COLLEGE GRADUATES GENDER GAP IN COLOMBIA, SHRINKING, GROWING OR EVOLVING?

Alejandra Rodas-Gaiter Felipe Lozano-Rojas Pedro Cerdan-Infante

Abstract

Numerous efforts to empower women´s financial stability have been promoted to reduce gender discrimination and violence against them. Although there has been a trend toward reducing the gender wage gap in recent years due to improving women's qualifications and job profiles, this difference in income persists. Furthermore, it is crucial to recognize the determinants of the gender wage gap in developing countries where women still being a vulnerable group. In these contexts, gender discrimination reduces women's access to education and increases their likelihood of being in poverty and facing direct violence. Therefore, it is relevant to understand the policy problem and its causes. This paper uses data from the formal labor market in Colombia from 2007 to 2012 to compare the income between women and men and identify significant differences. Using a regression analysis combined with a counterfactual Blinder Oaxaca model, it was possible to confirm a gender wage gap in Colombia among employees with college degrees. Moreover, this paper sought to explore the causes of this result. It found, for example, that even if women employees have the same characteristics as men regarding education, location, and social background, the gap remains at 51%. Additionally, the paper explores how the educational area of the college degree, and the quality of the academic programs influence the gender wage gap in this specific population.

1 Introduction

Despite the improvements in reducing gender discrimination and the increase of women´s representation in places of power in Colombia (Iregui-Bohórquez et al. 2021), in this country, the simple fact of being a woman increases the probability of poverty by 18 percentage points (Colombian Women's Observatory 2021). This situation of vulnerability due to gender identity is the consequence of multiple social and economic dynamics. Specifically, one of the determinants of the fragile condition of women in Colombian society is the access and coditions of the labor market (Martínez-Restrepo et al. 2015).

Different studies have shown that Colombia is the Latin American country with the most significant increase in the female participation rate in the labor force in the last decades (Amador Osuna et al. 2013; Elias and Ñopo 2010; Hoyos et al. 2010; Iregui-Bohórquez et al. 2021; Ramírez Bustamante et al. 2015). Moreover, the Colombian labor market had important changes from the gender perspective (hoyos2010). For example, the National Constitution includes equality and non-discrimination clauses, and the government enacted different laws following the International Labour Organization (ILO) trends protecting women from biological disparities such as pregnancy (Ramírez Bustamante et al. 2015). However, despite the efforts in the normative framework and the aggregate results, the reality is not the same for all the population.

The Colombian labor market shows gender differences that rely on the years of education and occupations. In this regard, women generally have more years of schooling, whereas men tend to accumulate more experience (Hoyos et al. 2010). Similarly, there are relevant differences in occupational choices (Hoyos et al. 2010). Therefore, despite Colombia being perceived as an "egalitarian" country and its legal framework explicitly forbids discrimination, there are still sizeable gender gaps (Hoyos et al. 2010).

One of the mentioned gender differences is the wage gap. Although since 2000, the average gender gap began a reduction trend (Hoyos et al. 2010), the job market conditions show little advance in terms of discrimination (Bernat Díaz et al. 2007). It is possible to observe evidence of the glass ceiling effect and the 'quicksand floor effect' (Badel and Peña 2010). In other words, the gender wage gap has a U shape along the earnings distribution, meaning that women at the top and bottom of the distribution present lower earnings than men workers (Badel and Peña 2010; Hoyos et al. 2010).

This paper aims to zoom in on one of the two major problems of the u-shape gender wage gap in Colombia: the top of the income distribution. Specifically, we use data from the formal labor market in Colombia from 2007 to 2012 to compare the income between college graduate women and men. Moreover, this database allows us to explore the difference within and across

educational areas of the college degree. In conclusion, this paper analyzes gender discrimination in a labor market in a country with specific anti-discrimination policies in a population with better endowments. This paper explores the evolution of the literature about the gender wage gap and the main explanations and results from the labor economics perspective. Next, it explains the Colombian context and why it is relevant to understand the gender discrimination phenomenon in developing countries where women still being a vulnerable group. Then, we describe the data and methods we used to answer the research questions. Finally, we conclude the paper with the results and their implications.

2 Literature Review

The study of the wage gap between gender is not new. Since Becker's work on the economics of discrimination in 1957, there have been many studies about this phenomenon, its evolution, and its implications for society (Black et al. 2008). Gender and labor market studies have focused on the longitudinal analysis of the behavior of this gap. They have found that perceived differences in the division of labor, racism, decades of formal and informal discrimination, power relations, and cultural stereotyping can place women and people of color in lower-paying and lower-status jobs (Feeney and Camarena 2021). In women's case, some of the most significant reasons for the differences that explain discrimination are core features of performance, such as human capital, labor supply, occupational segregation, and discrimination by employers (Mandel and Semyonov 2014).

Another aspect that generally influences the difference in average earnings between men and women is access to management positions. In 1970 arose the metaphor of the "glass ceiling" as a way to comprehend barriers and inequities that women faced in the public realm to access decision positions in the job market (Connell 2006). Today, women continue to be underrepresented in positions of political and administrative authority in a significant part of the countries and different areas of public sector activity (Connell 2006). This limitation of access to these positions translates simultaneously into a restriction of access to the best-paid jobs in each sector.

Other studies have found that gender differences in labor force participation patterns help explain wage differentials (Kunze 2008; Mincer and Polachek 1974). Specifically, Astrid Kunze (2008) describes those findings and suggests that differences in human capital accumulation might explain 25 to 50 percent of the wage gap. Furthermore, Mandel and Semyonov (2014) mention as a determinant of the income women's educational profiles and the number of hours worked. The latter refers to the amount of worked hours of women compared with men and how there is evidence of female employees systematically working fewer hours (Männasoo 2022; Weichselbaumer and

Winter-Ebmer 2005). In contrast, the former explains workers' training and formal education as a measure of their productivity and fit with the market. Nonetheless, later in this section, we will deepen on this aspect and its evolution over the recent decades.

Goldin and colleges (2017a) state that extensive literature focuses on identifying determinants of the difference between incomes between genders, emphasizing the role of children and its implications in the capacity to move between jobs and the greater demand for work amenities such as flexibility (Goldin et al. 2017a). In this sense, they found in the U.S context that 40 percent of the increased gender gap in earnings is related to the men's shift into higher-paying positions, and 60 percent is attributed to the lower women's capability to advance in their professional careers within firms (Goldin et al. 2017a). Their empirical research concludes that most of the gap widening occurs in the first seven years in the labor market, the ages when families usually are formed (Goldin et al. 2017a). Therefore, they suggest that personal decisions such as marital status and children help explain some income differences.

However, an interesting aspect of the gender wage gap as a social phenomenon is its capacity to evolve. In other words, the gender income gap is a changing statistic (Goldin and Katz 2016; Goldin et al. 2017b; Huber and Solovyeva 2020; Sloane et al. 2021), and its determinants also change over time. In 2017, Francine D. Blau and Lawrence M. Kahn sought to understand the gender wage gap extent, trends, and explanations (Blau and Kahn 2017). They used data from the U.S. labor market between 1980 and 2010. The authors compared the job market characteristics between the studied years and observed differences across time. One of their main conclusions is that in 2010, conventional human capital variables explained little of the gender wage gap in the U.S. context. However, they claimed that "gender differences in occupations and industries, as well as differences in gender roles and gender division of labor, remain important" (Blau and Kahn 2017, p. 789). Goldin and colleagues had a similar conclusion in 2014, where they identified crucial factors to define the gender wage gap: college degree selection, the opportunity to change jobs across the industry sector, and marital status (Goldin 2014).

In gender studies, occupational segregation and the difference in women's condition in the labor market due to this fact are called "glass walls" (Mandel and Semyonov 2014; Sabharwal 2015). This concept refers to the percentage of wage and working conditions differentials explained by the fact that women are sorted into low-paying jobs and occupations, either because they are denied access or self-selected (Mandel and Semyonov 2014; Sloane et al. 2021). In other words, male-dominated jobs such as chemical engineering usually have higher returns than female-dominated jobs (e.g., elementary education) (Black et al. 2008). In conclusion, as Black et al. (2008) stated, to have a comprehensive understanding of the gender wage gap, especially among well-educated workers, it is crucial to consider pre-labor market choices.

Finally, in the study of gender wage trends, it is possible to conclude that raw wage differentials worldwide have declined (Blau and Kahn 2017; Weichselbaumer and Winter-Ebmer 2005). Better labor market endowments can explain part of this improvement for female employees (Weichselbaumer and Winter-Ebmer 2005). Women have become better educated and trained to improve their profile and opportunities to be part of and attach to the market (Weichselbaumer and Winter-Ebmer 2005). As we mentioned before, human capital is one of the historical determinants of income differences between genders. However, the gender pay gap decrease is not constant across all types of workers. It has declined more slowly at the top of the wage distribution (Blau and Kahn 2017) and among college graduates (Goldin et al. 2017a; Black et al. 2008).

In conclusion, the gender pay gap has been decreasing since it was studied in the 1970s. Nonetheless, the process has not been homogeneous, and the gap is more significant among college graduates and in sectors known to penalize shorter hours, job flexibility, and time off (Goldin et al. 2017a). Besides, specific characteristics, such as human capital accumulation and social aspects, play a role in explaining the gender wa

3 Colombian Context

In recent decades, Colombia has had mixed progress regarding gender disparities. On the one hand, women are still a vulnerable group because being born as a woman in this country increases the probability of poverty by 18 percentage points (Colombian Women's Observatory 2021). In the same regard, 32.6% of women report not receiving income, in contrast to 17.5 % of men (Colombian Women's Observatory 2021). In consequence, the fact that women are more likely to be in extreme poverty underscores inequities in the labor market (Martínez-Restrepo et al. 2015)

However, on the other hand, the country leaders in Latin America in the increase of the female participation rate in the labor decades (Amador Osuna et al. 2013; Elias and Ñopo 2010; Hoyos et al. 2010; Iregui-Bohórquez et al. 2021; Ramírez Bustamante et al. 2015). Different factors help to explain this process. Ramírez Bustamante et al. (2015) stated that women's participation in the labor market is a consequence of women's entry to college education since 1934, the massive access to contraceptives, growing urbanization, and the increasing household need for dual earner.

Despite the fact, the female unemployment rate has always been a two-digit number and more considerable than the men's rate (DANE 2021; Hoyos et al. 2010; Sabogal 2009) today, the ratio of female to male participation is similar to that of developed countries (Iregui-Bohórquez et al. 2021; Olivetti and Petrongolo 2016). Additionally, there is evidence of a significant participation increase of women in historically male-dominated fields such as medicine, law, and engineering (Iregui-Bohórquez et al. 2021). Moreover, Colombia in 2019 was one of the eleven countries

with a National Cabinet with a female representation of at least 50% (Iregui-Bohórquez et al. 2021). Regardless, women-headed policy sectors are generally related to gender stereotypes, such as home care and educational and cultural activities (Archila 2014; Iregui-Bohórquez et al. 2021; Wills 2007).

However, the improvements in average labor market statistics hide some elements that are still relevant to understanding gender disparities. For example, women are less likely to be in the formal labor market when we explore the quality of employment. By 2021, only 38.6% of Colombian employees who are formally employed are women (DANE 2021). Additionally, official information about the informal labor market in the country in 2014 estimated that 52.4 percent of women held an informal job, whereas 46.7 percent of men were in the same unstable situation (Martínez-Restrepo et al. 2015) (Martinez Ramirez). This situation becomes worse when we explore the data for the population in extreme poverty in urban areas. By 2013, 90 percent of women in extreme poverty held informal jobs, and 32 percent participated in the formal market, 15 percent points less than the average Martínez-Restrepo et al. (2015).

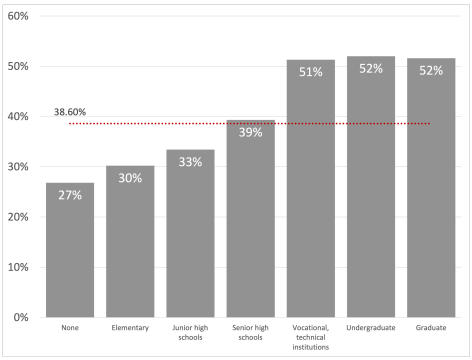
Different aspects help to explain this lack of equality regarding the quality of employment. For example, the level of education influences participation in the labor market, with women with graduate education having participation equivalent to their weight in society (see Figure 1). Nonetheless, there is evidence that gender-related factors such as pregnancy and house care responsibilities explain the discrimination in access and retention in the formal labor market (Ramírez Bustamante et al. 2015; Iregui-Bohórquez et al. 2021). For instance, the maternity leave Law increases the probability of being inactive, informality, and self-employment for high-fertility women relative to low-fertility women (Hoyos et al. 2010).

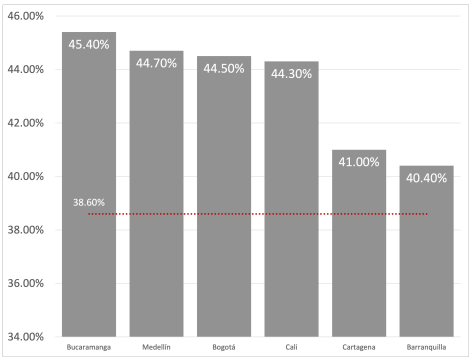
Another relevant factor to take into account to understand the differences between genders in the Colombian labor market is the regional dynamics. The data shows that women's participation is higher in the major cities compared with the average information, whereby it is possible to conclude that the urban dynamics play a role in the women's willingness to work and the available opportunities % (see Figure 1). Moreover, there are cities, such as Bucaramanga, Medellín, Bogotá, and Cali, whose rate is more than five perceptual points than the average national rate (DANE 2021). We highlighted before the improvement in the country in eliminating the "glass wall" and the tendency to observe more women in male-dominated fields. However, it is still possible to observe gender differences between occupations (Hoyos et al. 2010). Sectors such as construction, transportation, mining, and agriculture have less than 15% female participation (DANE 2021). In contrast, in sectors such as public education, arts, and services, women's share is greater than 60% (see figure 2) (DANE 2021). This situation highlights significant differences that may explain crises, such as the impact of COVID-19 in Colombian society (widespread lock-downs due to the

pandemic affected services the most).

Figure 1: Women's participation in labor market

Educational Level **Major Cities**





Notes: Authors' calculations using information from the National Department of Statistics (DANE, 2021).

A noteworthy aspect of the Colombian context is that there are differences that can be explained from the standpoint of gender roles. For example, 38.5% of Colombians believe women must take care of the household and family (Colombian Women's Observatory 2021). Additionally, women dedicate 7 hours and 12 minutes of unpaid domestic and care work, while men say only three and a half hours (Colombian Women's Observatory 2021). These examples and other gender norms are some women's constraints on their access to the labor market (Martínez-Restrepo et al. 2015). Therefore, it is not a surprise that despite the legal efforts to reduce discrimination in the labor market, there still rampant today.

A gender difference in the labor market that influences the likelihood of women in Colombia to be vulnerable is the gender wage gap. This income difference could also help explain the gender difference in the willingness to participate in the labor market due to the lower opportunity cost for women (Galor and Weil 1993; Iregui-Bohórquez et al. 2021) There is evidence that the difference in income between genders goes beyond access and quality of employment. Hoyos et al. (2010) conclude that there is a substantial gender wage gap which is mainly explained by the configuration of the labor market rather than by gender differences Hoyos et al. (2010).

As the women's labor market participation rate, the gender wage gap has had an average reduction since 2000 (Bernat Díaz et al. 2007; Hoyos et al. 2010). Nonetheless, it still affects Colombian women, specifically some sociodemographic groups. Badel and Peña (2010) found that the gender wage gap in the Colombian labor market has a U – shape along the earnings distribution (Badel and Peña 2010; Hoyos et al. 2010) In other words, the wage difference by gender is more considerable

when we compare employees with lower and higher incomes.

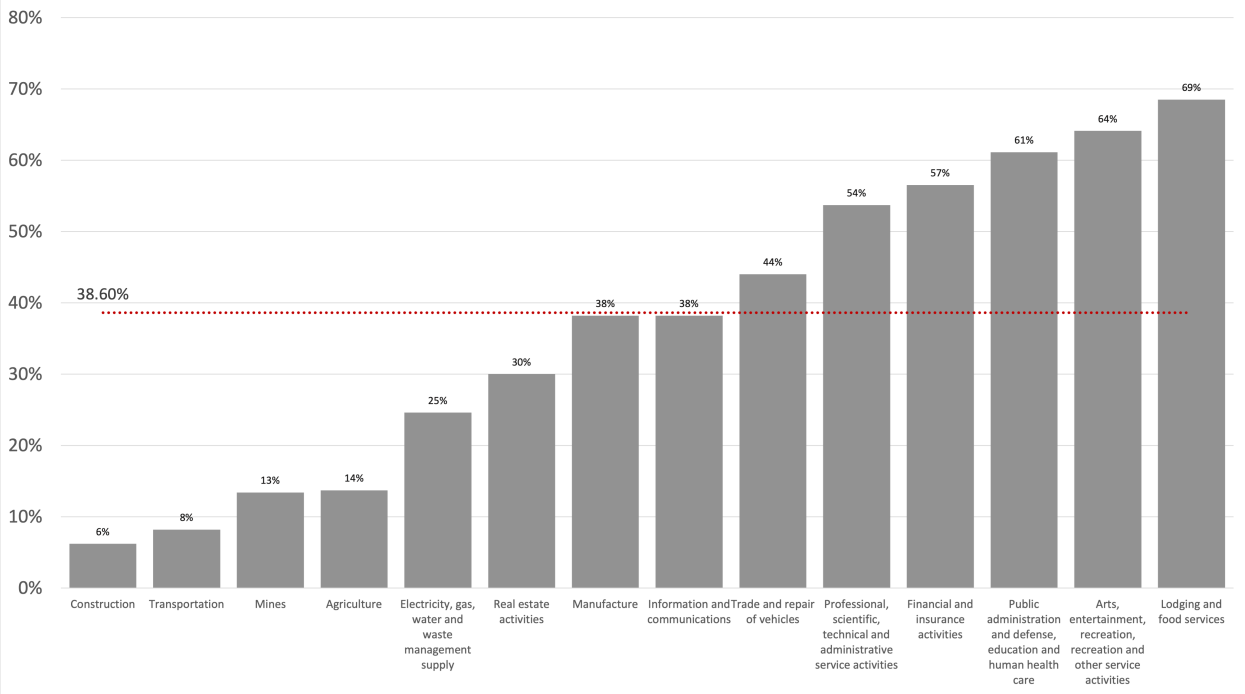
Traditionally, labor economics explains the differences in incomes using two aspects: education and experience. The widespread theory created by Mincer in 1958 helps to explain how characteristics related to the productivity of the employees and their inputs in the production process help to understand their incomes as a retribution for their job. In this regard, in recent years, women have had more years of education in Colombia, while men have more potential to accumulate experience (Hoyos et al. 2010). Whereas the second characteristic can be explained by the dynamics of the labor market we mentioned before, education has its own dynamics.

Looking at the education sector, there is currently no gender difference in the number of higher education enrolls and graduates. In some cases, there is even greater participation of women than men (Bertocchi and Bozzano 2020; Iregui-Bohórquez et al. 2021; Colombian Women's Observatory 2021; Bank Bank). This level of women's enrollment in higher education began in the 80s. In nearly three decades, the proportion of women enrolled in higher education in Colombia almost doubled, from 130 per ten thousand women to 230 per ten thousand in 2015 (Iregui-Bohórquez et al. 2021). Nonetheless this ratio still being low compare with the international context; the gender enrollment proportion is close to the ratio in countries like Spain and Finland, showing a generalized problem more than gender discrimination (Iregui-Bohórquez et al. 2021).

However, there are still differences in the type and quality of education and the areas of knowledge (Rettberg et al. 2021). When observing the performance of women and men in the OECD standardized tests (Pisa Test) that measure students' skills in schools, it is possible to find for the Colombian context that boys have a better performance in mathematics skills by 28 to 30 points (Iregui-Bohórquez et al. 2021; Rettberg et al. 2021). On the other hand, girls perform better in the language skills measures by nine points (Rettberg et al. 2021). It is relevant to mention that these differences increase as the years of education pass and might help to explain future decisions such as college majors.

In conclusion, there is an evident improvement in the formal opportunities for women in the labor market in Colombia. The legal framework not only prohibits gender discrimination; it promotes positive discrimination - using tax cuts as an incentive - and protects women from discrimination derived from biological differences – maternity leave law. The country has important milestones such as the increase of women's enrollment rate in higher education, better participation rates in labor marker, and symbolic events such as women's participation in the national cabinet. However, there is evidence that the increase of women in the labor market is more prevalent in groups with lower levels of education (Amador Osuna et al. 2013). Regarding the job quality, women face worse conditions than men, and on average, women's income is lower than men's.

Figure 2: Women's participation in labor market by occupation



Notes: Authors' calculations using information from the National Department of Statistics (DANE, 2021).

4 Data

Based on the literature review of the gender pay gap and the Colombian context in terms of gender equity, the research question of this paper is: What are the determinants of the gender wage gap in the Colombian formal labor market? To answer this question, we use different databases on individuals who are part of the formal labor market in Colombia between 2012 and 2017. The Colombian Ministry of Education is responsible for the first data-set and consists of a longitudinal follow-up of higher education graduates on the labor market through social security contributions enquire. This database provides data on the salary of the individuals (wage variable), their years of experience (experience variable), the educational establishments and the college major (area variables) from which they graduated, and the cities where they work at the time of the report (region). In addition, it allows knowing the year of each of the observations (years variables).

The second database is a panel from the student loan public program of applicants and funded undergrads students. This information came from the national public organization in charge of providing, administering, and collecting student loans and disbursements (Icetex). From these data, we identified the individuals that receive government assistance programs to study (Financial Aid variable). Finally, the third data-set is managed by the public organization (Icfes) that designs and implement the national standardized test. It has cross-sectional data about students' takers and their performance in the assessment (SABER test variable). The union of these databases results in 2.49 million observations (1.39 million women and 1.1 million men), of which correspond to 829,665 different individuals over seven years with a reported wage greater than zero.

With this information, the model's functional form is:

$$Ln(wage)_i = f(experience_i, prog\_quality_i, social_i, area_i, SABER\_test_i, region_i, year_i) \quad (1)$$

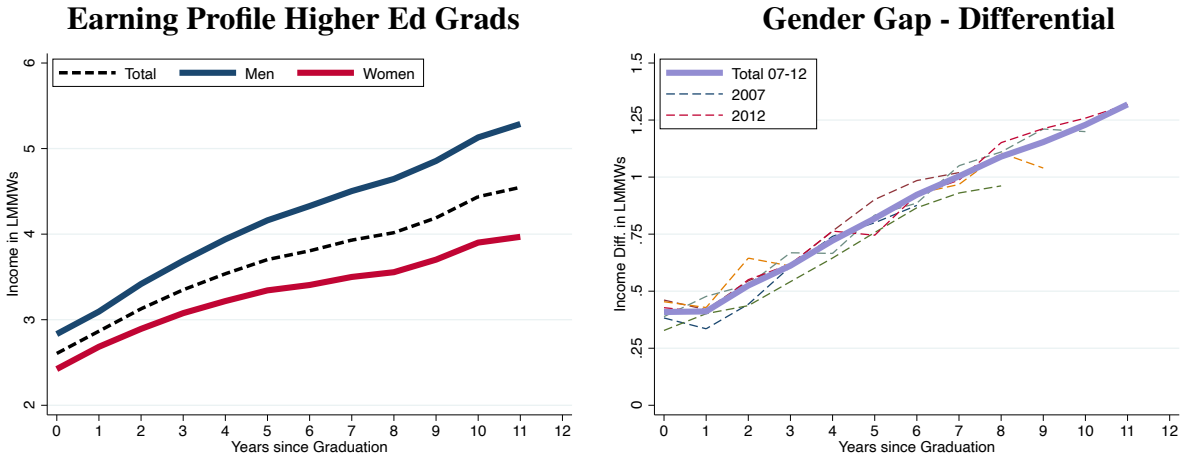
Where experience refers to the variables experience and experience squared and indicates the difference between the year of the report and the year of graduation. On the other hand, program quality (prog_quality) refers to a group of variables that measure aspects of program type and national standards. With the information of the higher educational institution, it was possible to cross-check the information with the Ministry's database to find out if they had high-quality academic accreditation (HEIs High Quality) and their educational methodology which could be in-person, remote or online (methodology).

Social refers to the social aspects of individuals are measured based on their social strata, mother's education, fathers' education, and whether they received state aid to access higher education programs. Meanwhile, area refers to how undergraduate academic programs (college majors) are grouped. The Colombian educational system is grouped into 13 categories: agronomy, fine

arts, education, health, economics and accounting, business, engineering, math and natural sciences, medicine, military or police, mineral and energy, and geology.

The other category of variables is the SABER test. This test is the standardized test conducted by the government to assess students' knowledge. At the same time, it is the examination used by undergraduate programs in their admission process. In this case, it is used the results of each individual and its polynomial forms of first, second, third, and fourth-degree are incorporated. Finally, region indicates the city where the individual is working. This paper uses the data of the country's five main cities. Table 1 has the descriptive statistics of each variable described above for the total population and each gender. In addition, it shows the test results to identify differences between the characteristics of men and women.

Figure 3: Earnings Profiles (OLE 2007-2012)



Notes: Authors' calculations using information from the Education Ministry (MEN) and the Graduate Labor Tracking System (OLE).

With the information described above, it is possible to observe trends in the data that signal the methodological approach. Figure 3 presents two graphs on the earnings profile of individuals. The first (left) shows the income profile of the graduates of higher education programs measured in minimum monthly salaries and its evolution over the years of experience. This result is discriminated by the average of the total number of individuals, women, and men. From this graph, it is possible to conclude that the average income of women is consistently below that of men and that the difference increases as experience grows.

Table 1: Descriptive Statistics: Complete Sample and Gender Split

Variable	General	Female	Male	p-value	Variable	General	Female	Male	p-value
Ln. Wage	1.0144	0.9409	1.1082	0.000	Area: Agronomy	0.0131	0.0080	0.0195	0.000
	(0.658)	(0.622)	(0.689)			(0.114)	(0.089)	(0.138)	
Experience	4.1299	4.1611	4.0901	0.000	Area: Fine Arts	0.0278	0.0266	0.0294	0.000
	(2.743)	(2.750)	(2.733)			(0.164)	(0.161)	(0.169)	
HEIs High Quality	0.2768	0.2572	0.3019	0.000	Education Sciences	0.1214	0.1458	0.0903	0.000
	(0.447)	(0.437)	(0.459)			(0.327)	(0.353)	(0.287)	
Methodology	1.0986	1.1153	1.0772	0.000	Health Sciences	0.0766	0.1135	0.0295	0.000
	0.298	(0.319)	(0.267)			(0.266)	(0.317)	(0.169)	
Social Strata	2.5668	2.5702	2.5618	0.000	Social Sciences	0.1331	0.1554	0.1045	0.000
	(0.914)	(0.908)	(0.923)			(0.340)	(0.362)	(0.306)	
Maternal Education	1.4886	1.4820	1.4984	0.000	Economics & Accounting	0.1248	0.1436	0.1010	0.000
	(0.501)	(0.501)	(0.501)			(0.331)	(0.351)	(0.301)	
Paternal Education	1.5833	1.5737	1.5975	0.000	Business	0.1751	0.1877	0.1590	0.000
	(0.494)	(0.495)	(0.491)			(0.380)	(0.390)	(0.366)	
Financial Aid	0.0411	0.0433	0.0384	0.000	Engineering	0.2686	0.1666	0.3986	0.000
	(0.284)	(0.291)	(0.274)			(0.443)	(0.373)	(0.490)	
SABER Test	260.25	282.88	226.66	0.000	Area: Math & Natural Sciences	0.0168	0.0167	0.0170	0.147
	(248.2)	(250.4)	(241.0)			(0.129)	(0.128)	(0.129)	
Reg: Bogota	0.3348	0.3286	0.3427	0.000	Area: Medicine	0.0347	0.0320	0.0381	0.000
	(0.472)	(0.470)	(0.475)			(0.183)	(0.176)	(0.192)	
Reg: Medellin	0.0911	0.0915	0.0905	0.005	Area: Military or Police	0.0000	0.0000	0.0000	0.000
	(0.288)	(0.288)	(0.287)			(0.004)	(0.002)	(0.005)	
Reg: Cali	0.0670	0.0689	0.0645	0.000	Area: Mineral & Energy	0.0063	0.0031	0.0105	0.000
	(0.250)	(0.253)	(0.246)			(0.079)	(0.055)	(0.102)	
Reg: Barranquilla	0.0368	0.0364	0.0374	0.000	Area: Geology	0.0017	0.0010	0.0025	0.000
	(0.188)	(0.187)	(0.190)			(0.041)	(0.032)	(0.050)	
Reg: Bucaramanga	0.0254	0.0263	0.0242	0.000					
	(0.157)	(0.160)	(0.154)						
Reg: Cartagena	0.0217	0.0219	0.0213	0.003					
	(0.146)	(0.146)	(0.145)						
Observations	2.49 Mlln	1.39 Mlln	1.1 Mlln		**Observations**	2.49 Mlln	1.39 Mlln	1.1 Mlln	

Note: The table presents the mean and standard deviations in parentheses of the variables included in the modeling. It reports the entire sample, as well as the gender splits' descriptive statistics. It also presents a comparison of means t-test across the two sub-samples.

On the other hand, the second graph (right) shows the difference in minimum wages of men's income minus women's and its evolution over time (describing work experience). This wage differential has a positive slope and grows as individuals' years of experience increase.

5 Methods

To answer the research question, we use two types of methodological approaches. The first analysis relies on the Mincer equation of income of the labor market following the functional form of equation 1. In this section, we run three log-lin regressions with fixed effects. In the first regression, we consider all the individuals, and we include a variable of gender to observe if it is significant to explain the individual income. The following regressions follow the characteristics for each gender group: women and men. These regressions seek to look at the determinants of individuals' wages and to observe whether there are differences between genders.

Subsequently, we did a counterfactual decomposition to observe the mean wage difference between the groups. Specifically, we use a two-stage Blinder Oaxaca decomposition (BO) to identify the wage difference and understand which part is explained by the group's differences in productivity characteristics. This paper uses the methodology established by Jann (2008) to decompose average log wage differentials (R) based on regression models using a counterfactual approach. Specifically, in this case, we are interested in the three-fold decomposition: endowments (E), coefficients (C), and interaction (I). With this information, the model's functional form is:

$$Y_M - Y_F = [E(X_M) - E(E_F)]'\beta_F + E(X_F)'(\beta_M - \beta_F) + [E(X_M) - E(E_F)]'(\beta_M - \beta_F) \quad (2)$$

where
$R = Y_M - Y_F$
$E = [E(X_M) - E(E_F)]'\beta_F$
$C = E(X_F)'(\beta_M - \beta_F)$
$I = [E(X_M) - E(E_F)]'(\beta_M - \beta_F)$

Using the three-fold Blinder Oaxaca, we can identify the groups' differences in the predictor (endowment effect), the contribution of differences in the coefficient (coefficient effect), and an interaction effect. This information allows us to identify the changes in the wage differences if women have the endowments and coefficient of men. Consequently, these results will enable us to determine what part of the average gender wage gap is explained by the endowments of each group and by labor market discrimination.

One of the drawbacks of the counterfactual decomposition method is that it uses aggregate measures and assumes homogeneous inequalities and socioeconomic characteristics (Bernat Díaz et al. 2007). Nonetheless, in this case, we are using a specific data set that includes information about Colombians in the formal labor market with higher education degrees. That implies that we focus on analyzing a particular population with some characteristics. As a result, the aggregate outcomes of this method answer our research question about the formal market and the gender wage gap in the higher-income population.

6 Results

In this section, we present the result of the two different analyses using the database we described before. Table 2 presents the results of the three log-lin regressions, which is the first approach to answer the research question. Confirming Mincer's theory, in the three regression models, the experience and the quality of education are statistically significant at 99% and have a positive relationship with individual income. Additionally, socioeconomic characteristics regarding the background of workers influence their salaries, such as social strata and the parent's educational level. Moreover, only when workers live in Bogotá or Medellín, the two largest cities in Colombia is their location statistically significant at 99% and positively impacts job remuneration. Otherwise, if they are in smaller cities such as Barranquilla and Bucaramanga, the effect on the income is negative regardless of the worker's gender.

Likewise, there are areas of undergraduate degrees that also influence remuneration. For example, agronomy, arts, education, economics, and business negatively affect salary. In contrast, engineering, military, mines, and geology increase compensation. These results show how college majors and occupations make a difference in the expected remuneration of individuals. Finally, it is essential to note that in the full model (column 1), the gender variable (female) is significant and negatively affects salary. In other words, with 99% confidence, it is possible to affirm that because an employee is a woman, she receives 6% less pay than a man. These results allow us to conclude that we have a gender difference in our database, and we can continue with the second analysis: counterfactual decomposition.

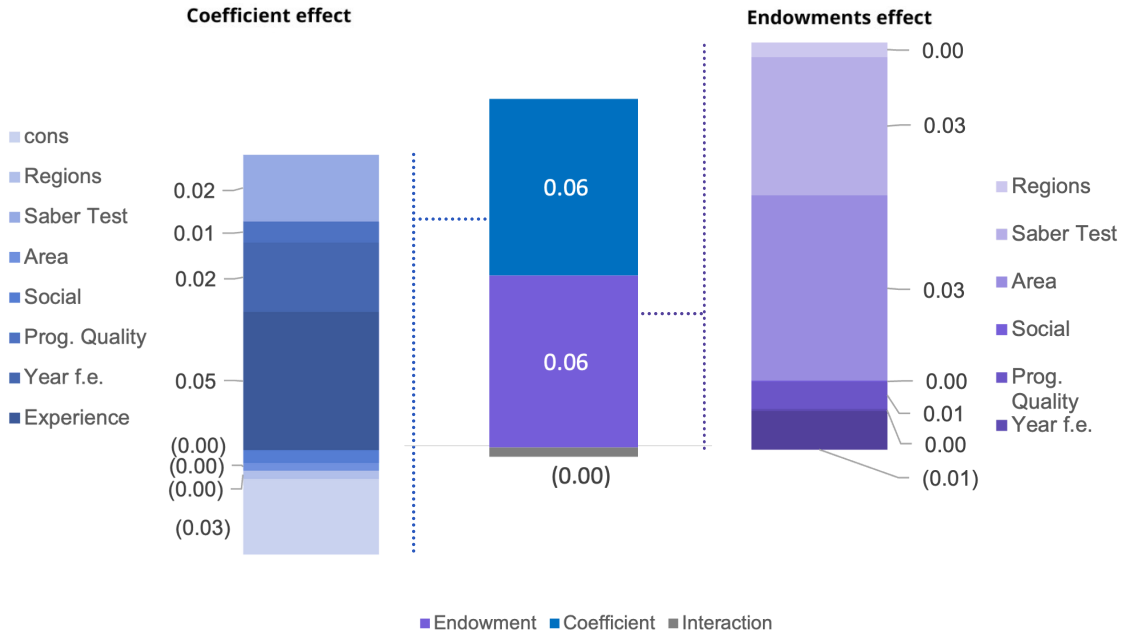
Table 2: Log Wages Regressions OLS: Complete Sample and Gender Split

	(1) Complete	(2) Female	(3) Male		(1) Complete	(2) Female	(3) Male
Experience	0.103***	0.0906***	0.120***	Area:Agronomy & Vet.	-0.283***	-0.308***	-0.268***
	(0.00856)	(0.00853)	(0.00826)		(0.0322)	(0.0399)	(0.0269)
Experience^2	-0.00589***	-0.00478***	-0.00724***	Area:Fine Arts	-0.280***	-0.242***	-0.314***
	(0.000815)	(0.000861)	(0.000730)		(0.0126)	(0.0143)	(0.0107)
HEIs High Quality	0.113***	0.114***	0.111***	Area:Education sciences	-0.309***	-0.297***	-0.278***
	(0.00916)	(0.0112)	(0.00747)		(0.0318)	(0.0320)	(0.0288)
Methodology	-0.0152	-0.0169*	-0.00940	Area:Health sciences	-0.0791***	-0.0498*	-0.0662**
	(0.0103)	(0.00910)	(0.0149)		(0.0249)	(0.0246)	(0.0248)
Social Strata	0.0508***	0.0485***	0.0541***	Area:Social sciences	-0.0967***	-0.0738***	-0.0907***
	(0.00498)	(0.00444)	(0.00605)		(0.0135)	(0.0126)	(0.0186)
Maternal Education	0.0123***	0.0205***	-0.00101	Area:Economics & Account.	-0.124***	-0.114***	-0.0972***
	(0.00335)	(0.00355)	(0.00378)		(0.0273)	(0.0265)	(0.0290)
Paternal Education	0.0129***	0.00964**	0.0176***	Area:Business	-0.0954***	-0.0860***	-0.0755***
	(0.00374)	(0.00379)	(0.00493)		(0.0148)	(0.0149)	(0.0147)
Financial Aid	0.00723*	0.00442	0.0114**	Area:Math & Nat. Science	-0.135***	-0.111***	-0.141***
	(0.00376)	(0.00387)	(0.00424)		(0.0289)	(0.0255)	(0.0343)
SABER Test	-0.00163***	-0.00172***	-0.00161***	Area:Medicine	0.292***	0.309***	0.295***
	(0.0000858)	(0.0000765)	(0.000114)		(0.0491)	(0.0454)	(0.0530)
SABER Test^2	0.00000486***	0.00000496***	0.00000515***	Area:Military or police	0.260	-0.188***	0.369*
	(0.000000335)	(0.000000296)	(0.000000460)		(0.265)	(0.00896)	(0.201)
SABER Test^3	-6.68e-09***	-6.63e-09***	-7.41e-09***	Area:Mineral and energy	0.704***	0.686***	0.709***
	(5.02e-10)	(4.52e-10)	(7.10e-10)		(0.0572)	(0.0724)	(0.0501)
SABER Test^4	3.29e-12***	3.19e-12***	3.71e-12***	Area:Geology	0.668***	0.671***	0.668***
	(2.50e-13)	(2.29e-13)	(3.63e-13)		(0.0334)	(0.0386)	(0.0328)
Reg:Bogota	0.130***	0.126***	0.136***				
	(0.0239)	(0.0237)	(0.0246)				
Reg:Medellin	0.102***	0.125***	0.0664***	Female	-0.0612***		
	(0.0226)	(0.0239)	(0.0215)		(0.00450)		
Reg:Cali	-0.0403*	-0.0302	-0.0543**				
	(0.0200)	(0.0199)	(0.0206)				
Reg:B/quilla	-0.128***	-0.131***	-0.122***				
	(0.0186)	(0.0183)	(0.0197)				
Reg:B/manga	-0.100***	-0.102***	-0.0946***				
	(0.0175)	(0.0168)	(0.0193)				
Reg:C/tagena	0.00647	-0.0100	0.0306*				
	(0.0173)	(0.0175)	(0.0175)				
R^2	0.198	0.191	0.192		0.198	0.191	0.192

Note: The table summarize the results of models for the entire sample and the splits by gender. All models include a constant and years fixed effects which are not reported. Standard errors clustered at the geographic department level in parentheses. Significance levels reported at: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4 show the counterfactual decomposition results using the Blinder Oaxaca method. The graph shows that there is indeed a wage differential explained by gender (middle bar in Figure 4). In this case, the first stage of the decomposition indicates that for every dollar received by a man, a woman gets on average 88 cents.

Figure 4: Blinder-Oaxaca Gender Wage Decomposition



Notes: Authors' calculations using information from the Education Ministry (MEN) and the Graduate Labor Tracking System (OLE).

We present the result of the second part of the analysis in two different ways. Figure 4 shows the counterfactual decomposition results using the Blinder Oaxaca method. The graph shows that there is indeed a wage differential explained by gender (middle bar in Figure 4). In this case, the first stage of the decomposition indicates that for every dollar a man receives, on average, a woman gets 88 cents. Additionally, Table 3 summarizes the results of the same method but in their exponential form, with their standard errors and p-values.

The second stage of the analysis shows the determinants of this wage differential. The first effect, endowments (right bar in Figure 4, explains the differences in income due to individual characteristics. It indicates whether employees' personal choices (e.g., area of study), socioeconomic conditions, or profile influence their salary. In this case, the endowment effect explains 58% of the wage differential. In other words, this percentage explains the mean increase in women's wages if they had the same characteristics as men. Moreover, it is possible to identify which variables have the most significant influence on this effect. In this case, the endowment effect of years

of experience negatively explains the gender wage gap. In contrast, the area of education explains 60% of the endowment effect, and standardized test scores explain 38%.

Table 3: Log Wages Oaxaca Decomposition: Complete Sample and Gender Split

	Total Differential		
	est	s.e.	p-value
Prediction male wage	2.6706	(0.0894)	0.0000
Prediction female wage	2.3658	(0.0698)	0.0000
Wage gap	1.1289	(0.0087)	0.0000

	Endowments			Coefficients			Interaction		
	est.	s.e.	p-value	est.	s.e.	p-value	est.	s.e.	p-value
Experience	0.9928	(0.0008)	0.0000	1.0467	(0.0038)	0.0000	0.9984	(0.0002)	0.0000
Area	1.0351	(0.0041)	0.0000	0.9975	(0.0045)	0.5760	1.0016	(0.0018)	0.3936
Saber Test	1.0260	(0.0014)	0.0000	1.0223	(0.0065)	0.0006	0.9963	(0.0008)	0.00001
Program Quality	1.0051	(0.0007)	0.0000	1.0069	(0.0115)	0.5442	0.9997	(0.0003)	0.4005
Strata	1.0001	(0.0007)	0.8610	0.9957	(0.0100)	0.6654	0.9997	(0.0001)	0.0176
Regions	1.0027	(0.0008)	0.0012	0.9974	(0.0071)	0.7134	1.0006	(0.0002)	0.0050
Year f.e.	1.0004	(0.0009)	0.6409	1.0232	(0.0133)	0.0762	1.0005	(0.0003)	0.0617
Constant				0.9753	(0.0179)	0.1725			
Total	1.0633	(0.0060)	0.0000	1.0651	(0.0078)	0.0000	0.9968	(0.0018)	0.0779

Note: The table summarize the results of the Oaxaca-Blinder decomposition model for the entire sample. The results are displayed in their exponential form.

The second effect is the coefficient effect (left bar in Figure 4), which explains discrimination in the labor market by identifying the wage gap attributed to gender alone. For this, the model explains the change in women's wages using the coefficients of men with women's characteristics. In the case of the Colombian data, this effect explains 40% of the wage gap. Specifically, it is seen that there are some negative values, such as social characteristics, cities, and a constant value. In contrast, the area of studies explains 119% of this effect, years of experience 69%, and quality of education 63%.

7 Discussion

As we mentioned, this paper has two parts analysis using the data to explore the gender wage gap and its determinants in the formal Colombian market between 2012 and 2017. We employed socio-demographic information, the individual's profile, and occupation to document and measure the gender wage gap in Colombia specific to higher education graduates. Mainly, we ran Blinder Oaxaca's two-stage decomposition model and established that for every dollar a man receives, a woman gets 88 cents. In addition, we document some of the most important determinants and how they weigh more heavily in explaining this gap.

One of the main contributions of this paper is the role of education as a mechanism to reduce gender differences. As we mentioned before, authors such as Weichselbaumer and Winter-Ebmer (2005), claimed that better labor market endowments reduce gender discrimination. The logic behind this statement is straightforward; if women become better educated and trained, they will improve their profile. Therefore, it is possible to conclude that they will get opportunities to be part of and attach to the market (Weichselbaumer and Winter-Ebmer 2005). Nonetheless, when we focus the analysis on workers with higher education degrees, the gender wage gap still is part of the labor market characteristics.

In Colombia, by 2012, the higher education enrollment rate was less than 44% (Ministry of Education 2022), and less than 50% of the labor force had a formal job (DANE 2012). Taking this information into account, we can conclude that we are observing a specific group of individuals with both characteristics in our database. Moreover, following the conclusions of the U-shape distribution of the gender wage gap in Colombia (Badel and Peña 2010; Hoyos et al. 2010), we can affirm that the population that is part of the formal job market and has higher education degree is more likely to have better incomes on average compare with the rest of population. Therefore, we can affirm that education did not completely eliminate the gender gap, which explains a significant percentage of the income difference in the Colombian case.

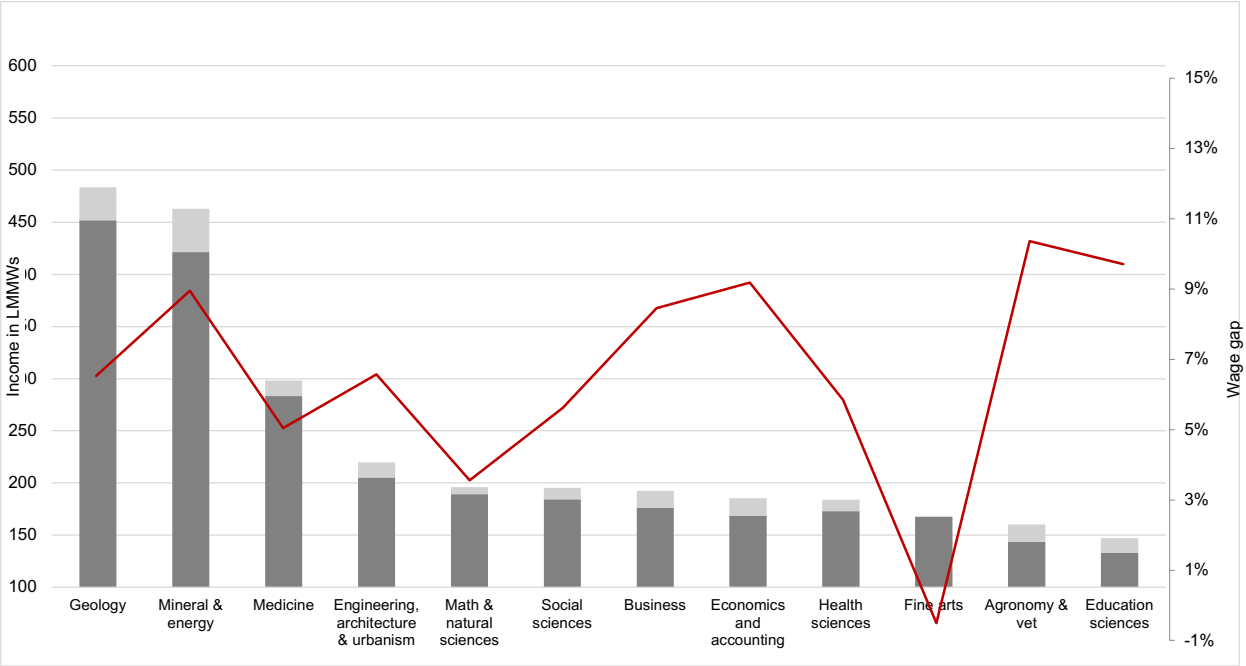
We found that 58% of this difference in income is due to employees' endowments. This result means that the experience and choices of women and men graduates drive the gap to that extent. It refers to the characteristics of each individual that lead to the existence of a difference. For example, it explains how areas of education account for 60% of the endowment-specific difference. This finding aligns with the theory regarding education's existing differences due to glass walls and gender assignments.

Following this result, we did an exercise to understand the differences between the academic areas. With the estimated coefficients of the Oaxaca-Binder decomposition, we forecast the women's wages with men's and women's coefficients (see figure 5). The outcome of this process allows us to compare the differences in wages across the areas due to the average endowments of each gender. More specifically, we calculated the year average income of women with women's estimated coefficient (darker grey area figure 5) and the gap compared with the same outcome with men's estimated coefficients (lighter grey area in figure 5). The red line in the figure present the average yearly gap explained by the different estimated coefficients.

As a result of this process, we found areas such as mineral and energy, business, agronomy, and education with a gender gap higher than 9%. In other words, only because of each gender's endowments can we observe some jobs that pay almost 10% less to women's workers. Conversely, areas such as math and fine arts have a gap of less than 4% or negative. Nonetheless, when

Figure 5: Women's wages forecast by area

(Five years experience)



Notes: Authors' calculations using information from the Education Ministry (MEN) and the Graduate Labor Tracking System (OLE).

we observe the number of legal minimum wages (LMW) per year, it is evidently the difference between the income magnitude between the areas. Whereas a professional in mineral and energy gains up to 476 LMW in a year, a professional in fine arts accumulates 36% of that income in the same period (172 LMW). For this reason, and to define more adjusted policy alternatives, it is crucial to identify whether these wage differences exist across sectors. Suppose it is confirmed that the sectors with the highest participation of women systematically have the best wages. In that case, governments could review whether there are conditions for market failures to intervene and reach optimal results.

On the other hand, the coefficients' effect explains the wage gap. This effect explains 40% of the wage gap and cannot be explained by differences in choices and endowments. We could think about it as a proxy measure of discrimination. With this result, it can be observed how being a woman reduces the salary, just for the fact of being a woman. Again, the study area is the most influential factor showing that the wage gap is also within sectors. According to our approximation, the second most important factor is experience after graduation, indicating that one more year in the working place for a man benefits him more than a woman. This situation is repeated with the quality of education, which disproportionately benefits men.

While our results are the first to shed light on the differences in wages across genders in a

developing country, our data has some limitations. First, we only observed data from formal employees and self-reported income for independent workers. After higher education, most of the graduates (75% in 2012 ((DANE 2012).) contribute to the social security system, so this is less of a concern in comparison to other analyses with Colombian data that look at administrative registers and different education levels. However, in our dataset, 3% declare to earn the monthly minimum wage, and 55% claim to make three or fewer monthly minimum wage (see figure 5). This result could constitute an under-report and hence, measurement error.

References

Amador Osuna, D., R. Bernal Salazar, X. Peña Parga, et al. (2013). El aumento en la participación laboral femenina en colombia:¿ fecundidad, estado civil o educación?

Archila, M. (2014). Aspectos sociales y políticos de las mujeres en colombia, siglos xx y xxi. *Repositorio Universidad nacional de Colombia*.

Badel, A. and X. Peña (2010). Decomposing the gender wage gap with sample selection adjustment: Evidence from colombia. *Documento CEDE* (2010-37).

Bank, W. World development report 2012 : Gender equality and development.

Bernat Díaz, L. F. et al. (2007). ¿ quienes son las mujeres discriminadas?: enfoque distributivo de las diferencias salariales por gnero. *Borradores de Economía y Finanzas; No. 13-Diciembre 2007* (13), 1–31.

Bertocchi, G. and M. Bozzano (2020). *Gender gaps in education*. Springer.

Black, D. A., A. M. Haviland, S. G. Sanders, and L. J. Taylor (2008). Gender wage disparities among the highly educated. *Journal of human resources 43*(3), 630–659.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature 55*(3), 789–865.

Colombian Women's Observatory (2021). Power and decision making.

Connell, R. (2006). Glass ceilings or gendered institutions? mapping the gender regimes of public sector worksites. *Public administration review 66*(6), 837–849.

DANE (2012). Empleo informal y seguridad social - históricos.

DANE (2021). Brecha salarial de genero en colombia. *Nota Estadistica*.

Elias, J. and H. Ñopo (2010). The increase in female labor force participation in latin america 1990-2004: Decomposing the changes. *Washington, DC, United States: Inter-American Development Bank. Mimeographed document*.

Feeney, M. K. and L. Camarena (2021). Gender, race, and diversity values among local government leaders. *Review of Public Personnel Administration 41*(1), 105–131.

Galor, O. and D. N. Weil (1993). The gender gap, fertility, and growth.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review 104*(4), 1091–1119.

Goldin, C. and L. F. Katz (2016). A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics 34*(3), 705–746.

Goldin, C., S. P. Kerr, C. Olivetti, and E. Barth (2017a). The expanding gender earnings gap: Evidence from the lehd-2000 census. *American Economic Review 107*(5), 110–14.

Goldin, C., S. P. Kerr, C. Olivetti, and E. Barth (2017b). The expanding gender earnings gap: Evidence from the LEHD-2000 census. *American Economic Review 107*(5), 110–14.

Hoyos, A., H. Ñopo, and X. Peña (2010). The persistent gender earnings gap in colombia, 1994-2006. *Documento CEDE* (2010-16).

Huber, M. and A. Solovyeva (2020). On the sensitivity of wage gap decompositions. *Journal of Labor Research 41*(1), 1–33.

Iregui-Bohórquez, A. M., L. A. Melo-Becerra, M. T. Ramírez-Giraldo, and A. M. Tribín-Uribe (2021). El camino hacia la igualdad de género en colombia: todavía hay mucho por hacer.

Jann, B. (2008). A stata implementation of the blinder-oaxaca decomposition. *Stata journal 8*(4), 453–479.

Kunze, A. (2008). Gender wage gap studies: consistency and decomposition. *Empirical Economics 35*(1), 63–76.

Mandel, H. and M. Semyonov (2014). Gender pay gap and employment sector: Sources of earnings disparities in the united states, 1970–2010. *Demography 51*(5), 1597–1618.

Männasoo, K. (2022). Working hours and gender wage differentials: Evidence from the american working conditions survey. *Labour Economics 76*, 102148.

Martínez-Restrepo, S., J. C. Mejía, E. Enríquez, et al. (2015). Women's empowerment among the extremely poor: Evidence from the impact evaluation of red unidos in colombia.

Mincer, J. and S. Polachek (1974). Family investments in human capital: Earnings of women. *Journal of political Economy 82*(2, Part 2), S76–S108.

Ministry of Education (2022).

Olivetti, C. and B. Petrongolo (2016). The evolution of gender gaps in industrialized countries. *Annual review of Economics 8*, 405–434.

Ramírez Bustamante, N., A. M. T. Uribe, C. O. Vargas, et al. (2015). *Maternity and Labor Markets: Impact of Legislation in Colombia*. Citeseer.

Rettberg, A., A. Acosta, M. d. P. López-URibe, P. Neira, L. Salazar Escalante, C. Sánchez León, and M. G. Vargas (2021). Balance de la igualdad de género en colombia: un panorama.

Sabharwal, M. (2015). From glass ceiling to glass cliff: Women in senior executive service. *Journal of Public Administration Research and Theory 25*(2), 399–426.

Sabogal, A. (2009). Brecha salarial por género y ciclo económico en colombia. *Bogota, Colombia: Universidad de los Andes. Mimeographed document*.

Sloane, C. M., E. G. Hurst, and D. A. Black (2021, November). College majors, occupations, and the gender wage gap. *Journal of Economic Perspectives 35*(4), 223–48.

Weichselbaumer, D. and R. Winter-Ebmer (2005). A meta-analysis of the international gender wage gap. *Journal of economic surveys 19*(3), 479–511.

Wills, M. E. (2007). *Inclusión sin representación: la irrupción política de las mujeres en Colombia (1970-2000)*. Editorial Norma.