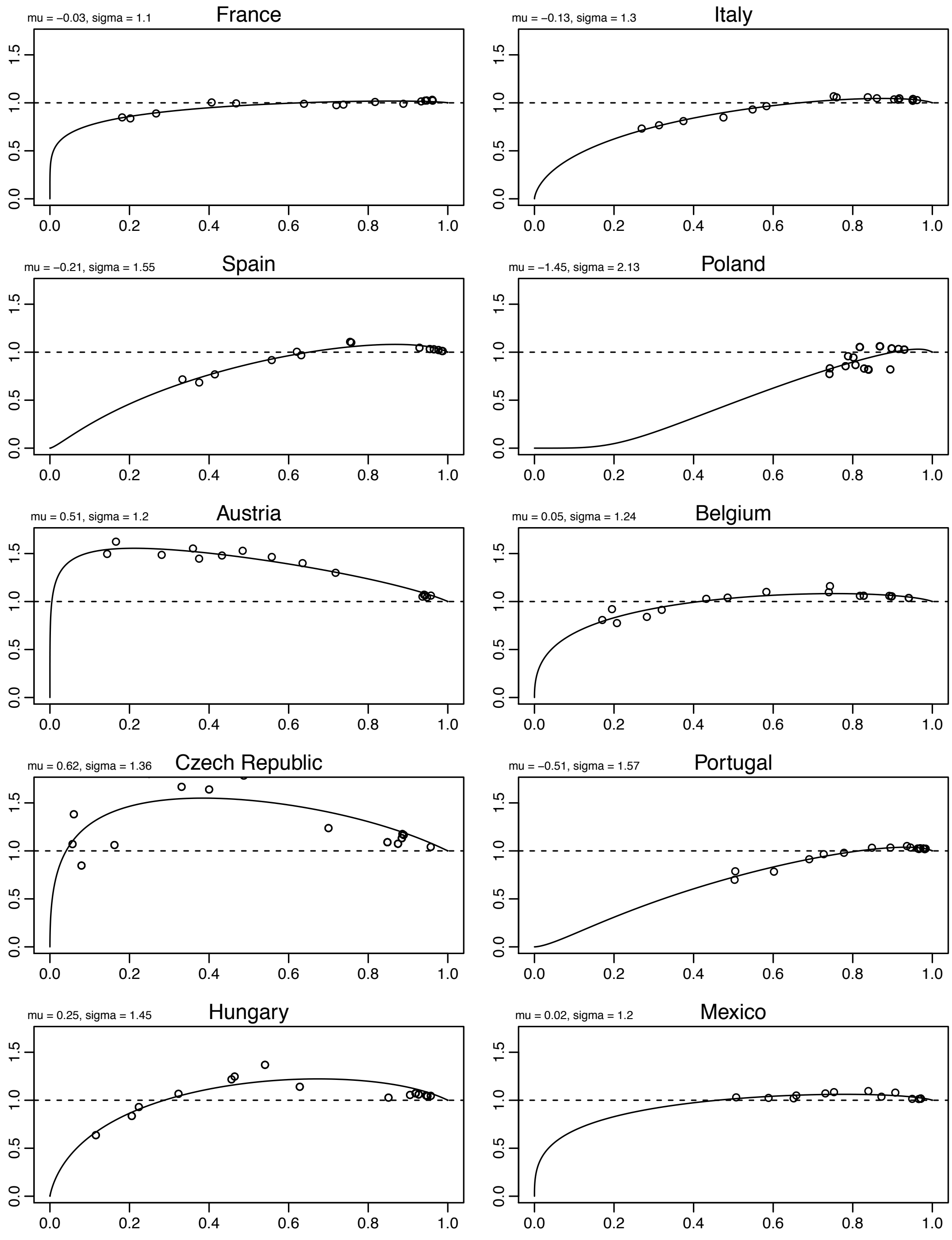


Notes. The x-axis measures the total gross enrollment ratio in tertiary education for country i. The y-axis measures the females-to-males ratio in tertiary education for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ *for country i from the UNESCO Institute of Statistics. The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

Figure 9: Model Fit - Gender ratio among Secondary School Non-Completers

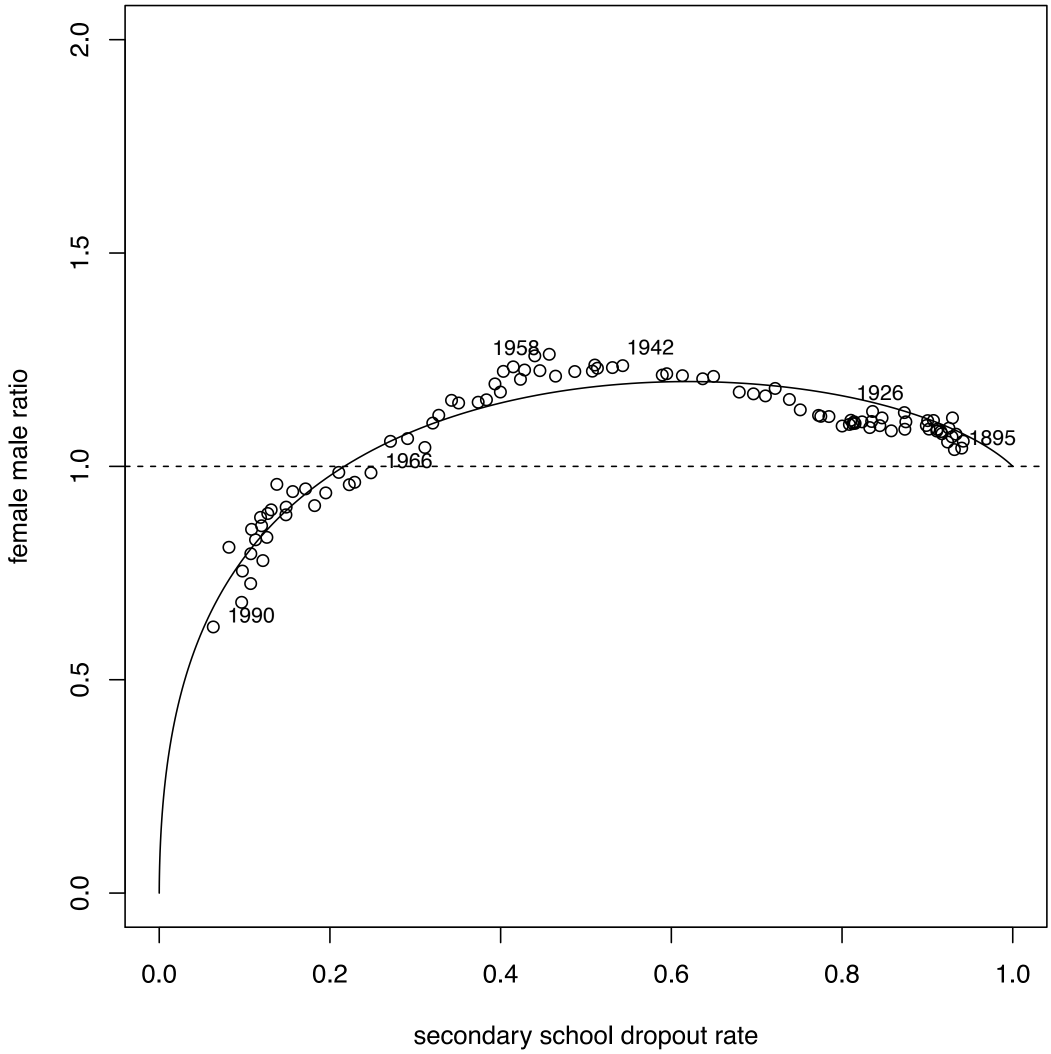


Notes. The x-axis measures the rate of secondary school non-completion for country i. The y-axis measures the females-to-males ratio among secondary school non-completers for country i. Each dot corresponds to a yearly observation of $\{x_i; y_i\}$ *for country i from Barro-Lee (2010). The full line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



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

Figure 10: Model Fit - Secondary School Non-Completers (France)

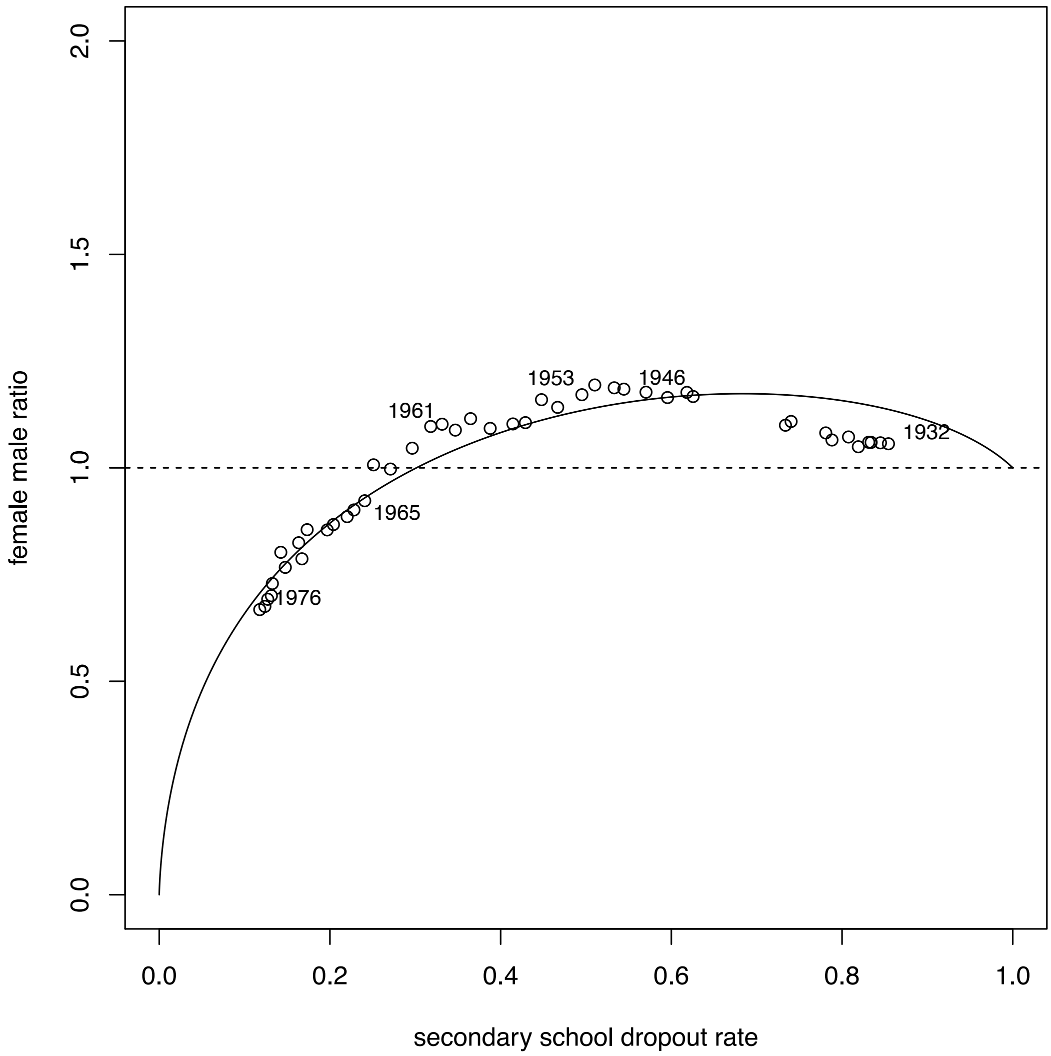


Source. French Labour Force Survey (1982 - 2010)

5.2 Matching our Estimates with PISA Distributions

Our model is: $\log y_t = \log R(E^{-1}(x_t, \mu, \sigma), \mu, \sigma) + \epsilon_t$. Given our estimates for μ and σ, the model predicts the gender ratio $\hat{y}$ among the enrolled, for a given value of the 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 both available in the UNESCO and PISA datasets. In total, 40 countries are common

Figure 11: Model Fit - Secondary School Non-Completers (Spain)



Source. Spanish Census (1990-2010)

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

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

Table 1 shows the correlation between the $\hat{y}_i$'s simulated from the two sets of $\{\hat{\mu}_i, \hat{\sigma}_i\}$ obtained from our model fit, and the PISA distributions. We use the PISA dataset from year 2000 to get as close to the median year of the UNESCO data to minimize any attenuation bias that might emerge if there is a systematic shift in the relative distributions of females and males. We report correlations for 3 different values of x, $x = 0.20$, $x = 0.50$ and $x = 0.70$. Another source of measurement error is the fact that we are comparing 15-year-olds to 18-year-olds. Any country-specific change in the relative distributions during that age gap will induce measurement errors. 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.

The fact that our model estimates does not perfectly correlates with gender differences found in PISA can be explained by several sources of measurement errors. 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, Lynn and Kanazawa (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 theoretical counterpart 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 envrionements 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 in our estimates for country specific-gender differences in ability distributions.

Table 1: Correlations between Predicted Gender Ratio by Quantile from our Estimates and PISA Estimates

	Predicted Gender Ratio - Model Fit		
	2nd decile	5th decile	7th decile
Predicted Gender Ratio - PISA Math	0.332*	0.335*	0.317*
	(0.059)	(0.057)	(0.072)
Predicted Gender Ratio - PISA Reading	0.488***	0.429**	0.362**
	(0.004)	(0.013)	(0.039)

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Sample size is 33 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. 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 explaining 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. (**?**), changes in social norms combined with higher returns to education for females can produce a reversal from male majority to female among university students. Second, a relative increase in females' mean test-taking ability over time can also generate a reversal in the college gender gap, as suggested by Cho (**?**).

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

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

Hypothesis 2, *Change in social norms hypothesis*: The lower bound of test-taking ability to enroll into tertiary education differs between men and women, i.e. there exist gender-specific $\bar{z}_f$ and $\bar{z}_m$ for females and males respectively while the test-taking ability distribution $G_z(\cdot)$ is common to both genders. In the past, $\bar{z}_f > \bar{z}_m$ before progressively converging and surpassing z_m over time

Hypothesis 3, *increase in females' mean performance hypothesis*: The female mean test-taking ability has increased and progressively surpassed the male men over time The variance of the ability distribution and the ability threshold for enrollment $\bar{z}$ are identical for both genders.

These three hypotheses could be combined into joint hypotheses. We analyze them separately to retain maximum simplicity and assess their respective explanatory powers.

6.1 The Change in Social Norms Hypothesis (Hypothesis 2)

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

$$E = 1 - F_z(\frac{\bar{s}}{b}) = G_z(\frac{\bar{s}}{b}) = G_z(\bar{z}) \tag{24}$$

We can formulate the change in social norms hypothesis in our framework by allowing b to differ and change over time across genders, with $G_z(.)$ being identical for males and females. In this setting, it is possible to generate the gender gap dynamics observed in the data if $b_f < b_m$ originally, before gradually converging and surpassing b_m over time. Optimal levels of investment in schooling are expressed separately for males and females as:

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

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

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

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

In this context, $b_f > b_m$ in most recent years is a necessary condition for the gender gap reversal in 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 topcoding 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

Let us now fit and assess Hypothesis 2, by allowing b_f and b_m to take different values. Hypothesis 2 is very flexible for fitting the data. Since the relative changes of $\bar{z}_f$ and $\bar{z}_m$ have not been constrained, we can calculate the values algebraically. The calculated values are depicted in Figure 12. The algebra is straightforward. For each time period, there is a system of two equations:

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

and

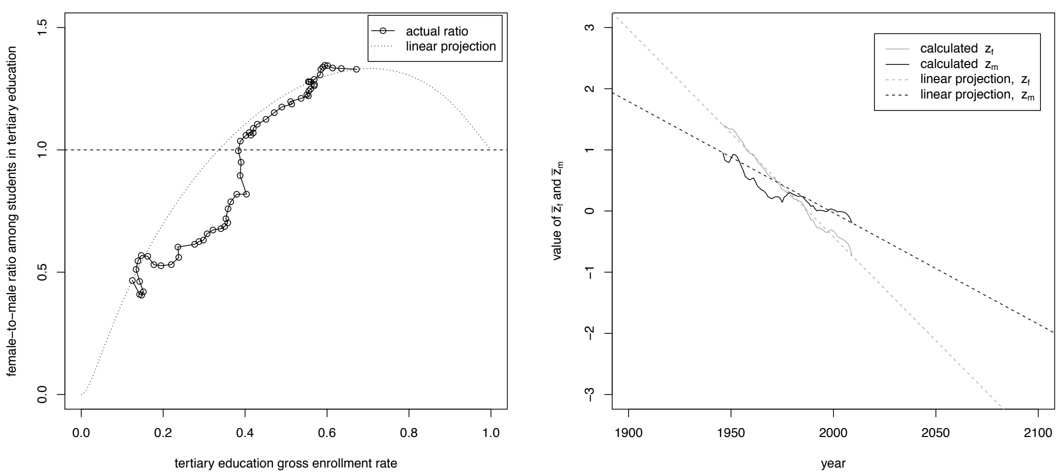
$$x_t = \frac{G_z(\bar{z}_{f,t}) + G_z(\bar{z}_{m,t})}{2},$$

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

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

We can easily solve numerically for the unknown and unique values of $\bar{z}_{f,t}$ and $\bar{z}_{m,t}$. To extrapolate outside the actual data range, we assume that $\bar{z}_m$ and $\bar{z}_f$ continue to change at the average estimated pace of change between the years 1946 and 2009. The estimated values of $\bar{z}_{f,t}$ and $\bar{z}_{m,t}$ and their change over time is depicted in Figure 12.

Figure 12: 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 sex ratio and gross enrollment rate in 1946 to 2009. 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.

6.1.1 Hypothesis 2 Versus Hypothesis 1: Test 1

Hypothesis 2 appears to do an excellent job at explaining the reversal, 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:

$$C(\bar{z}_f, \bar{z}_m) = \frac{G_{z_f}(\bar{z}_f) + G_{z_m}(\bar{z}_m)}{2} \tag{25}$$

and

$$E[z_f|z_f > \bar{z}_f] - E[z_m|z_m > \bar{z}_m], \tag{26}$$

which is the difference between the mean test-taking ability of females and males enrolled in tertiary education. This comes from the fact that Hypothesis 1 claims that z_f and z_m have different distributions and $\bar{z}_f = \bar{z}_m$, 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 the fraction of population taking the test increases. When the test takers are a representative sample of the whole population, the observed mean difference becomes an estimate of the mean difference of the whole population. To test these two opposite predictions against the data, we analyze data on both the average performance in cognitive test by gender, and the proportion of a given cohort taking the test.

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

As depicted in Figure 13, while Hypothesis 1 generates an increase in the average level of test-taking ability of females attending university relative to males over time, the change in social norms hypothesis implies an evolution of the opposite sign, until the sample restriction reaches a proportion of around 0.6 to 0.8 of the cohort taking the test.

Figure 13 shows that Hypothesis 1 performs well at predicting the relationship between the average gender in cognitive scores and the fraction of the population taking the test, implied by $\bar{z}$. It provides a good fit for both the shape of the relationship observed in the data, and the sign of the gender gap in average cognitive tests. The data shows that males do better relative to females in a restricted sample selected from the top of the distribution, than in the entire population. In PISA, which is a sample of the entire population of 15-year-olds, females obtain higher average test scores relative to males in reading. On the other hand, males perform on average better in the same discipline with the SAT test.

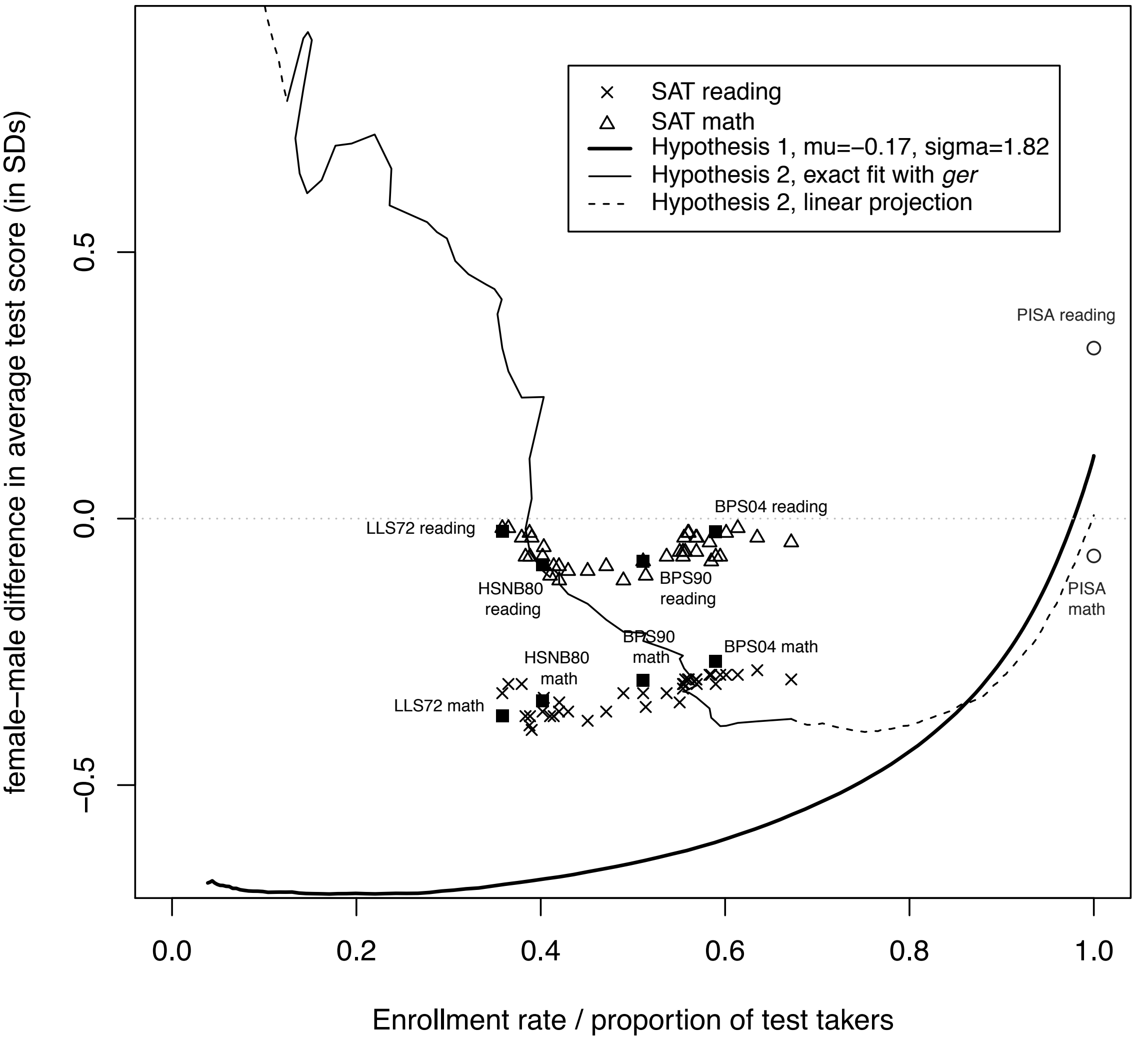
The higher male variability hypothesis (Hypothesis 1), although not a perfect fit, provides a simple explanation for the main facets of this puzzle, relying on the fact that test-takers are drawn from different ranges of the ability distribution in these tests. This matters for the observed average performance by gender, given different underlying distributions of test-taking ability by gender assumed in our model. On the other hand, the alternative hypothesis of change in social norms (Hypothesis 2) neither predicts the sign of the gender gap for the whole range, nor the sign of the relationship between the gender gap in average test performance and the proportion of population taking the test.

6.1.2 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 score and enrollment in tertiary education by gender. Hypothesis 1 assumes that test-taking ability is the only relevant choice variable for choosing to attend university. This implies that the empirical relationship between z and H should be identical for both genders, in every given cohort. In particular, there should not be any gender-specific change in the relationship over time between the test taking ability z and tertiary enrollment H. On the other hand, hypothesis 2 implies that females used to enroll less than males at university conditional on test scores ($b_f < b_m$), while they now have a higher propensity to enroll for a given test-taking ability, since $b_f > b_m$.

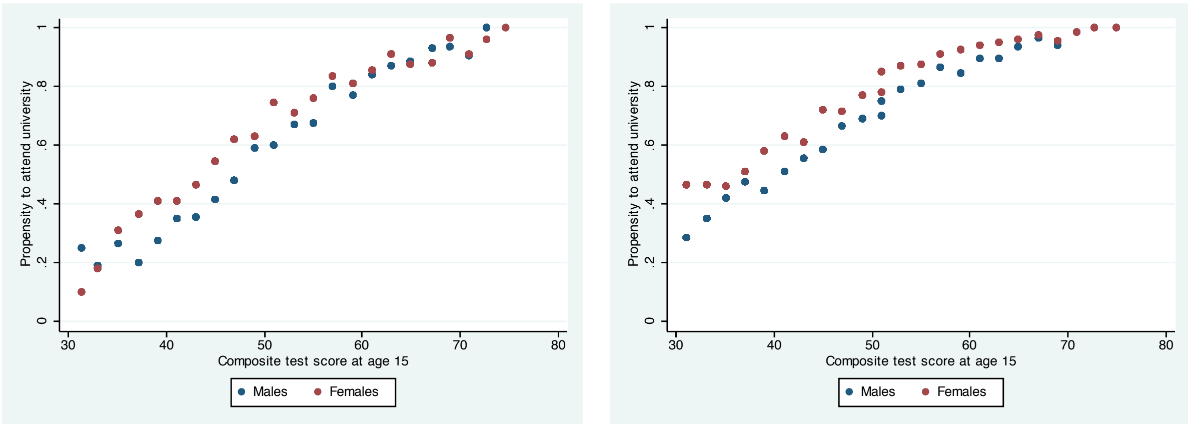
To assess these competing hypotheses, we use data from two longitudinal surveys conducted in the US in 1980 and 2002. The first of these surveys is the High School and Beyond 1980, which follows a cohort of 10th graders in 1980 until university studies. The second of these surveys is the US Educational Longitudinal Study 2002, which also follows 10th graders in 2002 until university studies. Both datasets survey a representative sample of the same age group in the US population, and provide test score information in 10th grade on a comparable scale. In addition, they both allow us to know whether individuals attended university

Figure 13: Fit for the empirical relationship between enrollment rate and gender gap in mean test score of the enrolled: Our model Vs the social norm hypothesis



Notes. The x-axis represents the proportion of the cohort taking a given cognitive test. The y-axis represents the female-to-male difference in means of the given cognitive test, expressed in standard deviations units. The thick line depicts the expected relationship between the gender difference in average cognitive score of university students as a function of total enrollment, as predicted by from model. The dashed lines represent the evolution of the same variable predicted by the discrimination hypothesis, when males and females cognitive score distributions are assumed to be the same, but women are initially facing a higher $\bar{z}$ relative to males that progressively converges to the males' level when university enrollment increases. The crosses and triangle dots represent the actual value of the gender difference in SAT test scores for the entire population of SAT test-takers, as a function of the fraction of SAT test-takers in the population. The squared-dots represent similar values computed from US post-secondary longitudinal surveys, for college students only.

Figure 14: The Relationship between test-taking ability z and tertiary enrollement H by Gender: 1980 Vs 2002



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

education, and to match this information with individual test scores in 10th grade. Using these two surveys, we constructed the empirical relationship between the test-taking ability z measured by test scores at age 15, and the propensity to enroll at university H at two different points in time: 1980 and 2002. We believe that the 22 years' time is sufficient to detect asymmetric changes by gender in the relationship between z and H, especially in a period during which enrollment at university increased dramatically in the US. The results of the analysis are depicted in Figure 14.

Figure 14 shows two main patterns. First, females have a higher propensity to enroll in tertiary education, conditional on test scores. Second, and more importantly in our context, this was already the case in 1980. In other words, we do not observe any change in the propensity of females to enroll at university conditional on test scores relative to males, which goes against the change in social norms hypothesis.

We further investigate changes in the relationship between z and H quantitatively, using regression analysis. The results are presented in Table 2 and Table 3. As shown in Table 2, rmales enroll significantly less at university than females, conditional on test scores, in both 1980 and 2002. IThe point estimate of this negative effect of the male dummuy is very stable between 1980 and 2002: -0.083 in 1980, against -0.078 in 2002. There is therefore no evidence for women being penalized in university enrollment given their test scores in earlier decades, as the change in social norm hypothesis would suggest. In addition, the effect of gender on university enrollment conditional on scores is fairly stable over time, which goes against the predictions of Hypothesis 2.

Table 3 brings further insights by running separate regressions by gender in 1980 and 2002. The intercept for both males and females increased sharply from 1980 to 2002, reflecting the fact that both genders have a higher propensity to enroll at university in 2002 than in 1980, conditional on their test scores. Therefore, although the propensity to enroll at university increased sharply over the period, there does not seem to be any asymmetrical change in the relationship between test scores and university enrollment: both genders enroll more at university conditional on their test scores, and males enroll less at university than females given their test score both in 1980 and 2002. Interestingly, the association between university enrollment and test scores have decreased for both genders over the period. This may reflect the multiplication of US postgraduate institutions in recent years implementing less strict criteria of admissions. Again, this evolution appears to be quite symmetric between genders: the association decreased from 0.25 to 0.19 for males, and from 0.21 to 0.16 for females. The interaction changes from positive and significant to negative and insignificant, but has only a distributional effect within boys, since the mean is very close to zero.

Table 2: The Relationship between test-taking ability z and enrollement H: 1980 Vs 2029 - graphical evidence

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

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

Table 3: The Relationship between test-taking ability z and enrollment H in 1980 and 2009 - Separate Regressions by Gender

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

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

6.2 Increase in Females' Mean Performance (Hypothesis 3)

The 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 assess 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 left relative to f_{z_m}, leading to an increase in E_f relative to E_m, without any change in $\bar{z}$. This possibility has been investigated by Cho (2007) and Fortin (2011) for the US. Fortin (2011) finds no relative increase in girls' self-reported grades over the period 1970–2010. Using data from the Monitoring the Future study, she reports an increase in high school grades for both boys and girls with parallel trends over the period, 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 caveat 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-taking ability distributions differs across genders. 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. 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. In the presence of a higher variance of test-taking ability for men in the entire population, sample restriction will therefore generate changes in the observed mean performance between genders over time, even if the mean of the underlying population distribution remained unchanged.

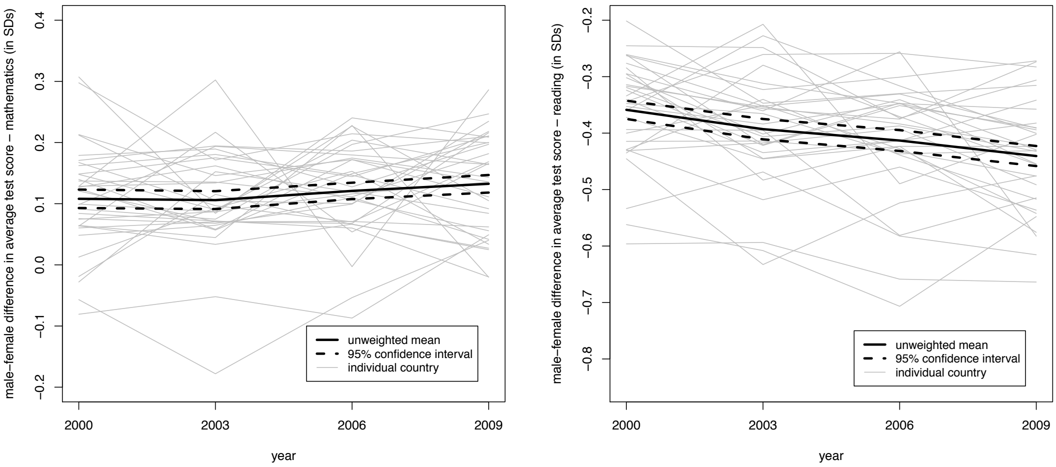
[FIGURE?]

His analysis therefore uses a selected sample of the population, and selection into high school attendance is likely to have changed over-time. In particular, enrollment in high school expanded over the period, and the simple comparison of males and females average test scores in high school includes therefore the confounding effect of changes in the sample of high school students. This is particularly the case if, as we argue, ability distributions differ by gender, and less and less talented individuals enroll into high school as the enrollment rate increases.

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

Given the short time-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 [school still compulsory at that age?), over the period 1980–2002 (Check 2009?). To this purpose, we use two nationally representative longitudinal surveys of secondary school students in the

Figure 15: Average PISA performance of males relatives to females in math and reading: 2000-2009



US conducted in 1980 and 2002. These surveys both contain information on test scores in mathematics and reading when individuals were in 10th grade. The results are depicted in Table 4. 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 school increased relative to boys over the period, they should be interpreted with care. First of all, there is the usual sample restriction issue since 10th graders in the US are already beyond the end of compulsory schooling. Therefore, if high school dropout rates differ between boys and girls and have changed between 1980 and 2002, which is likely, changes in average performance of 10th graders will give a biased estimate of the entire population. Second, one should keep in mind that such evidence is restricted to the US. Data constraints do not allow to repeat a similar exercise for other countries over the same period.

7 Conclusion

Building on a simple framework of optimal investment in human capital, our model is able to reconcile three internationally robust facts: the greater dispersion of males test score

Table 4: Gender difference in mean test score at 15 in the US: 1980 Vs 2002

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

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

distribution relative to females', the gap reversal form male majority to female majority in tertiary education attendance, and the reversal from female majority to male majority in secondary school non-completion. In addition to building the model framework, we contribute in this paper by showing that the second fact applies to a wider range of countries than shown previously. We are also the first to observe the third fact. Given data limitations and the inherent difficulty in measuring test-taking ability, our model provides a very satisfactory fit for the empirical relationship between total enrollment rate in tertiary education and the female-to male ratio. It is also the first theory to account for the reversal of the gender gap in secondary school non-completion rates. It turns out to be able to explain some patterns of the data that previous theories are inconsistent with. Also, the proposed model framework allows us to conveniently test different hypotheses put forth in the literature.

Beyond the particular question addressed in this paper, our findings suggest that gender differences in ability distribution might be relevant to account for other gender-related stylized facts in labour economics. Further research using this empirical fact to analyze other aspects of the gender gap in economic outcomes would therefore be of particular interest. Importantly, the larger variability of men's ability distribution observed empirically remains mostly unexplained, and stands as another promising area for future research, beyond the field of Economics.

8 References

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

Bossavie, L. and O. Kanninen. 2012: "Gender Inequality in Educational Outcomes: A Survey", European University Institute, unpublished manuscript

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

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

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.

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.

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

9 Appendix

9.1 Proof of Proposition 1 to 3

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

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

Building on empirical evidence on gender differences in test score distributions, we assume $\sigma_m^2 > \sigma_f^2$.

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

First, it is immediate to see that $\lim_{\bar{z}\to\infty} C(\bar{z}) = \frac{G_{z_f}(\bar{z})+G_{z_m}(\bar{z})}{2} = \frac{0+0}{2} = 0$. 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 females-to-males ratio $R(\bar{z})$ tends to one when the total enrollment rate $C(\bar{z})$ tends to one.*

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

Let us now study the behavior of $R(z)$ when 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* $C(\bar{z})$ *such that* $R(\bar{z}) = 1$*. This value is unique and always exists.*

Let us show now 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 our assumption $\sigma_m > \sigma_f$. In the particular case where $\mu_f = \mu_m = \mu$, the equation can be rewritten as

$$\bar{z} = \frac{\mu(\sigma_f - \sigma_m)}{\sigma_f - \sigma_m} = \mu$$

In the simple case in which the means are equal $R(\bar{z}) = 1$ when $\bar{z} = \mu(\bar{z})$.

Since the support of $C(\bar{z})$ is the whole real line, there always exists a value of $\bar{z}$ denoted $barz^*$ 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 $\{\mu_f, \mu_m, \sigma_f, \sigma_m\}$.