WPS8303

Policy Research Working Paper 8303

What Explains the Gender Gap Reversal in Education?

The Role of the Tail Hypothesis

Laurent Bossavie
Ohto Kanninen

WORLD BANK GROUP
Social Protection and Labor Global Practice Group
January 2018

Abstract

The gender gap reversal in educational attainment is ubiquitous in high-income countries, as well as in a growing share of low- and middle-income countries. To account for the reversal, this paper proposes a theoretical framework in which the interplay between the distributions of academic aptitudes and changes in the net benefits of schooling over time affect the gender composition of those getting more schooling. The framework is used to formulate and test alternative hypotheses to explain the reversal. The paper introduces the tail dynamics hypothesis, which builds on the lower dispersion of academic achievement among females observed empirically. It also studies the mean dynamics hypothesis, which is based on previous literature. Both hypotheses can explain the reversal in this framework. However, the assumption behind the tail hypothesis is better supported by the data. Its predictions are also consistent with gender differences in Scholastic Achievement Test score dynamics and in international test score distributions that cannot be explained by previous theories.

This paper is a product of the Social Protection and Labor Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at lbossavie@worldbank.org.

What Explains the Gender Gap Reversal in Education? The Role of the Tail Hypothesis*

Laurent Bossavie† and Ohto Kanninen

Keywords: Educational attainment, gender gap, test scores.
JEL-Codes: I20, J16

* We are grateful to Andrea Ichino and Luigi Guiso for constant advice and discussions. We also would like to thank Jerôme Adda, Christian Dustmann, Nicole Fortin, Peter Hansen, Steve Machin, Tommaso Naniccini, Tuomas Pekkarinnen, Kjell Salvanes and Olmo Silva as well as all participants to the micro-econometrics working group at the EUI for helpful comments and suggestions. Finally, we thank the participants to the SOLE/EALE 2015, EEA confer- ence 2012, Malaga, the EALE conference 2013, Turin and the 2nd Conference on Higher Education and Innovation Research 2012, Helsinki for their suggestions.
† Corresponding author. E-mail: lbossavie@worldbank.org. Address: The Word Bank, 1818 H Street NW, 20433 Washington DC, USA. Phone: 202-751-6478.

1 Introduction

The dramatic expansion of enrollment in education worldwide over the past decades was accompanied by a puzzling and striking phenomenon. As individuals acquired more schooling over time, females not only caught up with males, but surpassed them in educational attainment, a drastic shift from the historical standard. This phenomenon, sometimes referred as the gender gap reversal in education, is widespread globally. Some hypotheses have been proposed in the US literature to account for this reversal.[1] However, to the best of our knowledge, there has been no attempt to test these hypotheses empirically using a comprehensive framework.

Understanding the origins of the reversal is of interest in its own right, but also for efficiency purposes. Identifying its drivers can help understanding dynamics in other areas, particularly in the labor market. For efficiency purposes, it is relevant to determine whether differences in observable outcomes between genders originate from distortions such as discrimination or, instead, from optimizing behaviors based on gender differences in preferences or traits. In particular, learning about the origins of the reversal could potentially help identify areas of policy intervention.

We start by establishing that the gender gap reversal in education is widespread across high-income and lower-income countries, and is broader than just a reversal in participation to tertiary education. Using time-series on enrollment rates in education by gender in more than 140 countries, we evidence that the reversal occurred in virtually all high-income countries, but also in a rapidly growing proportion of lower-income countries. In addition, we show that the reversal occurred in secondary and primary school non-completion.[2] While males have historically constituted the mi-

[1] See Goldin et al. (2006), Cho (2007), Chiappori et al. (2009) or Fortin et al. (2015).

[2] The latter applies mostly to lower income countries.

nority of secondary and primary school non-completers, they became the majority over time.

The central contribution of the paper is to propose a framework to account for this fact and to introduce a novel explanation for the reversal. An important innovation of the theory lies in placing the distributions of academic aptitude at its core. The model uses the micro foundations of investment in schooling of Card (1994), before aggregating to the macro level. It has three building blocks. First, at the micro level, the optimal number of years of schooling is increasing in academic aptitude. Second, the benefits of investing in higher levels of schooling are allowed to vary over time and to differ between genders. Third, we introduce a distributional assumption which, together with an increase in the economy-wide incentives to invest in education, generates either a tail dynamics or a mean dynamics depending on the assumption made. We show that both dynamics can create the reversal observed empirically.

A key contribution of the paper is to introduce the tail dynamics hypothesis, or tail hypothesis, and to show that it can explain the gender gap reversal in education. The hypothesis is novel in the literature and builds on the findings of Machin and Pekkarinen (2008) who report that in an international assessment of 15 year-old students in OECD countries, girls exhibit a lower variance in test scores relative to boys in virtually all countries.[3] It also formalizes the intuition of Becker et al. (2010b), who suggest that the lower variance of non-cognitive skills among females could induce a higher elasticity of enrollment to returns to schooling relative to males. We show that a higher male variability alone can explain the reversal, independently of means.

Using the theoretical framework we developed and country-level time-series data on educational

[3]There exists a large and longstanding literature in psychology on this topic, such as Frasier (1919), Hedges and Nowell (1995), or Jacob (2002), among many others.

attainment by gender, we fit the tail hypothesis by estimating gender differences in the first two moments of test score distributions. The joint evolution of the enrollment rate and the gender ratio generated by the model reproduces the within-country time dynamics observed in our sample. In particular, it can generate the gender gap reversal observed among participants in post-secondary education and secondary school non-completers.

We model the mean dynamics hypothesis or mean hypothesis, suggested by previous literature, in two alternative ways. The first one is the mean benefit hypothesis (MBH), which posits that the net benefits of schooling for females have increased more rapidly over time than for males. The second way to model the mean hypothesis is through the mean performance hypothesis (MPH), which claims that the mean performance of females in achievement tests increased relative to males'. We show that both hypotheses can also produce a reversal in the education gender gap.

The mean hypotheses and the tail hypothesis rely on assumptions and generate predictions that are testable empirically. For the tail hypothesis, we correlate our model parameter estimates for gender differences in test score distributions with estimates from the Project for International Student Assessment (PISA) in 40 countries. We find a significant and positive correlation between parameters estimated from the model and estimates from the PISA reading exam data. This indicates that country dynamics in the educational gender gap resonate gender differences in test score distributions across countries, providing support for the tail dynamics hypothesis.

Regarding the mean hypothesis, we test whether the relationship between test scores and enrollment in tertiary education follows the predictions of the MBH. Estimates using cross-sectional data in 1980 and 2002 in the US cannot be reconciled with the assumptions of the MBH. For the MPH, we find some evidence that the relative academic performance of females has increased in the US, but international evidence is rather ambiguous. We also derive the mean and the tail hypotheses' predictions on the dynamics of the SAT scores in the US. This test again lends slightly more support for the tail hypothesis. Empirically, we mostly focus on tertiary education, but the logic of the model extends to secondary and even primary levels.

The paper is organized as follows. Section 2 presents the empirical facts motivating the paper. Section 3 lays out our theoretical framework at the micro and aggregate level. Section 4 presents the hypotheses for the gender gap reversal. Section 5 formulates the tail hypothesis in our theoretical framework, and shows that it can produce the gender gap reversal observed empirically. Section 6 performs a similar exercise for the mean dynamics hypotheses and tests their underlying assumptions. Section 7 assesses the validity of the predictions of the hypotheses against empirical data. Section 8 concludes.

2 Empirical Motivation

2.1 Data

The data on educational attainment used throughout the paper are from the Barro-Lee educational attainment dataset.[4] The database consists of harmonized data on educational attainment disaggregated by gender every five years from 1950 to 2010, for 146 countries. It provides information on the distribution of educational attainment of the adult population at seven different levels of

[4] See Barro and Lee (2013) and Barro and Lee (2016) for a complete description of the dataset construction and methodology.

schooling: no formal education, incomplete primary, complete primary, lower secondary, upper secondary, incomplete tertiary, and complete tertiary. This allows to compute the participation rate and completion rate at various levels of education for 5-year-band birth cohorts born from 1921 to 1990, by gender.[5] To describe time trends, we look at three levels of educational attainment: primary school non-completion, secondary school non-completion, and participation to tertiary education. To test the gender gap reversal hypotheses, we mostly focus on the tertiary level, but show some results for the secondary level. Our final sample includes 125, 135 and 115 countries for primary, secondary and post-secondary levels, respectively.[6]

2.2 The Gender Gap Reversal in Educational Attainment

Several contributions have reported a convergence, followed by a reversal in the number of females attending tertiary education relative to males for the US (Charles and Luoh (2003), Goldin et al. (2006), Chiappori et al. (2009), Becker et al. (2010b) or Autor and Wasserman (2013)).[7] Figure I shows that the gender gap reversal in participation to post-secondary education is becoming ubiquitous globally. In a large majority of the countries in the sample (∼90%), males in the earliest cohort outnumber females among participants to tertiary education. In the youngest cohort born in 1980, 64% of countries have seen their gender imbalances in post-secondary attainment revert over time (Table I). When averaged over all countries in the sample, the data show a reversal in the gender gap, reported in Figure I. The timing of the reversal varies across countries and regions.

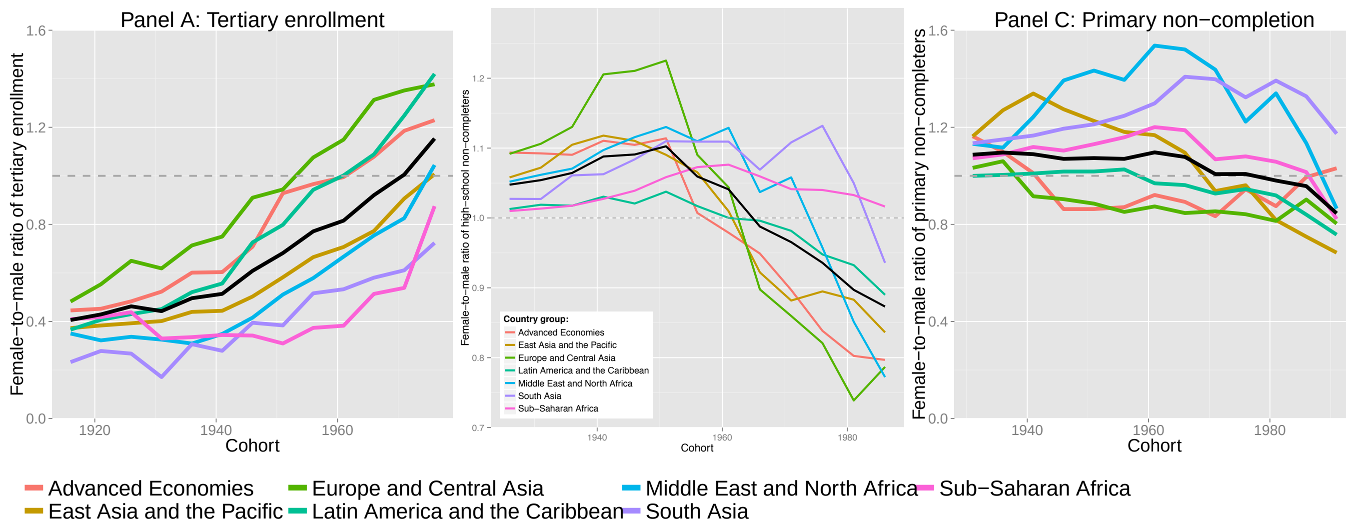
[5]For primary education, we compute non-completion rates for individuals born between 1936 to 1995, from 1931 to 1990 for secondary school non-completion and from 1921 to 1980 for tertiary education.

[6]Countries were dropped from the sample due to either having missing or zero cohort/country observations, a ratio of female/male cohort/country tertiary enrollment observation above 5, or a cohort/country male population below 10,000.

[7]Pekkarinen (2012) has also reported this phenomenon for Scandinavian countries.

Advanced economies, Latin America and Europe and Central Asia have seen their gender gaps revert to a high proportion, but other regions lag behind, such as Sub-Saharan Africa with only 8% of countries having experienced the reversal.

Figure I: The Gender Gap Reversal in Educational Attainment



Notes. The black line represents the unweighted average for all countries in the sample. Averages are unweighted by country size. In each panel, countries were dropped for having missing or zero cohort/country observations, a ratio of female to male ratio for tertiary enrollment observation above 5, or a cohort/country male population below 10,000. The total sample size is 115 (**Panel A**), 135 (**Panel B**) and 125 (**Panel C**) countries. See Table I for a breakdown of sample size by country group.

The gender gap reversal in education is not limited to tertiary education. The gender composition of secondary school non-completers has also reversed from female majority to male majority over time. Males now outnumber females among low educational achievers and this reversal is also ubiquitous globally. Females in the earliest cohort in the Barro-Lee database outnumber males as secondary-school non-completers in over 95 percent of countries in the sample. For the youngest cohort, the gender gap has reversed in 70% of countries. The reversal occurred in most regions and is also observed when statistics are averaged over all countries in the sample, as shown in Figure I and Table I. In Sub-Saharan Africa, only 21% of countries have experienced the reversal.

For primary school non-completion, a reversal of the gender gap from female majority to male majority also took place, although patterns are not as clear cut. Country-specific results are more unstable overall, as in many countries, particularly high income countries, primary school non-completion has been rare, even in older cohorts. Panel C of Figure I shows that, when educational attainment is aggregated by region, most regions experienced the reversal: while the number of females not completing primary school was higher than males in earlier cohorts, more males than females do not complete primary education among younger cohorts.

Table I: Share of Countries that Experienced the Gender Gap Reversal, by Birth Cohort

Panel A: Tertiary education

Birth cohort	1921	1926	1931	1936	1941	1946	1951	1956	1961	1966	1971	1976	N
Middle East and North Africa	0	0	0	0	0	0	0.10	0.10	0.20	0.20	0.20	0.30	10
Sub-Saharan Africa	0	0	0	0	0	0	0	0	0.04	0.04	0.04	0.04	24
Advanced Economies	0	0	0	0	0	0	0	0.04	0.48	0.52	0.70	0.78	23
Latin America and the Caribbean	0	0	0	0	0	0	0	0.10	0.35	0.40	0.45	0.70	20
South Asia	0	0	0	0	0	0	0	0	0.25	0.25	0.25	0.25	4
East Asia and the Pacific	0	0	0	0	0	0	0	0	0.07	0.07	0.21	0.29	14
Europe and Central Asia	0	0	0	0	0	0.05	0.25	0.35	0.65	0.70	0.85	0.90	20

Panel B: Secondary education

Birth cohort	1931	1936	1941	1946	1951	1956	1961	1966	1971	1976	1981	1986	N
Middle East and North Africa	0	0	0	0	0.07	0.07	0.14	0.14	0.29	0.36	0.50	0.57	14
Sub-Saharan Africa	0	0	0	0	0	0.03	0.06	0.09	0.12	0.15	0.18	0.21	33
Advanced Economies	0	0	0	0	0.09	0.13	0.26	0.48	0.65	0.70	0.70	0.78	23
Latin America and the Caribbean	0	0	0.09	0.09	0.17	0.22	0.35	0.48	0.52	0.57	0.65	0.65	23
South Asia	0	0	0	0	0	0	0.17	0.17	0.17	0.17	0.33	0.50	6
East Asia and the Pacific	0	0	0	0	0	0	0.19	0.31	0.50	0.50	0.50	0.50	16
Europe and Central Asia	0	0	0.10	0.10	0.15	0.15	0.30	0.30	0.45	0.60	0.70	0.80	20

Panel C: Primary education

Birth cohort	1936	1941	1946	1951	1956	1961	1966	1971	1976	1981	1986	1991	N
Middle East and North Africa	0	0	0	0	0	0	0	0	0	0	0.08	0.15	13
Sub-Saharan Africa	0	0	0	0	0	0	0	0	0	0.03	0.03	0.13	30
Advanced Economies	0	0	0	0	0	0	0.05	0.14	0.19	0.24	0.38	0.52	21
Latin America and the Caribbean	0	0	0	0	0	0	0	0	0	0.04	0.04	0.13	23
South Asia	0	0	0	0	0	0	0	0	0	0	0	0	5
East Asia and the Pacific	0	0	0	0	0	0	0	0	0	0.14	0.14	0.14	14
Europe and Central Asia	0	0	0	0	0	0	0.05	0.05	0.21	0.26	0.53	0.58	19

Notes. **Panel A.** A reversal is defined to happen when the female-to-male ratio among individuals that received some tertiary education increases above 1 and does not revert back to below 1. The total sample size is 115 countries. **Panel B.** A reversal is defined to take place, when the female-to-male ratio among those that did not complete secondary education falls below 1 and does not revert back to above 1. The total sample size is 135 countries. **Panel C.** A reversal is defined to take place, when the female-to-male ratio among those that did not complete primary education falls below 1 and does not revert back to above 1. The total sample size is 125 countries.
Source. Barro-Lee database 2013.

3 Theoretical Framework

3.1 Micro Foundations

The economy is assumed to be populated by individuals that differ in their academic aptitude z, which is continuous and perfectly observed by individuals.[8] For the sake of simplicity, a single-period model is assumed in which individuals receive the benefits of their investment in education in the same period as they invest. Individuals choose years of schooling s to maximize their utility U. Building on Card (1994), we express the utility function of individuals in the economy as:

$$U(s) = B(s) - C(s),$$

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, which is increasing and convex in s. The first-order conditions for the individual maximization problem read as:

$$B'(s) = C'(s),$$

where $B'(s)$ and $C'(s)$ are the marginal benefits and costs of schooling, respectively. Following Card (1994), we linearize the model by assuming that $B'(s)$ and $C'(s)$ are linear functions of s,

[8] We use the term academic aptitude in the theoretical framework, but we only observe achievement test scores empirically. 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 academic aptitude to test scores is beyond the scope of this paper, and we refer to the combination of cognitive and non-cognitive abilities captured by test scores as academic aptitude in the remainder of the paper.

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

$$B'(s) = z_j - k_1 s$$

$$C'(s) = k_2 s,$$

where $k_1 > 0$ and $k_2 > 0$. Intuitively, individuals with higher academic ability z_j receive higher marginal benefits from schooling.[9] In this framework, the optimal level of schooling s chosen by individual j is:

$$s_j^* = z_j \cdot b, \tag{1}$$

where $b \equiv \frac{1}{k_1+k_2}$ is an exogenous technology parameter, which we refer to as the structural net benefits of schooling (hereafter: net benefits), identical for all individuals in the economy. The net benefits capture the monetary benefits, non-monetary benefits and costs of schooling.

The optimal value of schooling chosen by individual j is therefore strictly increasing in individual academic aptitude z_j. In this framework, the minimum level of academic aptitude $\bar{z}$ so that individuals choose a given level of schooling $\bar{s}$, such as tertiary education, can be expressed as:

$$\bar{z} = \frac{\bar{s}}{b} \tag{2}$$

Equation 2 states that individuals whose academic aptitude is below the threshold $\bar{z}$ choose an optimal level of education below $\bar{s}$, while individuals whose ability is equal to or greater than $\bar{z}$

[9]In the original model of Card (1994), the expression for C'(s) also includes an individual-specific intercept, capturing individual-specific circumstances such as access to wealth and network or taste for education. For the sake of simplicity, we abstract from this distinction in our model.

choose $\bar{s}$ and above. It also implies that the ability threshold is determined by b, common to all individuals in a given cohort. An immediate implication of Equation 2 is:

$$\frac{\partial \bar{z}}{\partial b} \leq 0.$$

In words, the minimum level of ability required to attend a given level of schooling $\bar{s}$, such as tertiary education, decreases with the net benefits of investing in schooling in the economy.

3.2 Aggregate Enrollment Rate

We now assume that the economy is populated by successive cohorts. Each cohort comprises a continuum of agents that invest in schooling and differ in their level of academic aptitude z. Let $f_z(\bar{z})$ denote the probability density function of academic aptitude z in the population of a given cohort. The complementary cumulative distribution function of z is defined as:

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

All individuals belonging to the same cohort are exposed to the same value of the exogenous parameter $b_t \equiv \frac{1}{k_{1,t}+k_{2,t}}$, regardless of their level of ability. Given the micro properties of the model expressed in Equations (1) and (2), the share of individuals in the cohort choosing a level of schooling at least equal to $\bar{s}$ is:

$$P(\bar{z}) = 1 - F_z(\frac{\bar{s}}{b}) = G_z(\frac{\bar{s}}{b}) = G_z(\bar{z}). \tag{3}$$

Equation (3) states that the mass of individuals choosing a level of schooling of at least $\bar{s}$ or higher, for a given value of b, is made of all individuals whose ability is above the ability threshold $\bar{z}$. Exogenously to individual schooling decisions, b varies across cohorts. Equation (3) implies:

$$\frac{\partial P(\bar{z})}{\partial b} \geq 0.$$

The share of individuals choosing a level of schooling equal or higher than $\bar{s}$ increases with the net benefits of schooling b. One key implication at the aggregate level is that, as the net benefits of education in the economy rise, enrollment increases and the mean aptitude of individuals enrolled decreases. In Appendix A1.2, we show that this result is empirically supported, using US data.

4 Tail and Mean Dynamics: Hypotheses for the Gender Gap Reversal

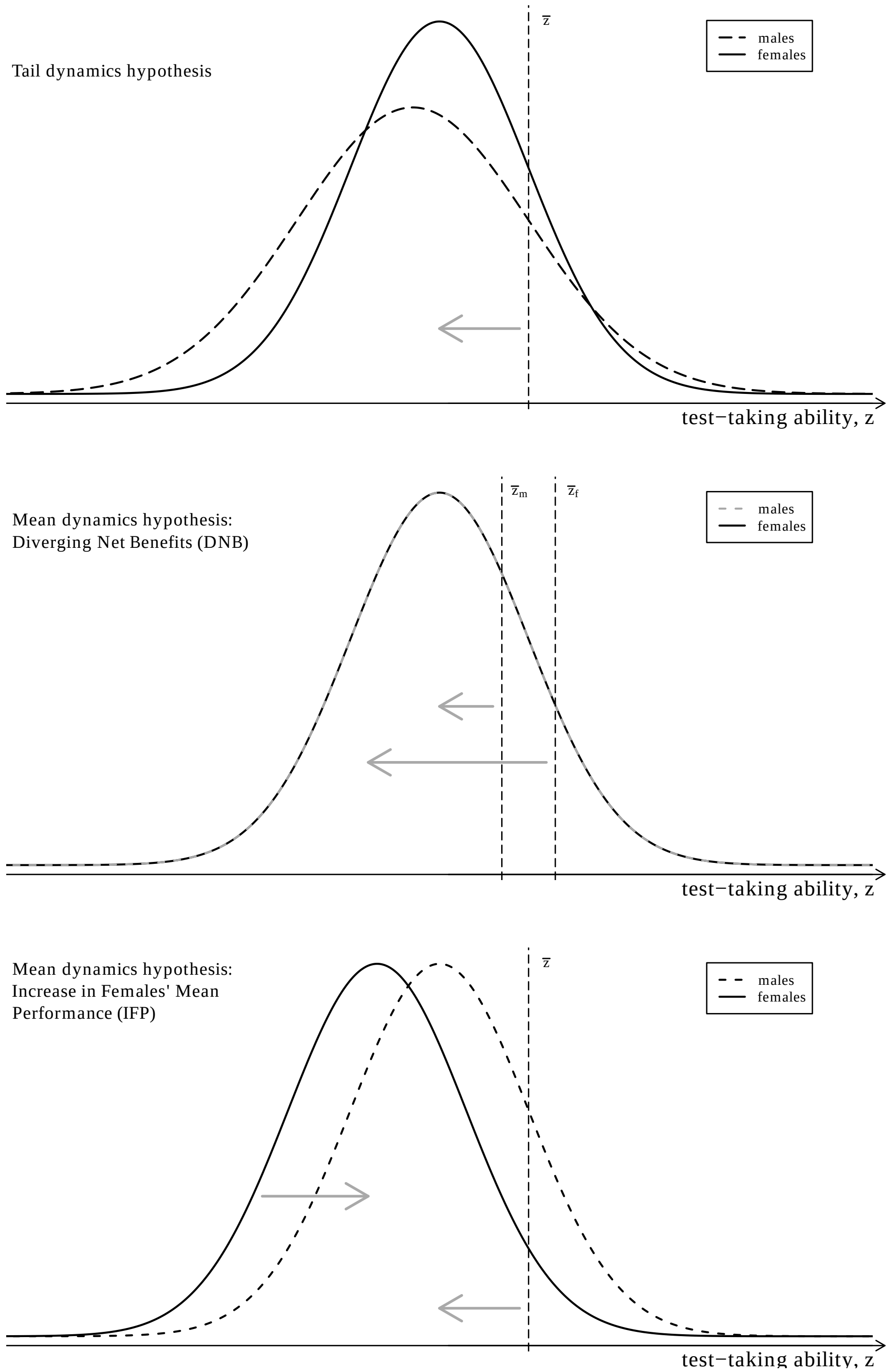
We formulate three hypotheses for the gender gap reversal, summarized in Table **??**. These hypotheses can be grouped into two broader categories: the tail dynamics hypothesis (or tail hypothesis) and the mean dynamics hypotheses (or mean hypotheses). The validity of the assumptions behind these hypotheses is discussed in sections 5 and 6 and their predictions are tested in section 7. We study the two mean hypotheses under the same category, since they are, as we demonstrate, very similar.

The tail hypothesis states that $G_z(\cdot)$ is gender specific with $Var[z_m] > Var[z_f]$. In words, academic aptitude for males and females, z_m and z_f respectively, have different distributions with a greater variance among males. The net benefits of education b are assumed to be identical between genders. The increase in b (or, equivalently, the decrease in the aptitude threshold $\bar{z}$) over time, combined with the greater dispersion of aptitude among males, produces the reversal.

The mean benefits hypothesis (MBH) claims that the net benefits of education differ between genders, i.e. there exist gender-specific b_f and b_m (or, equivalently, gender-specific ability thresholds $\bar{z}_f$ and $\bar{z}_m$) that have different time dynamics over time. Prior to the reversal, $b_f < b_m$ (or, equivalently $\bar{z}_f > \bar{z}_m$) before b_f progressively converges towards b_m and surpasses it over time, generating the reversal. Ability distributions $G_z(\cdot)$ are assumed to be identical for both genders.

The mean performance hypothesis (MPH) claims that the mean of academic aptitude for females $E[z_f]$ has increased over time and progressively surpassed $E[z_m]$, producing the gender gap reversal. The variance of z and the net benefits of education (or, equivalently, the academic aptitude threshold $\bar{z}$) are identical for both genders.

Figure II: Three Hypotheses for the Gender Gap Reversal in Education



Notes. **Panel A:** $\bar{z}$ is the threshold of academic aptitude above which individuals enroll at a given level of schooling $\bar{s}$, such as tertiary education. The grey arrow indicates the change in $\bar{z}$ over time. **Panel B:** The grey arrows indicate a faster decrease of $\bar{z}_f$ relative to $\bar{z}_m$, the female and male-specific ability thresholds, over time. The distribution of academic aptitude by gender are overlapping as they are identical under the mean benefits hypothesis. **Panel C:** The arrow pointing to the right indicate an increase in the mean of the ability distribution for females, μ_f, relative to males over time as $\bar{z}$ decreases over time.

5 The Tail Dynamics Hypothesis

5.1 The Relationship between Enrollment Rate in Education and Gender Composition of Individuals Enrolled

Machin and Pekkarinen (2008) have shown that the test score distribution of males exhibits a larger variance than females in a large sample of OECD countries. The authors use data from the 2003 Project for Student International Assessment (PISA) which tests nationally representative samples of 15 year old students from 40 countries in mathematics and reading. They report that males' test score variance is strictly greater than females' in 38 countries for mathematics and 39 countries for reading. The gender difference in variance is statistically significant in all but five countries with an average male-to-female variance ratio of 1.21 for reading and 1.20 for mathematics.

Although this evidence is recently known to economists, extensive and long-standing evidence on the larger variance in the distribution of some skills among males has been reported in the psychology literature.[10] In line with this finding, Becker et al. (2010a) and Becker et al. (2010b) suggested that a lower variance of non-cognitive skills among females could explain the gender gap reversal in education by inducing a higher elasticity of females' schooling to returns to education.

In our framework, let z_m and z_f be random variables that denote the academic aptitude for males and females, and $f_z(\bar{z}_m)$ and $f_z(\bar{z}_f)$ their density functions. The tail hypothesis assumes $Var[z_m] > Var[z_f]$, where the distribution of z_m and z_f is invariant over time. No assumption is imposed on the relative value of $E[z_m]$ and $E[z_f]$. We denote μ and σ^2 the mean and variance of the distri-

[10] See Frasier (1919), Hedges and Nowell (1995) or Jacob (2002), among many others.

butions, respectively. For empirical estimations and illustrations, we assume that z_m and z_f are normally distributed, as it is the closest approximation of empirical test score distributions.

In our framework, the enrollment rate in a given level of schooling $\bar{s}$ among the population of a given cohort is given by:

$$P^{TH}(\cdot) \equiv \frac{G_{z_f}(\bar{z}, \mu_f, \sigma_f^2) + G_{z_m}(\bar{z}, \mu_m, \sigma_m^2)}{2},$$

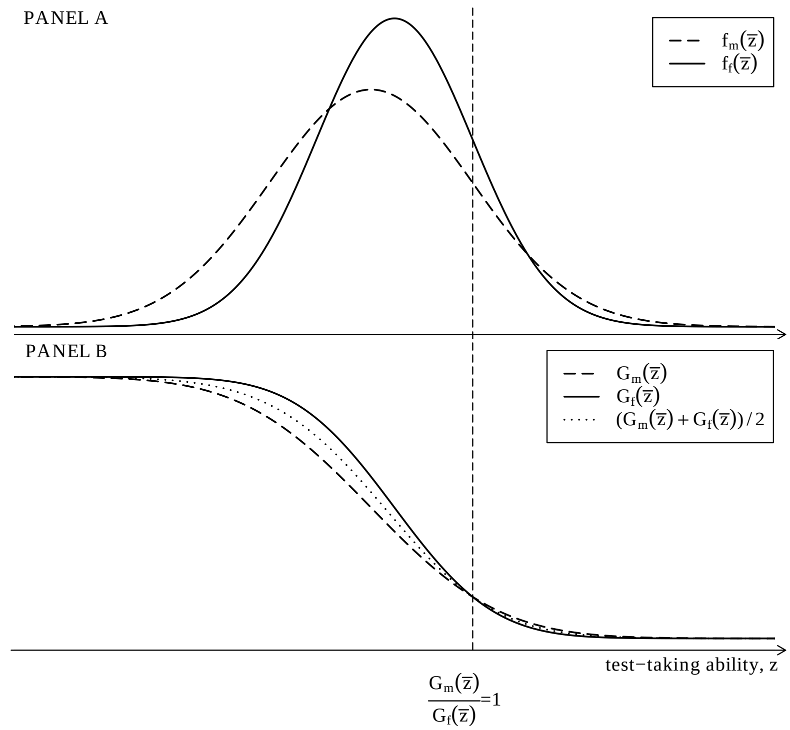
and the female-to-male ratio among individuals enrolled, denoted $R(\bar{z})$, can be expressed as:

$$R^{TH}(\cdot) \equiv \frac{G_{z_f}(\bar{z}, \mu_f, \sigma_f^2)}{G_{z_m}(\bar{z}, \mu_m, \sigma_m^2)}$$

Panel A of Figure III displays the two distributions of academic aptitudes, when $\sigma_m^2 > \sigma_f^2$ under the tail hypothesis, as well as the ability threshold $\bar{z}$ that truncates the distribution of those enrolling at higher levels of schooling. Individuals on the right side of the ability threshold enroll, while those on the left side do not. Panel B displays the two corresponding complementary cumulative distribution functions (CCDFs) under that assumption, where the reversal occurs when the two CCDFs cross. Figure IV illustrates the relationship between total enrollment rate and the gender ratio among individuals enrolled, where the reversal occurs when the ratio $\frac{G_{z_f}(\bar{z})}{G_{z_m}(\bar{z})}$ reaches 1.

A decrease in the ability threshold for enrolling, $\bar{z}$, associated with an increase in the net benefits of education b and leading to an increase in the enrollment rate P, is required for the reversal to occur under the tail hypothesis. In Appendix A1.1, we report evidence showing that the mean ability of individuals attending tertiary education indeed decreased over time, using US data.

Figure III: Distribution Functions of Test Scores by Gender under the Tail Hypothesis: Illustration



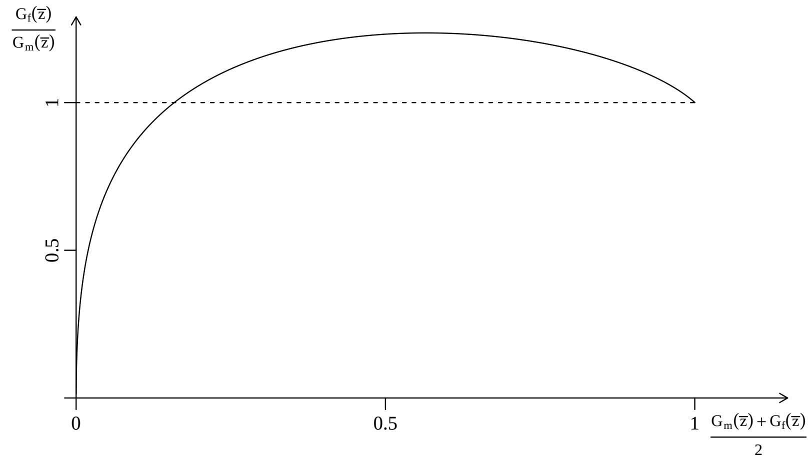
Notes. **Panel A:** The two curves show the probability density functions of academic aptitude, z, among males (full line) and females (dashed line), with $\sigma_m^2 > \sigma_f^2$ under the tail dynamics hypothesis and z being normally distributed. **Panel B:** The two curves show the complementary cumulative distribution function (CCDF), resulting from the integration from $+\infty$ to z of $f_{z_f}(z)$ and $f_{z_m}(z)$. The grey arrow indicates a decrease in $\bar{z}$. The gender gap reversal in enrollment occurs when the ratio of the two CCDFs reaches 1, as indicated in Panel B.

Under the assumption that $\sigma_m^2 > \sigma_f^2$, it can be shown that the relationship between the female-to-male ratio among individuals enrolled at a given level of schooling, denoted R^{TH}, and the enrolled rate P^{TH} has three notable properties:

Proposition 1. *The female-to-male ratio R^{TH} tends to zero when the total enrollment rate P^{TH} tends to zero.*

Proposition 2. *The female-to-male ratio R^{TH} tends to one when the total enrollment rate P^{TH} tends to one.*

Figure IV: Relationship between the Enrollment Rate in Tertiary Education and the Female-to-male Ratio among Individuals Enrolled under the Tail Hypothesis: Illustration



Notes. The x-axis reports the enrollment rate. The y-axis female-to-male ratio among individuals that enroll. The gender gap reversal occurs when the female-to-male ratio crosses y=1.

Proposition 3. *There exists a value of* $P^{TH} \in [0, 1[$ *such that* $R^{TH} = 1$*. This value is unique and always exists.*

Proof. See Appendix A4.

The distributional assumption $\sigma_m^2 > \sigma_f^2$ is necessary and sufficient for Proposition 1 to 3 to hold, without any condition imposed on the relative values of μ_m and μ_f. In addition, Propositions 1 to 3 also hold when z is assumed to follow two-parameter probability distribution functions other than the normal distribution.[11]

5.2 Empirical Estimation

From the Barro-Lee database, we observe the enrollment rate for a given level of education in country i and cohort t, denoted x_{it}, and the gender ratio among enrolled individuals, denoted y_{it}.

[11] Proofs for distributions other than the normal are available upon request.

In our model $P^{TH}(\cdot)$ and $R^{TH}(\cdot)$, the theoretical equivalents of x_{it} and y_{it}, are determined by the distributions of aptitude by gender, G_{z_f}, G_{z_m} and the ability threshold $\bar{z}$. Unlike $\bar{z}_t$, x_t is observable, and the two distributions are assumed to be fully characterized by the four-parameter vector $\{\mu_m,\, \sigma^2_m, \mu_f,\, \sigma^2_f\}$.

Without loss of generality, the parameters of the model can be reduced to two, by normalizing one of the two probability density functions. We standardize the female probability density function such that $f_{z_f}(\bar{z}_{it}) \sim N(0,1)$ and denote $(\mu_i,\, \sigma^2_i)$ the first two moments of the males' distribution of academic aptitudes relative to females in country i:

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

$$\sigma^2_i = \frac{\sigma^2_{i,m}}{\sigma^2_{i,f}} = \sigma^2_{i,m}.$$

The model under the tail hypothesis predicts a unique value $\hat{y}_{it}$, conditional on the triple $\{x_{it}, \mu_i, \sigma^2_i\}$. To fit the model, we estimate $\{\mu_i; \sigma^2_i\}$ for country i by maximum-likelihood estimation from the following equation:

$$y_{it} = \frac{G_{z_f}(\bar{z}_{it})}{G_{z_m}(\bar{z}_{it}, \mu_i, \sigma^2_i)} \cdot \exp(\varepsilon_{it}), \tag{4}$$

where $\varepsilon_{it} \sim N(0, \sigma^2_\varepsilon)$. Taking the logarithm of Equation 4 and substituting $\bar{z}_{it} = P^{-1}(x_{it}, \mu_i, \sigma^2_i)$ yields the final econometric model:

$$\log y_{it} = \log G_{z_f}(P^{-1}(x_{it}, \mu_i, \sigma^2_i)) - \log G_{z_m}(P^{-1}(x_{it}, \mu_i, \sigma^2_i), \mu_i, \sigma^2_i) + \varepsilon_{it}.$$

As 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. Thus, the model is a non-linear mapping from x to y defined by the parameters of the males' ability distribution relative to the females' distribution.

Figures A3 and A4 report the estimated relationship between the enrollment rate in tertiary education x and the female-to-male ratio among individuals enrolled y, when $\{\mu; \sigma^2\}$ are fitted with the model under the tail hypothesis. The model generates an accurate fit for the relationship between x and y observed in countries in the sample.[12] Figures A5 and A6 also shows our model fit for the relationship between the secondary school non-completion rate, and the gender ratio among non-completers.

6 The Mean Dynamics Hypotheses

6.1 The Mean Benefits Hypothesis (MBH)

Goldin et al. (2006) and Chiappori et al. (2009) suggested that the returns from education, including labor market and marriage market returns, may have risen more for women over the past decades, which could have driven the gender gap reversal. Both contributions invoke the progressive removal of female career barriers, as a driver of increased returns. Goldin et al. (2006) associate the removal of barriers to female employment to a progressive change in fertility and marriage patterns, driven by the access to reliable contraception (Goldin and Katz (2002)). The authors

[12] Due to space limitations, we only report results for a subset of countries in the sample. Results for the full set of countries are available upon request.

posit that this raised females' expectations regarding future labor market outcomes, increased labor force participation, and moved female employment out of traditionally female occupations. In a similar spirit, Chiappori et al. (2009) argue that technological progress has freed women from many domestic tasks, which disproportionally increased the labor market and marriage market returns of schooling for females. According to both contributions, the disproportionate increase in the benefits of education for females relative to males generated the reversal.

We formulate the mean benefits hypothesis in our framework by allowing b to differ between genders, where b_m and b_f denote the net benefits of education for males and females, respectively. b_f and b_m are allowed to have different dynamics over time. $G_z(.)$ is assumed to be identical for males and females. Under the MBH, the optimal level of schooling chosen by individuals is gender-specific:

$$s^* = z \cdot b_g,$$

or, equivalently, $\bar{z}$, is gender specific:

$$\bar{z}_g = \frac{\bar{s}}{b_g}, \tag{5}$$

where $g = \{m; f\}$, with m standing for males and f for females. Equation 5 states that there exists gender-specific ability thresholds such that males and females enroll in a given level of schooling $\bar{s}$, denoted $\bar{z}_m$ and $\bar{z}_f$ respectively. The enrollment rate in a given level of schooling $\bar{s}$ reads:

$$P^{MBH}(\bar{z}_f, \bar{z}_m) \equiv \frac{G_z(\bar{z}_m) + G_z(\bar{z}_f)}{2}, \tag{6}$$

Under the MBH, a shift from $b_m > b_f$ to $b_m < b_f$ over time, or equivalently, from $\bar{z}_m < \bar{z}_f$ to

$\bar{z}_m > \bar{z}_f$, is a necessary and sufficient condition for the gender gap reversal in education to occur. The gender ratio among individuals enrolled in a given level of education $\bar{s}$ is given by:

$$R^{MBH}(\bar{z}_f, \bar{z}_m) \equiv \frac{G_z(\bar{z}_f)}{G_z(\bar{z}_m)}, \tag{7}$$

and:

$$\begin{cases} R^{MBH} < 1 & \text{if } \bar{z}_m < \bar{z}_f \\ R^{MBH} = 1 & \text{if } \bar{z}_m = \bar{z}_f \\ R^{MBH} > 1 & \text{if } \bar{z}_m > \bar{z}_f. \end{cases}$$

Using the Barro-Lee data, we calculate the parameter values of the model by solving the system of two equations (6 and 7) and two unknowns $\{\bar{z}_{f_t}, \bar{z}_{m_t}\}$ for each cohort t, where $P^{MBH}(\bar{z}_{f_t}, \bar{z}_{m_t}) = y_t$ and $R^{MBH}(\bar{z}_{f_t}, \bar{z}_{m_t}) = x_t$. The estimated $\bar{z}_{f_t}$ and $\bar{z}_{m_t}$ for tertiary and secondary education, averaged by country group, are reported in Figure A7.

We now assess the validity of the assumptions of the MBH. A necessary and sufficient condition for the reversal in the education gender gap under the MBH is:

$$\begin{cases} b^{MBH}_{m,t_1} - b^{MBH}_{f,t_1} > 0 \\ b^{MBH}_{m,t_2} - b^{MBH}_{f,t_2} < 0, \end{cases}$$

where t_1 is a time prior to the reversal, and t_2 a time post reversal. Under the MBH, females used to enroll less than males conditional on test scores prior to the reversal, but now have a higher propensity to enroll. Under the tail hypothesis, on the other hand, there is no gender-specific time trend in the relationship between academic aptitude and enrollment.

A monetary component of the net benefits of education is the college wage premium, defined as the wage difference between individuals that attended tertiary education and those who did not. While Chiappori et al. (2009), Card and Lemieux (2000), Charles and Luoh (2003) and Trostel et al. (2002) find a higher college wage premium for women in recent years, this can only explain the reversal if it constitutes a shift compared to earlier years.[13] Cho (2007) shows that trends in the college premium have been very similar for men and women over the period 1970–2000, using data from the CPS. Becker et al. (2010b) also find little difference in the premiums for men and women since the 1980s, with both premiums growing at about the same rate over time. Considering the non-monetary benefits of education, the authors do not find clear evidence for a greater marriage market improvement for college-educated men relative to college educated-women. They also show that the effect of college on life expectancy remains greater for men despite a narrowing of the gender gap, which cannot justify that women now enroll more in college relative to men.

Testing these alternative hypotheses requires comparable data on test scores before and after the reversal that can be linked to enrollment decisions by gender in both periods. Surveys also need to be representative of the population of males and females in the country. According to the Barro-Lee time-series, the reversal in participation to tertiary education occurred in the mid-1980s in the US. We therefore need to observe the relationship between test scores and post-secondary enrollment before and after the mid 80s, with a sufficient time gap to capture time trends.

We exploit two nationally-representative longitudinal surveys for the US: the High School and Beyond (HS&B), starting in 1980, and the Education Longitudinal Survey (ELS), with base year in 2002. Both surveys contain information on standardized mathematics and reading tests in grade

[13] The methodology behind the findings of a higher college premium for females in the US has however been challenged by Hubbard (2011).

10, that can be linked to tertiary education attendance a few years later. These tests are 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 population of a specific age group.

Using these two surveys, we estimate the empirical relationship between academic aptitude proxied by achievement test scores, and enrollment in tertiary education, separately by gender. We estimate a binary version of Equation 1 in our model, where the dependent variable is tertiary education attendance, denoted H. H is observable and is a function of the latent variable s^* in the model:

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

We estimate β, the net benefits to tertiary education, from the linear probability model[14]:

$$H_j = \beta_g z_j + \epsilon_j,$$

where $g \in \{m, f\}$ and β_g is estimated in both 1980 and 2002, and z is measured by achievement test scores at age 15.

Table II reports our estimates for $(\beta_{m,t=1980}, \beta_{f,t=1980})$ and $(\beta_{m,t=2002}, \beta_{f,t=2002})$ using a linear probability model. The estimated $\hat{\beta_m} - \hat{\beta_f}$ is negative in both 1980 and 2002, meaning that for females, academic aptitude is a better predictor of tertiary education enrollment. The benefits of education

[14]This setting allows to study how well aptitude maps to educational enrollment. While our theoretical model produces a discrete jump in the probability to enroll from 0 to 1 above a given threshold of academic aptitude, many additional factors correlated or uncorrelated with aptitude are at play empirically. it is therefore as strongly simplified model of the determinants of enrollment in education. Empirically, we study actual probabilistic data and estimate a linear binary model where the likelihood to enroll increases with test scores in a continuous fashion. we believe the test is reasonable as we are primarily interested in general trends and signs or the relationships, rather than the exact magnitude of those relationships.

Table II: Model Estimates of the Relationship between Academic Aptitude and Participation to Tertiary Education in 1980 and 2002, by Gender

	Dependent Variable: Dummy for Tertiary Education Attendance					
	1980			2002		
	Males	Females	M-F Diff.	Males	Females	M-F Diff.
$\hat{\beta_m}$	0.255***			0.414***		
	(0.011)			(0.06)		
$\hat{\beta_f}$		0.329***			0.479***	
		(0.010)			(0.02)	
$\hat{\beta_f} - \hat{\beta_m}$			-0.074***			-0.065***
N. observations	6,066	6,427	-	6,023	6,515	-

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Standard errors are reported in parentheses. The regressions are linear probability models run by ordinary least squares. Regressions do not include an intercept. Test scores are standardized to have a mean of zero and a standard deviation of 1 in the full sample.
Sources. High School and Beyond, sophomore cohort: base year (1980) and 2nd follow-up (1984), US Educational Longitudinal Study, sophomore cohort: base year (2002) and 2nd follow-up (2006)

were higher already for females in 1980, prior to the reversal, and the estimated $\hat{\beta_m} - \hat{\beta_f}$ is very stable between 1980 and 2002. If anything, there was a slight increase in b for males relative to females from -0.74 to -0.65. This finding is not consistent with the MBH.

6.2 The Mean Performance Hypothesis (MPH)

A change in the distribution of z for females relative to males over time, in particular a relative increase in females' mean academic aptitude, can also generate the gender gap reversal in our framework. Such changes over time have been suggested as potential explanations for the reversal by Goldin et al. (2006) and Cho (2007) for the US. Goldin et al. (2006) use samples of high school graduating seniors from three US longitudinal surveys in 1957, 1972, 1988, among which the last two are nationally representative. Looking at achievement test scores, they find that girls reduced their disadvantage in math, and increased their advantage in reading between 1972 and 1992. Using three nationally representative longitudinal data sets of high school students, Cho (2007) also looks at the evolution of female performance in high school over time. He reports that females' mean high school test scores have increased more rapidly than males' over the past three decades. Using a simple Oaxaca decomposition, he finds that women's progress in high school achievement can account for more than half of the change in the college enrollment gender gap over the past thirty years. Fortin et al. (2015) use self-reported grades, rather than aptitude test scores, to look at the evolution of female high-school performance over time. Using a sample of 12th graders from the Monitoring the Future (MTF) study in 1976 and 1991, they report that the gender gap in the mean grades of high school seniors remained very stable since the 1970s. They however find that the mode of girls has shifted upwards over the period 1980–2010 compared to boys'.

We formalize the MPH in our framework as an increase in the mean aptitude for females relative to males over time, which, as we illustrate, can also generate the reversal. Under the MPH, as under the tail hypothesis, there exists a lower bound of academic aptitude $\bar{z}$ such that individuals choose a given level of schooling $\bar{s}$. Making a normalizing assumption of $\mu_{m_t} = 0$ for all t, the total enrollment rate in a given level of schooling $\bar{s}$ is given by:

$$P^{MPH}(\mu_{f,t}, \bar{z}_t) \equiv \frac{G_{z_m}(\bar{z}_t) + G_{z_{f,t}}(\mu_{f,t}, \bar{z}_t)}{2}. \tag{8}$$

A change from $E[z_{f,t}] < E[z_m]$ to $E[z_{f,t}] > E[z_m]$ is a necessary and sufficient condition for the gender gap reversal in education to occur, when the distributions have identical variance. Under the MPH, the gender ratio among individuals enrolled can be expressed as:

$$R^{MPH}(\mu_{f,t}, \bar{z}_t) \equiv \frac{G_{z_{f,t}}(\mu_{f,t}, \bar{z}_t)}{G_{z_m}(\bar{z}_t)}, \tag{9}$$

and:

$$\begin{cases} R^{MPH} < 1 & \text{when } E[z_{f,t}] < E[z_m] \\ R^{MPH} = 1 & \text{when } E[z_{f,t}] = E[z_m] \\ R^{MPH} > 1 & \text{when } E[z_{f,t}] > E[z_m] \end{cases}$$

We calculate the parameter values of the model by solving the system of two equations (8 and 9) and two unknowns $\{\mu_{f,t}, \bar{z}_t\}$ for each period t, where $P^{MPH}(\mu_{f,t}, \bar{z}_t) = y_t$ and $R^{MPH}(\mu_{f,t}, \bar{z}_t) = x_t$. The estimated $\mu_{f,t}$ and $\bar{z}_t$ for tertiary and secondary education, averaged by country group, are reported in Figure A8.

To evaluate whether the academic achievement of females increased over time relative to boys for the entire population, we need to use a representative sample of the population of a given age group. This precludes the use of test score data at an age at which schooling is no longer compulsory, in order to avoid gender-biased estimates due to early school dropouts.[15] In addition, one needs to rely on achievement test scores rather than grades, as grades can be subject to multiple biases when measuring scholastic achievement.

We first investigate this question in the US context, by looking at the evolution of the academic aptitude of 10th graders by gender over the period 1980–2002. We use two nationally-representative longitudinal surveys for the US, the High School and Beyond (HS&B) started in 1980 and the

[15]As pointed out by Fortin et al. (2015), the sample of test-takers is otherwise positively selected in a way that is likely to be gender specific, as we showed that males are more likely to be early dropouts than females.

Table III: Gender Gap in Average Achievement Test score in Mathematics Prior and Post-reversal, US Estimates

	1980				2002			
	Males	Females	M-F mean difference	M-to-F Var. ratio	Males	Females	M-F mean difference	M-to-F Var. ratio
Mean score	0.062	-0.059	0.134***	-	0.053	-0.053	0.106***	-
Test score sd	1.033	0.964	-	1.15***	0.959	1.037	-	1.17***
Observations	6,066	6,427	-	-	6,023	6,515	-	-

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Standard errors are reported in parentheses. Test scores are standardized to have a mean of zero and a standard deviation of 1 in the full sample.
Sources. High School and Beyond 1980 (base year) and US Educational Longitudinal Study 2002 (base year).

Education Longitudinal Study (ELS) initiated in 2002. Both surveys contain information on standardized test scores in mathematics in 10th grade that are comparable between the two years. The results are reported in Table III. The table shows that the girls' disadvantage decreased from -0.134 to -0.106 standard deviation from 1980 to 2002 which is statistically significant, suggesting an increase in mean z for females relative to males.

We also conduct a similar exercise for the UK at age 11, and for OECD countries using PISA test scores data over the past 15 years. Results are reported in Appendix A2 and show little change in the gender gap gap in mean achievement test score over time. Evidence about an increase in females' mean performance over time is therefore ambiguous overall.

6.3 The Similarity between the MBH and MPH

The MPH in our framework models the gender gap reversal such that females initially have lower participation rates in education, due to a lower $E[z]$. Conditional on z, the net benefits of schooling and the optional level of schooling chosen are identical for both genders: $E[s_f^*|z] = E[s_m^*|z]$. The

MPH and MBH thus differ in the interpretation of the source of the dynamics in the gender mean differences. They are, in the way they are modeled in our framework, algebraically very similar. It can be shown that if academic aptitude is normally distributed, they are actually identical within the intersection of the domains of the two distributions (See Appendix A5). As a result, the closer the CCDFs of z are to linear, the more similar the two hypotheses will be in our framework.

7 Testing Predictions of the Hypotheses

7.1 Tail Hypothesis Model Estimates against PISA Test Score Distribution Parameters

To assess the validity of the tail hypothesis, we predict the female-to-male ratio in tertiary enrollment with the mean and coefficient of variation parameter values obtained from our fit with the Barro–Lee data (Section 5). We predict the ratio for alternative values of the enrollment rate: 0.20, 0.50, 0.70 and 0.90.[16] We then conduct the same exercise by using instead the distribution parameters from the OECD PISA test scores. We then measure the correlation between the two vectors of predicted female-to-male ratios for countries that are in both the Barro–Lee and PISA datasets. In total, 56 countries are common to the two sources.

The intuition behind this exercise is to test whether the model under the tail hypothesis carries information on country-specific gender differences in test score distributions, where the PISA data are used as a benchmark to assess the tail hypothesis estimates. A strong positive correlation

[16] We use the coefficient of variation instead of variance to normalize the distributions.

Table IV: Correlations between the Gender Ratio $\hat{y}$ from the Model Estimates under the Tail Hypothesis and PISA Estimates, for Alternative Values of the Enrollment Rate x

	Predicted $\hat{y}$ - Model estimates			
	$x = 0.2$	$x = 0.5$	$x = 0.7$	$x = 0.9$
Predicted $\hat{y}$ - PISA Math	0.029	0.032	0.064	0.178
	(0.834)	(0.817)	(0.638)	(0.189)
Predicted $\hat{y}$ - PISA Reading	0.311**	0.349***	0.398***	0.304**
	(0.02)	(0.008)	(0.002)	(0.023)

Sources. PISA 2012 and Barro-Lee database 2013.
Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Sample size is 56 countries. p-values are reported into parentheses.

suggests that the dynamics of participation in education by gender carries information on gender differences in test scores by country, providing supportive evidence for the tail dynamics hypothesis.

Among available datasets, the PISA dataset provides a suitable benchmark to evaluate our estimates. The PISA sample was designed to be representative of the population of 15-year-olds in a given country. It surveys individuals in schools before the end of compulsory education in most countries, which is important in our setting to obtain estimates of the entire population of males and females in a given age group. In addition, PISA test score estimates of the population of 15-year-olds are comparable across a large sample of countries.

Table IV reports the correlation between the gender ratio predicted from the two sets of parameter estimates, for different values of the enrollment rate. Correlations between the parameter values predicted by the tail hypothesis and PISA reading test are positive and statistically significant. The correlation is approximately 0.30 and robust to alternative values of the enrollment rate. However, there is no significant correlation with the math scores. The model under the tail hypothesis appears to capture underlying parameters of academic aptitude distributions.

7.2 Dynamics of Mean Ability among the Enrolled

We test the implications of the tail and mean dynamics hypotheses by looking at the evolution of the mean ability of the enrolled by gender over time. As illustrated in Figure V, the tail and mean dynamics hypotheses generate different predictions about the relationship between the economy-wide enrollment rate for a given level of schooling, defined as:

$$P^{MBH}(\bar{z}_f, \bar{z}_m) = \frac{G_{z_f}(\bar{z}_f) + G_{z_m}(\bar{z}_m)}{2},$$

and the difference in mean academic aptitude among females and males who are enrolled:

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

This originates from the fact that the tail hypothesis posits that z_f and z_m have different distributions with $\bar{z}_f = \bar{z}_m$, and *vice versa* for the MBH. The MBH predicts that the average academic aptitude is initially higher for enrolled females than for enrolled males, and progressively converges towards it before taking lower values. According to the MBH, females are more heavily selected or self-selected to attend higher levels of schooling prior to the reversal, conditional on aptitude. In contrast, the tail hypothesis assumes that females are underrepresented in the tails of the test score distribution, but overrepresented around the mean. This implies that females' average gets higher relative to males as the enrollment rate at higher levels of schooling increases.[17]

To assess the validity of these predictions, we use data from several nationally representative

[17]When individuals enrolling at a given level of schooling are a representative sample of the full population, the observed mean difference becomes an estimate of the mean difference of the population.

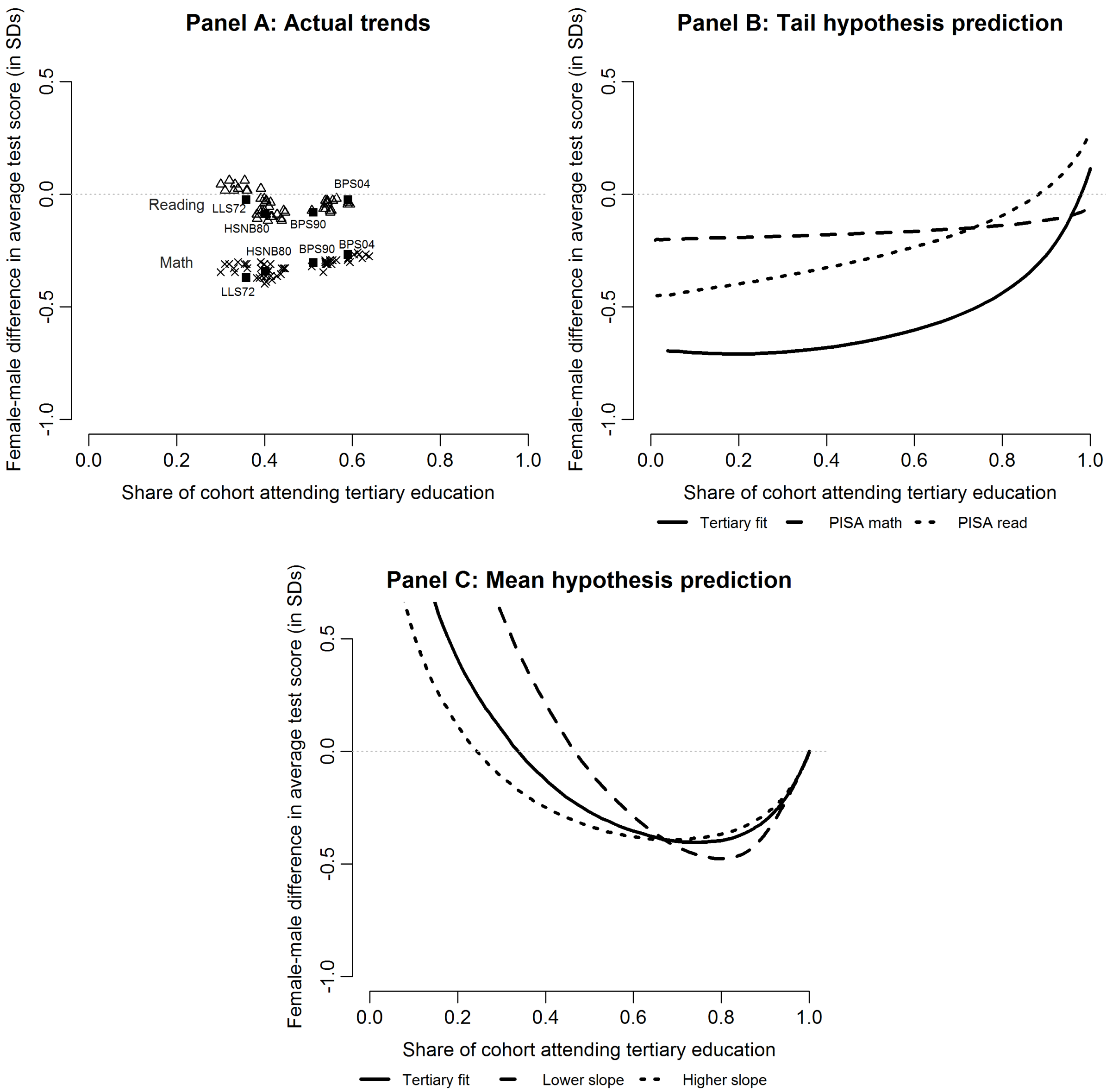
longitudinal surveys of students in the US conducted in 1972, 1980, 1990 and 2004.[18] These surveys contain information on test scores of students while they were in secondary school that can be linked to post-secondary education attendance a few years later. This allows to estimate $E[z_f|z_f > \bar{z}_f] - E[z_m|z_m > \bar{z}_m]$, the gender gap in average test-score at age 15 of post-secondary students, for different values of the enrollment rate.

[18]Those include the National Longitudinal Study 1972, the High School and Beyond 1980 and the Beginning Postsecondary Students Longitudinal Study (BPS) of 1990 and 2004.

We complement this information with yearly time-series data on mean test scores by gender from the Scholastic Achievement Test (SAT) from 1967 to 2010 in the US. Although these data are for the population of students taking the SAT test, and not only for individuals that would attend post-secondary education, taking the SAT is highly correlated with post-secondary education attendance. In addition, changes in gender-specific selection into SAT test-taking over time, as for enrollment into post-secondary education, would reflect potential gender-specific determinants of b and $\bar{z}$, which we want to test for. As illustrated in Figure V, longitudinal survey data and SAT time series report very similar trends on the evolution the the gender gap in mean test scores among the test-takers or among the enrolled in post-secondary education.

The predictions of the two hypotheses with alternative parameter values are reported in Figure V. Overall, the tail hypothesis appears to perform slightly better at predicting the relationship between the gender gap in test scores and the fraction of the population taking the test or enrolling into tertiary education. It predicts a slightly upward trend which is consistent with actual trends, except for reading test scores at low levels of enrollment. This portion shows a downward trend which is more consistent with the predictions of the mean dynamics hypothesis, before the trend reverts at higher levels of enrollment. Although evidence is not clear cut, the predictions of the tail hypothesis appear slightly more in line with data trends. Overall, our explanatory work conducted throughout the paper lends stronger support for the tail hypothesis, as summarized in Table V.

Figure V: Relationship between Participation to Tertiary Education and Gender Gap in Mean Test Score among Participants: Tail vs Mean Hypothesis (MBH)



Notes. **Panel A:** The cross and triangle dots report the gender difference in SAT mean scores for each cohort of SAT takers, as a function of the share of individuals in the cohort enrolling in tertiary education, from year 1967 to 2016. 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. **Panel B:** Predictions of the tail hypothesis were computed using parameter values from the least squares fit using the Barro-Lee data. The PISA math and read curves use parameter values from the PISA exam scores. **Panel C:** The tertiary fit depicts the prediction of the mean hypothesis given the time trend of the female aptitude threshold $\bar{z}_f$, estimated from time-series of the enrollment rate and the gender ratio in the US. Lower slope curve and higher slope curve depict the prediction if the time trend of $\bar{z}_f$ had developed at a 10% slower or 10% higher pace that the actual estimate, respectively.
Sources. College Board for SAT score data. National Longitudinal Study 1972, the High School and Beyond 1980 and the Beginning Postsecondary Students Longitudinal Study (BPS) of 1990 and 2004 for mean achievement test scores at age 15 of those enrolling in tertiary education. The proportion of individuals in the cohort attending post-secondary education is from the US Census.

Table V: The Validity of the Gender Gap Reversal Hypotheses

	Hypothesis prediction		
	Tail hypothesis	Mean hypotheses	
		MBH	MPH
Gender gap reversal	✓	✓	✓
Predictions			
PISA test score distributions	✓(reading, not math)	X	X
Mean test score dynamics of the enrolled in USA	✓(except early reading scores)	✓(early reading scores)	-
Assumptions			
Higher male variance in academic aptitude	✓	X	X
Non-increase in relative female net benefits to education in USA	✓	X	✓
Non-increase in relative female performance internationally	✓	✓	X

Notes. ✓indicates consistency of the prediction of assumption with what is observed in the data. X indicates inconsistency with what is observed in the data.

8 Conclusion

The origins of the gender gap reversal in education are not fully understood. This paper contributes to a better understanding of the forces behind this reversal, by modeling alternative hypotheses using a unifying theoretical framework. Importantly, the paper introduced a hypothesis for the reversal that has not been explored in the literature. Formulating the tail hypothesis in our framework allowed to reconcile two facts observed internationally: the larger variance of men's performance in achievement test scores found in virtually all OECD countries, and the gender gap reversal in educational attainment observed in virtually all high-income countries as well as in low and middle-income countries.

We also formalized two types of mean hypotheses, suggested by previous literature. The mean benefits hypothesis explained the gender gap reversal by allowing the female net benefits of education to first catch up and subsequently overtake those of males'. The mean performance hypothesis, in turn, posited that a similar dynamic in mean aptitude by gender can account for the reversal. We showed that these two mean hypotheses are very similar in nature.

Since the tail and mean hypotheses can both explain the reversal, we further assessed the underlying assumptions and predictions of the two hypotheses against the data. We found that the assumptions of the tail dynamics hypothesis are well supported, as gender differences in test score distribution by country estimated from our model correlate with estimates from the Project for International Student Assessment (PISA). Such correlations cannot be explained by previous hypotheses on the gender gap reversal. In contrast, we found limited support for the underlying assumptions and predictions of the mean dynamics hypotheses.

Those findings indicate that the lower variance in test score achievement among females was a driver of the observed gender gap reversal in education. The larger variability of men's achievement test scores observed empirically remains mostly unexplained, and could be explored by further research. Our findings outline that the over-representation of boys at the bottom end of school performance should be of concern among policy makers. Delving into the origins of the over-representation of boys at the bottom of the achievement distribution could help addressing boys' growing educational disadvantaged. This is important as early school dropouts have been linked to poor labor market performance, higher poverty incidence but also higher crime rates, particularly among males.

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. Fundamentally, when analyzing educational outcomes by gender, the researcher is always looking at truncated distributions. In such distributions, the mean is a function of the dispersion of the underlying distribution. Since evidence shows a higher dispersion of test scores among males, this effect needs to be accounted for, whenever educational achievement is discussed by gender. Our findings also suggest that the larger variance of traits among males may be relevant to explain gender differences in other areas. Building on this fact could be a direction for future research in labor economics or other fields of economics.

References

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

Autor, D. and M. Wasserman (2013). Wayward sons: The emerging gender gap in education and labor markets. Washington DC: Third Way.

Barro, R. and J. Lee (2016). Human capital on the long run. *Journal of Development Economics 122*, 147–169.

Barro, R. J. and J. Lee (2013). A new data set of educational attainment in the world, 1950Ð2010. *Journal of Development Economics 104*, 184–198.

Becker, G., W. Hubbard, and K. M. Murphy (2010a). Explaining the worldwide boom in higher education of women. *Journal of Human Capital 4*, 203–241.

Becker, G., W. Hubbard, and K. M. Murphy (2010b). The market for college graduates and the worldwide boom in higher education of women. *American Economic Review 100*, 229–233.

Card, D. (1994). Earnings, schooling, and ability revisited. NBER Working Paper.

Card, D. and T. Lemieux (2000). Can falling supply explain the rising return to college for younger men? a cohort-based analysis. NBER Working Paper.

Charles, K. and M. C. Luoh (2003). Gender differences in completed schooling. *Review of Economics and Statistics 85*(4).

Chiappori, P., M. Iyigun, and Y. Weiss (2009). Investment in schooling and the marriage market. *American Economic Review 99*(5).

Cho, D. (2007). The role of high school performance in explaining womenÍs rising college enrollment. *Economics of Education Review 26*.

Fortin, N., P. Oreopoulos, and S. Phipps (2015). Leaving boys behind: Gender disparities in high academic achievement. *Journal of Human Resources 50*(3), 549–579.

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

Goldin, C. and L. Katz (2002). The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of Political Economy 110*(4), 730–770.

Goldin, C. and L. Katz (2009). The race between education and technology: The evolution of us educational wage differentials, 1890 to 2005. NBER Working Paper.

Goldin, C., L. Katz, and I. Kuziemko (2006). The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic Perspectives 20*(4), 133–156.

Heckman, J. and T. Kautz (2012). Hard evidence on soft skills. *Labor Economics 19*(4), 451–464.

Hedges, L. 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. (2011). The phantom gender difference in the college wage premium. *Journal of Human Resources 46*(3), 568–586.

Jacob, B. 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.

Machin, S. and T. Pekkarinen (2008). Global sex differences in test score variability. *Science 322*, 1331–1332.

Pekkarinen, T. (2012). Gender differences in education. IZA Discussion Paper Series, 6390.

Trostel, P., I. Walker, and P. Woolley (2002). Estimates of the economic return to schooling for 28 countries. *Labour Economics 9*, 1–16.

Appendix

A1 Testing the Model's hypotheses

A1.1 Increase in b over time

The exogenous change in b over time is a key assumption of our model. To test this assumption empirically, we estimate a binary version of Equation 1, where the dependent variable is tertiary education attendance, denoted H. H is observable and is a function of the latent variable s^* in the model:

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

We estimate b, the net benefits to tertiary education in our model, from the linear probability model:

$$H_j = \beta z_j + \epsilon_j,$$

where β denotes the estimate of b. According to the Barro-Lee data, the gender gap reversal in participation to tertiary education occurred in the mid 1980s in the US. The data used for the estimation is from another longitudinal surveys of 10th graders, the High school and Beyond (HS&B), which started in 1980. As for the ELS 2002, this study collects 10th grade test scores that can be linked to tertiary education attendance a few years later, allowing to estimate changes in the value of b over the 20-year period in which the reversal occurred.

Table A1 reports our estimates for b in the US in 1980 and 2002 using Equation (1) from our

Table A1: Model Estimates of the Relationship between Academic Aptitude and Participation to Tertiary Education in 1980 and 2002, using US Data

	Dep. Variable: Tertiary Education Attendance	
	1980	2002
$\hat{\beta}$	0.278***	0.450***
	(0.008)	(0.005)
Number of Observations	12,493	12,538

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Standard errors are reported in parentheses. The regressions are linear probability models run by ordinary least squares. Regressions do not include an intercept. Test score are standardized to have a mean of 0 and a standard deviation of 1 in the full sample.
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).

framework. It shows that the estimated net benefits of tertiary education increased sharply from 1980 to 2002, providing empirical support to the second building block of our model. The null hypothesis of equality between b_{1980} and b_{2002} is strongly rejected by statistical tests. This indicates that individuals with the same level of academic aptitude are more likely to enroll in higher education in 2002 than they were in 1980, which is equivalent to higher net benefits of education in 2002 compared to 1980 in our framework.

This evidence is consistent with a large body of literature showing that returns to education, in particular returns to tertiary education, have increased over the past decades. Goldin and Katz (2009) or Acemoglu and Autor (1998) among others provide consistent evidence of a sharp increase of the college wage premium in the US since the beginning of the 1970s.[19] Card and Lemieux (2000) also report an important increase in the wage premium of university graduates relative to high school

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

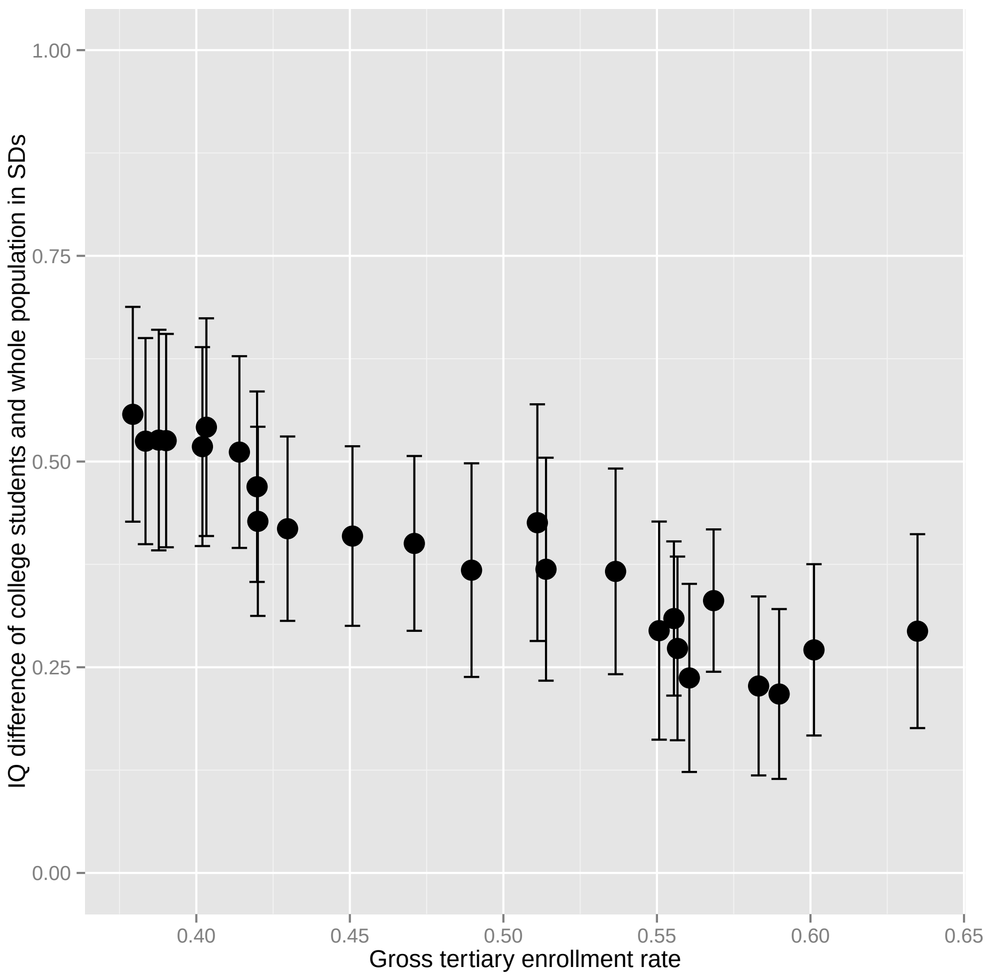
graduates in the UK and Canada over the same period.

A1.2 The Negative Association between the Enrollment Rate and Mean Ability of Individuals Enrolled

Given the positive association between individual ability z and optimal level of schooling s^* chosen by individuals, a rise in b leads to both an increase in enrollment, and a decrease in $\bar{z}$, the ability threshold for attending a given level of schooling $\bar{s}$. We do not observe $\bar{z}$ empirically, but one implication of a decrease in $\bar{z}$ with no accompanying increase in the mean of the distribution is a decrease in the mean ability of individuals enrolled, which can be tested. To do so, we estimate the average verbal test score of individuals attending tertiary education in the US over the period 1975–2010, using data from the General Social Survey (GSS). From 1974 onwards, the GSS includes a short 10-item multiple choice test assessing vocabulary knowledge of respondents. A measure of educational attainment in years is also reported, allowing to identify individuals that attended post-secondary education.

Figure A1 depicts the evolution of the average verbal score of post-secondary students relative to the entire population, from 1975 to 2010. In 1975, the average cognitive score of students attending post-secondary education was 0.60 standard deviation higher than the average cognitive score of the whole population. This relative difference decreased progressively until 2005 to reach approximately 0.30 standard deviations in 2010. Consistent with the assumptions of the model, this suggests that greater access to post-secondary education, an increase in P in our framework, was accompanied by a decrease in the average ability of individuals enrolling in tertiary education.

Figure A1: Empirical Relationship between Enrollment Rate in Tertiary Education and Average Verbal Score of Participants in Tertiary Education: 1975-2010



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

A2 Validity of the Mean Performance Hypothesis Assumptions: International Evidence

We also test the validity of the mean performance hypothesis assumptions in the UK, using two longitudinal surveys with base years in 1958 and 2000. The National Child Development Study

Table A2: Gender Gap in Average Achievement Test Score Prior and Post-reversal, UK Estimates

Panel B: UK, Reading test scores, age 11								
	Year 1969				Year 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	-	-

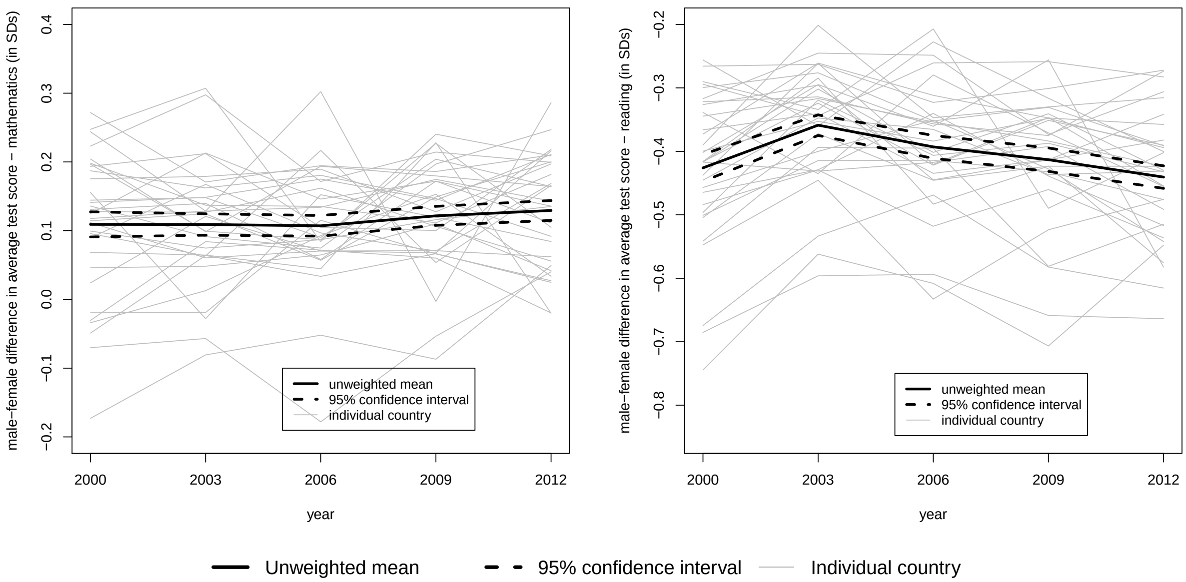
Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Standard errors are reported in parentheses. Test scores were standardized to have a mean of 0 and standard deviation of 1 in the full sample.
Sources. National Child Development Study (2nd follow-up, 1969) and Millennium Cohort Study (2nd follow-up, 2010).

(NCDS) was initiated in 1958 and follows a cohort of individuals born in the same week of 1958 over their lifetime. The third follow-up of the study in 1969 conducted when individual were 11 years old includes test score results in verbal and mathematics tests. In a similar way, the Millennium Cohort Study (MCS) initiated in 2000 follows a cohort of children born at the beginning of the 21st century, with a follow-up in 2010 at age 10 that reports individual scores in verbal achievement tests. Table A2 reports that the gender gap in mean verbal achievement at age 11 is very similar in 1969 and 2010, more than 40 years apart. While males were already lagging behind females in verbal achievement at age 10 in 1969, we still observe a similar gap in 2010. Evidence from the UK therefore does not seem to support the assumptions of the MPH.

For international evidence, the Project for International Student Assessment (PISA) surveys a representative sample of the 15-year-old population in more than 40 countries and test results have been designed to be comparable over time. Unfortunately, the data is only available from 2000 onwards, and therefore allows tracking relative changes in mean performance between genders only over the period 2000–2015. Figure A2 depicts the evolution of girls' mean test score relative to boys in reading and mathematics over the period 2000–2015, 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 2000–2015, females appear to do worse in mathematics relative to males in 2015 compared to 2000. Recent international evidence is therefore mostly inconclusive regarding the increase of female mean performance, but the short time span covered is a limitation.

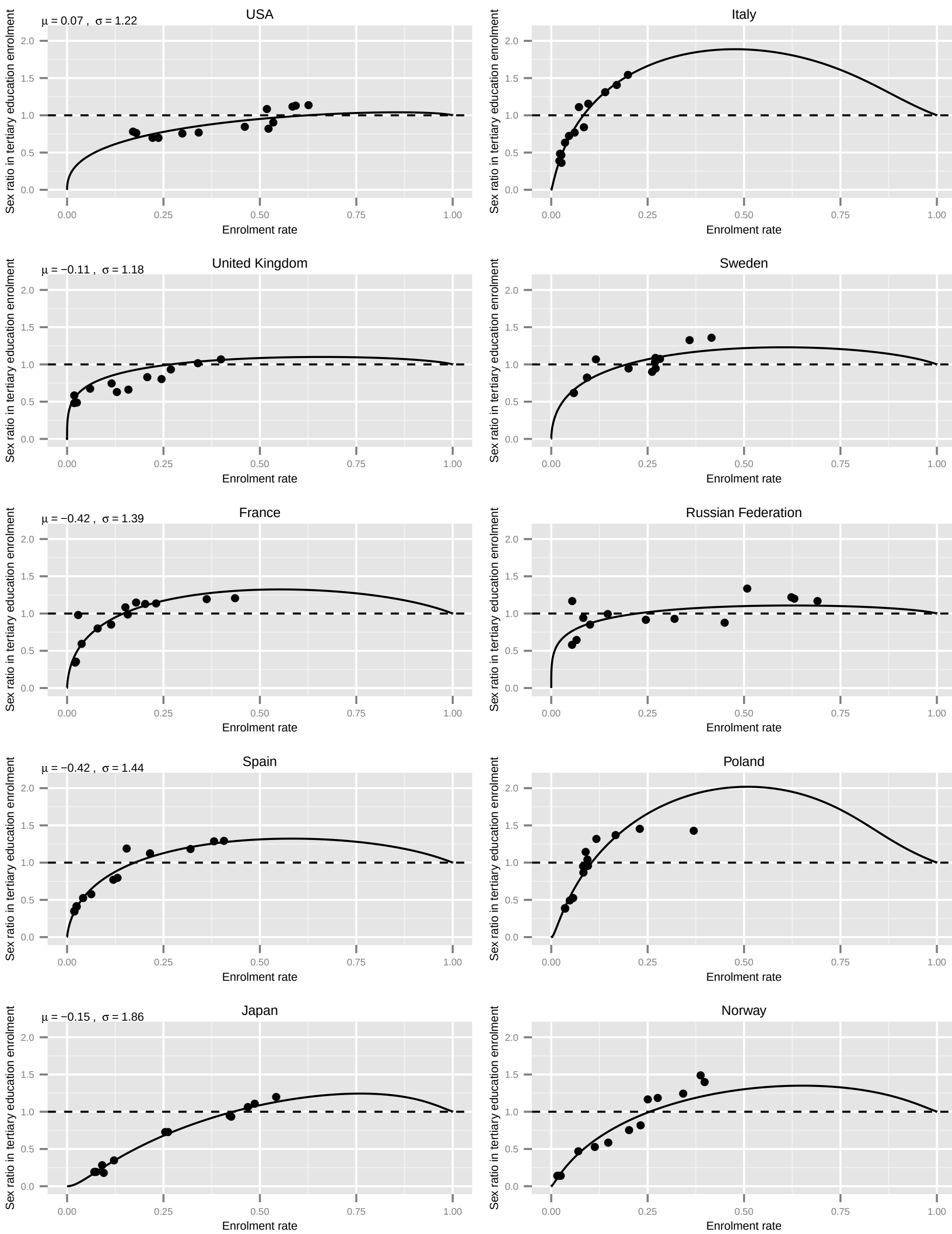
Figure A2: Relative Performance of Males and Females in Achievement Tests at Age 15, Over Time



Sources. PISA 2000, 2003, 2006, 2009, 2012 and 2015.
Notes. The sample includes 31 countries that are included in all PISA waves.

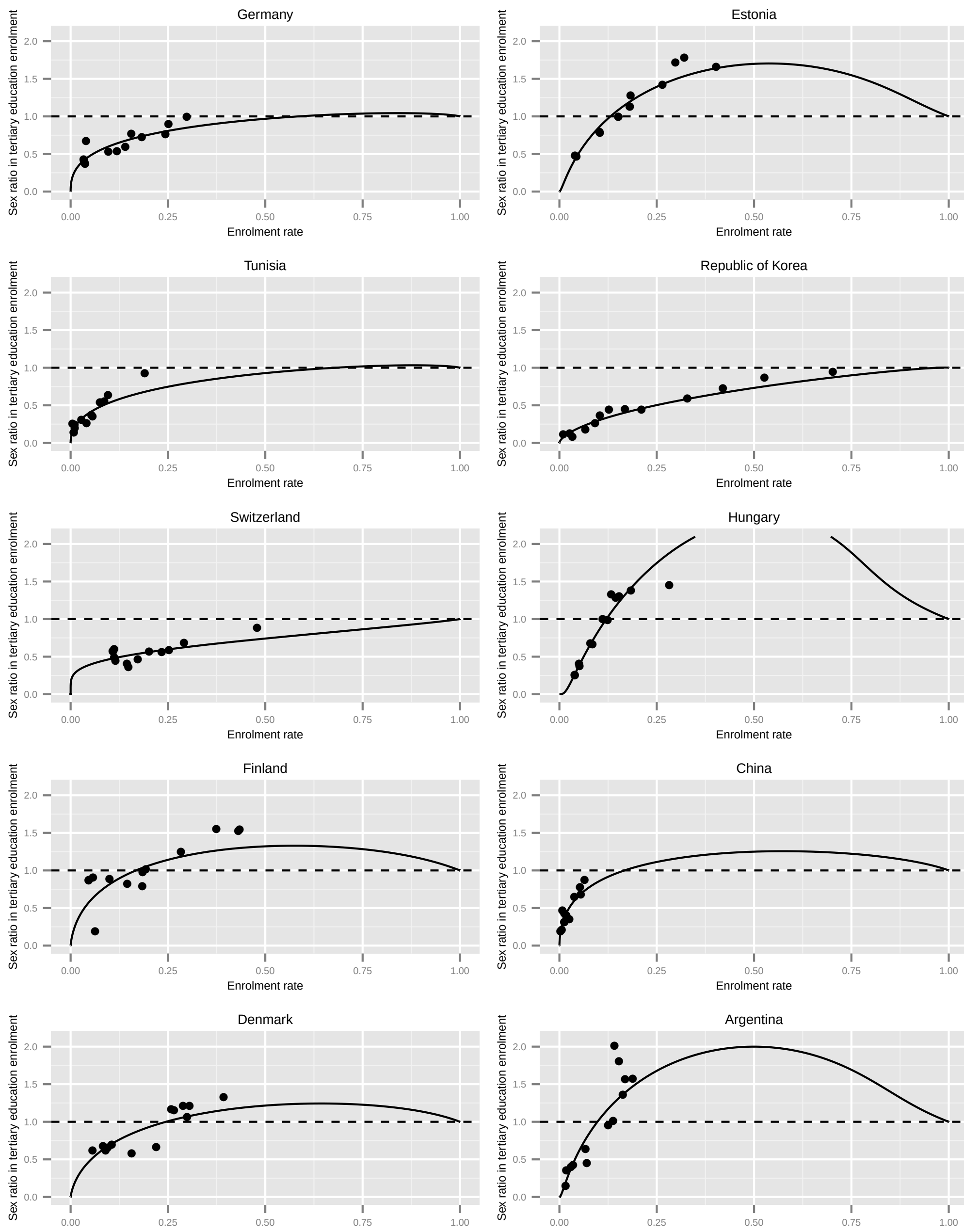
A3 Additional Tables and Figures

Figure A3: Model Fit under The Tail Hypothesis: Female-to-male Ratio among Participants to Tertiary Education



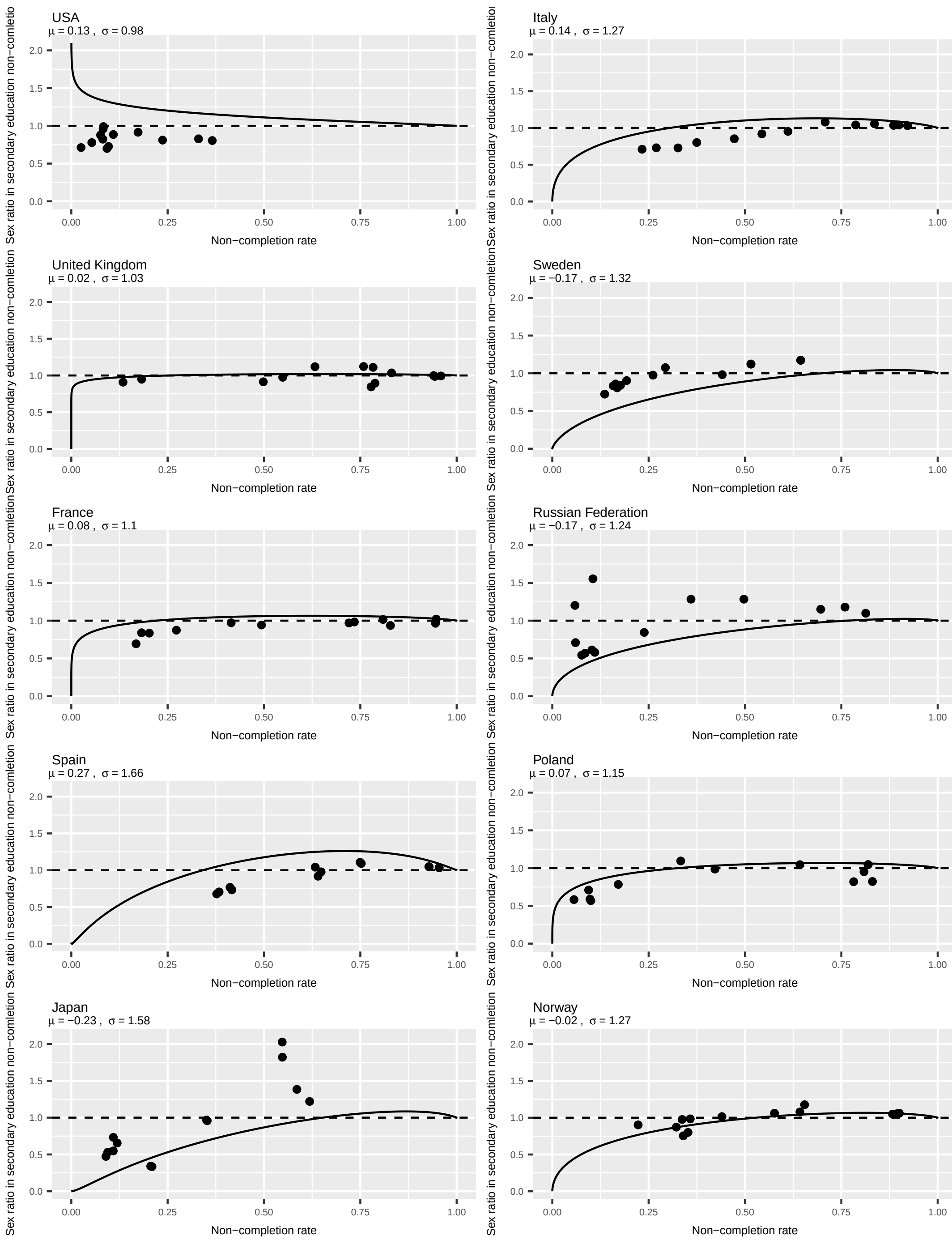
Notes. The x-axis measures the enrollment rate in tertiary education for country i. The y-axis measures the female-to-male ratio among individuals enrolled for country i. The full line depicts the estimated relationship between y and x from our model when the ability distribution parameters $\{\mu_i; \sigma_i^2\}$ are estimated by maximum likelihood.

Figure A4: Model Fit under The Tail Hypothesis: Female-to-male Ratio among Participants to Tertiary Education



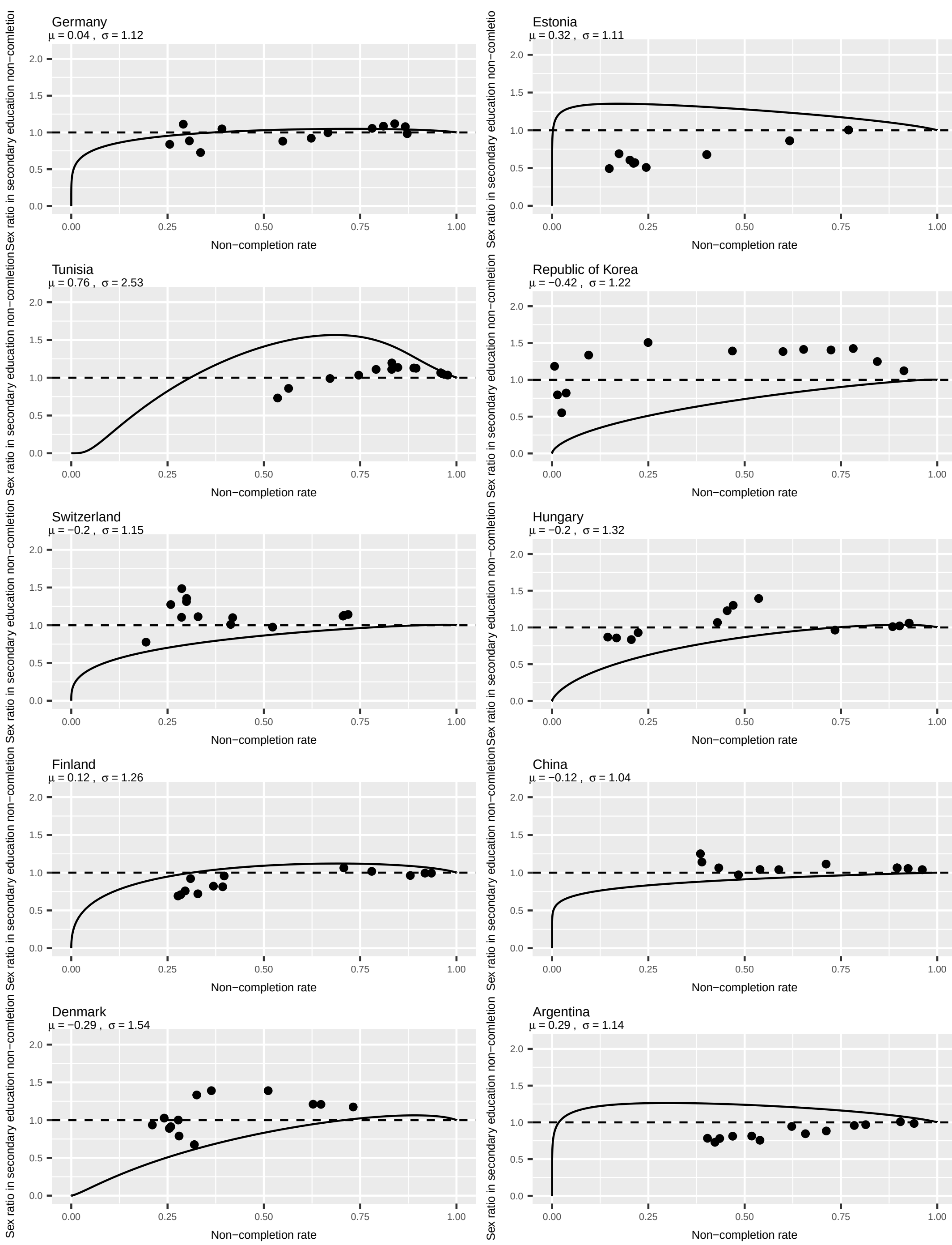
Notes. The x-axis measures the enrollment rate in tertiary education for country i. The y-axis measures the female-to-male ratio among individuals enrolled for country i. The full line depicts the estimated relationship between y and x from our model when the ability distribution parameters $\{\mu_i; \sigma_i^2\}$ are estimated by maximum likelihood.

Figure A5: Model Fit under the Tail Hypothesis: Gender Ratio among Secondary Non-completers



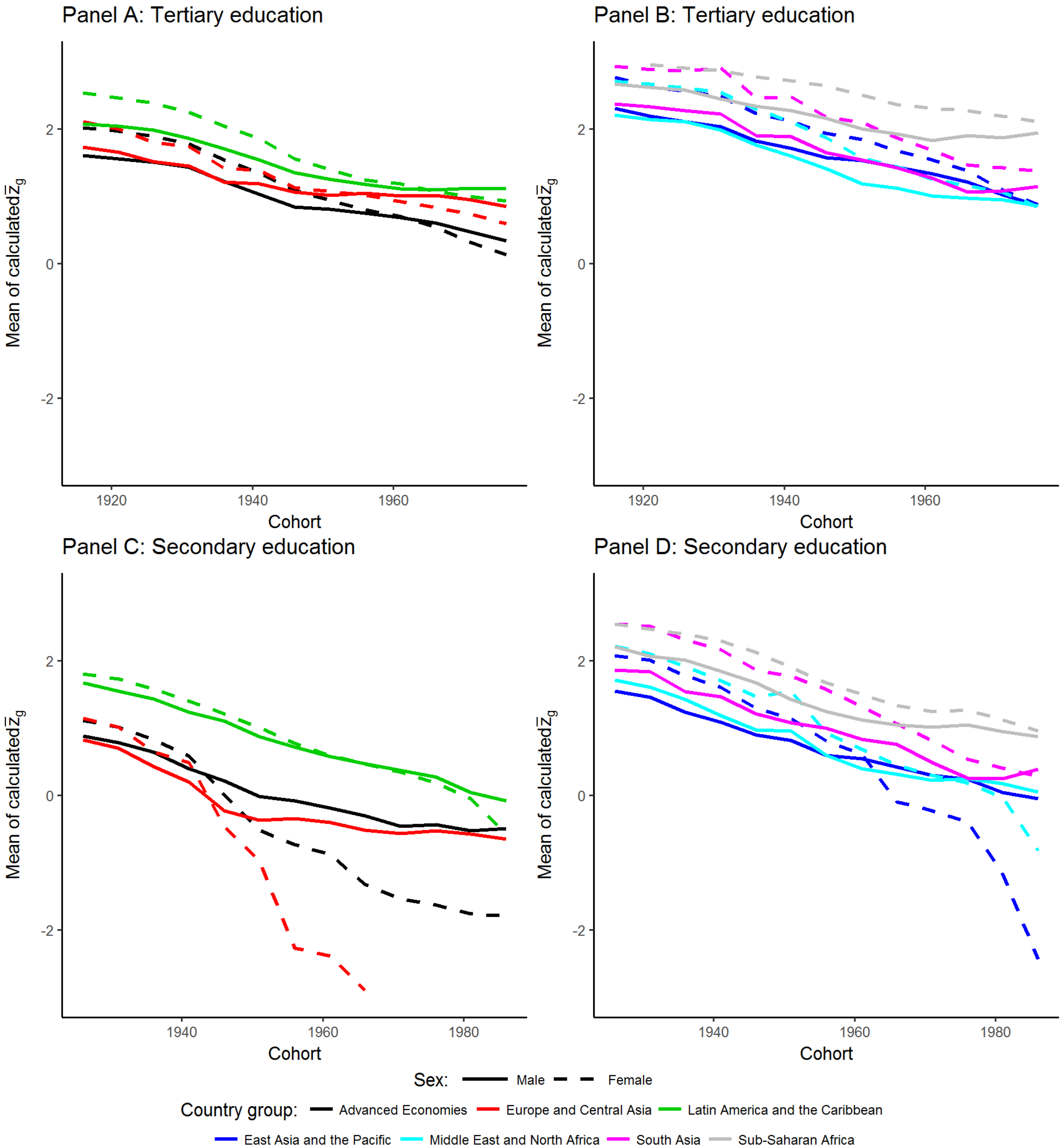
Notes. The x-axis measures the secondary school non-completion rate for country i. The y-axis measures the female-to-male ratio among secondary school non-completers for country i. The full line depicts the estimated relationship between y and x from our model when the ability distribution parameters $\{\mu_i; \sigma_i^2\}$ are estimated by maximum likelihood.

Figure A6: Model Fit under the Tail Hypothesis: Gender Ratio among Secondary Non-completers



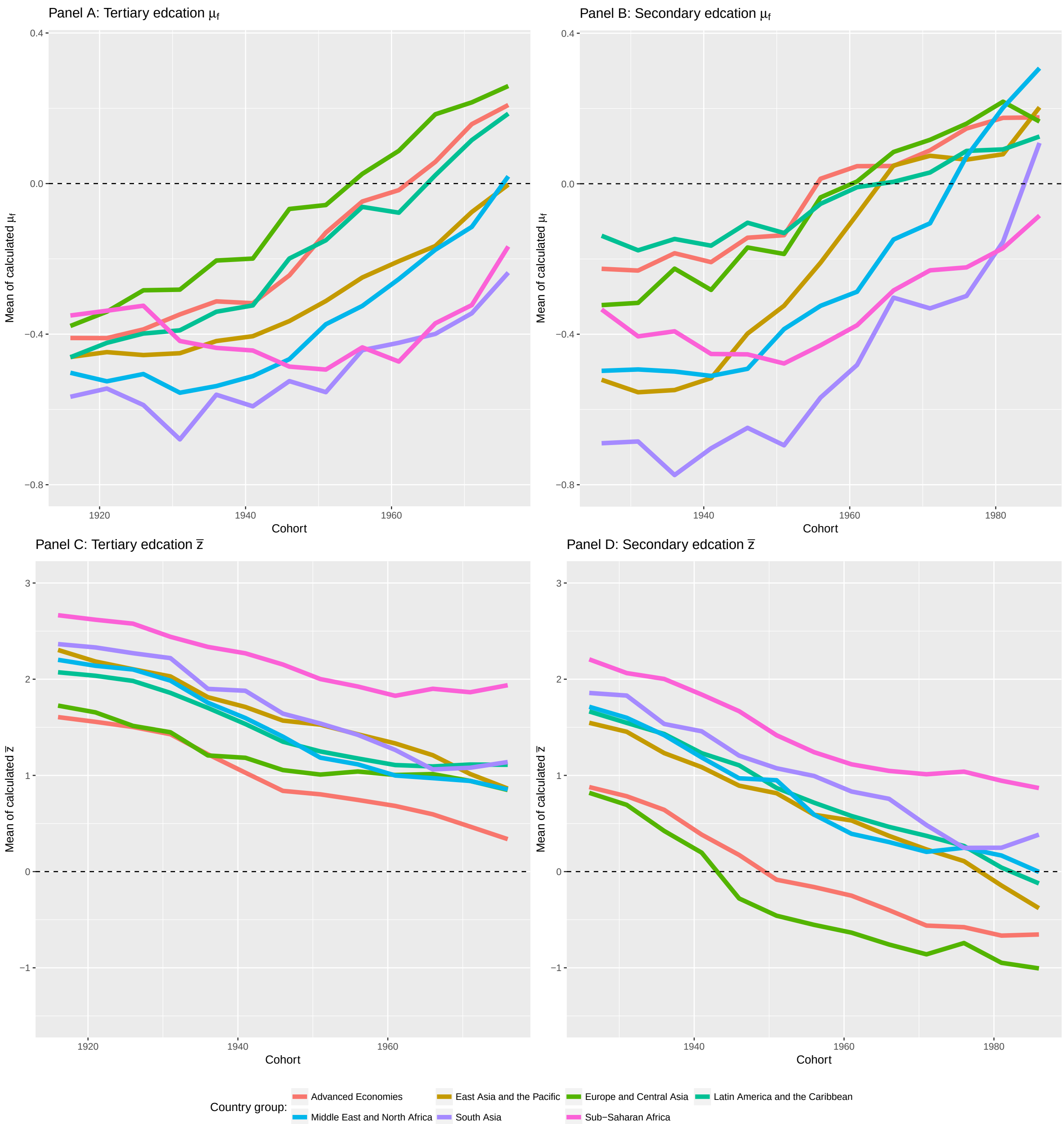
Notes. The x-axis measures the secondary school non-completion rate for country i. The y-axis measures the female-to-male ratio among secondary school non-completers for country i. The full line depicts the estimated relationship between y and x from our model when the ability distribution parameters $\{\mu_i; \sigma_i^2\}$ are estimated by maximum likelihood.

Figure A7: Estimated MBH Parameters, by Birth Cohort and Country Group



Notes. Each line represents the unweighted country group mean of the estimated model parameter value under the MBH. **Panels A and B** show the mean values of calculated $\bar{z}_f$ and $\bar{z}_m$ under the MBH, estimated from the fraction of individuals that attended tertiary education and the gender ratio among those who attended, for thirteen 5-year-band cohorts classified by country group. The total sample size is 115 countries. **Panels C and D** show the mean values of calculated $\bar{z}_f$ and $\bar{z}_m$ under the MBH, estimated from the fraction of individuals that did not complete secondary education and the gender ratio among non-completers, for thirteen 5-year-band cohorts. The total sample size is 135 countries.

Figure A8: Estimated MPH Parameter Values, by Birth Cohort and Country Group



Notes. Each line represents the unweighted country group mean of the estimated model parameter value under the MPH. **Panel A.** The graph shows the female mean $\hat{\mu}_f$, estimated from the gender ratio and enrollment rate in tertiary education for thirteen 5-year-band cohorts from the Barro-Lee dataset. Total sample size is 115 countries. **Panel B.** The graph shows the values of calculated $\bar{z}$, the ability threshold common to males and females. The total sample size is 115 countries. **Panel C.** The graph shows the values of $\hat{\mu}_f$, estimated from the gender ratio among non-completers and secondary school non-completion rates for thirteen 5-year-band cohorts from the Barro-Lee dataset. The total sample size is 135 countries. **Panel D.** The graph shows the values of $\bar{z}$, estimated from the same time-series as Panel C. Total sample size is 135 countries.

A4 Tail Hypothesis: Mathematical Proofs of Propositions 1 to 3

A4.1 Normal Distributions

Let $f_{zf}(z)$ and $f_{zm}(z)$ denote the probability distribution functions of test-taking ability 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 $\sigma_m^2 > \sigma_f^2$, according to the tail dynamics hypothesis.

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

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

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

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

Taking the integrals, one can express the ratio as:

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

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

Using the analytical expression of $R^{TH}(\bar{z})$, 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 $\lim_{\bar{z}\to\infty} erf(\bar{z}) = 1$.

Using the l'Hôpital rule, we take the derivative of the denominator and the numerator to get the following expression:

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

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

$$= \lim_{\bar{z}\to\infty} \frac{\sigma_m^2}{\sigma_f^2} exp\left\{\frac{\bar{z}}{\sigma_m^2\sigma_f^2}\left[\bar{z}\{\sigma_f^2 - \sigma_m^2\} - 2\mu_m\sigma_f^2 + 2\mu_f\sigma_m^2 + \frac{\mu_m^2\sigma_f^2}{\bar{z}} - \frac{\mu_f^2\sigma_m^2}{\bar{z}}\right]\right\} = 0,$$

since, by assumption under the tail dynamics hypothesis, $\sigma_m^2 > \sigma_f^2$, which are both positive by definition.

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

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

Let us now study the behavior of $R^{TH}(\bar{z})$ 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 $P^{TH}(\bar{z})$ such that $R^{TH}(\bar{z}) = 1$. 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:

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

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

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

This equation has a unique solution given $\sigma_m^2 > \sigma_f^2$. 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:

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

In addition, $\bar{z}^*$ is unique given the vector of exogenous parameters $\{\mu_f, \mu_m, \sigma_f^2, \sigma_m^2\}$.

A5 Proof that the MPH is equivalent to the MBH for uniform distribution within the intersection of their domains

MBH: For each time period t, let $b \equiv a + c$ and $z \sim unif(a, b) = unif(a, a + c)$. Then, $G_z(\bar{z}) = 1 - \frac{\bar{z}-a}{b-a} = 1 - \frac{\bar{z}-a}{a+c-a} = 1 - \frac{\bar{z}-a}{c}$. Also, let (*): $\bar{z}_m \equiv \bar{z}$ and (**): $\bar{z}_f \equiv \bar{z} + a - a_f$. Now, for $\bar{z}_f, \bar{z}_m \in \{a, a + c\} \Leftrightarrow \bar{z} + a - a_f, \bar{z} \in \{a, a + c\}$,

$$E^{MBH}(\bar{z}_f, \bar{z}_m) = \frac{G_z(\bar{z}_m) + G_z(\bar{z}_f)}{2} = \frac{1}{2}\left(1 - \frac{\bar{z}_m - a}{(a + c) - a} + 1 - \frac{\bar{z}_f - a}{(a + c) - a}\right) = \frac{1}{2}\left(2 - \frac{\bar{z}_m + \bar{z}_f - 2a}{2c}\right)$$

$$= 1 - \frac{\bar{z}_m + \bar{z}_f - 2a}{4c} \overset{(*,**)}{=} 1 - \frac{2\bar{z} - a - a_f}{4c}\; and$$

$$R^{MBH}(\bar{z}_f, \bar{z}_m) \equiv \frac{G_z(\bar{z}_f)}{G_z(\bar{z}_m)} = 1 - \frac{\bar{z}_f - a}{(a + c) - a}/1 - \frac{\bar{z}_m - a}{(a + c) - a} = \frac{c}{c} - \frac{\bar{z}_f - a}{c}/\frac{c}{c} - \frac{\bar{z}_m - a}{c}$$

$$= \frac{c - \bar{z}_f + a}{c - \bar{z}_m + a} \overset{(*,**)}{=} \frac{c - \bar{z} + a_f}{c - \bar{z} + a}$$

MPH: For $\bar{z} + a - a_f, \bar{z} \in \{a, a + c\}$,

$$E^{MPH}(\mu_f, \bar{z}) = \frac{G_{z_m}(\bar{z}) + G_{z_{f,t}}(\mu_f, \bar{z})}{2} = \frac{1}{2}\left(1 - \frac{\bar{z} - a}{(a + c) - a} + 1 - \frac{\bar{z} - a_f}{(a_f + c) - a_f}\right)$$

$$= \frac{1}{2}\left(2 - \frac{2\bar{z} - a - a_f}{2c}\right) = 1 - \frac{2\bar{z} - a - a_f}{4c} = E^{MBH}(\bar{z}_f, \bar{z}_m)\, and$$

$$R^{MPH}(\mu_f, \bar{z}) \equiv \frac{G_{z_{f,t}}(\mu_f, \bar{z})}{G_{z_m}(\bar{z})} = 1 - \frac{\bar{z} - a_f}{(a_f + c) - a_f} / 1 - \frac{\bar{z} - a}{(a + c) - a} = \frac{c}{c} - \frac{\bar{z} - a_f}{c} / \frac{c}{c} - \frac{\bar{z} - a}{c}$$

$$= \frac{c - \bar{z} + a_f}{c - \bar{z} + a} = R^{MBH}(\bar{z}_f, \bar{z}_m)$$

Thus, the MBH and MPH are equivalent when test-taking ability is uniformly distributed and when $\bar{z} + a - a_f, \bar{z} \in \{a, a + c\}$. The difference between the two hypotheses comes from non-linearities in $G_z(\cdot)$.