

A Work Project presented as part of the requirements for the Award of a Master's degree in  
Economics from the Nova School of Business and Economics.

REGIONAL DISPARITIES IN SOCIOECONOMIC-BASED  
GENDER ACHIEVEMENT GAPS

SARA MADEIRA CAL

Work project carried out under the supervision of:

Luís Catela Nunes

Pedro Freitas

16-12-2022

**Abstract:**

We define Socioeconomic-Based Gender Achievement Gap as the difference between females' and males' achievement conditional on socioeconomic background and present estimates for 278 municipalities in Portugal. Achievement is based on national exams scores in Portuguese Language and Mathematics. When considering the proportion of students who achieve a passing score, the pooled subjects gaps are consistently larger for students from less privileged backgrounds. Regardless of socioeconomic background, the gaps tend to be larger in higher grades and when considering Portuguese Language, and smaller in Mathematics. We examine regional disparities in the gaps and explore their association with regional factors. The gaps are, on average, larger in more religious, less densely populated and less educated municipalities, and in municipalities where the wage difference between males and females is lower.

**Keywords:** Educational Achievement; Gender Achievement Gaps; Regional Disparities; Socioeconomic Background

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209). We are thankful to Direção-Geral de Estatísticas da Educação e Ciência (DGEEC) and Gabinete de Estratégia e Planeamento (GEP) for providing access to the data used in this study.

---

**Acknowledgements:** I would like to thank first and foremost Professor Luís Catela Nunes and Pedro Freitas for their time, guidance, and comments which enormously improved this work. I would also like to thank Miguel Nunes, whose research was fundamental to this thesis. Finally, I want to express my greatest thanks to my parents for their unconditional support.

## 1. Introduction

Gender differences in the labour market have been widely studied in the Literature. Studies report that women earn, on average, less than men per hour: in 2021, the Gender Pay Gap in OECD countries was -12% (OECD 2021). Moreover, women are underrepresented in the labour market, particularly in STEM<sup>2</sup> areas and in decision-making positions, and overrepresented in part-time jobs (Kahn and Ginther 2017; EIGE 2022). Combining employment and wage effects, we observe that, in 2019, the global female labor income share was 35.1% (Neef and Robilliard 2021).

However, the gender pattern in educational outcomes is quite different, as Literature reports females tend to outperform their male peers, particularly in Language courses. This pattern has been widely documented in several countries (e.g. Buchmann, DiPrete, and McDaniel 2008; OECD 2015; Delaney and Devereux 2021). Besides gender, students' socioeconomic background has also been identified, by a large body of Literature, as an important determinant of educational outcomes (e.g. Coleman 1966, Carneiro 2008, Hanushek et al. 2019). However, few studies have focused on how the crossing between gender and socioeconomic inequalities explains educational outcomes.

In this paper, we study how gender differences in educational achievement vary with students' socioeconomic background, at the national and regional levels, in Portugal. For that purpose, standardized national exams data are combined with student-level socioeconomic data.

We define Socioeconomic-Based<sup>3</sup> Gender Achievement Gap as the difference between females' achievement and males' achievement, conditional on socioeconomic background. We find that, when considering the proportion of students who achieve a passing grade, female-

---

<sup>2</sup> STEM is an abbreviation for "Science, Technology, Engineering and Mathematics".

<sup>3</sup> Henceforth, referred to as "SE".

favouring SE-Based Gender Achievement Gaps are consistently larger for students from a less privileged background, in a pooled subjects analysis.

We also report that, in the subject Portuguese Language and in the pooled subjects analysis, girls outperform boys in the majority of the municipalities and at the national level, and that this is the case for students from all socioeconomic backgrounds. Moreover, we find that the size of the SE-Based Gender Achievement Gaps tends to be larger in higher grades. In Mathematics, in the 4<sup>th</sup> grade, regardless of socioeconomic background, boys outperform girls in most municipalities and at the national level, but, in higher grades, this pattern is reversed.

When mapping the SE-Based Gender Achievement Gaps, we report a sizable variation across municipalities, and no clear geographical pattern is observed. Finally, we correlate the SE-Based Gender Achievement Gaps at the regional level with several municipality-level characteristics: the population's education and income, populational density, political outcomes, cultural habits, and gender differences in the labour market.

The remainder of the paper is organized as follows. Section 2 provides a literature review on gender differences in educational performance. Section 3 briefly describes the main datasets used in this study. Section 4 explains the methodological approach. Section 5 reports the SE-Based Gender Achievement Gaps at the national and municipality levels, characterizes the municipalities according to such measures and discusses the results. Section 6 concludes.

## **2. Literature Review**

Literature has identified socioeconomic background as an important determinant of educational outcomes (Coleman et al. 1966; Leibowitz 1977; Becker 1998). Differences in parental investment driven by inequality in parental inputs are recognized as a major source of educational inequality, leading students from more privileged social groups to have, on average, better performance at school.

Some authors have found variations across countries and within countries in the association between socioeconomic background and educational outcomes (Blanden, Doepke, and Stuhler 2022). Using national exam grades from Portugal, Nunes et al. (2021) computed socioeconomic-based achievement measures and documented how educational inequality by family socioeconomic conditions varies across municipalities in Portugal. These results are consistent with the Literature that suggests that the place where children are raised and attend school shapes educational outcomes, at the neighbourhood (Chetty, Hendren, and Katz 2016; Chetty and Hendren 2018a; Chetty and Hendren 2018b; List, Momeni, and Zenou 2020) and regional levels.

Gender has also been recognized as an important determinant of educational performance, as various studies report gender differences in educational outcomes in a wide number of countries (Buchmann, DiPrete, and McDaniel 2008; OECD 2015; Delaney and Devereux 2021). Portugal is no exception and various studies provide evidence of gendered patterns in educational outcomes (e.g. Fernandes et al. 2018). National reports suggest that female students outperform males in most subjects and grades. Using 2014/2015 academic year data, Pereira and Baptista (2017) report that the female passing rate is higher for all subjects except Physical Education. In a report by DGEEC (Direção Geral de Estatísticas de Educação e Ciência) and JNE (Júri Nacional de Exames) (2018), it is estimated that the percentage of female students who conclude Lower Secondary Education within the expected time and achieve a passing score in the 9<sup>th</sup> grade exams is 10 p.p. higher in comparison to male students, both in 2015/2016 and 2016/2017 academic years.

When considering the subject Portuguese Language, the proportion of females who achieve a “good” or “very good” score is significantly higher than the proportion of male students. Pereira and Baptista (2017) estimate, using teacher scores from the 2014/2015 academic year, that the gender gap in passing rates of Portuguese Language is approximately 8 p.p. in the 9<sup>th</sup>

grade. The same study reports no significant gender gap in Mathematics in the 9<sup>th</sup> grade, illustrating a widely documented pattern that the relative performance of females is better in Languages and Humanities than in Mathematics (e.g. OECD 2015). In line with these results, in PISA 2018, 15-year-old female students scored, on average, on the PISA scale, more 24 points on the Reading test and less 9 points on the Mathematics test, when compared to their male peers (Lourenço et al. 2019).

Another branch of Literature has studied the mechanism leading to Gender Achievement Gaps. Some observational and interview-based studies suggest that gender stereotypes are reflected in a more detached attitude of boys toward school, as some report that they feel it is “inappropriate” to show interest in school, particularly in reading (Muntoni, Wagner, and Retelsdorf 2000; Warrington, Younger, and Williams 2000). There is indeed evidence that there is a gender difference in engagement at school and attitudes toward learning. Using 2012 data, it was estimated that, in Portugal, 15-year-old boys spend less 1.2 hours per week than girls doing homework and that boys are more likely to arrive late for school (OECD 2015). In the HBSC (Health Behaviour in School-aged Children) study, which is an international study on the health and well-being of adolescents across countries in Europe and North America, 76.8% of the female respondents reports enjoying school, compared to 69.0% of the male respondents (Matos et al. 2015). Additionally, studies have recognized that gender stereotypes influence females’ self-perception about their competencies in Mathematics, while recognizing that one important factor that explains girls’ underperformance in Mathematics is the lack of confidence in their own abilities in that subject (e.g. OECD 2015; Ganley and Lubienski 2016). Differences in parental expectations and investments depending on the child’s sex are also identified as a factor that explains the Gender Achievement Gap (e.g. Warrington and Younger 2000; OECD 2015; Lei and Lundberg 2020).

Other reasons have been brought up by scholars to explain gender gaps in achievement, namely higher return to education for females (Becker, Hubbard, and Murphy 2010), feminization of the teaching profession (e.g. Carrington and McPhee 2008; Drudy 2008), and biological differences (e.g. Geist and King 2008).

Some authors have studied how socioeconomic background interacts with gender to explain educational outcomes. Studies using US data show that gender differences in achievement are more pronounced amongst children from families with lower socioeconomic status (DiPrete and Buchmann 2013; Reardon et al. 2019; Autor et al. 2020; Lei and Lundberg 2020). OECD (2015) finds no significant difference between Gender Achievement Gaps among disadvantaged and advantaged households but estimates that males may be at a particular disadvantage when their parents have low levels of education in OECD countries.

Other authors have studied how Gender Achievement Gaps vary at the regional level. Studies using data from the US suggest that Gender Achievement Gaps computed at the national level may hide statistically significant disparities across regions. Hyde and Mertz (2009) and Pope and Sydnor (2010) report state-level gender gaps and Reardon and coauthors (2019) estimate and describe geographical patterns in Gender Achievement Gaps at the school-district level, using data on Math and English Language test scores. The two latter papers suggest that variation in gender educational outcomes across regions may be partly explained by the extent to which gender stereotypes prevail in a community and, more broadly, to the extent to which gender disparities in various areas of life exist in that region. Thereby, Reardon et al. (2019) explore, for approximately 10,000 school districts, the association between the Gender Achievement Gaps and local gender disparities. They find that male students' performance in Math relative to their female peers is better in socioeconomically advantaged regions and in regions where gender disparities are more pronounced.

Other authors have found evidence of an association between Gender Achievement Gaps and national gender disparities (Baker and Jones 1993; San Román and Rica 2012). Using 2003 PISA data, Guiso et al. (2008) find that the gender gap in Mathematics scores disappears in more gender-equal societies, measured by the Gender Gap Index. Similarly, OECD (2015) finds, using PISA 2009 data, that female students perform better in both Mathematics and Reading in more gender-equal societies.

### 3. Data

In this paper, we focus our analysis on the period of 2007-2018<sup>4</sup> and on the 278 municipalities in mainland Portugal. We combine four data sources: standardized national exams data from *JNE*; student-level socioeconomic data of students attending public schools from *MISI*; *PISA* scores and student-level socioeconomic data from *PISA*; and municipality-level demographic and socioeconomic data from *INE*. All student-level data are anonymized. The subsections below detail these datasets.

#### 3.1. National Exams Scores Data

Data on standardized national exams scores are provided by *Júri Nacional de Exames* (JNE), which is the entity responsible for the organization and implementation of the national exams in Portugal. This database, managed by the Portuguese Ministry of Education, includes students' scores on national exams, as well as their gender, age, and school attended.

The Portuguese Educational system is divided into four cycles (Appendix Table 1) and, since 2012, education is compulsory until the 12<sup>th</sup> grade or until the age of 18. We limit our analysis to the compulsory exam-year grades, that is, the 4<sup>th</sup>, 6<sup>th</sup>, and 9<sup>th</sup> grades.<sup>5</sup> Until the academic year of 2014/2015, students of the 4<sup>th</sup>, 6<sup>th</sup>, and 9<sup>th</sup> grades performed national exams for Mathematics

---

<sup>4</sup> We exclude data on the 2006-2007 academic year, because of misreporting issues on the socioeconomic data.

<sup>5</sup> We focus on these grades, since, after the 9<sup>th</sup> grade, students may choose to follow a vocational track where national exams are not required. In the academic year of 2020/2021, 40% of the students in Upper Secondary Education chose such track (DGEEC 2022).

and Portuguese Language. Thereafter, there are only exam scores data regarding the 9<sup>th</sup> grade, as the national exams for 4<sup>th</sup> and 6<sup>th</sup> grades were abolished (Appendix Tables 2 and 3).

National exams scores are reported on a scale of 1 to 5 for the 4<sup>th</sup> and 6<sup>th</sup> grades and on a scale of 0 to 100 for the 9<sup>th</sup> grade. Based on the conversion criteria of the Portuguese Ministry of Education<sup>6</sup>, we convert the 9<sup>th</sup> grade national exams to a scale from 1 to 5, so that these scores are comparable with those from the 4<sup>th</sup> and 6<sup>th</sup> grades. We keep the exam scores that correspond to the first time students are taking that exam, for each subject and grade, to avoid the bias that could arise if double counting repeaters scores.

### **3.2 Students' Socioeconomic Data**

Data at the individual level on the socioeconomic characteristics of students who were enrolled in the Portuguese Public Education system is provided by MISI. MISI is an administrative dataset managed by DGEEC<sup>7</sup> that also includes information on students' area of residence and the school in which they are enrolled.

Additionally, the index of students' socioeconomic conditions makes use of 2009, 2012, 2015 and 2018 data from *PISA*, which is the OECD's Programme for International Student Assessment. *PISA* contains data on the socioeconomic characteristics and scores of a representative sample of fifteen-year-old Portuguese students that took the *PISA*'s tests on Reading, Mathematics, and Science.

### **3.3 Municipalities' Socioeconomic and Demographic Data**

Socioeconomic and demographic data at the municipality level is provided by *Statistics Portugal (INE)*, which is a central body dedicated to the production, analysis, and diffusion of official statistics. The majority of the municipalities' socioeconomic and demographic data was

---

<sup>6</sup> Score 1-2 (failing score): score less than 50; Score 3: between 50 and 69; Score 4: between 70 and 89; Score 5: greater or equal to 90.

<sup>7</sup> DGEEC (Direção Geral de Estatísticas de Educação e Ciência) is a public body dedicated to the production and analysis of statistical content regarding Education and Science, in Portugal.

collected in the context of the Portuguese Census, which is a survey answered by all Portuguese population. This survey is conducted every ten years and the results for 2001, 2011, and 2021 are available for all the chosen variables. As *JNE* and *MISI* datasets regard 2007-2018, we use 2011 data, as this is the only Census year included in the referred timeframe. We highlight that the chosen sociodemographic variables, due to their nature, are unlikely to present abrupt changes over the considered timeframe.

## **4. Methodological Approach**

### **4.1 Achievement Outcomes**

In our empirical analysis, to understand how Socioeconomic-Based Gender Achievement Gaps change across achievement levels, we consider two score thresholds: (i) achieving a score higher than 2, that is, achieving a passing score and (ii) achieving a score higher than 3, scores that are “good” or “very good” scores<sup>8</sup>.

### **4.2 Socioeconomic Index**

In order to quantify students’ socioeconomic condition, we use the Socioeconomic (SE) Index suggested by Nunes et al. (2021). We choose to aggregate various socioeconomic characteristics rather than using a single one, to have a more complete characterization of the socioeconomic status.

This index is based on a first principal component from a factor analysis of student-level socioeconomic data on parental education, professional situation and nationality, eligibility for free or reduced-price lunch, and computer possession. The SE index, which uses the factor loadings derived using middle-year data<sup>9</sup>, is computed for each grade individually, thereby allowing for comparability across time and municipalities.

---

<sup>8</sup> The scores scale is presented in Decree-Law nr 176/2012 of August 2 of Ministry of Education and Science.

<sup>9</sup> We use the academic year 2013/2014 for the 9<sup>th</sup> grade and the academic year 2011/2012 for the 4<sup>th</sup> and 6<sup>th</sup> grades.

The Socioeconomic Index is computed for public school students with data from MISI, which provides various binary and categorical variables that allow us to access students' socioeconomic conditions. In MISI, student-level data on socioeconomic characteristics is not available for private school students. Thereby, it is established a correspondence between the PISA's socioeconomic index (ESCS) and the SE index computed previously. Then, using private school students' ESCS values, SE index values are randomly inputted for private school students, for each grade individually<sup>10</sup> (Appendix Figure 1).

### 4.3 SE-Based Gender Achievement Gaps

We estimate SE-Based Gender Achievement Gaps for the subjects Mathematics and Portuguese Language, (i) pooled and separately, (ii) for the 4<sup>th</sup>, 6<sup>th</sup> and 9<sup>th</sup> grades, (iii) for the students below or at the median of the SE index distribution (low-SE students) and for the students in the upper half of the SE index distribution (high-SE students)<sup>11</sup>, and (iv) for the two previously described score thresholds. We pool all the years with available data and consider the 278 municipalities in mainland Portugal. We compute the SE-Based Gender Achievement Gaps as follows:

$$GG_{crys} = Pfemale_{crys} - Pmale_{crys} \Leftrightarrow GG_{crys} = \frac{n_{fcrys|t}}{n_{fcrys}} - \frac{n_{mcrys|t}}{n_{mcrys}} \quad (1)$$

where  $Pfemale_{crys} = \frac{n_{fcrys|t}}{n_{fcrys}}$ ,  $Pmale_{crys} = \frac{n_{mcrys|t}}{n_{mcrys}}$ ,  $n_{gcrys|t}$  stands for the number of students of gender  $g$ , socioeconomic background  $c$ , who live in municipality  $r$ , in grade  $y$  and achieved a score higher than  $t$  in the national exam of subject  $s$ ; and  $g = \mathbf{f}, \mathbf{m}$ <sup>12</sup>;  $c = \mathbf{h}, \mathbf{l}$ <sup>13</sup>;  $r = \mathbf{1}, \mathbf{2}, \dots, \mathbf{278}$ <sup>14</sup>;  $y = \mathbf{4}, \mathbf{6}, \mathbf{9}$ ;  $t = \mathbf{2}, \mathbf{3}$ ;  $s = \mathbf{p}, \mathbf{m}$ <sup>15</sup>.

<sup>10</sup> For a more detailed description of the methodology used to estimate the SE Index, see Nunes et al. (2021).

<sup>11</sup> We consider the SE threshold to be the 50<sup>th</sup> percentile, rather than the 25<sup>th</sup> and 75<sup>th</sup> percentiles used in Nunes et al. (2021), due to the existence of a small number of observations in various municipalities when considering the latter percentiles.

<sup>12</sup>  $\mathbf{f}$  stands for female and  $\mathbf{m}$  stands for male.

<sup>13</sup>  $\mathbf{h}$  and  $\mathbf{l}$  stand for high and low economic status, respectively.

<sup>14</sup>  $\mathbf{r}$  denotes the municipalities in mainland Portugal.

<sup>15</sup>  $\mathbf{p}$  stands for Portuguese Language and  $\mathbf{m}$  stands for Mathematics.

#### 4.4 Regional Disparities in SE-Based Gender Achievement Gaps

In the second part, we aim to assess how regional disparities in SE-Based Gender Achievement Gaps relate to differences in regional characteristics. First, we regress the municipality-level SE-Based Gender Achievement Gaps on each socioeconomic and demographic variable measured at the municipality level, using simple linear regressions with robust standard errors. Then, we group the independent variables by fields and run a multiple linear regression with robust standard errors of municipality-level SE-Based Gender Achievement Gaps on each group of variables  $X_r$ . Descriptive statistics of these variables are presented in Appendix Table 4 and in Appendix Figures 5 to 8.

$$GG_r = \alpha + \beta X_r + \epsilon_r \quad (1)$$

We consider socioeconomic and demographic variables at the municipality level that may be grouped into five dimensions:

*Gender Differences in the Labour Market:* We use labour market data disaggregated by municipality and gender. We use the municipality's average gross monthly wage of full-time workers employed in the private sector and workers who have individual contracts of employment in the central, regional, and local administration and public institutes. The unemployment rate is defined as the number of people who reported being unemployed in the 2011 Census divided by the active population. The inactivity rate is computed as the number of people who are older than 15 years old and who reported not being employed nor unemployed in the 2011 Census divided by the working-age population. We aim to assess how gender differences in the labour market correlate with the SE-Based Gender Achievement Gaps. Thereby, we compute, for each municipality, the difference between (i) females' mean wage and males' mean wage, (ii) female unemployment rate and male unemployment rate, and (iii) female inactivity rate and male inactivity rate.

*Education and Income:* We include two variables regarding education: one measures the population education level, while the other measures the gender difference in education. The former is defined as the number of people without Upper Secondary Education who are aged 15 or older divided by the number of people aged 15 or older. The latter is computed as the difference between the proportion of women without Upper Secondary Education and the proportion of men without Upper Secondary Education. Data was collected in Census and regards the year 2011. We also include in our analysis the average wage per municipality in 2011.

*Election Outcomes:* We add to our analysis the abstention rate in the elections for the National Parliament of 2011, defined by the number of votes divided by the number of registered voters, and the proportion of votes in left-wing parties<sup>16</sup>, which is computed as the ratio between votes in left-wing parties divided by the total number of votes.

*Populational Density:* We consider the population density of each municipality in 2011, which is defined as the number of inhabitants divided by the land area.

*Cultural Habits and Social Norms:* Using 2011 Census data, we compute the degree of religiosity of each municipality, by defining a variable as the number of people who identify as “religious” divided by the total population, and the mean family size, which is computed as the total population divided by the number of households.

We also include the relative number of marriages per municipality in 2011, which is defined by the number of marriages divided by the number of people aged between 20 and 44. In this indicator, we do not consider the municipalities’ total population but, instead, a narrower age interval, to avoid biased results in municipalities characterized by population ageing. According

---

<sup>16</sup> In the considered timeframe, the left-wing parties with seats in the national parliament in Portugal were the Socialist Party (PS), the Left Bloc (BE), and the Portuguese Communist Party (PCP) (Carvalho 2020).

to *Statistics Portugal*, in 2011, the average age at the first wedding was 29,5 for females and 31,1 for males and we choose an age interval wide enough to avoid excluding relevant observations. We also include a variable for divorces, which is computed by dividing the number of divorces in 2011 by the number of weddings.

To capture the degree to which gender stereotypes are embedded in municipalities, we also consider the degree to which fathers request paternity leave and pay. We define this variable by the sum of the amounts of the first transfer of parental pay divided by the number of live births. As no data is available for 2011, we use 2014 data, which is the closest year available data.

## **5. Results**

In the first part of the empirical analysis, we estimate the SE-Based Gender Achievement Gaps at the national and municipality levels, separately for low-SE and high-SE students. Using 4<sup>th</sup>, 6<sup>th</sup>, and 9<sup>th</sup> grade students' data, we provide estimates that consider each subject separately and estimates for pooled subjects, for the score thresholds "Passing Score" and "Good or Very Good Score".

In the second part, where we aim to explore which municipalities' characteristics are associated with regional disparities in gender achievement, we regress the SE-Based Gender Achievement Gaps on socioeconomic and demographic regional characteristics.

### **5.1.1 SE-Based Gender Achievement Gap estimates at the National level**

Following the methodology described in the previous section, we compute the SE-Based Gender Achievement Gaps, at the national level (excluding Portuguese islands). Our results for mainland Portugal align with the cited literature on Gender Achievement Gaps.

We first present the results of the pooled subjects analysis. When considering both subjects simultaneously (Table 1), we observe that, for both low-SE and high-SE students, the SE-Based

Gender Achievement Gap is always positive, that is, on average, female students have a better performance when compared to males, for all grades, and for both score thresholds. Moreover, we observe that the magnitude of the SE-Based Gender Achievement Gaps increases with the grade, in all cases except one (*GG<sub>193</sub>*). Indeed, in the 4<sup>th</sup> grade, the SE-Based Gender Achievement Gaps have a modest size, varying between 0.5 p.p.<sup>17</sup> and 1.7 p.p., depending on the threshold and SE background. In the 9<sup>th</sup> grade, the magnitude of the SE-Based Gender Achievement Gaps is 3 to 9 times larger, ranging between 5.0 p.p. and 7.6 p.p.

Still considering the two subjects, we observe that socioeconomic background seems to affect the Gender Achievement Gaps: when considering the probability of achieving a passing score, the SE-Based Gender Achievement Gaps are larger for low-SE students. However, when the threshold chosen is achieving a good or very good score (score higher than 3), the SE-Based Gender Achievement Gaps are larger for high-SE students in the 6<sup>th</sup> and 9<sup>th</sup> grades.

Similar conclusions may be drawn from the analysis of results for the subject Portuguese Language (Table 2). The absolute performance of female students improves in all the considered cases and the magnitude of the SE-Based Gender Achievement Gaps is considerably higher, being from 1.7 to 5.2 times larger than the size of the gaps registered in the pooled subjects analysis. As in the subject pooled results, the SE-Based Gender Achievement Gaps increase with grade and are affected by students' socioeconomic background.

The results of the analysis that considers solely the scores of the subject Mathematics present different statistical patterns from the ones described previously (Table 3). In the 4<sup>th</sup> grade, male students outperform female students, for all score thresholds and for both high-SE and low-SE students. As the grade increases, the performance of girls in Mathematics relative to boys improves, and, from the 6<sup>th</sup> grade on, the SE-Based Gender Achievement Gaps become positive,

---

<sup>17</sup> Our percentage points refer to the probabilities of obtaining a passing score or a good or very good score.

except in one case ( $GG_{h63m}$ ). No clear pattern arises regarding how the SE-Based Gender Achievement Gap differs with changes in the score thresholds and socioeconomic conditions.

It is important to highlight that, although gender seems to play a relevant role in the determination of students' achievement, the socioeconomic condition seems to be a much stronger determinant of the students' performance. For all grades, for both genders and score thresholds, students from the bottom 50% of the SE index distribution perform worse than the upper 50%. In the pooled subjects analysis, the difference between the achievement measures of high-SE and low-SE students ranges from 12.9 p.p. to 25.9 p.p. for females and from 14.0 p.p. and 25.6 p.p., for males (Table 1).

### **5.1.2 SE-Based Gender Achievement Gap estimates at the Regional level**

In this subsection, we first report, at the municipality level, the proportion of students achieving a grade higher than 2 and higher than 3, by gender and by socioeconomic condition ( $P_{female_{cryts}}$  and  $P_{male_{cryts}}$ ). We present these results in choropleth maps (Appendix Figure 2). We divide the observations into seven quantiles<sup>18</sup> and the regions belonging to the lowest septile, where the lowest values of the achievement measures are reported, are colored in red.

In line with the findings of Nunes et al. (2021), we observe a pattern in these choropleth maps: there seems to be a geographical concentration of municipalities that belong to the lowest septile of the socioeconomic-based achievement measure distribution around *Área Metropolitana de Lisboa, Algarve e Alentejo*. This is the case for female and male students and for both low-SE and high-SE students, although, when considering low-SE students, some municipalities in Northern Portugal also present low values of the socioeconomic-based achievement measure.

---

<sup>18</sup> We divide the observations into 7 quantiles, to allow for comparability with the paper by Nunes et al. (2021).

In Figure 1, we map the SE-Based Gender Achievement Gaps at the municipality level, using both subjects' scores. Following the same methodology, we divide the observations into seven quantiles and the municipalities where the municipalities with higher female-favouring SE-Based Gender Achievement Gaps are colored in blue.

We observe in Figure 1 that, in most of the municipalities, female students outperform males and this is true for low-SE and high-SE students (Table 4). As the grade increases, the number of municipalities where females outperform males never decreases, for all score thresholds and for both low-SE and high-SE students, in the pooled subjects analysis.

Contrarily to Nunes et al. (2021), who identified a geographic pattern regarding the SE-Based Achievement Gap in Portugal, we do not find a clear spatial pattern of SE-Based Gender Achievement Gaps, neither at the municipality nor at the NUTS 3 level (Appendix Figure 3). We report, however, a sizeable variability in the magnitude of such gaps across municipalities. The difference between the highest and the lowest value of the SE-Based Gender Achievement Gap across municipalities lies between 22,6 p.p. and 30.8 p.p.

When comparing the choropleth maps for low-SE students and high-SE students, we observe differences, but we do not find a consistent geographical pattern. However, in line with the results at the national level, when focusing on the proportion of students with passing scores, we find that the number of municipalities with a positive SE-Based Gender Achievement Gap is higher when considering the low-SE students (Table 4). There is not a consistent pattern regarding the other score threshold.

In Appendix Figure 4, choropleth maps by subject are shown and, when compared to the subject pooled results, we find that the number of municipalities where females perform better than males is higher when considering solely the subject Portuguese Language and lower when considering only the subject Mathematics, for both low-SE and high-SE students and for all

score thresholds and grades (Table 4). Mathematics and Portuguese SE-Based Gender Achievement Gaps at the municipality level present a positive correlation that lies between 0.23 and 0.57, values substantially lower than the ones found by Reardon et al. (2019), using a similar methodology.

## **5.2. Regional Correlates of SE-Based Gender Achievement Gap**

In the second part of the empirical analysis, we use solely educational data of 9<sup>th</sup> graders, because: (i) As shown in the previous subsections, there is consistent evidence, both at the national and municipality levels, of the general trend of SE-Based Gender Achievement Gaps widening as grade increases. Thus, the magnitudes of the SE-Based Gender Achievement Gaps in the 9<sup>th</sup> grade suggest that increasing our knowledge about this disparity would be particularly relevant; (ii) As learning is a cumulative process, 9<sup>th</sup> grade SE-Based Gender Achievement Gaps are likely to capture more persistent patterns; (iii) It is closer to the moment of decision of the track of studies in the Upper Secondary Education.

We choose municipal variables that the literature has associated with gender inequality, as well as other demographic and socioeconomic variables that characterize the regions. We, then, estimate equation (1) using standardized variables and, thereby, coefficients shall be interpreted as follows: an increase of 1 standard deviation in the sociodemographic variable  $X$  is associated with an increase of  $\beta$  standard deviations in the SE-Based Gender Achievement Gap. We report the estimated coefficients in Figure 2 (and in Appendix Tables 5 to 10) and highlight that causality cannot be inferred from this analysis. We focus on the results of the score threshold “Passing Score”, but similar conclusions shall be drawn from the other score threshold results (Appendix Figure 11).

*Education and Income:* We start by analyzing the association of SE-Based Gender Achievement Gaps with populations’ literacy and income. We find that, in the single (multiple) linear regression, a decrease of one SD in the percentage of the population without a high school

diploma is associated, for low-SE and high-SE students, with a 0.37 SD (0.24 SD) and 0.22 SD (0.28 SD) decrease in the SE-Based Gender Achievement Gap, respectively. This result is not surprising, as social and non-market benefits from education have been widely recognized in literature and, particularly, its contribution to human rights (Haveman and Wolfe 2002; McMahon 2010).

The mean salary in the single linear regression analysis always presents negative coefficients (-0.35 for low SE and -0.16 for high SE students), as well as in the multiple linear regression for low-SE students, which suggests that, *ceteris paribus*, richer regions have lower SE-Based Gender Achievement Gaps. In the multiple linear regression, the coefficient associated with high-SE students is positive (0.04), although it is not statistically significant.

The coefficient associated with the difference between the percentage of females without Upper Secondary Education and the percentage of males without Upper Secondary Education is negative in both regressions and for both high-SE and low-SE students, implying that, on average, in regions where women are less educated when compared to men, girls' performance in national exams worsens, when compared to boys'. However, the coefficient is not statistically significant at 5% significance level in the multiple linear regression analysis.

*Populational Density:* We also find, in line with the geographical patterns observed in the choropleth maps, that the coefficients associated with populational density, which are -0.19 and -0.10 for low-SE and high-SE students, respectively, are negative. This suggests that SE-Based Gender Achievement Gaps are larger in rural areas than in urban areas.

*Gender Differences in the Labour Market:* When analyzing the results from the simple linear regression, we observe that the gender difference in the Unemployment Rates and the gender difference in Inactivity Rates have positive coefficients, for low-SE and high-SE students. This suggests that there may be higher SE-Based Gender Achievement Gaps in regions where there

is more gender inequality in the labour market. In the multiple linear regression, the latter variable is not significant, while the former is statistically significant at a 10% significance level. We also consider a variable for Gender Pay Gap (defined as females' mean earnings minus males' mean earnings), which presents a positive coefficient in the single linear regressions, as well as in the multiple linear regression for low-SE students (0.24). This result suggests that, in regions where women's mean wage (in comparison to men's) is higher, girls' performance improves (in comparison to boys').

*Cultural Habits and Gender Roles:* In the second to last group of variables, we include variables that aim to capture the regions' degree of conservatism and commitment to traditional values and gender roles. All five variables are statistically significant in the simple linear regression, except the Divorce Rate, which is not significant when considering the high-SE students. We find that religiosity, family size, and the number of marriages are positively associated with SE-Based Gender Achievement Gaps, while divorces and paternity leave variables are found to have a negative association with the dependent variable. In the multiple linear regression, the religiosity variable is statistically significant, and we find that 1 SD increase in the proportion of religious population is associated with an increase in the SE-Based Gender Achievement Gap of 0.39 SD and 0.11 SD, for low-SE students and high-SE students, respectively.

Moreover, in the multiple linear regression, for high-SE students, the paternity leave variable has a statistically significant negative coefficient (-0.17). The sign of the coefficient is consistent with the results of Farré et al. (2022), who find evidence that the introduction of paternity leave in Spain led children to display more gender-egalitarian norms.

*Election Outcomes:* In the single (multiple) linear regression, the association between the left-wing parties voting share and the SE-Based Gender Achievement Gap of low-SE students is statistically significant and we estimate that a 1 SD increase in the voting share of left-wing parties is associated with 0,19 SD (0,18 SD) decrease in the SE-Based Gender Achievement

Gap for low-SE students. This sign is consonant with the papers reviewed under Section 2, since, as left parties are more associated with the defense of equality between women and men (PPDM 2019), it is likely that, in regions with a high share of left voters, gender stereotypes are less prevalent. We also find that the abstention rate is statistically significant when considering high-SE students, in both regressions. In the multiple linear regression analysis, a 1 SD increase in abstention rate is associated with a 0.24 SD increase in the SE-Based Gender Achievement Gap.

When comparing the results from low-SE and high-SE students, we observe that the signs of all coefficients are the same with only two exceptions. However, as it is possible to observe in Figure 2, the magnitudes are, in general, higher in absolute terms for the low-SE students. The results of the second part of the analysis by subject are presented in Appendix Figures 9 and 10. Except for two, all the coefficients whose sign differs from the pooled analysis are not statistically significant.

## **6. Conclusion**

This paper provides a cross-sectional analysis of achievement in standardized national exams in Portugal by gender and socioeconomic background. We report SE-Based Gender Achievement Gaps at the national and municipality levels, considering two different achievement measures (achieving (i) a passing score and (ii) a “good” or “very good” score). Below, we summarize our main findings. We find that, independently of socioeconomic background, on average, females tend to outperform males, when considering Mathematics and Portuguese Language together and the latter separately. Females’ gender-relative performance improves with grade and, in the 9<sup>th</sup> grade, the Mathematics SE-Based Gender Achievement Gap becomes positive for the majority of the municipalities. We find that, when considering the proportion of students who achieve a passing grade, SE-Based Gender Achievement Gaps are consistently larger for students from a less privileged background, in a pooled subjects analysis.

These findings suggest that boys, particularly those from a more disadvantaged socioeconomic background, may not be achieving their full potential in school, which has negative consequences for their future, but also for the whole society (OECD 2015). Girls' underperformance in Mathematics in the earlier grades seems to be consistent with a pattern that later emerges in the labour market, where women are underrepresented in STEM areas (ILO 2021).

In the regional analysis, when considering indicators for educational performance for each gender, we observe a similar geographical pattern as the one described by Nunes et al. (2021). We also observe a sizable variation of SE-Based Gender Achievement Gaps across municipalities, although we do not observe a clear geographical pattern. Then, we explain the regional differences in SE-Based Gender Achievement Gaps in five dimensions: population's education and income, populational density, political outcomes, cultural habits, and gender differences in the labour market. We find that SE-based Gender Achievement Gaps have a positive association with religiosity and a negative association with populational density and education. Additionally, SE-Based Gender Achievement Gaps also tend to be higher in municipalities where the wage difference between males and females is lower.

Although these findings give us some insights into the characterization of regions with higher SE-Based Gender Achievement Gaps, we cannot identify the causal mechanism of such disparities. Another limitation of this study relates to the difficulty of obtaining accurate measures of regional cultural and social norms. Possible extensions of this study could consider (i) teacher scores (as grading systems based on exams tend to favor boys (Ângelo and Reis 2021), the computed SE-based Gender Achievement Gaps may be underestimated); (ii) the choice of track and achievement in Upper Secondary Education; and (iii) the enrolment and choice of studies in Higher Education. Further research on this topic should also focus on exploring exogenous local shocks, in order to estimate causal effects.

## Figures and Tables

Table 1: Pooled Subjects SE-Based Gender Achievement Gaps at the National Level

Grade		Passing Score ( $> 2$ )		Good or Very Good Score ( $> 3$ )	
		Low-SE	High-SE	Low-SE	High-SE
		4	Females	0.742	0.871
	Males	0.726	0.866	0.298	0.518
	Gender Gap	0.016	0.005	0.017	0.016
6	Females	0.681	0.854	0.223	0.481
	Males	0.627	0.824	0.167	0.423
	Gender Gap	0.054	0.030	0.055	0.058
9	Females	0.540	0.768	0.180	0.425
	Males	0.473	0.719	0.131	0.349
	Gender Gap	0.066	0.050	0.050	0.076

Notes: This table displays the educational performance of female and male students conditional on the socioeconomic background, and the SE-Based Achievement Gaps, for the 4<sup>th</sup>, 6<sup>th</sup> and 9<sup>th</sup> grades, considering the subjects Portuguese Language and Mathematics, at the national level. We consider two score thresholds: "Passing Score" ( $> 2$ ) and "Good or Very Good Score" ( $> 3$ ).

Table 2: Portuguese SE-Based Gender Achievement Gaps at the National Level

Grade		Passing Score ( $> 2$ )		Good or Very Good Score ( $> 3$ )	
		Low-SE	High-SE	Low-SE	High-SE
		4	Females	0.804	0.914
	Males	0.739	0.880	0.269	0.494
	Gender Gap	0.065	0.034	0.067	0.083
6	Females	0.799	0.919	0.274	0.532
	Males	0.703	0.865	0.167	0.406
	Gender Gap	0.096	0.054	0.107	0.126
9	Females	0.684	0.859	0.198	0.432
	Males	0.557	0.769	0.115	0.302
	Gender Gap	0.127	0.090	0.083	0.130

Notes: This table displays the educational performance of female and male students conditional on the socioeconomic background, and the SE-Based Achievement Gaps, for the 4<sup>th</sup>, 6<sup>th</sup> and 9<sup>th</sup> grades, considering the subject Portuguese Language, at the national level. We consider two score thresholds: "Passing Score" ( $> 2$ ) and "Good or Very Good Score" ( $> 3$ ).

Table 3: Mathematics SE-Based Gender Achievement Gaps at the National Level

Grade		Passing Score ( $> 2$ )		Good or Very Good Score ( $> 3$ )	
		Low-SE	High-SE	Low-SE	High-SE
		4	Females	0.680	0.828
	Males	0.713	0.852	0.328	0.541
	Gender Gap	-0.033	-0.024	-0.039	-0.051
6	Females	0.562	0.789	0.171	0.430
	Males	0.551	0.783	0.167	0.440
	Gender Gap	0.011	0.006	0.004	-0.010
9	Females	0.400	0.678	0.163	0.418
	Males	0.390	0.668	0.147	0.395
	Gender Gap	0.009	0.009	0.016	0.023

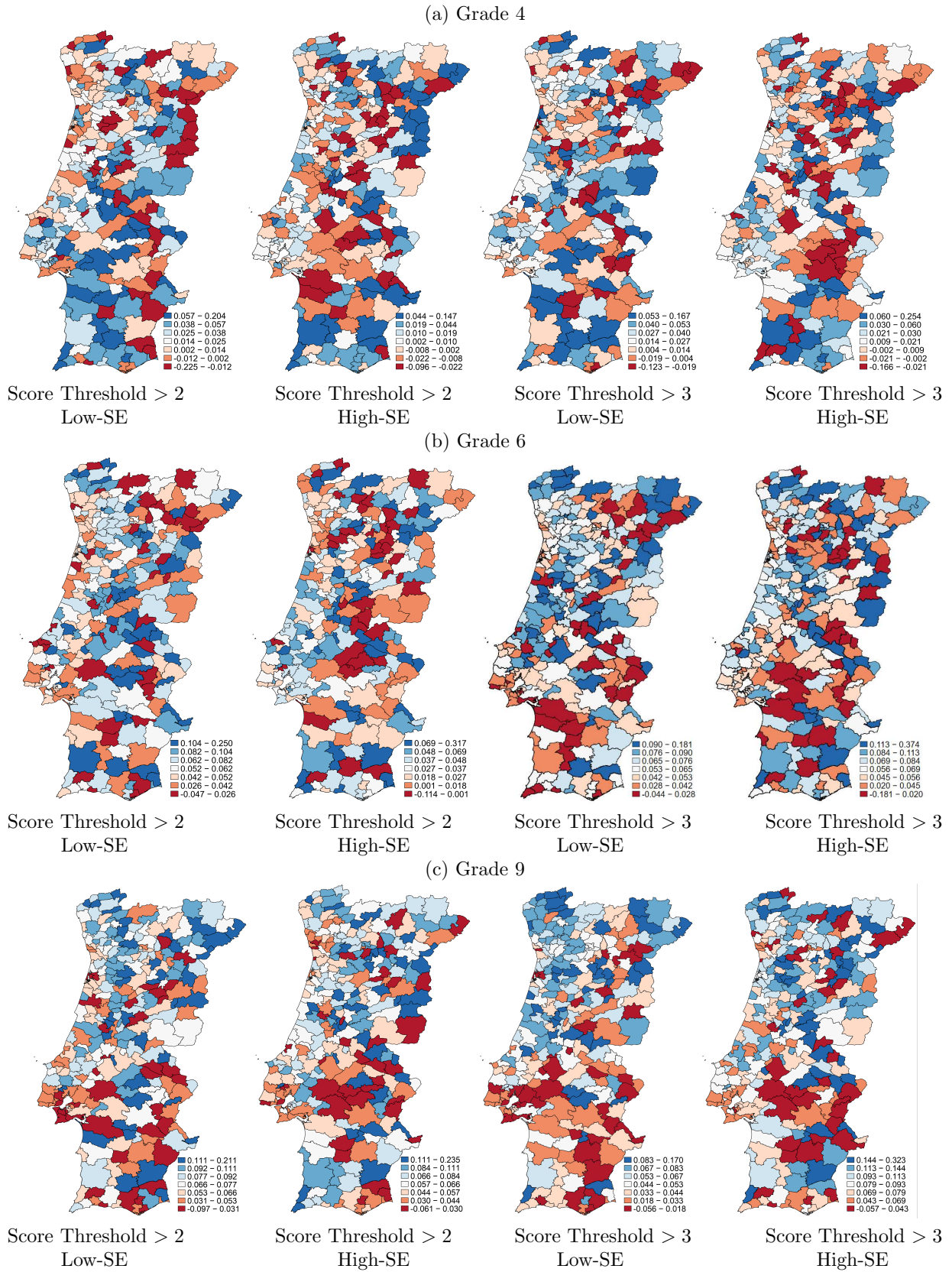
Notes: This table displays the educational performance of female and male students conditional on the socioeconomic background, and the SE-Based Achievement Gaps, for the 4<sup>th</sup>, 6<sup>th</sup> and 9<sup>th</sup> grades, considering the subject Mathematics, at the national level. We consider two score thresholds: "Passing Score" ( $> 2$ ) and "Good or Very Good Score" ( $> 3$ ).

Table 4: Percentage of Municipalities where SE-Based Gender Achievement Gap is positive

Grade		Passing Score ( $> 2$ )		Good or Very Good Score ( $> 3$ )	
		Low-SE	High-SE	Low-SE	High-SE
		4	Pooled	0.763	0.604
	Mathematics	0.245	0.219	0.205	0.158
	Portuguese Language	0.946	0.871	0.939	0.935
6	Pooled	0.968	0.860	0.957	0.935
	Mathematics	0.655	0.579	0.615	0.396
	Portuguese Language	0.989	0.942	0.986	0.989
9	Pooled	0.968	0.946	0.957	0.975
	Mathematics	0.550	0.673	0.694	0.773
	Portuguese Language	0.996	0.986	0.989	0.996

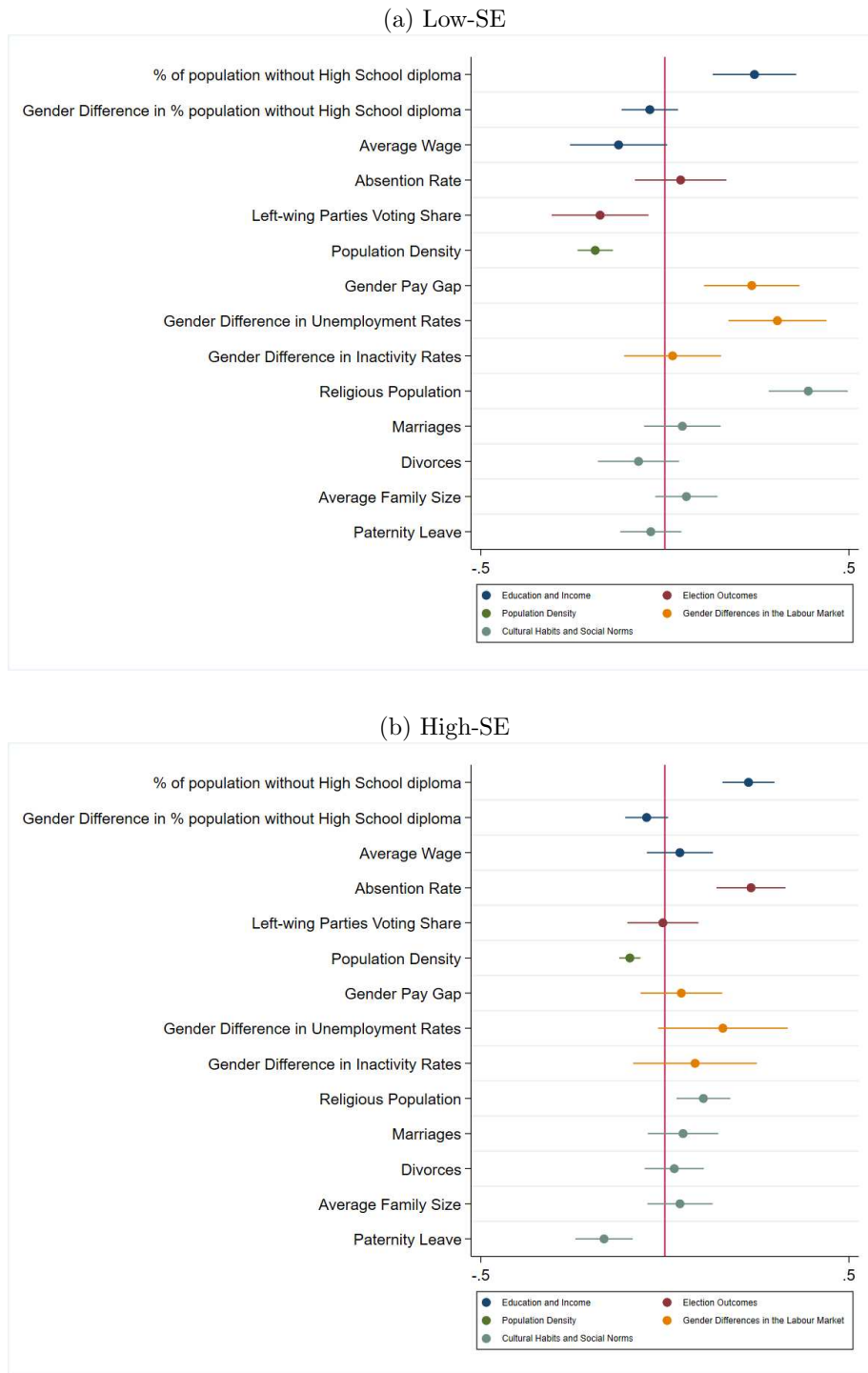
Notes: This table displays the percentage of municipalities where the SE-Based Gender Achievement Gap is positive, that is, where females outperform males, for the 4<sup>th</sup>, 6<sup>th</sup> and 9<sup>th</sup> grades, considering the each subject separately and the pooled subjects case. We consider two score thresholds: "Passing Score" ( $> 2$ ) and "Good or Very Good Score" ( $> 3$ ).

Figure 1: The Geography of SE-Based Gender Achievement Gaps



Notes: This figure displays choropleth maps of SE-Based Gender Achievement Gaps in pooled subjects for low-SE and high-SE students, for both score thresholds and for the 4<sup>th</sup>, 6<sup>th</sup> and 9<sup>th</sup> grades, at the municipality level, using data on all years of observation. The higher female-favouring SE-Based Gender Achievement Gaps are coloured in blue.

Figure 2: SE-Based Gender Achievement Gaps and Socioeconomic and Demographic variables



Notes: This figure displays the estimated coefficients from OLS multiple linear regressions, as in equation (1), of pooled subjects SE-Based Achievement Gaps on the socioeconomic and demographic variables, grouped in five dimensions. We consider the score threshold "Passing Score" ( $> 2$ ) and 9<sup>th</sup> grade students from high-SE and low-SE backgrounds. Coefficients shall be interpreted as follows: an increase of 1 standard deviation in the socioeconomic variable is associated to an increase of  $\beta$  standard deviations in SE-Based Gender Achievement Gap.

## References

- Ângelo, Catarina and Ana Balcão Reis. 2021. "Gender gaps in different grading systems"  
*Education Economics* 29(1): 105-119. doi: 10.1080/09645292.2020.1853681
- Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2020. "Males at the Tails: How Socioeconomic Status Shapes the Gender Gap." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3609648>
- Baker, David, and Deborah Perkins Jones. 1993. "Creating Gender Equality: Cross-National Gender Stratification and Mathematical Performance." *Sociology of Education* 66 (2): 91–103. <https://doi.org/10.2307/2112795>.
- Becker, Gary. 1998. *A Treatise on the Family*. Cambridge, Massachusetts: Harvard University Press.
- Becker, Gary, William Hubbard, and Kevin Murphy. 2010. "Explaining the Worldwide Boom in Higher Education of Women." *Journal of Human Capital* 4 (3): 203-241. <https://doi.org/10.1086/657914>
- Blanden, Jo, Matthias Doepke, and Jan Stuhler. 2022. "Educational Inequality\*." *NBER Working Paper No. 29979*. doi:10.3386/w29979
- Buchmann, Claudia, Thomas A. DiPrete, and Anne McDaniel. 2008. "Gender Inequalities in Education." *Annual Review of Sociology* 34 (1): 319-337  
10.1146/annurev.soc.34.040507.134719
- Carneiro, Pedro. 2008. "Equality of opportunity and educational achievement in Portugal". *Port. Econ. J.* 7: 17–41. <https://doi.org/10.1007/s10258-007-0023-z>

- Carrington, Bruc, and Alastair McPhee. 2008. "Boys' 'underachievement' and the feminization of teaching." *Journal of Education for Teaching* 34 (2): 109-120. doi: 10.1080/02607470801979558
- Carvalho, Bruno. 2020. "Campaign spending on local elections: the more the merrier" (Tech. Rep.). *Working paper*. <https://lagv2019.sciencesconf.org/251138/document>
- Chetty, Raj, and Nathaniel Hendren. 2018a. "The Impacts of Neighborhoods on Intergenerational Mobility i: Childhood Exposure Effects." *The Quarterly Journal of Economics* 133 (3): 1107–62. <https://doi.org/10.1093/qje/qjy007>
- Chetty, Raj, and Nathaniel Hendren. 2018b. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *The Quarterly Journal of Economics* 133 (3): 1163–1228. <https://doi.org/10.1093/qje/qjy006>
- Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4): 855–902. <https://doi.org/10.1257/aer.20150572>
- Coleman, James, Ernest Campbell, Carol Hobson, James McPartland, Alexander Mood, Frederic Weinfeld, and Robert York. 1966. "Equality of Educational Opportunity." *National Center for Educational Statistics*.
- Decree-Law nr 176/2012 of August 2 of Ministry of Education and Science. *Official Gazette of the Portuguese Republic nr.149/2012, Series I of 2012-08-02*: 4068 – 4071 Accessed December 2, 2022. <https://dre.pt/dre/detalhe/decreto-lei/176-2012-179057>

Decree-Law nr 17/2016 of April 4 of Ministry of Education and Science. *Official Gazette of the Portuguese Republic nr.66, Series II of 2016-04-05*: 11440-(10) Accessed December 3, 2022. <https://dre.pt/dre/detalhe/decreto-lei/176-2012-179057>

Delaney, Judith, and Paul J. Devereux. 2021. "The Economics of Gender and Educational Achievement: Stylized Facts and Causal Evidence." *Oxford Research Encyclopedia of Economics and Finance*.  
<https://doi.org/10.1093/acrefore/9780190625979.013.663>.

DGEEC. 2022. *Estatísticas Da Educação 2020/2021: Síntese De Resultados*. Lisboa: DGEEC. <https://www.dgeec.mec.pt/np4/1372.html>

DGEEC and JNE. 2018. *Provas Finais E Exames Nacionais - Principais Indicadores - Ensino Básico e Secundário 2017*. Lisboa: DGEEC and JNE  
<https://www.dgeec.mec.pt/np4/441>

DiPrete, Thomas, and Claudia Buchmann. 2013. *Rise of Women, The Growing Gender Gap in Education and What It Means for American Schools*. Russell Sage Foundation. <http://www.jstor.org/stable/10.7758/9781610448000>.

Drudy, Sheelagh. 2008. "Gender balance/gender bias: the teaching profession and the impact of feminization." *Gender and Education* 20 (4): 309-323.  
doi: 10.1080/09540250802190156

EIGE. 2022. *Gender Equality Index 2022: The COVID-19 pandemic and care*. Luxembourg: Publications Office of the European Union. doi: 10.2839/035888

Farré, Lúdia, Christina Felfe, Libertad González, and Patrick Schneider. 2022. "Changing Gender Norms across Generations: Evidence from a Paternity Leave

Reform.” *BSE Working Paper 1310*

[https://bse.eu/sites/default/files/working\\_paper\\_pdfs/1310\\_0.pdf](https://bse.eu/sites/default/files/working_paper_pdfs/1310_0.pdf)

Fernandes, Susana, Patrícia Pereira, Joana Duarte, and Luísa Canto e Castro. 2018.

*Estudantes à entrada do Ensino Secundário em 2016/17*. Lisboa: DGEEC

<https://www.dgeec.mec.pt/np4/47/>

Ganley, Colleen, and Sarah Lubienski. 2016. “Mathematics confidence, interest, and

performance: Examining gender patterns and reciprocal relations.” *Learning and*

*Individual Differences* 47: 182-193. <https://doi.org/10.1016/j.lindif.2016.01.002>.

Geist, Eugene, and Margaret King. 2008. “Different, not better: Gender differences in

mathematics learning and achievement.” *Journal of Instructional Psychology*,

35(1): 43–52. <https://eric.ed.gov/?id=EJ813305>

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. “Diversity.

Culture, gender, and math” *Science* 320 (5880): 1164-1165

[doi:10.1126/science.1154094](https://doi.org/10.1126/science.1154094)

Hanushek, Eric, Paul Peterson, Laura Talpey, and Ludger Woessmann. 2019. “The

Achievement Gap Fails to Close: Half Century of testing shows persistent divide between haves and have-nots.” *Education Next* 19 (3): 8-17.

<https://www.educationnext.org/achievement-gap-fails-close-half-century-testing-shows-persistent-divide/>

Haveman, Robert, and Barbara Wolfe. 2002. "Social and nonmarket benefits from

education in an advanced economy," *Conference Series; [Proceedings], Federal*

*Reserve Bank of Boston*, 47(Jun): 97-142.

- Hyde, Janet, and Janet E. Mertz. 2009. "Gender, culture, and mathematics performance" *PNAS*. 106 (22):8801-8807 <https://doi.org/10.1073/pnas.0901265106>
- ILO. 2021. *Report on the Gender Pay Gap in Portugal*. Geneva: PRODOC of the ILO. ISBN 97-89-22035690-6 (Web PDF) [ISBN]
- Kahn, Shulamit, and Donna Ginther. 2017. "Women and STEM." *NBER Working Paper No. 23525*. doi: 10.3386/w23525
- Lei, Ziteng, and Shelly J. Lundberg. 2020. "Vulnerable Boys: Short-Term and Long-Term Gender Differences in the Impacts of Adolescent Disadvantage." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3534492>.
- Leibowitz, Arleen. 1977. "Parental Inputs and Children's Achievement." *The Journal of Human Resources* 12 (2): 242. <https://doi.org/10.2307/145387>.
- List, John, Fatemeh Momeni, and Yves Zenou. 2020. "The Social Side of Early Human Capital Formation: Using a Field Experiment to Estimate the Causal Impact of Neighborhoods." *NBER Working Paper No. w28283*, Available at SSRN: <https://ssrn.com/abstract=3756315>
- Lourenço, Vanda, Alexandra Duarte, Alexandra Nunes, Ana Amaral, Conceição Gonçalves, Madalena Mota, and Rosário Mendes. 2019. *PISA 2018 – PORTUGAL. National Report*. Lisboa: IAVE
- Matos, Margarida, Celeste Simões, Inês Camacho, Marta Reis. 2015. *A saúde dos adolescentes portugueses em tempos de recessão: Dados Nacionais HBSC/OMS 2014*. Lisboa: FMH/UL. <https://www.dge.mec.pt/sites/default/files/Esauade/hbsc2014.pdf>

- McMahon, Walter. 2010. "The External Benefits of Education." *International Encyclopedia of Education* (3): 260–271. <https://doi.org/10.1016/B978-0-08-044894-7.01226-4>
- Muntoni, Francesca, Jenny Wagner, and Jan Retelsdorf. 2020. "Beware of Stereotypes: Are Classmates' Stereotypes Associated With Students' Reading Outcomes?" *Child Development* 92 (1) <https://doi.org/10.1111/cdev.13359>
- Neef, Theresa, and Anne–Sophie Robilliard. 2021. "Half the Sky? The Female Labor Income Share in a Global Perspective." *World Inequality Lab – Working Paper N° 2021/22*. <https://ideas.repec.org/p/hal/wpaper/halshs-03693182.html>
- Nunes, Miguel, Luís Catela Nunes, Pedro Carneiro, and Pedro Freitas. 2021. "Regional Disparities in Socioeconomic-Based Achievement Measures".
- OECD. 2015. *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. PISA: OECD Publishing.  
<http://dx.doi.org/10.1787/9789264229945-en>
- OECD. 2021. "Gender wage gap (indicator)". OECD Data. Accessed November 30, 2022. doi: 10.1787/7cee77aa-en
- Pereira, Patrícia, and João Baptista. 2017. *Resultados escolares por disciplina, 3.º ciclo do ensino público, 2014/2015*. Lisboa: DGEEC  
[https://static.publico.pt/DOCS/ResultadosDisciplinas\\_3\\_CEB.pdf](https://static.publico.pt/DOCS/ResultadosDisciplinas_3_CEB.pdf)
- Pope, Devin, and Justin Sydnor. 2010. "Geographic Variation in the Gender Differences in Test Scores." *Journal of Economic Perspectives*, 24 (2): 95-108.  
doi:10.1257/jep.24.2.95

- PPDM. 2019. *Guia Feminista: eleições legislativas 2019*. Lisboa: PPDM  
[https://plataformamulheres.org.pt/site/wp-content/ficheiros/2019/10/Mini\\_Guia\\_Feminista-Legislativas2019.pdf](https://plataformamulheres.org.pt/site/wp-content/ficheiros/2019/10/Mini_Guia_Feminista-Legislativas2019.pdf)
- Reardon, Sean, Erin Fahle, Demetra Kalogrides, Anne Podolsky, and Rosalía C. Zárate. 2019. “Gender Achievement Gaps in U.S. School Districts”. *American Educational Research Journal* 56 (6): 2474–2508. <https://doi.org/10.3102/0002831219843824>
- San Román, Ainara, and Sara de la Rica. 2012. “Gender Gaps in PISA Test Scores: “The Impact of Social Norms and the Mother's Transmission of Role Attitudes” *SSRN Electronic Journal* 34. doi:10.2139/ssrn.2007329
- Warrington, Molly, and Michael Younger. 2000. *The Other Side of the Gender Gap*, *Gender and Education* 12 (4): 493-508. doi: 10.1080/09540250020004126
- Warrington, Molly, Michael Younger, and John Williams. 2000. “Student Attitudes, Image and the Gender Gap.” *British Educational Research Journal* 26 (3)

## Appendix

Table 1: Portuguese Educational System

	Grades	Typical Ages
Pre-School		3 to 5
Elementary Education	1 <sup>st</sup> to 6 <sup>th</sup>	6 to 11
Lower Secondary Education	7 <sup>th</sup> to 9 <sup>th</sup>	12 to 14
Upper Secondary Education	10 <sup>th</sup> to 12 <sup>th</sup>	15 to 18
(Students may choose between Science-humanities courses or Technological courses)		

Notes: This table reports information on the 4 cycles of the Portuguese Educational System.

Table 2: Number of Exams used in the analysis for each academic year and grade

Grade	Academic Year											Total
	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017	2017/2018	
4	193,555	194,924	198,413	182,721	177,115	183,912	171,102	172,405	0	0	0	1,474,147
6	205,200	198,759	199,877	197,064	195,367	188,008	183,579	176,206	0	0	0	1,544,060
9	171,784	157,275	156,210	156,877	160,224	166,375	162,744	166,773	161,492	164,954	160,200	1,784,908
Total	570,539	550,958	554,500	536,662	532,706	538,295	517,425	515,384	161,492	164,954	160,200	4,803,115

Notes: This table reports the number of exams used in the analysis for each academic year and grade. After the academic year of 2014/2015, there are only exam scores data regarding the 9<sup>th</sup> grade, as the National Exams for 4<sup>th</sup> and 6<sup>th</sup> grades were abolished.

Table 3: Number of Exams used in the analysis for each subject and grade

Grade	Mathematics	Portuguese Language
4	736,709	737,438
6	772,227	771,833
9	893,588	891,320

Notes: This table reports the number of exams used in the analysis for each subject and grade.

Table 4: Descriptive statistics on Socioeconomic and Demographic Municipality-level variables

Variable	Mean	Standard Deviation	Minimum	Maximum	Percentile		
					25	50	75
% of Population without High School Diploma	0,763	0,071	0,470	0,886	0,725	0,777	0,815
Gender Difference in % of Population without High School Diploma	-0,038	0,017	-0,080	0,029	-0,049	-0,038	-0,027
Average Wage	876,909	158,761	673,100	1 721,200	770,200	841,900	932,000
Abstention Rate	0,433	0,056	0,300	0,627	0,395	0,427	0,468
Left-wing Parties Voting Share	0,443	0,125	0,140	0,796	0,349	0,435	0,516
Population Density	311,491	865,988	5,100	7 363,400	25,200	70,200	174,900
Gender Pay Gap	-184,073	131,181	-1 460,200	38,300	-221,900	-165,400	-110,900
Gender Difference in Unemployment Rate	0,030	0,033	-0,050	0,168	0,008	0,026	0,046
Gender Difference in Inactivity Rate	0,115	0,033	0,052	0,234	0,093	0,111	0,137
Religious Population	0,942	0,050	0,767	0,993	0,914	0,962	0,982
Marriages	0,343	0,114	0,051	0,783	0,273	0,333	0,414
Divorces	0,739	0,360	0,000	2,500	0,505	0,672	0,900
Average Family Size	2,542	0,203	2,100	3,100	2,400	2,500	2,600
Paternity Leave	0,554	0,176	0,182	1,072	0,416	0,545	0,676

Notes: This table reports descriptive statistics on socioeconomic and demographic municipality-level variables.

Data concerns the year 2011, except Paternity Leave, which regards the year 2014. The data source is Statistics Portugal (INE).

Table 5: Single Linear Regressions of SE-Based Gender Achievement Gaps on Socioeconomic and Demographic variables

Variable	Passing Score (> 2)	
	Low-SE (1)	High-SE (2)
Percentage of Population Without High School Diploma	0.367*** (0.037)	0.223*** (0.027)
Gender Difference in the Percentage of Population Without High School Diploma	-0.246*** (0.048)	-0.146*** (0.021)
Average Wage	-0.354*** (0.038)	-0.161*** (0.030)
Abstention Rate	0.079 (0.063)	0.235*** (0.044)
Left.wing Parties Voting Share	-0.188*** (0.065)	-0.048 (0.046)
Population Density	-0.189*** (0.024)	-0.095*** (0.015)
Gender Pay Gap	0.388*** (0.072)	0.119* (0.066)
Gender Difference in Unemployment Rates	0.409*** (0.051)	0.230*** (0.055)
Gender Difference in Inactivity Rates	0.309*** (0.046)	0.205*** (0.045)
Religious Population	0.473*** (0.042)	0.187*** (0.040)
Marriages	0.181** (0.081)	0.129** (0.060)
Divorces	-0.254*** (0.084)	-0.008 (0.076)
Average Family Size	0.296*** (0.044)	0.154*** (0.049)
Paternity Leave	-0.200** (0.078)	-0.229*** (0.046)

Notes: This table reports the coefficients and standard errors of OLS simple regressions of SE-Based Gender Achievement Gaps on the socioeconomic and demographic variables. We consider the score threshold "Passing Score" (>2) and we run a regression for each socioeconomic group of students.

Significance Levels: \*p<0.10, \*\*p<0.05, \*\*\*p< 0.01 .

Table 6: Multiple Linear Regression of SE-Based Gender Achievement Gaps on Education and Income variables

Variable	Passing Score (> 2)	
	Low-SE	High-SE
	(1)	(2)
Percentage of Population Without High School Diploma	0.243***	0.227***
	(0.058)	(0.036)
Gender Difference in the Percentage of Population Without High School Diploma	-0.041	-0.050*
	(0.039)	(0.030)
Average Wage	-0.126*	0.041
	(0.067)	(0.046)
Constant	0.066	-0.101**
	(0.043)	(0.039)
Observations	278	278
$R^2$	0.374	0.302

Notes: This table reports the coefficients and standard errors of an OLS multiple regression of SE-Based Gender Achievement Gaps on the variables that belong to the group "Education and Income". We consider the score threshold "Passing Score" (>2) and we run a regression for each socioeconomic group of students. Significance Levels: \*p<0.10, \*\*p<0.05, \*\*\*p< 0.01 .

Table 7: Single Linear Regression of SE-Based Gender Achievement Gaps on Population Density

Variable	Passing Score (> 2)	
	Low-SE	High-SE
	(1)	(2)
Population Density	-0.189***	-0.095***
	(0.024)	(0.015)
Constant	0.001	-0.250***
	(0.055)	(0.037)
Observations	278	278
$R^2$	0.207	0.180

Notes: This table reports the coefficients and standard errors of an OLS simple regression of SE-Based Gender Achievement Gaps on the Population Density Variable. We consider the score threshold "Passing Score" (>2) and we run a regression for each socioeconomic group of students.

Significance Levels: \*p<0.10, \*\*p<0.05, \*\*\*p< 0.01 .

Table 8: Multiple Linear Regression of SE-Based Gender Achievement Gaps on Gender Differences in the Labour Market variables

Variable	Passing Score (> 2)	
	Low-SE (1)	High-SE (2)
Gender Pay Gap	0.236*** (0.066)	0.045 (0.056)
Gender Difference in Unemployment Rates	0.306*** (0.068)	0.157* (0.089)
Gender Difference in Inactivity Rates	0.021 (0.067)	0.082 (0.085)
Constant	-0.035 (0.044)	-0.202*** (0.048)
Observations	278	278
$R^2$	0.3242	0.1401

Notes: This table reports the coefficients and standard errors of an OLS multiple regression of SE-Based Gender Achievement Gaps on the variables that belong to the group "Gender Differences in the Labour Market". We consider the score threshold "Passing Score" (>2) and we run a regression for each socioeconomic group of students. Significance Levels: \*p<0.10, \*\*p<0.05, \*\*\*p< 0.01 .

Table 9: Multiple Linear Regression of SE-Based Gender Achievement Gaps on Cultural Habits and Social Norms variables

Variable	Passing Score (> 2)	
	Low-SE (1)	High-SE (2)
Religious Population	0.389*** (0.054)	0.105*** (0.037)
Marriages	0.047 (0.053)	0.049 (0.049)
Divorces	-0.072 (0.056)	0.026 (0.041)
Average Family Size	0.058 (0.043)	0.041 (0.045)
Paternity Leave	-0.038 (0.042)	-0.165*** (0.040)
Constant	-0.115** (0.045)	-0.190*** (0.048)
Observations	277	277
$R^2$	0.414	0.228

Notes: This table reports the coefficients and standard errors of an OLS multiple regression of SE-Based Gender Achievement Gaps on the variables that belong to the group "Cultural Habits and Social Norms". We consider the score threshold "Passing Score" (>2) and we run a regression for each socioeconomic group of students.

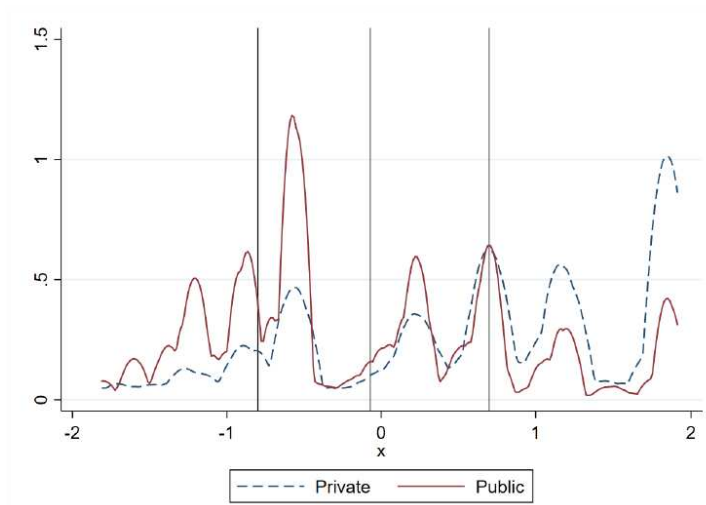
Significance Levels: \*p<0.10, \*\*p<0.05, \*\*\*p< 0.01 .

Table 10: Multiple Linear Regression of SE-Based Gender Achievement Gaps on Election Outcomes variables

Variable	Passing Score (> 2)	
	Low-SE	High-SE
	(1)	(2)
Abstention Rate	0.043 (0.063)	0.234*** (0.048)
Left-Wing Parties Voting Share	-0.176*** (0.067)	-0.005 (0.049)
Constant	-0.110* (0.062)	-0.229*** (0.042)
Observations	278	278
$R^2$	0.042	0.122

Notes: This table reports the coefficients and standard errors of an OLS multiple regression of SE-Based Gender Achievement Gaps on the variables that belong to the group "Election Outcomes". We consider the score threshold "Passing Score" (>2) and we run a regression for each socioeconomic group of students. Significance Levels: \*p<0.10, \*\*p<0.05, \*\*\*p< 0.01 .

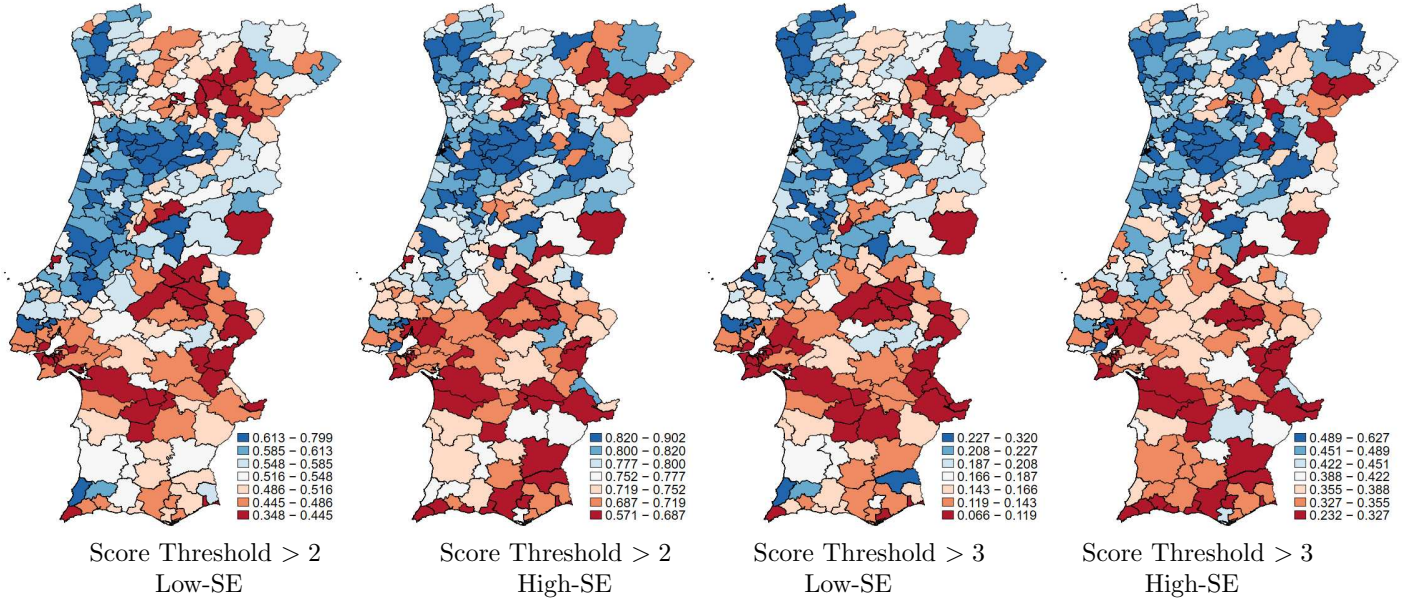
Figure 1: Socioeconomic Index distribution



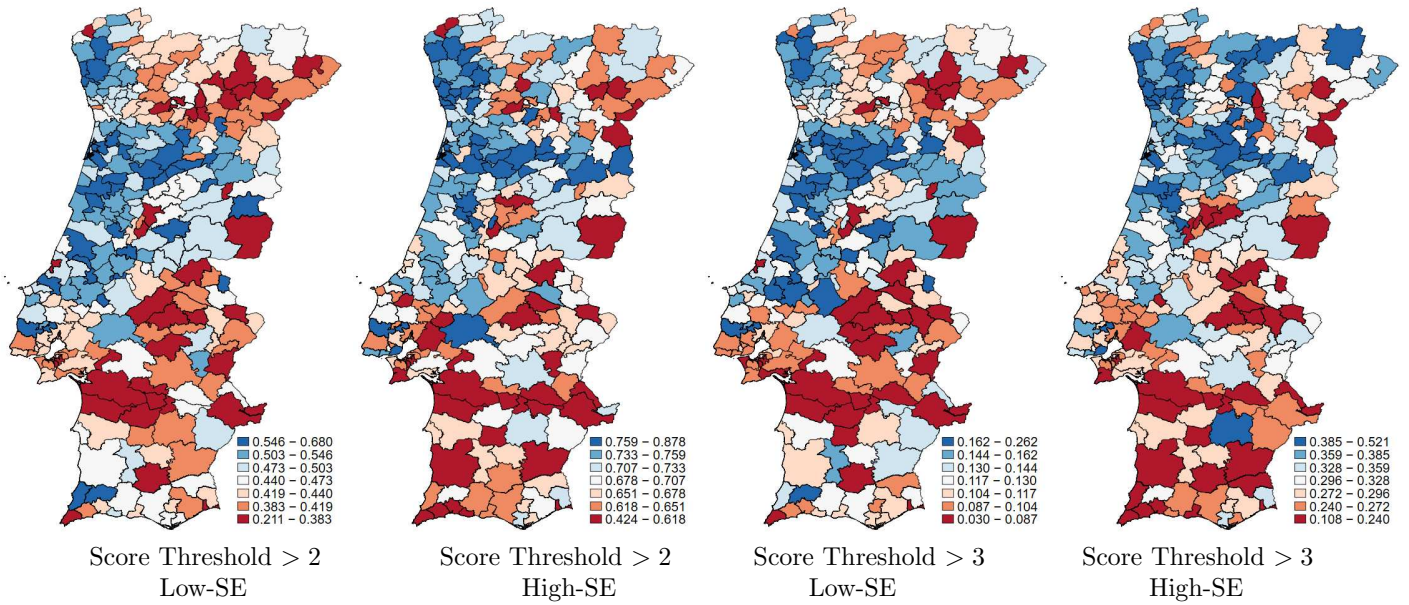
Notes: This figure displays the distribution of the SE Index, for public and private students. The x-axis represents the SE Index.

Figure 2: The Geography of the Performance of Low-SE and High-SE Students

(a) Female

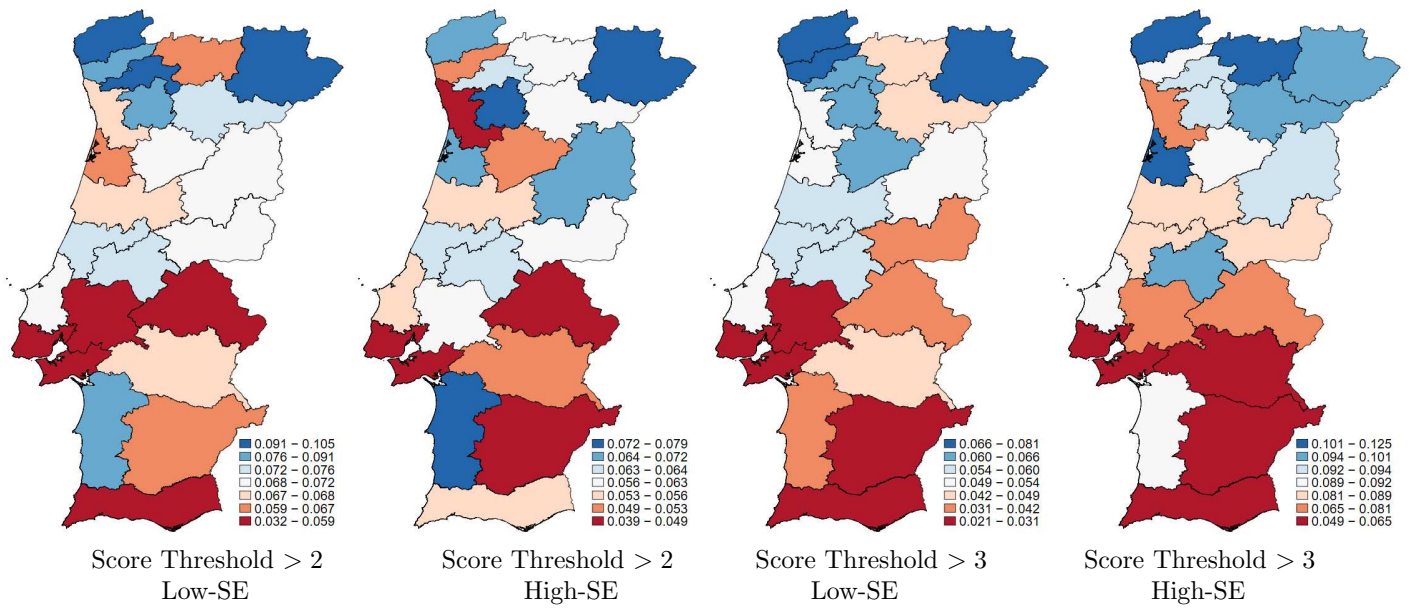


(b) Male



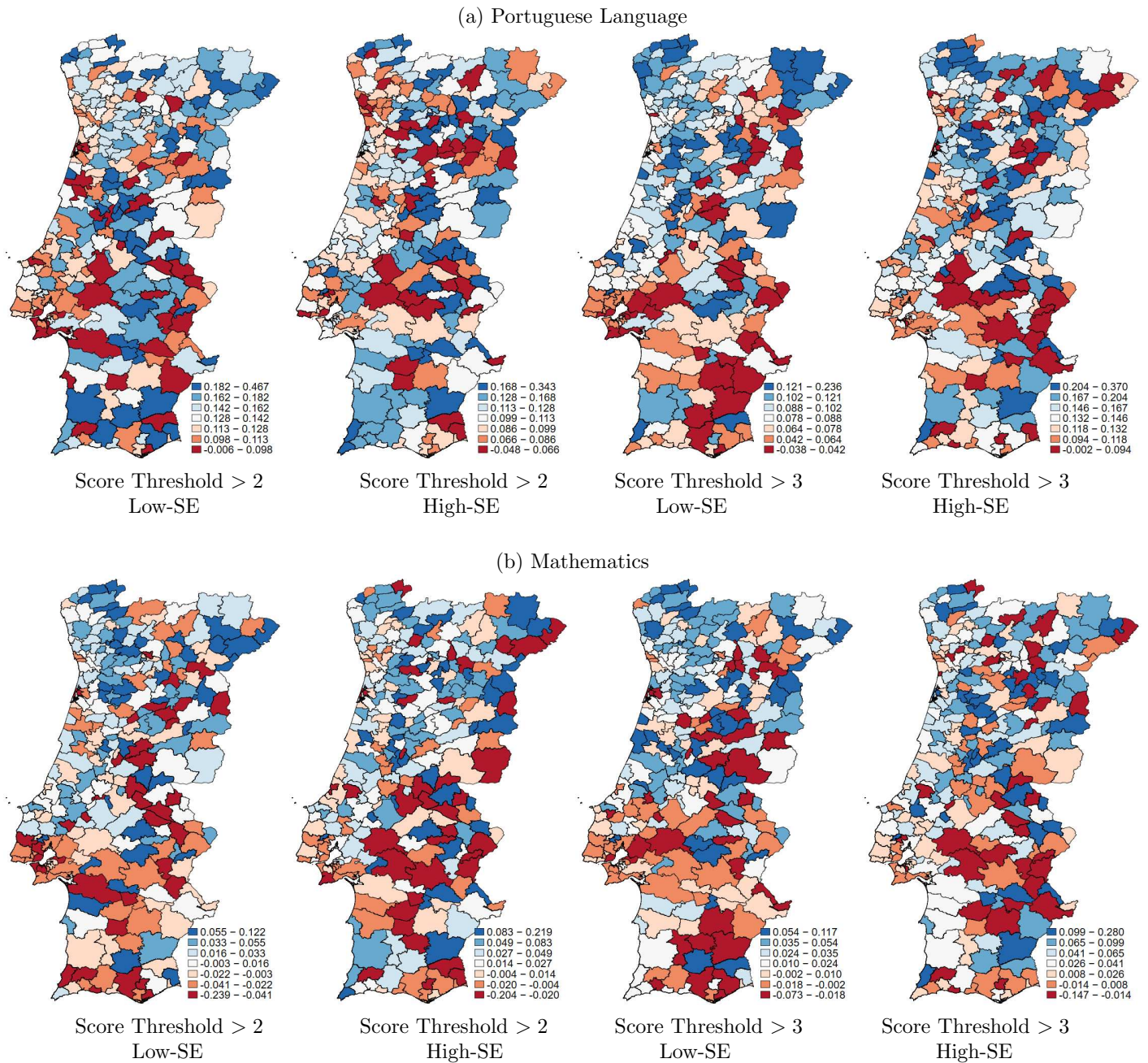
Notes: This figure displays choropleth maps of the performance of low-SE and high-SE students by gender, for pooled subjects, for both score thresholds and for the 9<sup>th</sup> grade, at the municipality level. The lowest values of the achievement measures are colored in red.

Figure 3: The Geography of SE-Based Gender Achievement Gaps by NUTS III



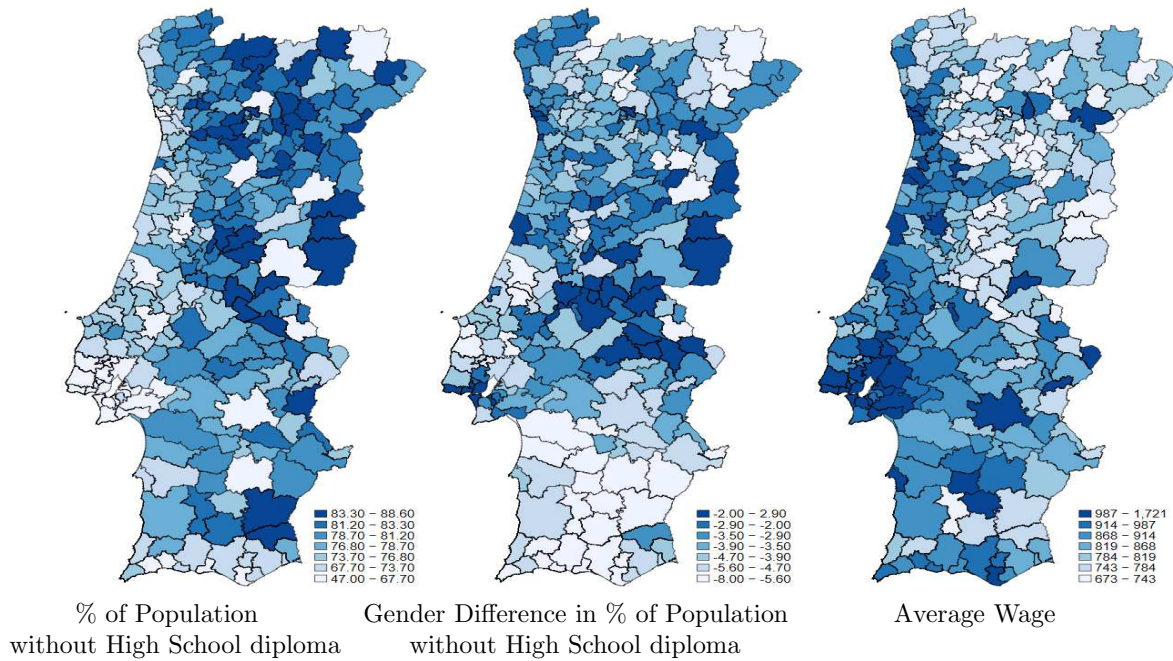
Notes: This figure displays choropleth maps of SE-Based Gender Achievement Gaps for low-SE and high-SE students, for pooled subjects, for both score thresholds and for the 9<sup>th</sup> grade, at the NUTS III level. The higher female-favouring SE-Based Gender Achievement Gaps are coloured in blue.

Figure 4: The Geography of SE-Based Gender Achievement Gaps by Subject



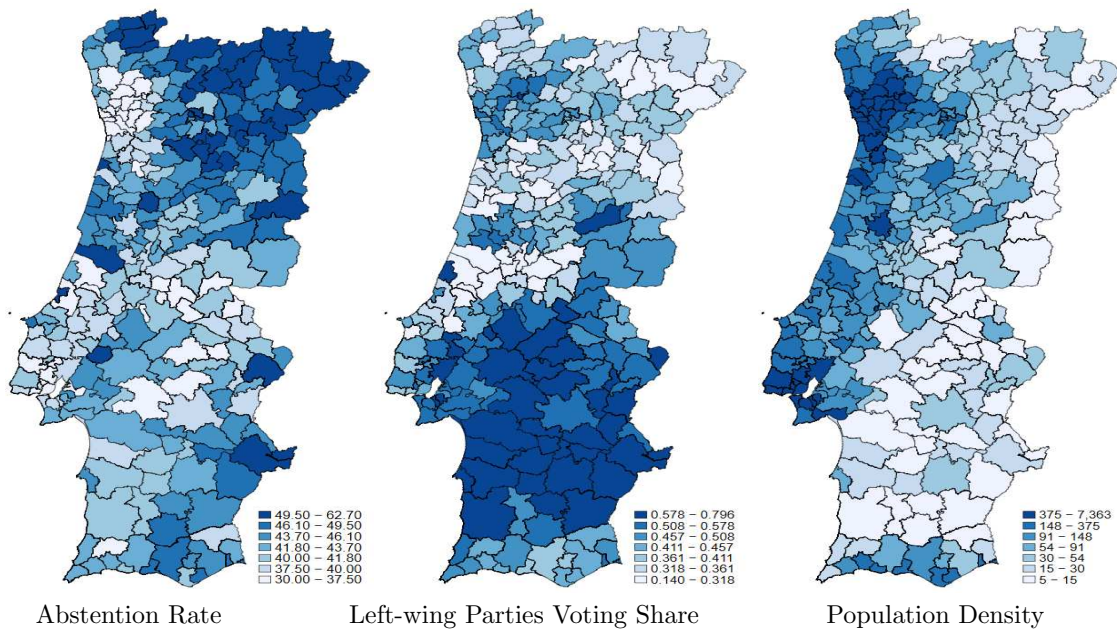
Notes: This figure displays choropleth maps of the SE-Based Gender Achievement Gaps for the 9<sup>th</sup> grade, for low-SE and high-SE students, for both score thresholds and for the subjects Portuguese Language and Mathematics separately. The higher female-favouring SE-Based Gender Achievement Gaps are coloured in blue.

Figure 5: Education and Income



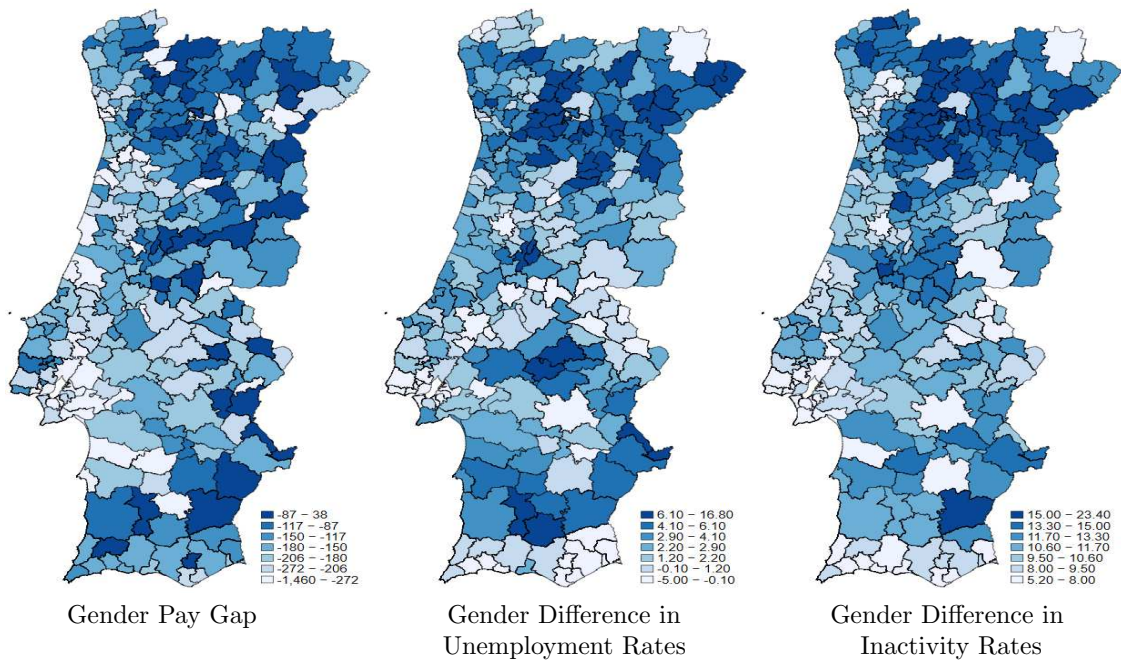
Notes: This figure displays the geographic distribution of (i) the number of people aged 15 or older without Upper Secondary Education divided by the number of people aged 15 or older in 2011; (ii) the gender difference in the former measure; and (iii) the average wage of full-time workers in the private sector in 2011. The data source is Statistics Portugal (INE).

Figure 6: Election Outcomes and Population Density



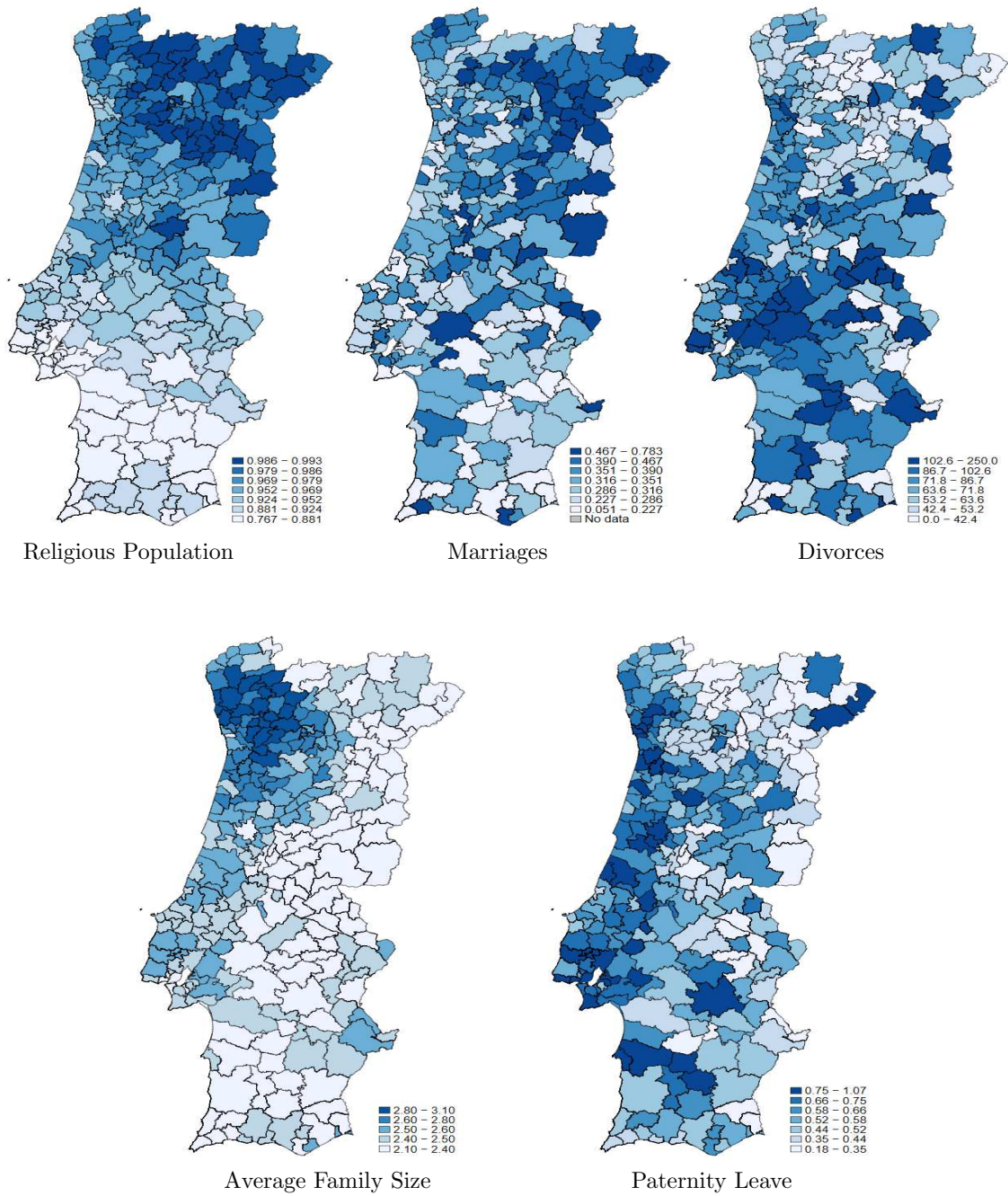
Notes: This figure displays the geographic distribution of (i) the abstention rate in the 2011 Elections; (ii) the left-wing parties voting share in 2011 Elections; and (iii) the population density in 2011. The data source is Statistics Portugal (INE).

Figure 7: Gender Differences in the Labour Market



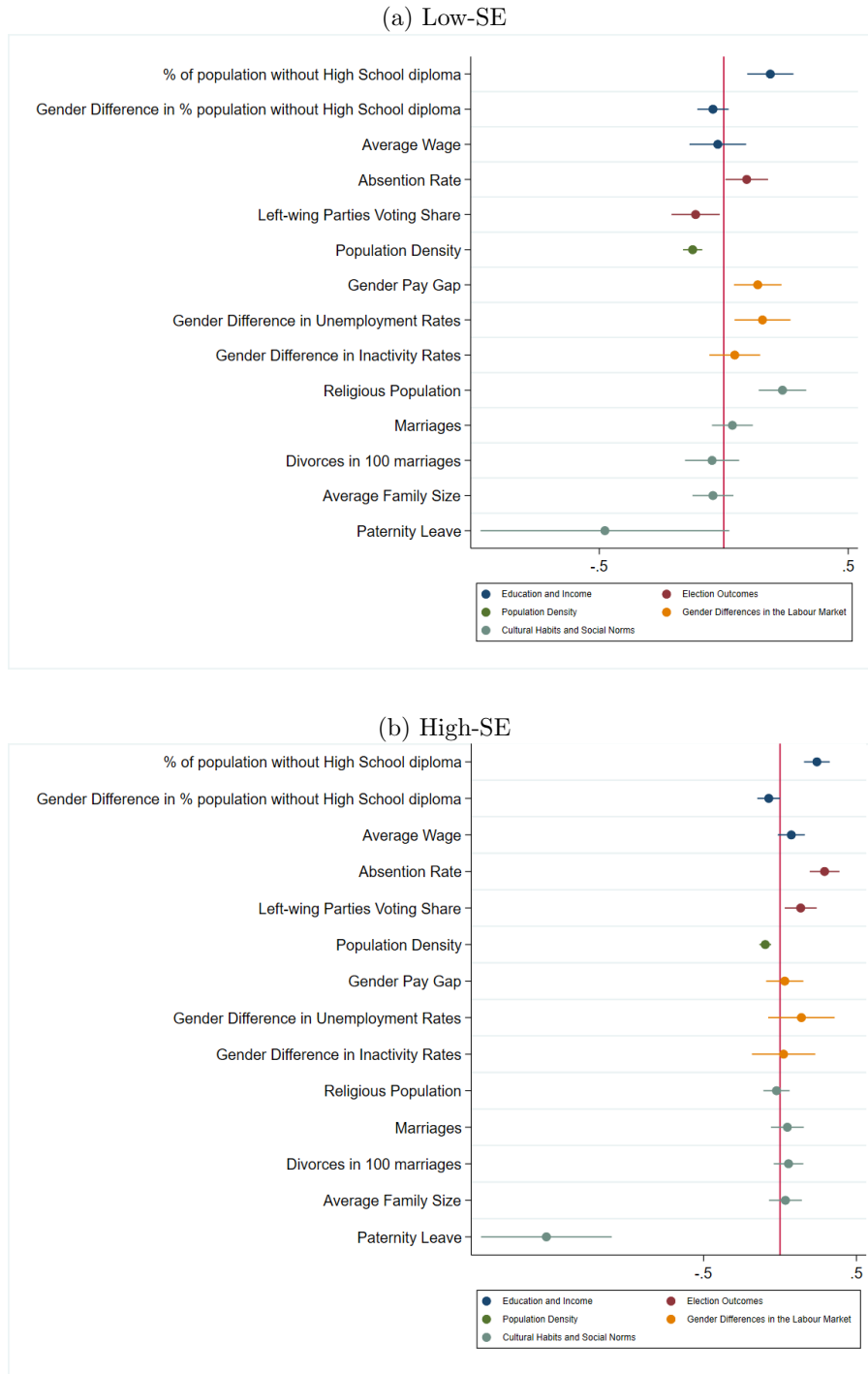
Notes: This figure displays the geographic distribution of (i) the gender difference in gross monthly wage of full-time workers in the private sector in 2011; (ii) the gender difference in unemployment rates in 2011; and (iii) the gender difference in inactivity rates in 2011. The data source is Statistics Portugal (INE).

Figure 8: Cultural Habits and Social Norms



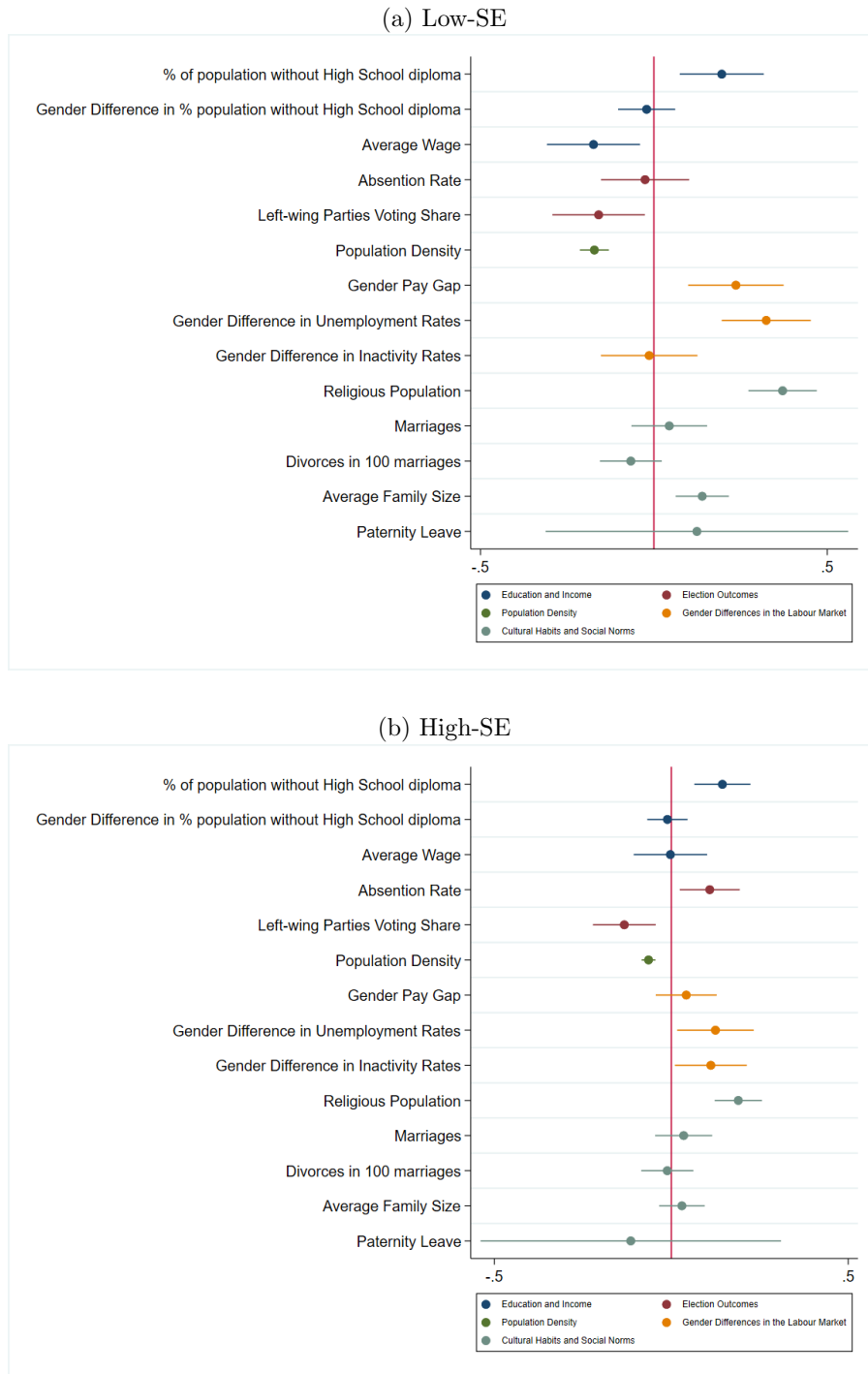
Notes: This figure displays the geographic distribution of (i) the percentage of the religious Portuguese population in 2011; (ii) the number of marriages in 2011 divided by number of people aged between 20 and 44 in that year; (iii) the number of divorces in 2011 per 100 weddings in 2011; (iv) the average family size in 2011; and (v) the paternity leave amount per birth in 2014. The data source is Statistics Portugal (INE).

Figure 9: Portuguese Language SE-Based Gender Achievement Gaps and Socioeconomic and Demographic variables



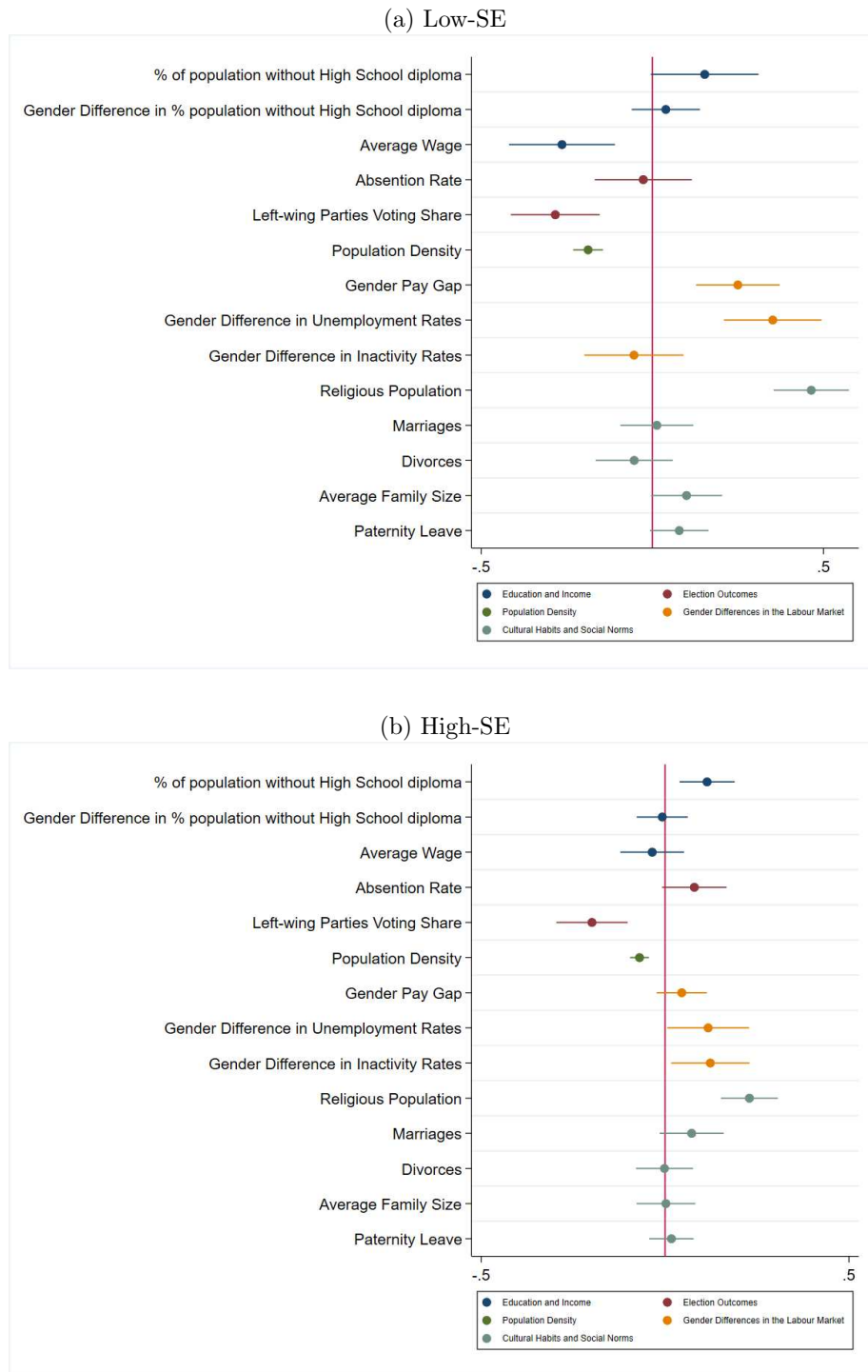
Notes: This figure displays the estimated coefficients from OLS multiple linear regressions, as in equation (1), of Portuguese Language SE-Based Achievement Gaps on the socioeconomic and demographic variables, grouped in five dimensions. We consider the score threshold "Passing Score" ( $> 2$ ) and 9<sup>th</sup> grade students from high-SE and low-SE backgrounds. Coefficients shall be interpreted as follows: an increase of 1 standard deviation in the socioeconomic variable is associated to an increase of  $\beta$  standard deviations in SE-Based Gender Achievement Gap.

Figure 10: Mathematics SE-Based Gender Achievement Gaps and Socioeconomic and Demographic variables



Notes: This figure displays the estimated coefficients from OLS multiple linear regressions, as in equation (1), of Mathematics SE-Based Achievement Gaps on the socioeconomic and demographic variables, grouped in five dimensions. We consider the score threshold "Passing Score" ( $> 2$ ) and 9<sup>th</sup> grade students from high-SE and low-SE backgrounds. Coefficients shall be interpreted as follows: an increase of 1 standard deviation in the socioeconomic variable is associated to an increase of  $\beta$  standard deviations in SE-Based Gender Achievement Gap.

Figure 11: SE-Based Gender Achievement Gaps and Socioeconomic and Demographic variables



Notes: This figure displays the estimated coefficients from OLS multiple linear regressions, as in equation (1), of pooled subjects SE-Based Achievement Gaps on the socioeconomic and demographic variables, grouped in five dimensions. We consider the score threshold "Good or Very Good Score" ( $> 3$ ) and 9<sup>th</sup> grade students from high-SE and low-SE backgrounds. Coefficients shall be interpreted as follows: an increase of 1 standard deviation in the socioeconomic variable is associated to an increase of  $\beta$  standard deviations in SE-Based Gender Achievement Gap.