I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11260

Explaining the Gender Test Score Gap in Mathematics: The Role of Gender Inequality

Z. Eylem Gevrek
Christian Neumeier
Deniz Gevrek

JANUARY 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11260

Explaining the Gender Test Score Gap in Mathematics: The Role of Gender Inequality

Z. Eylem Gevrek
Católica Porto Business School

Christian Neumeier
University of Konstanz

Deniz Gevrek
Texas A&M University–Corpus Christi

JANUARY 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Explaining the Gender Test Score Gap in Mathematics: The Role of Gender Inequality

Using data from the 2012 PISA across 56 countries, this study examines the link between societal gender inequalities and the gender test score gap in mathematics. We employ a novel two-stage empirical strategy in which the first stage involves decomposing the gender mathematics gap into a part that is explained by gender differences in observable characteristics and a part that remains unexplained. We use a semiparametric Oaxaca-Blinder (OB) decomposition to analyze the gap in each country individually. In the second stage, we investigate whether the decomposition components of the gap are systematically related to country-level gender inequality measures. The results indicate that the gap is not statistically significantly associated with the indicators of gender inequality, but the unexplained part of the gap is. In more gender-equal countries, the unexplained part of the gap favoring boys appears smaller. Moreover, we find that the relationship between the unexplained part of the gap and the societal gender inequality varies within the test score distribution, and tends to become less pronounced at the upper end of the distribution.

JEL Classification: C14, I24, I25, J16

Keywords: gender math gap, Semiparametric Oaxaca Blinder decomposition, culture

Corresponding author:
Z. Eylem Gevrek
Catolica Porto Business School
R. de Diogo Botelho 1327
4169-005 Porto
Portugal

E-mail: egevrek@porto.ucp.pt

1 Introduction

What accounts for the gender achievement gap in mathematics? The literature offers two main explanations for the existence of the gender gap in mathematics achievement. The proponents of biological theories argue that gender differences in brain composition (Cahill, 2005), hormone levels (Davison and Susman, 2001), or spatial ability (Kucian et al., 2005) produce the gap.[1] The other body of research attributes the gender gap in mathematics achievement to a complex variety of sociocultural factors rather than biological differences. Societal or cultural explanations for the gender gap focus on two possibilities. First, boys and girls might differ in terms of their educational inputs that affect mathematics performance. Second, although girls have the same observed characteristics as boys, they might be affected by those characteristics differently (Dickerson et al., 2015). For instance, boys may benefit from family and school resources more than girls due to the stereotype that mathematics is a male domain.

Using data from the 2012 Program for International Student Assessment (PISA) across 56 countries, this study aims to examine the link between societal gender inequalities and the gender test score gap in mathematics. To this end, we employ a novel two-stage empirical strategy in which the first stage involves decomposing the gender mathematics gap into a part that is explained by gender differences in observable characteristics and a part that remains unexplained. We use a semiparametric Oaxaca-Blinder (OB) decomposition to analyze the gap in each country individually. In the second stage, we investigate whether the decomposition components of the gap are systematically related to country-level gender inequality measures.

The decomposition method employed in the first stage allows us to explore the following three questions: (i) Can the gap be explained by differences in observed characteristics across genders? If so, we can conclude that males and females with comparable characteristics are equally likely to obtain similar mathematics test scores. (ii) To what extent can differences in returns to these characteristics between females and males account for the gender gap? Girls may have a particular disadvantage with converting educational inputs into higher mathematics test scores. For example, parents may treat their sons and daughters differentially. This differential treatment could arise from stereotypical thinking such as believing that "girls just cannot do mathematics, language is for girls and mathematics is for boys". Tiedemann (2000) and Jacobs and Eccles (1992) provide evidence that parents' gender-stereotyped beliefs about their children's competence in mathematics may influence children's self-perceptions of ability in mathematics and hence their mathematics achievement. Differential teacher attention to boys and girls in the classroom might serve as a second example. It could, for instance, be the case that mathematics teachers discriminate against girls due to stereotypes about female inferiority in mathematics. The last question we explore is: (iii) Does the relationship between societal level female empowerment and the international variation in the gender gap vary across the test score distribution? For example, is this relationship more (or less) pronounced among high achievers? We apply a propensity score matching method to decompose the gender gap into an explained and an unexplained part for each country. This semiparametric decomposition does not specify any functional form assumptions on outcome equations, imposes a common support restriction, and makes it possible to account for heterogeneity across individuals by estimating the counterfactual outcome for each individual separately (Heckman et al., 1999; Imbens, 2004).[2]

The semiparametric OB decomposition results indicate that the mean gender test score gap in mathematics is statistically significant in 42 out of 56 countries. Girls do not significantly outperform boys in any country while the statistically significant gap in favor of boys varies considerably from 6.82 points in Indonesia to 30.75 points in Austria. The explained part of the gap is statistically insignificant in 34 countries, suggesting that the gap cannot be accounted for by gender differences in observed characteristics in most countries. The unexplained part of the gap exhibits a different pattern. It is statistically significant in 42 countries, providing evidence that boys and girls, exposed to the same family, school, and societal influences are affected by those factors differently. The second-stage results of our empirical analysis can be summarized as follows. Contrary to several previous studies that examine to what extent can cross-country variation in the gender gap in mathematics be predicted by indicators of societal gender equity (Baker and Jones, 1993; Riegle-Crumb, 2005; Guiso et al., 2008), we find that the gap is not statistically significantly associated with the indicators of gender inequality, but the unexplained part of the gap is. In more gender-equal countries, the unexplained part of the gap favoring boys appears smaller. This finding demonstrates the importance of decomposing the gap into an explained and an unexplained part. Moreover, we find that the relationship between the unexplained part of the gap and the societal gender inequality varies within the test score distribution, and tends to become less pronounced at the upper end of the distribution.

The remainder of this paper is organized as follows. The next section gives an overview of the relevant literature. Section 3 describes the data and variables used in the empirical analysis. Section 4 introduces the semiparametric OB decomposition employed to investigate the gender PISA test score gap in mathematics across countries. Section 5 presents results from the empirical analysis with a discussion on robustness checks and Section 6 concludes.

2 Literature review

A number of studies investigate the nature of the relationship between gender equality and the variation in the gender mathematics test score gap across countries to illuminate the role of societal factors in explaining gender differences in academic achievement. Using the Second International Mathematics Study (SIMS) microdata on 77,000 eighth-graders from 19 countries, Baker and Jones (1993) examine the link between the gender mathematics score gap and cross-country differences in gender equity. The authors argue that childhood academic performance might be shaped by anticipated future occupational and educational opportunities. A female student who does not foresee high future returns to mathematics competency may not invest in mathematics skills. Baker and Jones (1993) find that the cross-country gender mathematics test score gap is correlated with the level of gender equity in higher education and labor market.[3]

In their well-publicized study, Guiso et al. (2008) examine the gender mathematics gap using the 2003 Programme for International Student Assessment (PISA) data set on 276,165 fifteen year-olds from 40 countries. Their findings indicate that the mean gender gap in mathematics is significantly correlated with country measures of gender equality. To assess gender equality, they use the following four indicators: (i) the World Economic Forum's Gender Gap Index (GGI), which reflects economic and political opportunities, education and well-being for women; (ii) the index for cultural societal attitudes for women based on the World Values Surveys (WVSs); (iii) female economic activity rate as measured by various labor supply indicators for women aged 15 or older; and (iv) the political empowerment index computed by the World Economic Forum. Guiso et al. (2008) provide evidence that gender differences in mathematics achievement are smaller in more gender-equal countries. Employing the same empirical strategy as in Guiso et al.

(2008), Fryer and Levitt (2010), on the other hand, show that the correlation between the gender mathematics gap and the GGI does not hold across countries taking part in the 2003 Trends in International Mathematics and Science Study (TIMSS).

Using an epidemiological approach, Nollenberger et al. (2016) confirm the findings of Guiso et al. (2008).[4] Nollenberger et al. (2016) focus on second-generation immigrant children who were born and raised in their respective host countries. These second-generation immigrants share host country institutions and environment; however, they differ in terms of their ancestry. Using the PISA (2003, 2006, 2009, and 2012) data sets on 12,027 second-generation immigrant children from 45 different countries who reside in 12 host countries, Nollenberger et al. (2016) find that lower gender equality in the country of ancestry is associated with a higher gender mathematics gap for second-generation immigrant children, suggesting that culture matters for the gender mathematics test score gap.

In the developing country context, Dickerson et al. (2015) examine the determinants of the gender mathematics test score gap across 19 African countries. They find that almost half of the cross-country variation in the gap is explained by differences in women's role in society as proxied by the fertility rate in a country.

Using the National Assessment Educational Process (NAEP) data from the United States, Pope and Sydnor (2010) indicate that there is a large and statistically significant variation in gender gaps in standardized test scores for eighth-graders scoring at the top percentiles across states and census divisions. The authors find that in areas where men and women are viewed as more equal, gender disparities are smaller in both stereotypically male-dominated tests of mathematics and science and stereotypically female-dominated tests of reading.

3 Data and Descriptive Statistics

We use data from the 2012 Programme for International Student Assessment (PISA) to explore the gender test score gap in mathematics across 56 countries. PISA is a standardized international achievement test in reading, mathematics, and science for 15-year-old students, implemented on a 3-year cycle by the Organization for Economic Co-operation and Development (OECD). Employing item response theory, the PISA 2012 reports five plausible values to present each student's mathematics performance. We take the average of those five plausible values to measure the students' mathematics achievement. The test scores in mathematics are standardized to a mean of 500 points and a standard deviation of 100 points across the OECD countries. The PISA 2012 follows a stratified two-stage sample design. Schools having 15-year-old students are sampled as the first stage unit. In the second stage, students are randomly selected with equal probability within schools. In the econometric analysis, we use the final student weights that incorporate both the school weights and the within-school student weights. OECD (2013) provides detailed information on the technical characteristics of the PISA 2012. Our final sample consists of 220,333 students from 56 different countries.

The second column of Table A.1 indicates that there is large cross country variation in the mean gender test score gap in mathematics, which is defined as the average girls' score minus the average boys' score.[5] There are 14 countries in which there is no significant difference in the average mathematics test scores. In all remaining 42 countries, boys perform better than girls in mathematics. The mean gap in favor of boys varies considerably, from 30.75 points in Austria to 6.82 points in Indonesia.

<< **Table 1 here** >>

The PISA 2012 provides information on a wide range of student and family characteristics that may affect mathematics achievement. Table 1 presents summary statistics for the student-level variables used in our empirical analysis. We control for the grade level in which students are enrolled as it may capture students' cognitive development and grade repetition effects.[6] The index of self-efficacy is a measure of students' general level of belief in their academic abilities in mathematics. To assess self-efficacy in mathematics, students are asked about their level of confidence in tackling several mathematics tasks. The list of tasks is shown in Table A.2. The higher the index is, the more confident a student feels about her/his academic abilities in mathematics. To control for students' motivation with respect to mathematics, we utilize two indices. The index of enjoyment of mathematics measures how much students enjoy learning mathematics topics and acquiring new knowledge in mathematics. The index of instrumental motivation in mathematics is based on a series of questions about the usefulness of mathematics for students and their future careers. We create a new index of motivation in mathematics by adding up these two indices. The higher the index is, the more motivated a student is to do well in mathematics.

We use parents' educational attainment and occupational status to control for the socioeconomic status of the family. Parental education is classified into three categories: i) at most primary education; ii) secondary education; and iii) tertiary education. The index of the highest parental occupational status, based on the International Socio-Economic Index of Occupational Status (ISEI), corresponds to the higher ISEI score of either parent or to the only available parent's ISEI score. Higher ISEI scores are associated with occupations that have higher returns to education. The PISA 2012 also asks students to report how much they agree with the following statement: "My parents believe that mathematics is important for my career". Accounting for parents' attitudes about the importance of studying mathematics, the indicator variable 'math is important' takes the value of one if the student's level of agreement with this statement is 'strongly agree' or 'agree', and zero otherwise. We include the number of books at home and the index of home education resources which is derived from the availability of various school items such as a study room, a computer for school work, and educational software. Moreover, we control for school fixed effects to deal with possible non-random selection into different schools across genders. For example, boys would be exposed to favorable school inputs if parents sent their sons to better schools than their daughters.

We use the Gender Gap Index (GGI) taken from the Global Gender Gap Report (2012) to measure societal gender equity (Hausmann, Tyson, and Zahidi, 2012).[7] The GGI, which is composed of four sub-indices, assesses national gender equity in educational, economic, political and health domains. The GGI and its four sub-indices range from 0 to 1, with larger values indicating greater female empowerment in society. Table A.3 presents the components of each sub-index. The education sub-index includes female-to-male ratios in four areas: literacy, primary enrollment, secondary enrollment, and tertiary enrollment. The economic sub-index captures female-to-male ratios in labor force participation rate, labor market earnings, share of legislators, senior officials and managers, and share of professional and technical positions. The political sub-index is based on the ratios of women to men in minister-level positions, parliamentary seats, and the number of years in executive office (prime minister or president) for the last 50 years. The health and survival sub-index captures the sex ratio at birth and the gap between men's and women's life expectancy. The GGI is the unweighted mean of the four sub-indices.[8] Columns 4-8 of Table A.1 present the 2012 GGI and its sub-indices values for all the 56 countries. The GGI ranges between 0.601 in Turkey and 0.864 in Iceland.

Luxembourg (Iceland) holds the top spot in the economic sub-index (political sub-index) while Turkey (Qatar) is in the last place. The education and health sub-indices exhibit very small variation across countries. In addition, to account for economic conditions across countries, we use the 2012 real per capita Gross Domestic Product (GDP), taken from the Global Gender Gap Report 2012.

4 Econometric Model

The Semiparametric OB Decomposition

We use a propensity score matching method to decompose the mean gender mathematics test score gap into an explained and an unexplained part in all 56 countries. Although propensity score matching (PSM) is extensively employed in the evaluation literature to estimate average treatment effects (Rosenbaum and Rubin, 1983; Black and Smith, 2004; Frolich, 2004; Sianesi, 2004; Smith and Todd, 2005), it is also applied outside the treatment evaluation context to decompose the gender wage gap (Frolich, 2007), the test score gap across genders (Gevrek and Seiberlich, 2014), and the test score gap across countries (Botezat and Seiberlich, 2013). Following Frolich (2007), we implement a semiparametric version of the twofold Oaxaca-Blinder (OB) decomposition to investigate the gender PISA test score gap in mathematics.[9] This decomposition method has distinct advantages over the standard OB decomposition. First, it relaxes the parametric assumptions of the standard OB decomposition. Second, unlike the standard OB decomposition, it computes counterfactual outcomes only for the common support sub-population. The semiparametric matching method allows us to estimate the counterfactual outcomes for each individual separately, accounting for heterogeneity across individuals (Heckman et al., 1999; Imbens, 2004). Moreover, this method makes it possible to explore the gender mathematics gap at different quantiles of the test score distribution.

We first estimate the propensity score by a logit regression of G on a set of explanatory variables, i.e., $p = \Pr[G = 1|X = x] = F(x'\beta)$, where G is a binary variable indicating whether the student is female and $F(\cdot)$ represents the cumulative logistic distribution. Figure A.1 presents the densities of the estimated propensity scores for females and males in each country. The region of common support is defined as follows:

$$S : p_i \in [\min(p_j), \max(p_j)] \tag{1}$$

where p_i denotes the estimated propensity scores for females while p_j denotes the estimated propensity scores for males. All observations with an estimated propensity score that is larger than the maximum propensity score of males and smaller than the minimum estimated propensity score of males are discarded from the analysis. The gender test score gap over the common support sub-population can be written as follows:

$$\begin{aligned} \Delta_S &= E^S[Y_1|G=1] - E^S[Y_0|G=0] \\ &= \int\limits_S E[Y|P(X)=p, G=1] f_1^S(p)\,\mathrm{dp} - \int\limits_S E[Y|P(X)=p, G=0] f_0^S(p)\mathrm{dp} \end{aligned} \tag{2}$$

where Y_0 and Y_1 denote the potential outcomes. $f_1^S(p)$ and $f_0^S(p)$ represent the distributions of the propensity score $p = P(X)$ over the common support sub-population S for females ($G = 1$) and males ($G = 0$), respectively.[10] To simplify notation, we henceforth use $m_0(p)$ and $m_1(p)$ to denote $E[Y|P(X) = p, G = 0]$ and $E[Y|P(X) = p, G = 1]$, respectively.

The counterfactual outcome that represents the expected test score that females would have if they had the same returns to educational inputs as males is identified as follows:

$$E^S[Y^0|G=1] = \int_S m_0(p) f_1^S(p) \text{ dp} \tag{3}$$

After adding and subtracting the counterfactual outcome in Equation 3, we can decompose the gender test score gap for the common support sub-population in Equation 2 into two parts:

$$\Delta_S = \underbrace{\int_S m_0(p) \left[f_1^S(p) - f_0^S(p)\right] \text{ dp}}_{\text{Explained part: } \Delta_e} + \underbrace{\int_S \left[m_1(p) - m_0(p)\right] f_1^S(p) \text{ dp}}_{\text{Unexplained part: } \Delta_u} \tag{4}$$

The first term can be attributed to gender differences in the distribution of propensity scores. It would vanish if females had the same characteristics as males. The second term is due to differences in returns to these characteristics. It would vanish if females had the same returns to educational inputs as males.[11]

The main identifying assumption requires that conditional on the distributions of observable characteristics, the potential outcomes are stochastically independent of gender: $Y_0, Y_1 \perp G|P(X)$. To justify this assumption, we control for a large set of covariates available in the PISA data set, including subjective measures of ability and motivation in mathematics, student and family background characteristics, and school fixed effects (Gevrek and Seiberlich, 2014). The second assumption we rely on for the identification of the explained and unexplained parts is called the overlap assumption (Rosenbaum and Rubin, 1983; Heckman, Ichimura, and Todd, 1997; Hahn, 1998; Wooldridge, 2002; Imbens, 2004). It rules out the phenomenon of perfect predictability of gender given X, i.e., 0 <Pr$(G = 1|X) < 1$. To satisfy the overlap assumption, we restrict the estimation of the explained and unexplained parts to the common support sub-population.

The counterfactual outcome for each female is estimated with the following Kernel regression (Nadaraya, 1964; Watson, 1964):

$$\hat{m}_0(p) = \frac{\sum_{j=1}^{n_0} w_j \cdot K(p_j - p) \cdot Y_j}{\sum_{j=1}^{n_0} w_j \cdot K(p_j - p)} \tag{5}$$

where K is the Gaussian kernel with a bandwidth that is calculated according to Silverman's rule Silverman (1986), p_i represents the estimated propensity score for each female i where $i = 1, .., n_1$. p_j represents the estimated propensity score for each male j where $j = 1, .., n_0$. w_i and w_j denote sampling weights for females and males, respectively.

The weighted average of the estimated counterfactual outcomes is equal to:

$$\hat{E}_S[Y_0|G = 1] = \frac{\sum_{i=1}^{n_1} w_i \cdot \hat{m}_0(p_i)}{\sum_{i=1}^{n_1} w_i} \tag{6}$$

This semiparametric method allows us to explore how the gender gap varies across the test score distribution. We are able to decompose the gender test score gap for the common support sub-population into the explained and unexplained parts at different quantiles. The two parts of the gap at quantile τ are estimated as follows:

$$\Delta_u^\tau = F^{-1}_{y_1|G=1,S}(\tau) - F^{-1}_{y_0|G=1,S}(\tau)$$

$$\Delta_e^\tau = F^{-1}_{y_0|G=1,S}(\tau) - F^{-1}_{y_0|G=0,S}(\tau)$$

$$\Delta_S^\tau = \Delta_u^\tau + \Delta_e^\tau = F^{-1}_{y_1|G=1,S}(\tau) - F^{-1}_{y_0|G=0,S}(\tau)$$

The adjusted quantiles are obtained indirectly through inverting the adjusted distribution functions.[12] As it is not possible to estimate standard errors for the decomposition components analytically, we employ bootstrapping with replacement as suggested by Efron (1979). 250 bootstrap replications are used to calculate the standard errors.

Cross-country analysis

After decomposing the gender mathematics test score gap in each country individually, we investigate whether differences in the indicators of gender inequality can account for variations in the gender gap and its components across countries. To illuminate the role of societal factors in explaining gender differences in mathematics achievement, we estimate the following simple model:

$$Y_c = \beta_0 + \beta_1 Z_c + \beta_2 \ln(GDP_c) + \varepsilon_c \tag{7}$$

where Y_c is the mean gender mathematics test score gap (or its decomposition components) in country c. The vector Z_c denotes gender equality measures for country c in 2012. We use the Gender Gap Index (GGI) and its sub-indices to assess societal level female empowerment. GDP_c represents GDP per capita for country c in 2012. ε_c is the error term.

Our parameter of interest, β_1 is meant to capture the relationship between a country's measures of gender inequity and the mathematics gender gap at the mean. Taking advantage of the fact that our decomposition method allows us to explore the gap, not only at the mean, but also across the test score distribution, we also examine how this relationship evolves across the distribution. In this case, the dependent variable is equal to the gender test score gap (or its decomposition components) at the relevant quantile.

5 Results

The Semiparametric OB Decomposition

Before interpreting the results from the semiparametric OB decomposition, we investigate whether the propensity score matching procedure has the ability to balance the distributions of the covariates for females and males in each country. To check the matching quality, we use the test for standardized bias (SB) proposed by Rosenbaum and Rubin (1985).[13] Rosenbaum (2002) considers a covariate balanced if the SB after matching in absolute value is smaller than 0.2. Based on the rule of thumb proposed by Rosenbaum (2002), Table A.4 indicates that there are no countries with unbalanced covariates after matching.[14]

<< **Table 2 here** >>

Table 2 presents the results from the semiparametric OB decomposition of the mean gender gap for all countries. The mean gap is statistically insignificant in 14 out of 56 countries.[15] When the mean gap is statistically significant, it is always negative, suggesting that boys, on average, have significantly higher mathematics test scores than girls. The gap in favor of boys varies widely, from 30.75 score points in Austria to 6.82 score points in Indonesia. As indicated in the second column of Table 2, the explained part of the gap is statistically insignificant in 34 out of 56 countries, suggesting that observable factors that affect mathematics test scores do not vary systematically between girls and boys in most countries. The explained part of the gap is positive (negative) and statistically significant in 6 (16) countries and ranges between 21.48 score points in favor of boys in New Zealand and 17.96 score points in favor of girls in Turkey. A negative and statistically significant explained part implies that gender differences in observed characteristics predict an advantage for males over females in the average mathematics scores.

The third column of Table 2 shows that the unexplained part of the gap is statistically significant in 42 countries. Only four out of the 42 countries exhibit positive unexplained parts, providing evidence that females are more efficient in transforming educational inputs into higher mathematics test scores in Finland, Iceland, Norway, and Singapore. In the remaining 38 countries, unexplained parts are negative, suggesting that males have a particular advantage with converting educational inputs into better mathematics test scores in most countries.

The number of females who are in (out of) the common support is shown in column 5 (6) of Table 2. The fourth column of Table 2 shows the component of the gap that corresponds to the non-overlapping support. In this case, the whole gap is decomposed into three components: $\Delta_y = \Delta_1 + \Delta_e + \Delta_u$. It is important to note that we restrict the estimation of the explained (Δ_e) and unexplained (Δ_u) parts to the common support sub-population. The first component of the gap, Δ_1, captures the difference in mean test scores between females who cannot be matched with males and those who can. As shown in the fourth column of Table 2, Δ_1 is statistically significant only in the following six countries: Croatia, France, Hungary, Italy, Japan, Slovenia and the United Arab Emirates. The negative and statistically significant value of Δ_1 in these countries indicate that females who are in the common support perform better than those who are out of the common support.

Cross-country analysis

Table 3 shows the results from the country-level analysis. Following Guiso et al. (2008), we first examine the relationship between the mean gender test score gap in mathematics and gender inequality as measured by the World Economic Forum's Gender Gap Index (GGI). Unlike Guiso et al. (2008), we find that the GGI and its sub-indices are not significantly correlated with the size of the mean gender mathematics gap.[16] Contrary to previous studies investigating the link between the gender mathematics gap and indicators of gender equity in society, we employ a novel empirical strategy, in which we use a semiparametric Oaxaca-Blinder (OB) decomposition method to decompose the gender mathematics test score gap into a part that is explained by gender differences in observable characteristics and a part that remains unexplained in each country. After isolating the contribution of the explained part, which is due to differences in observable characteristics between females and males, to the overall gender gap, we investigate whether the variation in gender inequality indicators can help explain variation in the unexplained part of the gap. The results presented in columns 4-7 of Table 3 indicate that gender equality measures significantly correlate with the unexplained part of the gap across nations. The more gender-equal a country, the larger the unexplained part of the gap. We find that a one standard deviation increase in the GGI is associated with a 0.41 standard deviation increase in the unexplained part of the gap.

<< **Table 3 here** >>

The estimated coefficient of the GGI implies that if Turkey, the lowest ranked country in terms of the GGI index in our sample, had the same degree of gender equality as Iceland, the highest ranked one, the unexplained part of the gap in Turkey would increase in favor of females by 20.8 score points, which corresponds to a 92% increase in the unexplained part of the gap in favor of girls.[17] Our finding that girls' comparative disadvantage in transforming educational inputs into higher mathematics test scores tends to decrease in nations with greater gender equality points to the importance of decomposing the gap into an explained and unexplained part. The results presented in column 6 of Table 3 also indicate that cross-country variation in the unexplained part of the gap is significantly related to differences in GDP per capita when we use the political sub-index as a measure of societal gender equity. The richer the country, the larger the unexplained part.

<< **Figure 1 here** >>

Moreover, we explore the relationship between the unexplained part of the gender test score gap and the GGI at different quantiles of the mathematics test score distribution. Figure 1 demonstrates the estimated coefficients of the GGI index obtained from running regressions where the dependent variable is the unexplained part of the gap at a specific quantile. As indicated in Figure 1, the positive correlation between societal level female empowerment as measured by the GGI index and international variation in the unexplained part of the gap weakens as we move towards the upper tail of the distribution.

Robustness Checks

We conduct several robustness checks to test the validity and strength of our estimates. Table A.5 shows the results of the robustness tests. Panel A in Table A.5 replicates the baseline estimates presented in Table 3. Following Guiso et al. (2008), we restrict our sample to those who come from the upper half of the index of economic, social, and cultural status in each country.[18] The decomposition analysis we conduct in the first stage is based on a sample consisting of 110,292 observations. As students from low socioeconomic status are more likely to drop out of school, removing those students from the sample would alleviate the potential bias that may arise from differential drop-out rates across genders. Panel B indicates that the GGI and its sub-indices are not statistically significantly associated with the mean gender mathematics gap while GGI and the economic sub-index statistically significantly correlate with the unexplained part of the gap.

As traditional gender stereotypes might be stronger in rural areas, we also restrict our sample to the students attending schools in villages and small towns. This sub-sample consists of 68,315 students from 53 countries.[19] We conjecture that the link between societal level female empowerment and international variation in the unexplained part of the gap may be more pronounced in this sub-sample. Consistent with our conjecture, Panel C shows that the estimated coefficients of the GGI and the economic sub-index increase in magnitude and remain statistically significant at the conventional significance levels.

We check the sensitivity of the results to a change in the procedure used to determine the region of common support. Following Lechner (2002), we use the tenth smallest and tenth largest estimated propensity scores for males to define the region of common support. Panel D suggests that the results are robust to the alternative definition of the common support. It is worth noting that the number of females that fall outside the common support region increases in all countries when we replace the minima and maxima estimated propensity scores given in Equation 1 with the tenth smallest and tenth largest ones to define the region of common support.[20]

To assess the matching quality, we use a stricter decision criterion that considers a covariate balanced if the SB after matching in absolute value is smaller than 0.1. Table A.4 shows that under the new cut-off, there are nine countries that have at least one unbalanced covariate after matching. We check whether the results are sensitive to the exclusion of those countries from the analysis. Panel E reveals that excluding these countries does not change our results.

6 Conclusion

This study extends the literature on the role of gender inequality in explaining the cross-country variation in the gender mathematics test score gap by applying a novel two-stage empirical strategy. In the first-stage, we use a semiparametric Oaxaca-Blinder (OB) method to decompose the gender mathematics test score gap into a part that is explained by gender differences in observable characteristics and a part that remains unexplained for 56 countries that participated in the 2012 PISA. In the second-stage, we examine the relationship between the decomposition components of the gap and country-specific gender inequality measures. The semiparametric OB decomposition results show that the mean gender mathematics test score gap is statistically significant and in favor of boys in 42 out of 56 countries. The explained part of the gap is not statistically significant in 34 countries, suggesting that gender differences in observable characteristics do not predict an advantage for boys over girls in most countries. However, the unexplained part of the gap is statistically significant in 42 countries and significantly negative in 38 countries, implying that boys have a particular advantage with converting educational inputs into better mathematics test scores. These findings provide evidence that even if girls had the same observable characteristics as boys, the gender test score gap in mathematics would not disappear because societal and cultural factors affect boys and girls differently.

The second-stage results show that indicators of societal gender inequity predict the cross-country variation in the unexplained part of the gap. In more gender-equal countries, the unexplained part of the gap favoring boys becomes smaller. Moreover, the relationship between the unexplained part of the gap and the societal gender inequality exhibits a heterogeneous pattern across the test score distribution, and tends to become less pronounced at the upper tail of the distribution.

The link between gender inequity and the gender mathematics achievement gap might be reciprocal such that the disproportionate under-representation of females compared to males in the economic, political, and social domains may partly result from the fact that girls lag behind boys in mathematics. Although our data do not allow us to establish any causal relationships, we present evidence that policy initiatives aiming at bolstering female empowerment could serve as powerful tools to improve girls' mathematics achievement.

Notes

[1]Ceci, Williams, and Barnett (2009) provide an extensive review of studies that focus on biological explanations. They point out that biological evidence in this domain is inconsistent and sometimes contradictory.

[2]Gevrek and Seiberlich (2014) employ this method to examine the gender PISA test score gap in mathematics/science in Turkey. They discuss the advantages of the semiparametric OB decomposition over the standard OB decomposition in detail.

[3]Baker and Jones (1993) use the following indicators of gender equity: percentage of females in higher education; ratio of female university to non-university higher education programs; percentage of females in the labor force (as a percentage of the total labor force); percentage of females in the industrial, service, agricultural sectors of the labor force; occupational segregation of women (computed as the natural logarithm of the odds ratio of the sexes belonging to high-or low-status occupations).

[4]Please see Fernandez and Fogli (2009) and Gevrek et al. (2013) for more information on the epidemiological approach.

[5]A negative mean gender test score gap implies that boys, on average, outperform girls in mathematics.

[6]The PISA 2012 assesses mathematics achievement of 15-years-olds enrolled in Grades 7 or above.

[7]As the GGI is not available for Liechtenstein, Montenegro, Taiwan, Hong Kong, Macao, Shanghai, we exclude those countries from the analysis. Moreover, we use the 2011 GGI value for Tunisia due to data limitations.

[8]For detailed information on the computation of the GGI, see Hausmann (2012).

[9]The OB decomposition methodology (Oaxaca, 1973; Blinder, 1973), which is primarily used to investigate discrimination in the labor market, has been recently applied in the economics of education. It has been implemented to examine the test score gap between countries (Ammermueller, 2007), schools –private versus public– (Duncan and Sandy, 2007; Krieg and Storer, 2006), boys and girls (Sohn, 2012; Gevrek and Seiberlich, 2014), and ethnic groups –indigenous versus non-indigenous– (Sakellariou, 2008; McEwan, 2004).

[10]$f_g^S(p) = \frac{f_g(p)}{\mu_{S|G=g}}$ is scaled such that the integral integrates to one, where $\mu_{S|G=g}$ is the empirical probability of being in the common support conditional on having gender g.

[11]Following Nopo (2008), we decompose the whole gap Δ_y into three parts: $\Delta_y = \Delta_1 + \Delta_e + \Delta_u$, where Δ_1 accounts for differences between two groups of females: those who can be matched with males and those who remain out of the common support, weighted by the empirical fraction of females who are out of the common support. A positive (negative) value of Δ_1 indicates that female students, who are out of the common support, perform better (worse) than their counterparts, who are in the common

support. The decomposition of Δ_y can be written as follows:

$$\Delta_y = \underbrace{\mu_{\bar{S}|G=1}\left[\int_{\bar{S}} m_1(p) f_1^{\bar{S}}(p)\ \mathrm{dp} - \int_{S} m_1(p) f_1^{S}(p)\ \mathrm{dp}\right]}_{\Delta_1}$$
$$+ \underbrace{\int_{S} m_1(p) f_1^{S}(p)\ \mathrm{dp} - \int_{S} m_0(p) f_0^{S}(p)\ \mathrm{dp}}_{\Delta_S}$$
$$+ \underbrace{\mu_{\bar{S}|G=0}\left[\int_{S} m_0(p) f_0^{S}(p)\ \mathrm{dp} - \int_{\bar{S}} m_0(p) f_0^{\bar{S}}(p)\ \mathrm{dp}\right]}_{\Delta_0}$$

where $\bar{S}$ denotes out of the common support and $\mu_{\bar{S}|G=g}$ is the empirical probability of being out of the common support conditional on having gender g. The second summand represents the gender test score gap over the common support sub-population and can be decomposed into the explained part Δ_e and the unexplained part Δ_u. Due to our definition of the common support, $\mu_{\bar{S}|G=0}$ is equal to zero. Therefore, Δ_0 turns out to be zero in our estimation.

[12]The empirical distribution function used to estimate the distribution functions is defined as follows: $\hat{F}(\tau) = \frac{1}{n}\sum_{i=1}^{n} 1\{p_i \leq \tau\}$, where $1\{p_i \leq \tau\} = \begin{cases} 1 & if \quad p_i \leq \tau \\ 0 & otherwise. \end{cases}$

[13]The standardized bias for a covariate X is defined as follows:

$$\mathrm{SB} = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{0.5 \cdot (s_1^2 + s_0^2)}}$$

where $\bar{X}_i$ is the mean value and s_i^2 the estimated variance of X for group $i = 0, 1$. For the standardized bias before matching, we use females as group 1, for the standardized bias after matching, we use the counterfactual control group as group 1. In both cases, we use males as the comparison group 0.

[14]In a robustness check, we use a stricter cut-off as our decision criterion to assess the matching quality (i.e., the SB in absolute value is smaller than 0.1).

[15]The gap is positive (negative) but not statistically significant in Iceland, Latvia, Malaysia, Qatar, Russia, Singapore and Thailand (Finland, Kazakhstan, Norway, Romania, Sweden, Turkey and United Arab Emirates)

[16]Because the cross-country variation in the education and health sub-indices is very small, we do

not use the specifications that separately include those indices in Table 3. Across countries, the average education attainment (health and survival) sub-index is 0.991 (0.974) with a standard deviation of 0.014 (0.007).

[17]The unexplained part of the gap is 22.76 score points in favor of boys in Turkey while the GGI values for Turkey and Iceland are 0.601 and 0.864 respectively. Using the estimated coefficient of the GGI index (β_1=79.26) presented in column 4 of Table 3, we calculate the counterfactual outcome for Turkey as follows: $(GGI_{\text{Iceland}}\text{-}GGI_{\text{Turkey}}) \cdot \beta_1 = 0.263 \cdot 79.26 = 20.85$.

[18]The index of economic, social and cultural status (IESCS) is based on the following variables: the international socio-economic index of occupational status; the highest level of parental education, and an index of home possessions related to family wealth, home educational resources and possessions related to "classical" culture in the family home such as works of classical literature, books of poetry, and works of art. OECD (2013) provides detailed information on the construction of the IESCS.

[19]We have to drop Japan, Singapore, and the United Arab Emirates in this robustness check because the data contain no observations on the common support in each of those countries.

[20]The percentage of females who are out of common support ranges from 0.1% in Mexico to 13.1% in the United Arab Emirates.

References

Baker, D. P., and Jones, D. P. (1993). Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociology of Education*, 66(2): 91-103.

Black, D. A., and Smith, J. A. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics*, 121(1): 99-124.

Botezat, A., and Seiberlich, R. (2013). Educational performance gaps in Eastern Europe. *Economics of Transition*, 21(4): 731-756.

Cahill, L. (2005). His brain, her brain. *Scientific American*, 292(5): 40-47.

Davison, K. K., and Susman, E. J. (2001). Are hormone levels and cognitive ability related during early adolescence? *International Journal of Behavioral Development*, 25(5): 416-428.

Dickerson, A., McIntosh, S., and Valente, C. (2015). Do the maths: An analysis of the gender gap in mathematics in Africa. *Economics of Education Review*, 46: 1-22.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1): 1-26.

Frolich, M. (2004). Finite-sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics*, 86(1): 77-90.

Frolich, M. (2007). Propensity score matching without conditional independence assumption with an application to the gender wage gap in the United Kingdom. *Econometrics Journal*, 10(2): 359-407.

Fryer, R. G., and Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2): 210-240.

Gevrek, Z. E., and Seiberlich, R. R. (2014). Semiparametric decomposition of the gender achievement gap: An application for Turkey. *Labour Economics*, 31: 27-44.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Diversity. Culture, gender, and math. *Science*, 320: 1164-1165.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2): 315-331.

Hausmann, R., Tyson, L. D., and Zahidi, S. (2012). *The Global Gender Gap Report 2012*. World Economic Forum.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64(4): 605-654.

Heckman, J. J., LaLonde, R. J., and Smith, J. A. (1999). The economics and econometrics of active labor market programs. *Handbook of Labor Economics*, 3: 1865-2097.

Imbens, G. W. G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1): 4-29.

Jacobs, J. E., and Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers and childrens ability perceptions. *Journal of Personality and Social Psychology*, 63(6): 932-944.

Kucian, K., Loenneker, T., Dietrich, T., Martin, E., and von Aster M. (2005). Gender differences in brain activation patterns during mental rotation and number related cognitive tasks. *Psychology Science*, 47(1): 112-131.

Lechner, M. (2002). Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods. *Journal of the Royal Statistical Society:*

Series A (Statistics in Society), 165(1): 59-82.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability and Its Applications*, 9(1): 141-142.

Nollenberger, N., Rodriguez-Planas, N., and Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, 106(5): 257-261.

OECD. (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy.* OECD Publishing.

Pope, D. G., and Sydnor, J. R. (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives*, 24(2): 95-108.

Riegle-Crumb, C. (2005). The cross-national context of the gender gap in math and science. *In: The social organization of schooling (pp. 227-243).* New York, NY: Russell Sage Foundation.

Rosenbaum, P. R. (2002). *Observational studies.* Springer.

Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1): 41-55.

Rosenbaum, P. R., and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1): 33-38.

Sianesi, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics*, 86(1): 133-155.

Silverman, B. (1986). *Density estimation for statistics and data analysis.* CRC press.

Smith, J. A., and Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2): 305-353.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, 92(1): 144-151.

Watson, G. (1964). Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A*, 26(4): 359-372.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

Table 1: Descriptive statistics: Student-level.

		Boys		Girls	
Variable	Description	Mean	St. Dv.	Mean	St. Dv.
Outcome variables					
test score	Average of the five plausible values in mathematics	489.14	(97.94)	474.08	(93.10)
Student characteristics					
grade8	Student is in Grade 8 or lower	0.064	(0.25)	0.042	(0.20)
grade9	Student is in Grade 9	0.383	(0.49)	0.364	(0.48)
grade10	Student is in Grade 10	0.475	(0.50)	0.514	(0.50)
grade11	Student is in Grade 11 or higher	0.077	(0.27)	0.080	(0.27)
self-efficacy	Index of self-efficacy in mathematics	0.129	(1.02)	-0.161	(0.88)
motivation	Sum of two indices: Enjoyment and instrumental motivation in mathematics	0.441	(1.79)	0.121	(1.79)
Family characteristics					
math is important	Parents think that studying mathematics is important	0.853	(0.35)	0.789	(0.41)
meduc1	Mother has at most primary education	0.098	(0.30)	0.115	(0.32)
meduc2	Mother has at most secondary education	0.476	(0.50)	0.489	(0.50)
meduc3	Mother has at most tertiary education	0.426	(0.49)	0.396	(0.49)
feduc1	Father has at most primary education	0.093	(0.29)	0.106	(0.31)
feduc2	Father has at most secondary education	0.489	(0.50)	0.510	(0.50)
feduc3	Father has at most tertiary education	0.418	(0.49)	0.384	(0.49)
book10	0 - 10 books at home	0.183	(0.39)	0.156	(0.36)
book25	11 - 25 books at home	0.202	(0.40)	0.197	(0.40)
book100	26 - 100 books at home	0.298	(0.46)	0.300	(0.46)
book101	More than 100 books at home	0.317	(0.47)	0.347	(0.48)
parent's occupational status	Index of highest parental occupational status	50.05	(22.39)	48.87	(22.32)
home education resources	Index of home education resources	-0.084	(1.05)	-0.073	(1.02)
n	Number of Students	107115		113789	

Table 2: Semiparametric OB decomposition of the mean gender mathematics test score gap.

	Δ_y	Δ_e	Δ_u	Δ_1	On	Off
Argentina	-20.49**	-3.69	-16.71**	-0.08	1,462	4
Australia	-17.44**	-19.33**	1.94	-0.05	3,287	3
Austria	-30.75**	-9.79	-20.80**	-0.16	1,280	9
Belgium	-15.99**	-2.51	-13.45**	-0.03	2,335	1
Brazil	-16.67**	4.76**	-21.42**	-0.01	5,569	1
Bulgaria	-10.97**	2.70	-12.46**	-1.21	1,368	25
Canada	-14.88**	-11.77**	-3.11	-0.01	6,203	1
Chile	-21.04**	2.80	-23.23**	-0.61	1,693	13
Colombia	-25.67**	2.06	-28.11**	0.38	2,216	25
Costa Rica	-22.79**	2.84	-25.46**	-0.17	1,190	14
Croatia	-15.84**	-0.77	-13.85**	-1.22*	1,463	25
Czech Republic	-18.24**	2.83	-20.69**	-0.38	1,325	5
Denmark	-15.08**	-13.95**	-1.13	0	1,913	0
Estonia	-9.52**	-5.49	-3.95	-0.08	1,387	1
Finland	-2.21	-11.12**	9.00**	-0.08	2,581	2
France	-18.69**	-6.65	-11.39**	-0.65*	1,243	17
Germany	-16.22**	-7.13	-8.36*	-0.74	909	9
Greece	-15.03**	-1.59	-13.44**	0	1,575	0
Hungary	-20.29**	1.44	-21.33**	-0.40*	1,402	14
Iceland	0.46	-13.71**	14.62**	-0.44	927	5
Indonesia	-6.82*	1.65	-8.49**	0.03	1,561	1
Ireland	-22.17**	-8.10*	-13.93**	-0.14	806	1
Israel	-25.30**	-14.81**	-9.68*	-0.81	1,124	7
Italy	-25.71**	-9.76**	-15.06**	-0.89*	8,454	29
Japan	-22.33**	-14.06**	-7.02**	-1.26*	1,390	15
Kazakhstan	-0.80	3.47	-4.30*	0.03	1,637	1
Korea	-10.06*	-7.38	-2.32	-0.36	804	5
Latvia	6.29	4.89	2.39	-0.99	1,182	22
Lithuania	-7.35*	-5.10	-1.78	-0.46	1,288	5
Luxembourg	-22.19**	-0.45	-21.69**	-0.05	1,136	1
Malaysia	1.36	4.28	-2.99	0.07	1,589	2
Mexico	-12.92**	-2.14*	-10.76**	-0.02	9,998	1
Netherlands	-15.49**	-3.93	-11.59**	0.02	1,121	3
New Zealand	-29.80**	-21.49**	-7.68*	-0.63	613	5
Norway	-2.13	-9.82**	7.73**	-0.04	1,246	4
Peru	-25.66**	1.40	-27.62**	0.57	1,473	8
Poland	-9.28**	-2.70	-6.54*	-0.03	1,421	1
Portugal	-22.52**	-4.24	-18.24**	-0.04	1,560	1
Qatar	1.79	3.02	-1.23	0	796	0
Romania	-3.38	8.47**	-11.77**	-0.08	1,471	5
Russia	0.38	-4.48	5.05	-0.19	1,473	7
Serbia	-10.92**	7.54*	-18.30**	-0.16	1,303	4
Singapore	3.04	-3.73	6.76**	0	1,350	0
Slovakia	-19.39**	1.07	-20.27**	-0.19	1,272	17
Slovenia	-22.60**	-0.14	-22.21**	-0.25*	1,472	6
Spain	-21.48**	-3.87*	-17.61**	0	7,691	0
Sweden	-1.94	-5.12	3.99	-0.81	1,363	46
Switzerland	-16.40**	-11.69**	-4.61	-0.11	3,251	4
Thailand	0.71	5.84*	-6.70*	1.57	2,087	47
Tunisia	-20.70**	5.12	-25.46**	-0.36	1,255	13
Turkey	-6.24	17.96**	-22.76**	-1.44	1,089	26
United Arab Emirates	-6.88	3.63	-9.59	-0.92*	277	6
United Kingdom	-12.63**	-10.51**	-1.67	-0.46	2,892	4
United States	-14.04**	-5.84*	-8.13**	-0.07	1,387	1
Uruguay	-14.02**	4.06	-18.04**	-0.04	1,496	2
Vietnam	-10.48**	7.31**	-18.15**	0.35	1,652	7

Notes: Δ_y represents the mean gender mathematics test score gap. The mean gender mathematics test score gap of the common support sub-population is decomposed into an explained part, which is denoted by Δ_e, and an unexplained part, which is denoted by Δ_u. Δ_1 shows the differences between two groups of females: those who can be matched with males and those who remain out of the common support. On/Off denote the number of females who are on/off the common support, respectively. Significance was calculated using 250 bootstrap replications. * significant at 5% (97.5% and 2.5% quantile of the bootstrap distribution have the same signs) and ** significant at 1% (99.5% and 0.5% quantile of the bootstrap distribution have the same signs).

Table 3: The relationship between the gender mathematics gap and gender inequality.

	Δ_y	Δ_y	Δ_y	Δ_u	Δ_u	Δ_u
Gender Gap Index (GGI)	20.41			79.26**		
	(24.93)			(24.87)		
Economic Sub-index		19.40			45.57**	
		(13.76)			(13.10)	
Political Sub-index			3.99			20.48*
			(9.04)			(9.32)
LOG(GDP)	-1.43	-1.76	-1.05	3.12	3.18	4.41*
	(2.06)	(1.99)	(2.01)	(1.93)	(1.86)	(1.92)
Constant	-13.87	-9.13	-3.90	-97.89**	-72.77**	-58.52**
	(21.01)	(18.61)	(19.41)	(19.04)	(16.61)	(18.30)
Observations	56	56	56	56	56	56

Notes: Standard errors in parentheses. * significant at 5% and ** significant at 1%. In the first three columns, the dependent variable is the mean gender mathematics test score gap, denoted by Δ_y. In the last three columns, the dependent variable is the unexplained part of the mean gender mathematics test score gap, denoted by Δ_u. Country weights that are inversely proportional to the estimated standard errors of Δ_u are used when the dependent variable is Δ_u.

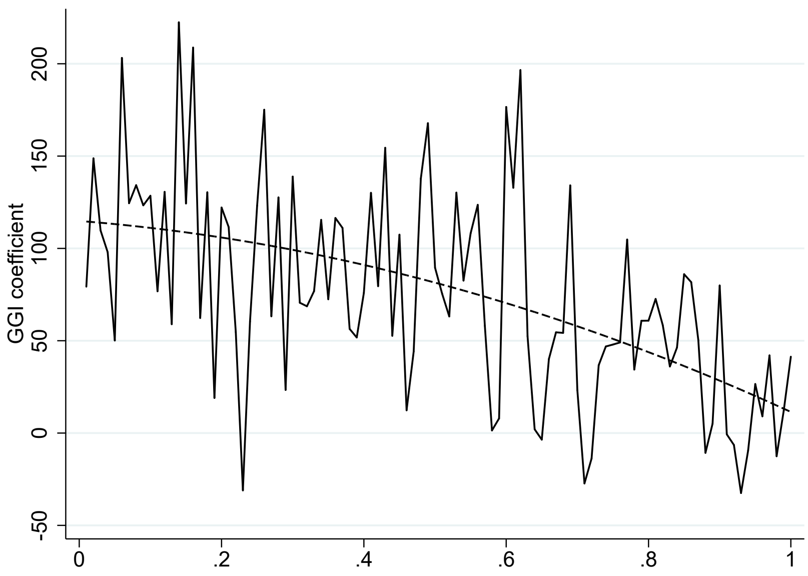

Figure 1: Relationship between the unexplained part of the mean gender math test score gap Δ_u and the GGI at different quantiles of the math test score distribution. The dashed line is a quadratic prediction.

7 Appendix

Table A.1: Descriptive statistics: Country-level.

	n	Δ_y	GDP	GGI	Eco. part.	Pol. emp.	Educ.	Health
Argentina	2,749	-20.49**	13,040	0.721	0.607	0.302	0.996	0.980
Australia	6,754	-17.44**	67,646	0.729	0.759	0.185	1.000	0.974
Austria	2,511	-30.75**	48,334	0.739	0.652	0.332	0.995	0.979
Belgium	4,240	-15.99**	44,741	0.765	0.724	0.366	0.992	0.979
Brazil	10,314	-16.67**	12,157	0.691	0.650	0.134	1.000	0.980
Bulgaria	2,674	-10.97**	7,378	0.702	0.696	0.141	0.992	0.979
Canada	12,078	-14.88**	52,495	0.738	0.788	0.196	0.991	0.978
Chile	3,415	-21.04**	15,253	0.668	0.547	0.145	0.999	0.980
Colombia	4,375	-25.67**	7,885	0.690	0.621	0.166	0.994	0.979
Costa Rica	2,348	-22.79**	9,985	0.722	0.599	0.316	1.000	0.975
Croatia	2,777	-15.84**	13,236	0.705	0.669	0.178	0.995	0.979
Czech Republic	2,566	-18.24**	19,730	0.677	0.603	0.125	1.000	0.979
Denmark	3,693	-15.08**	58,125	0.778	0.772	0.364	1.000	0.974
Estonia	2,759	-9.52**	17,422	0.698	0.719	0.099	0.994	0.979
Finland	5,081	-2.21	47,416	0.845	0.785	0.616	1.000	0.980
France	2,348	-18.69**	40,838	0.698	0.669	0.145	1.000	0.980
Germany	1,863	-16.22**	44,065	0.763	0.740	0.349	0.985	0.978
Greece	3,119	-15.03**	22,243	0.672	0.633	0.086	0.994	0.974
Hungary	2,556	-20.29**	12,834	0.672	0.659	0.057	0.992	0.979
Iceland	1,834	0.46	44,259	0.864	0.754	0.733	1.000	0.970
Indonesia	2,999	-6.82*	3,701	0.659	0.565	0.132	0.973	0.966
Ireland	1,678	-22.17**	49,231	0.784	0.751	0.412	0.999	0.974
Israel	2,132	-25.30**	32,570	0.699	0.682	0.156	0.987	0.970
Italy	16,394	-25.71**	34,814	0.673	0.591	0.135	0.992	0.973
Japan	2,940	-22.33**	48,629	0.653	0.576	0.070	0.987	0.979
Kazakhstan	3,225	-0.80	12,387	0.721	0.768	0.146	0.992	0.980
Korea	1,742	-10.06*	24,454	0.636	0.509	0.101	0.959	0.973
Latvia	2,341	6.29	13,799	0.757	0.762	0.288	1.000	0.980
Lithuania	2,586	-7.35**	14,343	0.719	0.755	0.147	0.995	0.979
Luxembourg	2,543	-22.19**	105,447	0.744	0.815	0.193	0.996	0.972
Malaysia	2,962	1.36	10,835	0.654	0.599	0.053	0.991	0.973
Mexico	19,257	-12.92**	9,721	0.671	0.538	0.176	0.991	0.980
Netherlands	2,290	-15.49**	49,475	0.766	0.758	0.336	1.000	0.970
New Zealand	1,249	-29.80 **	40,067	0.781	0.782	0.370	1.000	0.970
Norway	2,508	-2.13	101,564	0.840	0.830	0.562	1.000	0.970
Peru	2,977	-25.66**	6,386	0.674	0.610	0.142	0.980	0.966
Poland	2,693	-9.28**	13,145	0.702	0.650	0.179	0.998	0.979
Portugal	3,034	-22.52**	20,577	0.707	0.679	0.183	0.994	0.972
Qatar	1,691	1.79	92,693	0.626	0.556	0.000	0.998	0.952
Romania	2,887	-3.38	8,558	0.686	0.681	0.089	0.994	0.979
Russia	2,893	0.38	15,042	0.698	0.720	0.095	0.998	0.979
Serbia	2,473	-10.92**	5,659	0.704	0.660	0.192	0.993	0.970
Singapore	2,776	3.04	54,451	0.699	0.788	0.095	0.941	0.972
Slovakia	2,507	-19.39**	17,275	0.682	0.628	0.122	1.000	0.980
Slovenia	2,629	-22.60**	22,486	0.713	0.714	0.168	0.998	0.973
Spain	15,167	-21.48 **	28,648	0.727	0.646	0.284	0.997	0.979
Sweden	2,663	-1.94	57,134	0.816	0.796	0.498	0.997	0.974
Switzerland	6,427	-16.40**	83,164	0.767	0.752	0.353	0.991	0.974
Thailand	3,635	0.71	5,915	0.689	0.699	0.090	0.989	0.980
Tunisia	2,319	-20.70**	4,179	0.625	0.444	0.128	0.966	0.964
Turkey	2,278	-6.24	10,539	0.601	0.414	0.087	0.930	0.976
United Arab Emirates	656	-6.88	41,712	0.639	0.475	0.121	1.000	0.961
United Kingdom	5,813	-12.63**	41,538	0.743	0.730	0.274	0.999	0.970
United States	2,735	-14.04**	51,433	0.737	0.814	0.156	1.000	0.979
Uruguay	2,739	-14.02**	15,092	0.675	0.660	0.062	0.997	0.980
Vietnam	3,012	-10.48**	1,755	0.687	0.710	0.125	0.968	0.944

Notes: * significant at 5% and ** significant at 1%. Δ_y is the mean gender mathematics test score gap. GDP is the 2012 real per capita Gross domestic product. GGI is the Gender Gap Index. Eco. part., Pol. emp., Educ and Health are the Economic Participation, Political Empowerment, Educational Attainment, and Health and Survival sub-indices, respectively. Both GDP and GGI are from the Global Gender Gap Report (2012).

Table A.2: Construction of the indices of self-efficacy in mathematics, enjoyment of mathematics and instrumental motivation in mathematics.

The index	Construction of the index	Interpretation
1. Index of self-efficacy in mathematics	It was created by using students' responses to the following eight statements 1. Using a train timetable to work out how long it would take to get from one place to another 2. Calculating how much cheaper a TV would be after a 30% discount 3. Calculating how many square metres of tiles you need to cover a floor 4. Understanding graphs presented in newspapers 5. Solving an equation like 3x+5=17 6. Finding the actual distance between two places on a map with a 1:10 000 scale 7. Solving an equation like 2(x+3)=(x+3)(x-3) 8. Calculating the petrol-consumption rate of a car Response options for each statement were: I could do this easily, I could do this with a bit of effort, I would struggle to do this on my own and I couldn't do this.	Higher values for the index mean that students self-assess their mathematics abilities higher.
2. Index of enjoyment of mathematics	It was created by using students' responses to the following four statements 1. I enjoy reading about mathematics 2. I look forward to my mathematics sessions 3. I do mathematics because I enjoy it 4. I am interested in the things I learn in mathematics Response options for each statement were: I strongly agree, I agree, I disagree and I strongly disagree.	Higher values for the index mean that students have a higher enjoyment of mathematics.
3. Index of instrumental motivation in mathematics	It was created by using students' responses to the following four statements 1. Making an effort in mathematics is worth it because it will help me in the work that I want to do later on 2. Learning mathematics is worthwhile for me because it will improve my career prospects and chances 3. Mathematics is an important subject for me because I need it for what I want to study later on 4. I will learn many things in mathematics that will help me get a job Response options for each statement were: I strongly agree, I agree, I disagree and I strongly disagree.	Higher values for the index mean that students have a higher motivation in mathematics.

Table A.3: Structure of the Global Gender Gap Index. Source is the Global Gender Gap Report 2012 (Hausmann, 2012).

Sub-index	Variable	Source
Economic Participation and Opportunity	Ratio: female labor force participation over male value	International Labour Organisation, *Key Indicators of the Labour Market (KILM), 2009*
	Wage equality between women and men for similar work (converted to female-over-male ratio)	World Economic Forum, *Executive Opinion Survey, 2012*
	Ratio: estimated female earned income over male value	World Economic Forum, calculations based on the United Nations Development Programme methodology (refer to Human Development Report 2009)
	Ratio: female legislators, senior officials and managers over male value	International Labour Organisation, *LABORSTA Internet*, online database, 2008 or latest data available; United Nations Development Programme, *Human Development Report 2009*, the most recent year available between 1999 and 2007
	Ratio: female professional and technical workers over male value	International Labour Organisation, *LABORSTA Internet*, online database, 2008 or latest data available; United Nations Development Programme, *Human Development Report 2009*, the most recent year available between 1999 and 2007
Educational Attainment	Ratio: female literacy rate over male value	UNESCO Institute for Statistics, *Education Indicators*, 2011 or latest data available; *World Banks World Development Indicators & Global Development Finance online Database*, 2010 or latest available data; United Nations Development Programme, *Human Development Report 2009*, the most recent year available between 1997 and 2007
	Ratio: female net primary level enrolment over male value	UNESCO Institute for Statistics, *Education Indicators*, 2011 or latest data available; *World Banks World Development Indicators & Global Development Finance online database*, 2011 or latest available data
	Ratio: female net secondary level enrolment over male value	UNESCO Institute for Statistics, *Education Indicators*, 2011 or latest data available; *World Banks World Development Indicators & Global Development Finance online database*, 2011 or latest available data
	Ratio: female gross tertiary level enrolment over male value	UNESCO Institute for Statistics, *Education Indicators*, 2011 or latest data available; *World Banks World Development Indicators & Global Development Finance online database*, 2011 or latest available data
Health and Survival	Sex ratio at birth (converted to female-over-male ratio)	Central Intelligence Agency, *The CIA World Factbook*, data updated weekly, 2012
	Ratio: female healthy life expectancy over male value	World Health Organisation, *Global Health Observatory database*, data from 2007
Political Empowerment	Ratio: females with seats in parliament over male value	Inter-Parliamentary Union, *Women in Politics: 2012*, reflecting elections/ appointments up to 1 January 2012
	Ratio: females at ministerial level over male value	Inter-Parliamentary Union, *Women in Politics: 2012*, reflecting elections/ appointments up to 1 January 2012
	Ratio: number of years of a female head of state or government (last 50 years) over male value	World Economic Forum calculations, 30 June 2012

Table A.4: Number of characteristics with SB larger than 0.1 and 0.2 before and after PSM.

	$\|SB\| > 0.1$		$\|SB\| > 0.2$			$\|SB\| > 0.1$		$\|SB\| > 0.2$	
	Before	After	Before	After		Before	After	Before	After
Argentina	14	0	2	0	Lithuania	10	0	2	0
Australia	5	0	3	0	Luxembourg	12	0	4	0
Austria	46	1	3	0	Malaysia	6	0	0	0
Belgium	15	0	3	0	Mexico	4	0	1	0
Brazil	7	0	2	0	Netherlands	6	0	3	0
Bulgaria	14	1	1	0	New Zealand	6	0	3	0
Canada	3	1	2	0	Norway	5	0	2	0
Chile	13	0	2	0	Peru	2	0	0	0
Colombia	9	1	1	0	Poland	11	0	1	0
Costa Rica	8	0	3	0	Portugal	4	0	1	0
Croatia	21	0	1	0	Qatar	7	0	2	0
Czech Republic	26	3	5	0	Romania	8	0	0	0
Denmark	7	0	4	0	Russia	5	0	3	0
Estonia	7	0	2	0	Serbia	24	0	3	0
Finland	8	0	2	0	Singapore	2	0	0	0
France	15	0	3	0	Slovakia	27	0	3	0
Germany	11	1	4	0	Slovenia	26	1	3	0
Greece	8	0	2	0	Spain	4	0	1	0
Hungary	17	0	3	0	Sweden	5	0	3	0
Iceland	3	0	1	0	Switzerland	14	2	4	0
Indonesia	16	0	1	0	Thailand	10	0	1	0
Ireland	8	0	1	0	Tunisia	9	0	1	0
Israel	11	0	1	0	Turkey	26	0	1	0
Italy	6	3	1	0	United Arab Emirates	26	0	8	0
Japan	13	0	2	0	United Kingdom	5	0	2	0
Kazakhstan	7	0	0	0	United States	6	0	1	0
Korea	17	0	0	0	Uruguay	14	0	2	0
Latvia	10	0	3	0	Vietnam	10	0	3	0

Table A.5: Robustness checks.

Panel A: Baseline specification	Δ_y	Δ_y	Δ_y	Δ_u	Δ_u	Δ_u
Gender Gap Index (GGI)	20.41			79.26**		
	(24.93)			(24.87)		
Economic Sub-index		19.40			45.57**	
		(13.76)			(13.10)	
Political Sub-index			3.99			20.48*
			(9.04)			(9.32)
n	56	56	56	56	56	56
Panel B: Only upper half from socio-economic status	Δ_y	Δ_y	Δ_y	Δ_u	Δ_u	Δ_u
Gender Gap Index (GGI)	29.78			64.72**		
	(26.04)			(23.75)		
Economic Sub-index		24.65			40.82*	
		(14.33)			(12.55)	
Political Sub-index			6.93			15.82
			(9.47)			(8.88)
n	56	56	56	56	56	56
Panel C: Only students from rural areas	Δ_y	Δ_y	Δ_y	Δ_u	Δ_u	Δ_u
Gender Gap Index (GGI)	8.12			91.31**		
	(32.30)			(32.34)		
Economic Sub-index		2.88			47.49*	
		(18.47)			(17.75)	
Political Sub-index			3.69			22.61
			(11.36)			(11.40)
n	53	53	53	53	53	53
Panel D: Trimming of the common support	Δ_y	Δ_y	Δ_y	Δ_u	Δ_u	Δ_u
Gender Gap Index (GGI)	20.41			74.13**		
	(24.93)			(24.72)		
Economic Sub-index		19.40			42.87**	
		(13.76)			(13.08)	
Political Sub-index			3.99			19.40*
			(9.04)			(9.23)
n	56	56	56	56	56	56
Panel E: No countries with imbalanced characteristics	Δ_y	Δ_y	Δ_y	Δ_u	Δ_u	Δ_u
Gender Gap Index (GGI)	17.98			76.97**		
	(25.55)			(26.03)		
Economic Sub-index		15.94			43.47**	
		(14.36)			(14.06)	
Political Sub-index			4.14			20.38*
			(9.27)			(9.71)
n	47	47	47	47	47	47

Notes: Standard errors in parentheses. * significant at 5% and ** significant at 1%. In the first three columns, the dependent variable is the mean gender mathematics test score gap, denoted by Δ_y. In the last three columns, the dependent variable is the unexplained part of the mean gender mathematics test score gap, denoted by Δ_u. Country weights that are inversely proportional to the estimated standard errors of Δ_u are used when the dependent variable is Δ_u.

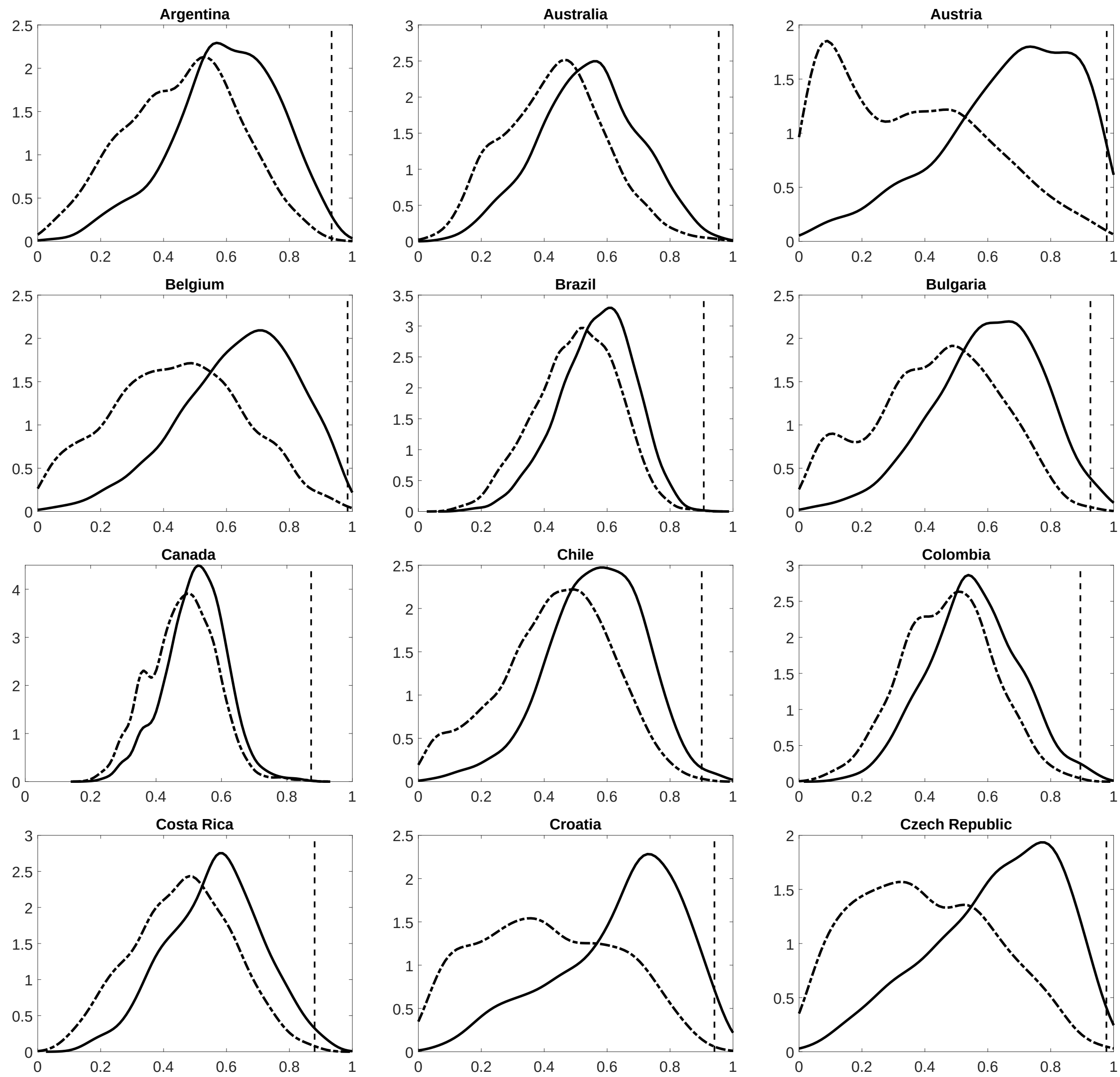

Figure A.1: Gaussian kernel density estimates for females (solid line) and males (dashed). The vertical line denotes the bound for the common support.

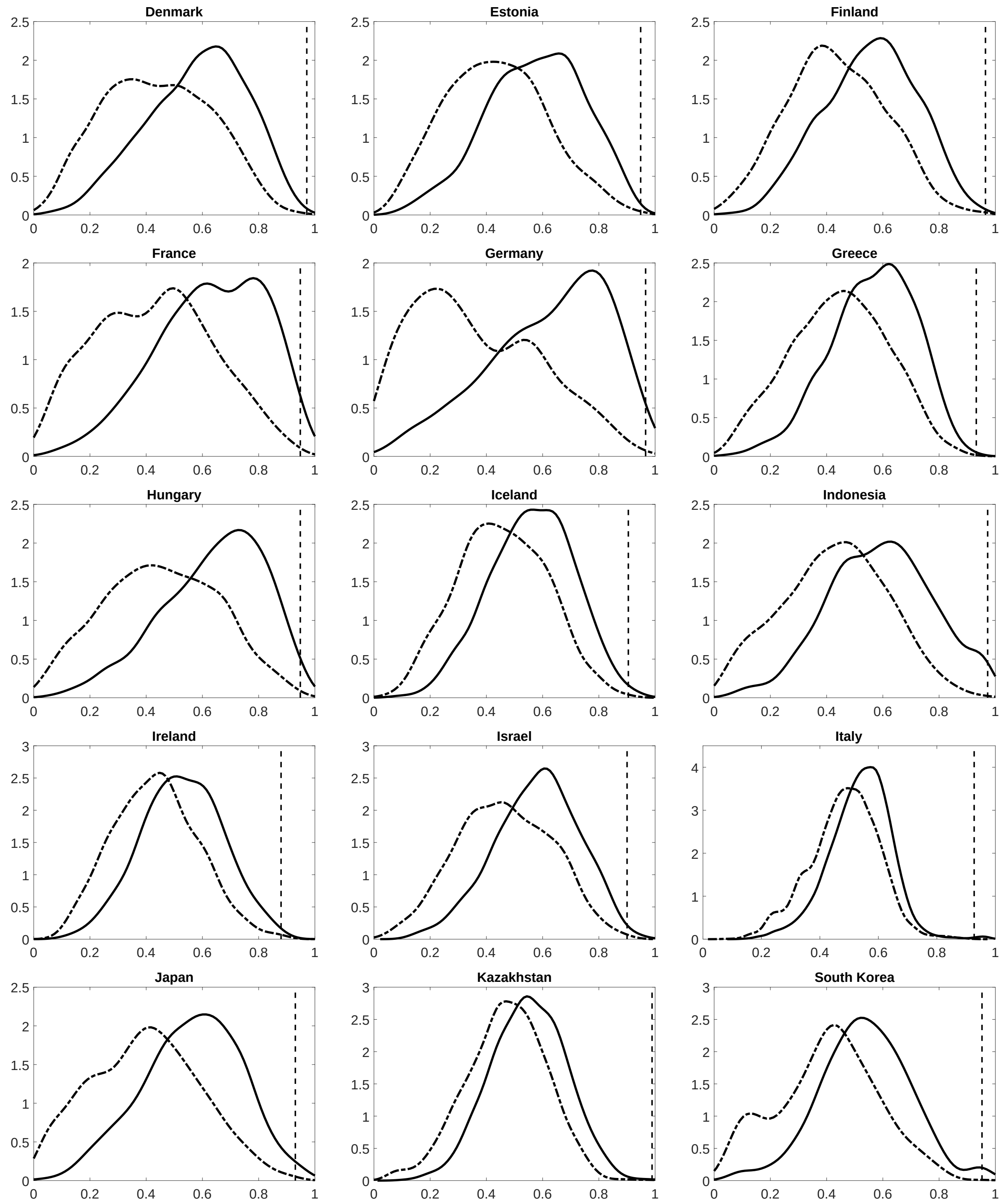

Figure A.1 (continued).

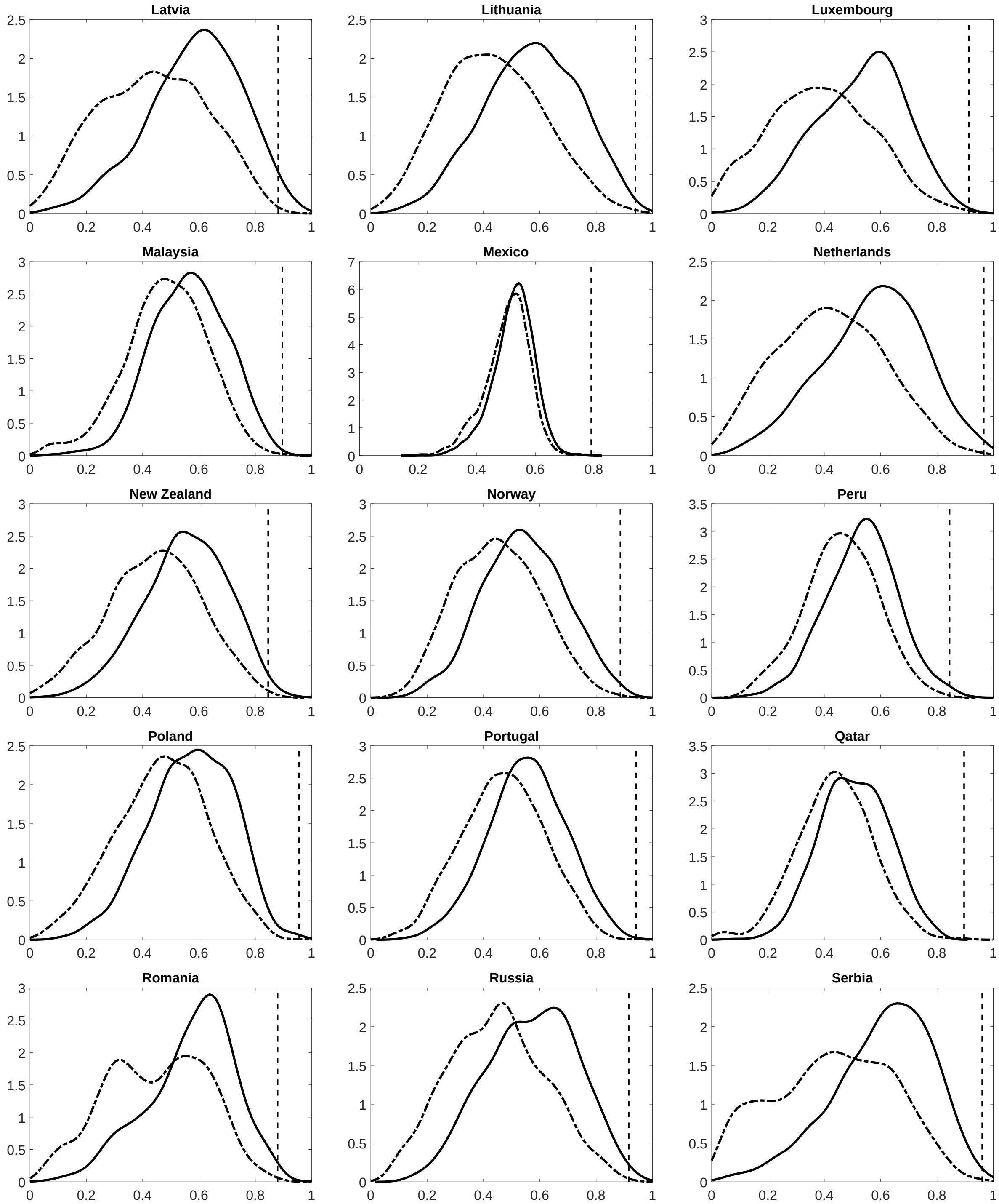

Figure A.1 (continued).

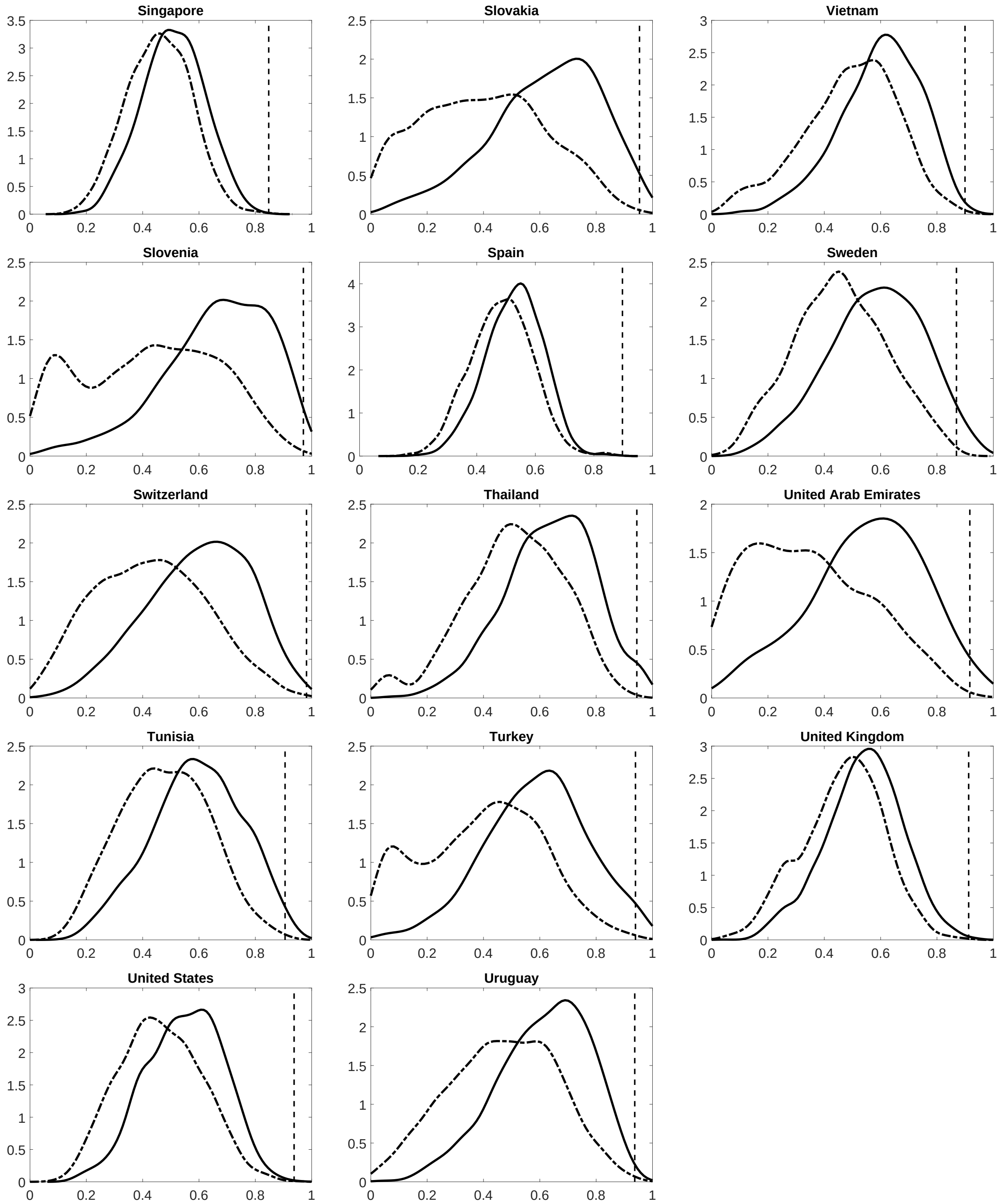

Figure A.1 (continued).