Agosto 2021

WP2-2021-AG1
N° de serie

DOCUMENTO DE TRABAJO

Decomposing the Gender Pay Gap in Colombia: Do Industry and Occupation Matter?

Autores

Tania Camila Lamprea   Andrés García-Suaza

COLOMBIA CIENTÍFICA
Conocimiento Global para el Desarrollo

ALIANZA EFI
Economía Formal e Inclusiva

Universidad del Rosario | Facultad de Economía

SERIE DOCUMENTOS DE TRABAJO

No. 272

Agosto de 2021

Decomposing the Gender Pay Gap in Colombia: Do Industry and Occupation Matter?

Tania Camila Lamprea-Barragan

Andrés García-Suaza

Universidad del Rosario, Facultad de Economía
Cl. 12c #4-59. Bogotá, D. C., Colombia
PBX: (031) 2970200 ext. 4163

facultadeconomia@urosario.edu.co
https://www.urosario.edu.co/Facultad-de-Economia/Inicio/
https://repository.urosario.edu.co/handle/10336/20291

DECOMPOSING THE GENDER PAY GAP IN COLOMBIA: DO INDUSTRY AND OCCUPATION MATTER?

Tania Camila Lamprea-Barragan♣ Andrés García-Suaza♦

Abstract

This paper aims to quantify at which extent industry and occupation characteristics explain the gender pay gap in Colombia. To quantify the role of these factors we perform counterfactual decomposition methods that allow to split the total gap into the contribution of the gender share of employment at the industry level, the demographic composition and the characteristics pay premia. This is possible by adapting the classical Oaxaca-Blinder decomposition to a two-step procedure, which is illustrated through Monte Carlo simulations. Using Colombian data for 2019 and exploiting the heterogeneity at the industry and the occupation level, findings suggest that the three components shape the gender pay gap. While differences in returns are the main force driver of the existing gap, the gender employment share, and the composition of workers across industries and occupations provide mixed results.

Keywords: Gender pay gap; decomposition methods, sorting.

JEL code: J16, J31, J24.

♣ Advisor, Economics Study Direction, National Planning Department. E-mail: tania.lamprea01@gmail.com

♦ Professor, School of Economics, Universidad del Rosario. E-mail: andres.garcia@urosario.edu.co.

We thank the participants at 1er Congreso Internacional en Empleabilidad y Mercado laboral de la Alianza del Pacífico: La Importancia de la Integración for their comments and suggestions. This working paper is funded by the Colombia Científica-Alianza EFI Research Program, with code 60185 and contract number FP44842-220-2018, funded by The World Bank through the call Scientific Ecosystems, managed by the Colombian Ministry of Science, Technology and Innovation. Errors, opinions and omissions are ours and do not compromise the institutions.

1. INTRODUCTION

Although the gender pay gap has declined in the past two decades, gender equality in the labor market is still one of the main challenges for policy makers. The Global Wage Report (ILO, 2018), based on data for salaried workers in 73 countries, estimates that the global gender gap stands at around 16%, and shows important differences across countries. The drivers of the pay gap have been widely studied in literature. For instance, from the human capital perspective, variables such as schooling and tenure have resulted important for explaining gender differentials (Mincer & Polachek, 1974; O'Neill & Polachek, 1993; Blau, 1997; and Altonji & Blank, 1999). Nevertheless, such variables are not enough to have a complete picture of the total earnings differences between men and women. In fact, an important portion of the gap remains unexplained.

Recent evidence has shown that, beyond human capital, job characteristics such as type of occupation (e.g., being salaried worker or self-employed) and industry play a relevant role in accounting for gender pay gap (Blau & Khan, 2017). For that matter, previous work has pointed out that gender gaps vary across occupations (see, e.g., Eastough & Miller, 2004; Lechmann & Schnabel, 2012), i.e., earnings are different by gender within salaried workers and self-employed. Besides, studies devoted to investigating the importance of industry have estimated that it explains an important portion (more than one-third, in general) of the total gender gaps (Blau & Khan, 2017; Sin et al, 2020). Therefore, due to the fact that women are concentrated in small businesses, activities as retail and services and informal jobs (see Morton et. al, 2014 for a further discussion), which reinforce disparities in earnings, discussion about the role of gender sorting into industry and occupations has become a relevant.

Hence, this paper aims to study the gender pay gaps for the case of Colombia and quantify the contribution of occupation and the within and between industry patterns. To do so, using information from household surveys 2019, we analyze earnings differences for salaried workers and self-employed separately and implement decomposition techniques which allow us to analyze the contribution of each factor to the total gap. Colombia is an interesting case of study. An important reduction of the gender gap has been observed in Colombia, e.g., Ribero & Meza (1997) report a reduction from 36% to 21% between 1970s and 1995, while Fernandez (2006) estimates that in 13% using data for 2003. In turn, Tenjo & Bernat (2017) conclude that the gap has been diminishing in the recent years, but its magnitude is still remarkable (7,05% for the main cities in Colombia in 2017). Although this decreasing trend, the Global Gender Gap Report 2021 (WEF, 2021) presents a multidimensional index of gender gaps at the country levels which shows that Colombia was the country highest drop comparing the two most recent rankings, 37 positions. This remarks the relevance of deeply understand the causes and consequences of the gender gaps, which are mainly expressed in the labor market outcomes.

Furthermore, both a higher participation of women among self-employment and differences in the gender share across industries are prevalent in the Colombian labor market. According to the data from the household surveys (GEIH for its acronyms in Spanish), more than half of the total female employment is in the services sector, while building is a male-dominated industry. It may have implications in gender earning gaps, because women who work in services are consistently paid less than men, while the low proportion of women in building industry are paid better than men.

This paper relates to the literature on gender sorting into occupations and industry. In this line, the fact that men and women work in different jobs has been explained as a consequence of differences in preferences to perform activities at work, comparative advantages with respect to the occupation or the jobs amenities[1] (e.g., flexibility in working hours) that individuals' value differently (Blau et al. 2013; Goldin, 2014; Baker & Cornelson 2018; Cortes & Pan, 2018; Gallen et al., 2019; Das & Kotikula; 2019). For instance, Goldin (2014) argues that the fact that men are employed into higher paying industries, where women continue to be underrepresented is a crucial cause of the gender gap. This is consistent with the idea that firms prefer long working hours rather than human capital, so that women pay a penalty because their high preferences for flexibility, and so decide to work comparatively fewer hours and as self-employed (Nightingale, 2019).

[1] Other channels such as wage bargaining, labor turnover, self-selection and job search frictions has been also proposed as drivers of the persistence in wage gender gaps (see c.f. Krueger & Summers, 1988; and Gibbons & Katz, 1992; Babcock & Laschever, 2003).

Although some studies have documented that self-employment might become in a female employment alternative with more equal earnings, the pay gap is still present as women value more flexible arrangements that facilitate to balance family-work responsibilities, which tend to pay less (Budig, 2006; Leung, 2006; Craig, 2012; Leuze & Strauß, 2016). In fact, Gallen et al. (2019) present evidence on the women sorting into lower-hours workplace, while Clain (2000), Hundley (2001) and Walker (2009) report a strong sorting across occupations by comparing socioeconomic characteristics of self-employed women and men. For the case of Latin American countries, Atal et al. (2010) found that occupation status plays an important role into the unexplained gap. This turns out in that the gender pay gap is higher among self-employed, which are mainly driven by hours worked and family background[2] (see Eastough & Miller, 2004; Lechmann & Schnabel, 2012). Hence, the reasons why workers choose to be self-employed are rather than diverse. Workers might decide rationally to work as self-employed according to their abilities and capacities of generating income, but also, they can be forced since of salaried workers sector is rationed and do not provide enough alternative. These two factors, exclusion and exit, are actually the main causes of informality in developing countries (Perry et al. 2007).

Regarding the influence of industry on gender gaps, two factors can be considered. First, the gender composition varies between industries, and so a pay gap is observed even if there are no gender differences in earnings within industries[3]. Secondly, industries also might pay differently by gender, not only as a consequence of discrimination but also because of the segregation in particular tasks, which generates an additional source of gender gap. Allen & Sanders (2002) address this question and find that social services, commerce and hotels and business services industries are typical female-dominated jobs.

Regarding these possible sources of earnings differentials, Hodson & England (1986), Fields & Wolff (1995) and Gannon et. al. (2007) provide estimates of the contribution of industry to the gender gap by using counterfactual decomposition techniques. The general results indicate that, both employment distribution and idiosyncratic industry's pay

[2] Further research has shown that the observed gap among self-employed is also related to capital market imperfections as women face stronger liquidity constraints (see Clain, 2000; and Hundley, 2001).

[3] For instance, in an economy with two industries that pay equally to men and women, but one of the industries is more intensive in male jobs and pays higher salaries, there will be observed a gender gap at the aggregate level.

structure, are important components. These two components together explain up to one-third of the total gap. Importantly, this relation between inter-industry wage structure and gender gaps has been documented for the case of the US and European countries, but the evidence for developing countries is still limited, and country level diagnosis is required to inform policy decisions (Das & Kotikula, 2019).

We use a similar approach by implementing counterfactual decompositions (see Fortin et. al., 2011, for a general introduction to decomposition methods in economics). In particular, we adapt the classical Oaxaca-Blinder decompositions (Oaxaca, 1973; Blinder, 1973) to a two-step method that allows to disentangle the sources of gender gap associated to occupation and industry, and in addition, avoids identification issues related to the omitted group problem (Oaxaca and Ransom, 1999). This identification problem appears as a consequence of the choice of a reference category, which generates interpretation issues of the decomposition results. This problem has been largely discussed in the case of inter-industry gender wage differentials (see Haisken-DeNew & Schmidt, 1997; Reilly & Zanchi, 2003; Yun; 2006; and Lin, 2007), where the conventional practice is to normalize the categories or imposing coefficient restrictions.

The first step of our empirical strategy consists of an aggregated decomposition that represents the total gender gap as a weighted sum of the differences of gender composition of employment and the average industry pay gap. In the second step, we implement Oaxaca-Blinder (OB) decompositions on the average pay gap for each industry. From this, we provide a detailed decomposition that considers three components: the effect of the female-dominated patterns across industries, the effect of the population composition and the characteristics' premia. For the two latter, we also exploit how the heterogeneity across industries and occupations contribute to the total gap. To illustrate how our proposed method account by the omitted group problem, we provide different scenarios using Monte Carlo simulations.

As control variables, we consider socioeconomic and job characteristics such as age, education, hours worked, informality status, among others. Besides, to make proper statistical inference of resulting components, confidence intervals are estimated using bootstrapping techniques as suggested by Chernozhukov et. al. (2013). Our estimates, suggest that there are important differences in the gender pay gap between salaried workers and self-employed. The magnitude of the total gap is determined by a higher gap for self-employed (63.6%). Moreover, when interacting occupation and industry to explain the gap, we obtain that the gender employment share plays mixed role as it tends to reduce the gap for salary workers and expand the gap for self-employed. Overall, the difference in the premia, i.e., the structure effect, is the main driver of the gender gap. Hence, dropping this component, a pay gap of 4% in favor of women would be observed for salaried workers, and that for self-employed would be reduced to 27.25%.

The heterogeneity across industries is also an important source of earning differences. Eliminating the differences in the population composition across industries would decrease the gap for salaried workers, but that would increase for self-employed, implying that sorting into industries to this group contribute to have a lower gap. This means that the flexibility provided by self-employed jobs has an important cost in term of pay gap, which is partially vanish by the heterogeneous workers composition in socioeconomic characteristics. In the case of the salaried workers, these heterogeneities play a minor role albeit differences in composition and return across industries tend to increase the gap. In sum, these findings contribute to the policy making discussion regarding the relevance reducing differential treatment of men and women but also of more flexible arrangements, vocational programs, and care economy policies as strategies to reduce the gender gap.

The rest of the paper is organized as follows. The second section describes the data and presents the general context of gender gaps by industry and occupation. The third section briefly introduces the Oaxaca-Blinder type decomposition inter-industry gender wage differentials, presents the proposed decomposition method and some simulation showing possible scenarios of study. The fourth section discusses the main results while the fifth considers some concluding remarks.

2. DATA AND DESCRIPTIVE STATISTICS

To measure the gender pay gap and their main components, we use the Colombian household survey, named the Great Integrated Household Survey (GEIH as its acronym in Spanish), for 2019. It is a monthly-based nationwide survey that provides information about labor conditions, family structure, household income including labor income and health conditions. This survey is representative for both urban metropolitan areas and rural areas. We focus on urban labor market, considering workers located in 13 main metropolitan areas. This is equivalent to 150,977 observations, which accounts for 9'762,372 employees, a half of the total employment in Colombia. In this sample, 60.4% of individuals are salaried workers and 39.6% correspond to self-employed.

A descriptive analysis reveals important differences in the labor market outcomes by gender. In 2019, the unemployment rate for men was 9.6%, 3.4 p.p. lower than the observed for women[4]. In terms of labor income (see Table 1), the average men's earnings are 1.46 million Colombian currency (around 450 USD[5]), 13.9% higher than for women. Comparing with data from ILO (2018)[6], this result is slightly lower than the global average gender gap (16.5%), and close to the reported values for Ireland and Bulgaria, and higher than other countries of the region as Uruguay, Mexico, Brazil and Peru. Table 1 also presents other statistics of the labor earnings distribution by gender that show how the gender gap varies over earning distribution. For instance, the gender gap in the median is 7.0% and it decreases over earnings distribution from 59.3% in the bottom to 0.3% in the upper tail, in the quantile 90%. Interestingly, the gap in the hourly income is also close to zero, which indicates that number of hours is an important component of the gender gap.

Table 1. Descriptive statistics of the gender pay gap

Variables	Total	Men	Women	Gap
Average	1,362.78	1,452.20	1,250.10	13.9%
Median	945.00	1,002.51	932.06	7.0%
Quantile 10%	270.97	405.31	165.09	59.3%
Quantile 90%	2,520.89	2,526.53	2,518.68	0.3%
Standard deviation	607.93	892.98	782.90	12.3%
Hourly average	8.43	8.45	8.41	0.5%

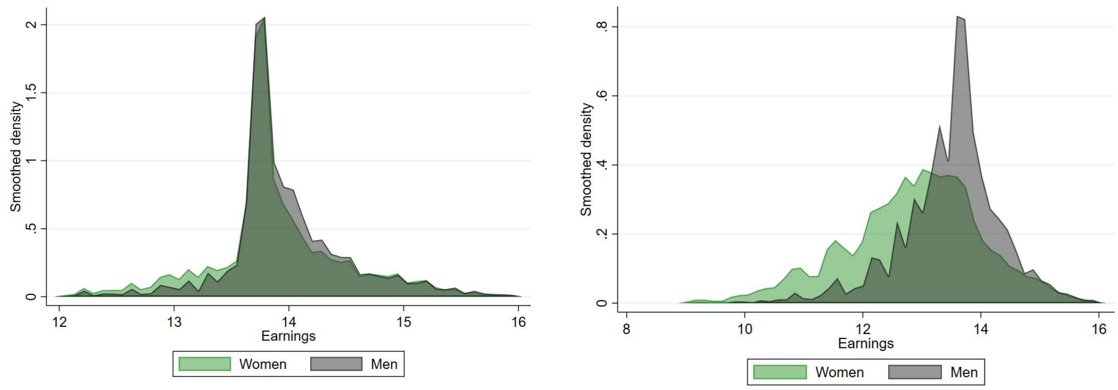
Source: GEIH 2019. Own calculations. Earnings values expressed in thousand Colombian pesos

[4] As consequence of the Covid-19 pandemic, this gap has increase importantly.
[5] To have some reference, exchange rate in 2019 was 3,277.14 Colombian pesos as for December 31th.
[6] It is important to notice that this report use information only for salaried workers.

There are also patterns of gender gap related to the occupation, as shown in the kernel densities in Figure 1, where higher differences are observed for self-employed. The gap for self-employed is 31.7% which significantly higher than that for salaried workers (5.5%). This is consistent with the fact that there is an important participation of women among low paid self-employed (see Table 2). In the bottom 25% less paid self-employed, women account for 67.4% of the total workers, while in the top 25% of the same group, the participation is 35.2%. Table 2 also shows that the share of men and women seems to be more stable across the distribution of labor income of salaried workers. While for the self-employed, the bottom part of earnings distribution is female-dominated, but the upper part of the distribution of this occupation is male-dominated.

Figure 1. Earnings distribution by job type

a. Salaried workers b. Self-employed



Source: GEIH 2019. Own calculations.

Table 2. Gender share across earnings distribution by occupation. Gender share across earnings distribution by occupation

Variables	**Salaried workers**			**Self-employed**		
	Total	Bottom 25%	Top 25%	Total	Bottom 25%	Top 25%
Women	43.70%	48.2%	44.3%	44.4%	67.4%	35.2%
Men	56.30%	51.8%	55.7%	55.6%	32.6%	64.8%
Difference	12.6%	3.6%	11.5%	11.2%	-34.8%	29.6%

Source: GEIH 2019. Own calculations. Each value corresponds to the percentage of men or women in a particular quartile of the earnings distribution.

Similarly, we also examine how industry relates to the gender gap. To this end, workers are classified into 7 group of industries[7]: Manufacturing (including services and utilities); Building; Commerce; Transport and communication; Professional services[8]; Public administration[9] and Other services (including entertainment, artistic, recreation, and similar activities). First, estimating the industry size in terms of employment, we observe that Commerce and Manufacturing provide almost half of the total employment (29.1% and 15.9%, respectively); while the contribution of the other industries is below 15%, i.e., Public administration (14.5%); Professional services (14.6%); Transport and communication (10.7%); Building (7.8%) and Other services (7.4%). Industries also differ in the composition between salaried workers and self-employed. For instance, Commerce has the greatest proportion of workers in both cases. Nonetheless self-employed percentage in Other services and Transport and communication is above salaried workers. On the other hand, Building has a very low participation in workers for both cases.

Gender composition within industries is also quite dissimilar (see Figure 2). Among salaried workers, all industries are male-dominated except Public administration and Other services where women are 62.0% and 51.2%, respectively. For the rest of the industries is remarkable that Building (88.8%) and Transport and communication (72.9%) are male-dominated. In the case of self-employed, the composition is more balanced, women are more representative in three sectors: Public administration; Professional services; and Other services. In turn, the proportion of men is more pronounced in Building (97.7%) and in Transport and communication (93.3%). This is a strong evidence of the sorting into industry and occupation.

To summarize how gender gap is affected by industry and occupation, we compute the raw labor earnings difference across the combination of these characteristics. Figure 7 (see Annex) implies that the gender gap has a high variation, although it is lower for salaried workers. Indeed, the higher gap for this group is 23.0% in Public administration,

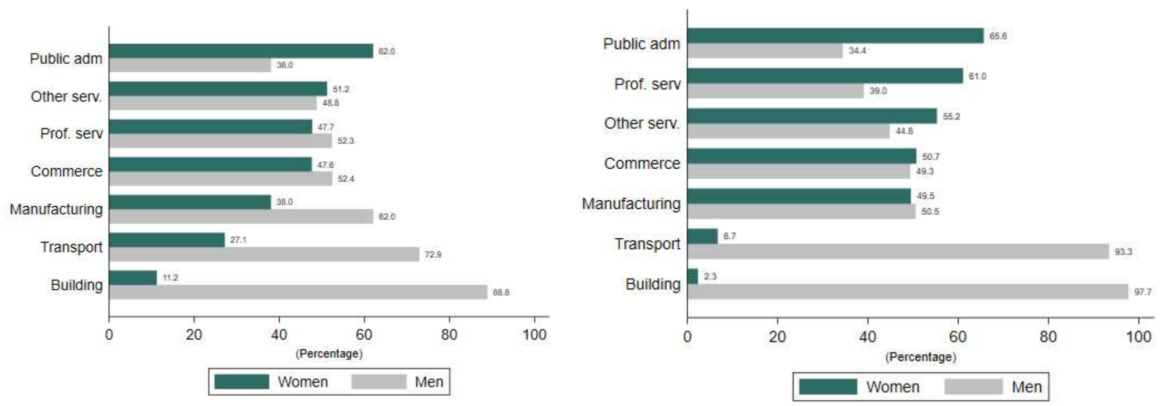
[7] Agriculture, and mines and quarries, were not included in this analysis given their low representativeness in urban areas.
[8] The industry is composed by professional and scientific activities, finance activities and real estate activities.
[9] The industry is composed by public and defense administration, education, services for the health.

while among self-employed the higher gap is 55.0%, in the case of Professional services. There are some similarities in the ordering of the industries, mainly for those where the gender gap is the lowest. Specifically, Building is the only industry with negative gap, however it is also the industry with lowest female participation. In turn, Transport, a male-dominated industry, has a negative gap for salaried workers, and is the second lowest among self-employed. Overall, it is observed that industries with a high participation of women employment are those with the higher gender gap.

Figure 2. Employment distribution across industries and occupations

a. Salaried workers

b. Self-employed



Source: GEIH 2019. Own calculation.

We investigate additional features of the workplace that might be simultaneously correlated to industry and occupation. This analysis reveals that for most of the industries, men tend to work as salaried workers, report more working hours, and are in larger firms[10]. In turn, informality rate[11] is higher for men, especially in industries with lower women employment participation (see Table 3). Gender differences in these characteristics are more important among self-employed. For instance, comparing with results in Table 4, it seems that there is a gap in working hours which is more pronounced for self-employed. While salaried workers report more worked hours on average than self-employed, the

[10] Firm's size is defined as follows: small firms are firms with five workers or less, medium firms are firms between 6 and 50 workers, and large firms are firms with more than 50 workers.
[11] Defines as workers who are affiliated to health and contribute to a pension or are pensioners.

gender gap in this item is higher for self-employed. Accordingly, the gender gap in working hours is 17 hours for salaried workers and 48 hours for self-employed.

Table 3. Workplace characteristics for salaried workers

Variables	Manufacturing		Building		Commerce		Transport		Prof. serv.		Public adm		Other serv.	
	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men
%	38.0	62.0	11.2	88.8	47.6	52.4	27.1	72.9	47.7	52.3	62.0	38.0	51.2	48.8
Informality														
Yes	26.1	20.7	14.3	35.7	44.9	39.1	15.9	15.3	9.8	10.7	8.8	5.6	35.9	34.6
Hours worked														
Average	184.5	192.9	178.4	192.3	185.1	199.6	180.0	204.3	174.6	193.7	169.0	189.0	175.7	181.6
Size of firm														
Small	14.3	12.5	7.3	27.0	37.4	30.7	12.7	14.4	11.3	14.3	4.2	1.2	22.9	21.9
Medium	31.6	29.4	34.9	36.7	32.9	36.5	18.7	18.9	23.9	30.6	18.3	9.6	33.8	32.1
Large	54.1	58.1	57.8	36.4	29.7	32.8	68.7	66.7	64.8	55.2	77.5	89.2	43.4	46.0

Source: GEIH 2019. Own calculations. Working hours correspond to the monthly-equivalent average.

Table 4. Workplace characteristics of self-employed

Variables	Manufacturing		Building		Commerce		Transport		Prof. serv.		Public adm		Other serv.	
	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men
All	38.0	62.0	11.2	88.8	47.6	52.4	27.1	72.9	47.7	52.3	62.0	38.0	51.2	48.8
Informality														
Yes	26.1	20.7	14.3	35.7	44.9	39.1	15.9	15.3	9.8	10.7	8.8	5.6	35.9	34.6
Hours worked														
Average hours	184.5	192.9	178.4	192.3	185.1	199.6	180.0	204.3	174.6	193.7	169.0	189.0	175.7	181.6
Size of firm														
Small	14.3	12.5	7.3	27.0	37.4	30.7	12.7	14.4	11.3	14.3	4.2	1.2	22.9	21.9
Medium	31.6	29.4	34.9	36.7	32.9	36.5	18.7	18.9	23.9	30.6	18.3	9.6	33.8	32.1
Large	54.1	58.1	57.8	36.4	29.7	32.8	68.7	66.7	64.8	55.2	77.5	89.2	43.4	46.0

Source: GEIH 2019. Own calculations

Thus, there is also a sorting of women into jobs with lower working hours. Women in Public administration are the group of salaried workers with less working hours, 169 in average. This magnitude is similar for women self-employed with the highest working hours level (175,3 in Transport). In all other industries, self-employed women work significantly fewer hours. Besides, other expected results are that self-employed are informal, and work in small firms. These characteristics present the highest difference when compare self-employed in Building.

Finally, we study socioeconomic characteristics that will be considered to control differences across gender, industries and occupations. Table 8 shows that there are also differences in the socioeconomic composition of employment. Women are higher educated than men, especially in Building industry where women have 13 years of education and men have only 8. It is also noticeable that on average, salaried workers have more years of education than self-employed. There is also a remarkable trend across industries when comparing household head position of women between occupations. In most of the cases, the proportion of household head women is higher for self-employed. This is also true for married women. These two facts are consistent with the findings regarding the motivation of women to work as self-employed since these jobs facilitates the family-work balance.

3. DECOMPOSITION METHOD

3.1 OAXACA-BLINDER DECOMPOSITION

Counterfactual decomposition methods are the most popular technique to investigate the sources of the gender pay gaps. This method consists of computing two components known as composition effect and structure effect. The first is the part of the earnings gap due to the differences in the population composition of men and women, e.g., the difference cause by the fact that women have higher schooling level than men. This is also known as explained effect. The second is the remainder which is interpreted as a structure effect (or unexplained component). The latter relates to the labor earnings schedule which depends, e.g., on the industry wage premia. The implementation of these methods, in general, require specifying the relation between labor earnings and workers characteristics which is estimated using regression techniques (see Cherhoznukov et al., 2013, for a broad discussion)[12].

[12] This approach provides an aggregated decomposition for the difference the average pay, but it can be extended in order to assess the quantitative relevance of workers characteristics individually or studying the difference in other functionals of the labor earnings distribution (Machado & Mata, 2005; Melly, 2005; Rothe, 2010).

The starting point is the total average difference of labor earnings, Δ^{μ}, that is defined as follows:

$$\Delta^{\mu} = \mu^{m} - \mu^{w}$$

where μ^{m} and μ^{w} are the average earnings of men and women. Under a linear specification of the labor earnings conditional mean and additional assumptions of common support and ignorability (see Section 2 in Fortin et al., 2011, for a detailed discussion), Oaxaca (1973) and Blinder (1973) propose to write this difference as follows:

$$\Delta^{\mu} = \mu'^{m}_{X}\beta^{m} - \mu'^{w}_{X}\beta^{w} = \left(\mu'^{m}_{X} - \mu'^{w}_{X}\right)\beta^{m} + \mu'^{w}_{X}(\beta^{m} - \beta^{w}) \qquad (1)$$

where the first term in the right-hand side corresponds to the composition effect (CE) and relates to the difference in the average worker characteristics vector $\left(\mu'^{m}_{X} - \mu'^{w}_{X}\right)$, while the second corresponds to structure effect (SE) as depends on the gender specific earnings structure. The term $\mu'^{w}_{X}\beta^{m}$ is a counterfactual outcome representing the average earnings that men would obtain if they had same characteristics than women, which is not observable.

These two terms compose the aggregated decomposition known as Oaxaca-Blinder decomposition. The estimation is usually based on linear regression methods, and the inference can be performed either using asymptotic standard errors (Jann, 2008) or bootstrapping methods (Cherhoznukov et al., 2013). Previous work studying the quantitative importance of industry on the gender gap has implemented this approach including the gender employment shares by industry as well as dummies at the industry level as regressors. The individual contribution of each factor is estimated by extracting the results related to these variables from the CE and the SE, (see Fields & Wolff, 1995; Gannon et. al., 2007). Jones (1983) and Oaxaca & Ransom (1999) have pointed out that this approach suffers the so-called omitted group identification issue, which is related to the sensitivity of the choice of the reference category as the coefficient of that industry is not distinguishable from constant term. Alternative solutions to the problem can be found in Oaxaca & Ransom (1999), Horrace & Oaxaca (2001) and Yun (2005). These solutions consist of normalizing the regressor, imposing restriction on the coefficients or netting out the effect of the omitted group contained in the constant terms.

A second shortcoming of this approach is the absence of standard errors, which are crucial for determining the statistical meaningful of the results. In this line, Horrace & Oaxaca (2001) and Lin (2007) have also proposed corrected standard errors for statistical inference. We propose an alternative approach which exploit the heterogeneity within and between industries in order to add to the knowledge on the relevance of industry in three dimensions: the gender composition of employment, the within population composition and the difference of characteristics premia at the industry level.

3.2 A TWO-STEP DECOMPOSITION METHOD

To overcome the possible identification and interpretability issues in inferring the contribution of categorical variables, we propose an alternative decomposition method based on a two-step procedure. The first step consists of estimating the contribution of the gender employment shares by industry, while the second implements OB decompositions to quantify the importance of the earnings schedule at industry level.

Particularly, in the first step the average of labor earnings is expressed as a weighted average of the industry labor earnings, that is:

$$\Delta^{\mu} = \mu^{m} - \mu^{w} = \sum_{j=1}^{J} \left(s_j^{m} \mu_j^{m} - s_j^{w} \mu_j^{w} \right)$$

where s_j^m represents men's share employment in industry j, μ_j^m is the corresponding average earnings, and J the total number of industries. Re-arranging terms, the total difference can be written as:

$$\Delta^{\mu} = \sum_{j=1}^{J} s_j^{m} \left(\mu_j^{m} - \mu_j^{w} \right) + \sum_{j=1}^{J} \mu_j^{w} \left(s_j^{m} - s_j^{w} \right). \qquad (2)$$

The first term captures the part of the gender gap due to the difference in average earnings, while the second measures the component on behalf of differences in the employment shares. Note that this decomposition is possible to estimate without further parametric assumption.

To explore the part of the gender gap related to the idiosyncratic industry's pay structure, we decompose the terms $\mu_j^m - \mu_j^w$ using OB decompositions. In such way, using Equation (1) into Equation (2), the final decomposition is given by:

$$\Delta^{\mu} = \sum_{j=1}^{J} s_j^m \left({\mu'}_{jX}^{m} - {\mu'}_{jX}^{w} \right) \beta_j^m + \sum_{j=1}^{J} s_j^m {\mu'}_{jX}^{w} \left(\beta_j^m - \beta_j^w \right) + \sum_{j=1}^{J} \mu_j^w \left(s_j^m - s_j^w \right) \qquad (3)$$

where the first term is an aggregated composition effect (ACE), which is computed as the weighted average of the CE at the industry level. Similarly, the second is related to the industries' earnings structure, named aggregated structure effect (ASE), and the third is the employment-share effect (ESE).

This decomposition allows to directly quantify the importance of gender employment allocation avoiding the omitted group issue. Besides, the heterogeneity in the returns of workers characteristics, that is capture by variation in $s_j^m {\mu'}_{jX}^{w} \left(\beta_j^m - \beta_j^w \right)$, reflects the differences in the production technology, and others industry features shaping the pay schedule. Similarly, variability of $s_j^m \left({\mu'}_{jX}^{m} - {\mu'}_{jX}^{w} \right) \beta_j^m$ across industry makes possible to study the contribution of the sorting in employment composition on the pay gap.

From this decomposition some quantities of interest can be studied. For instance, the sum of ACE and ESE would be the observed gender gap in absence of gender and industry differences in premia, i.e., if industries equally pay men and women characteristics. In turn, the sum of the terms ACE and ASE quantifies the gender gap that one would observe if men and women workers in the same proportion at each industry, which is the total affect due the variation in the average income (say, Total Pay Effect, TPE). Therefore, considering these three components, it is possible to build counterfactuals related to the contribution of industry to the total gap. First, we can measure the gap assuming that any of the components are zero, which means dropping composition or returns differences between men and women. And secondly, we can assess how the gap change when the heterogeneity between industries is eliminated, i.e., assuming that composition and structure effect behaves as the average.

Interestingly, this approach avoids the omitted group identification problem and enables to capture the interacting effect between industry and characteristic premia, which is neglected when the pay structure influence is defined as a single industry dummy variable, i.e., the coefficient of the industry dummies in the OB decomposition. For instance, it enables to measure whether financial activities pay higher on average, but this is not considering the fact that the return for a marginal education year might contribute more to the gender gap in one industry respect to other. Moreover, the OB decomposition estimates the share effect through the part of the composition effect related to the industry dummies, but this is not capturing that an industry has a gender employment share or socioeconomic characteristics composition that differ from the average. Later, we compare the proposed method and the classical OB using Monte Carlo simulations.

Lastly, the estimation of the three components, specifically for the second step, is possible under similar assumptions to those required by the OB decomposition. In the case of the ignorability condition, since the estimation is conditional on industry j, it is less restrictive than the ignorability condition over the whole sample. In turn, the common support assumption might become problematic if J is large and therefore estimates are made on small data sets.

3.3 MONTE CARLO SIMULATION

Monte Carlo simulation exercises are performed to illustrate how the proposed method works. These exercises present different scenarios of the decomposition components and make comparisons with the results obtained using the OB method to illustrate possible confounders associated with the omitted group problem. For this, we assume data generating processes that specify the outcome of interest, denoted by y, for two groups subject to comparison $g = \{0,1\}$, where individuals from each group can belong to a particular segment, denoted by $j = \{0,1\}$, which in our case is the industry. To include variation in the population composition and heterogeneity at the segment (industry) level, we consider two additional covariates: a dummy variable denoted by z and a continuous variable denoted by x.

Therefore, the outcome for the observation i of the group g will be given by:

$$y_i^g = \delta_0^g + \delta_1^g x_i^g + \delta_2^g z_i + \delta_3^g d_i^g + \delta_4^g d_i^g x_i^g + \delta_5^g d_i^g z_i + \varepsilon_i^g$$

where d_i indicates whether the observation i belongs to the segment 1. The covariates x and d are indexed by g since different underline distribution can be specified to generate composition effects and different share compositions across segments. In the benchmark simulation we define the following parameters: $\delta_0^0 = \delta_0^1 = 10$, $\delta_1^0 = \delta_1^1 = 2$, $\delta_2^0 = \delta_2^1 = 0.5$, $\delta_3^0 = \delta_3^1 = 2$. The rest of the parameter are used to control the presence of composition and structure effects. In the benchmark simulation z follows a Bernoulli distribution with parameter 0.7 and x is uniform in the interval between 0 and 1.

We perform three simulations under the following scenarios: i. difference in the share of groups across segments, ii. difference in the share and difference in the premia based on the covariate z_i, iii. difference in the share, population composition through the variable x by group, and difference in the return in the same variable[13]. To generate these scenarios, we run 1,000 simulations using the parameters in Table 5, that also reports the theoretical value of the components. Accordingly, we compute OB decomposition and the three components of the two-step decomposition as well as the individual OB by segment, which is the bases of the second step.

Table 5. Scenarios for Monte Carlo simulations

Scenario	Parameters	Theoretical components		
		ESE	**ACE**	**ASE**
1	$d^0 \sim B(0.3)$, $d^1 \sim B(0.7)$	-0.8	0	0
2	$d^0 \sim B(0.3)$, $d^1 \sim B(0.7)$, $\delta_5^1 = 5$	-2.2	0	-1.05
3	$d^0 \sim B(0.3)$, $d^1 \sim B(0.7)$, $x^1 \sim U(0,2)$, $\delta_4^1 = 2$	-1.6	-1	-0.6

Note: $B(p)$ denotes Bernoulli distribution and $U(a,b)$ the uniform distribution.

Result for scenario 1 are presented in the box plots in Figure 4, where it is observed that the OB decomposition consistently captures the share effect through the part of the composition effect that relates to d, which coincides with the ESE in the two-step decomposition (see Figure 4 panel b). Note that the OB decompositions at the segment level, as expected, do not report any variation. Under this specification, the OB decomposition completely characterize the gender gap in a more parsimonious way.

Regarding the scenario 2, which consider that the distribution of groups across segments is uneven, most important differences between the two approaches are revealed (see **Figure 5**). Respect to the total gap, that is -2 for group 0 and -5 for group 1 which generate

[13] Additional scenarios were simulated, e.g. the case considering only composition effect in x, and the results are available upon request.

a higher contribution of the share effect, since the proportion of observation belonging to segment 1 into de group 1 is higher (30% and 70%, respectively). Results for the two-step procedure are close to the theoretical values, and the TPE is also correctly estimated and captures the difference in the premia corresponding to the variable z.

The OB decomposition shows the same estimate for the composition effect as before, but it does not correspond to the effect of the group share (see Table 5). Indeed, this result is not reflecting the fact that the higher gap between segments the higher impact of the share. The estimates for the SE are also different from the theoretical values, in particular the corresponding component to the segment's dummy and the variable z are both negative. And it seems that the constant absorbs the rest of the gap to adjust the total gap. This is actually related to the omitted group problem. An interesting feature of the two-step lies on the possibility of analyzing the specific source of the gap. Indeed, the second step shows that the gap comes from the SE of the variable z of the group 1.

In the last simulation exercise, we aim to combine the three sources of group gap. The box plots in Figure 6 highlight the differences of the two approaches. The CE of the OB decomposition contains the effect of the dummy variable of the segment (-0.8) and the composition effect due to the variable x (-1). However, the former is not the aggregated effect of the share, which is equal to -1.6. In the case of the SE, similar results to scenario 2 are obtained. In the two-step decomposition, it is observed that the CE depending on x is relevant in both segments while the premium is prevalent in segment 1, which correctly estimates the proposed data generating process. This suggests that the two-step decomposition might provide a richer approach to captures the influence of a group variable on the gap.

4. THE CONTRIBUTION OF INDUSTRY AND OCCUPATION TO GENDER GAPS

Using the proposed method, we perform a decomposition analysis for salaried workers and self-employed, separately. We first estimate the contribution of the gender employment shares and average pay gap at the industry level on the total gender gap. Subsequently, the latter component is decomposed into composition and structure effects. By estimating the raw differentials in labor income, we observed that gender gap varies importantly across job types. Particularly, while the gender gap among salaried workers is 4.2%, the estimates among self-employed is 63.63%[14].

The specification used for the decomposition includes the following controls: age, education, marital status, whether the individual is household head, number of working hours (in logs), firm size, informality status and job qualification (job position as professional or director). Confidence intervals at 95% are computed using 1,000 bootstrapping simulations. Results for the aggregate decomposition between share effect and pay effect point out strong patterns related to occupations. In particular, in the case of salaried workers, the gender gap seems to be mostly explained by the differences in average pay across industries (see Table 6). In turn, the share effect is negative and statistically significant, but it is not quantitatively relevant. This indicates that the allocation of female and male employment across industries tends to slightly narrow the gap. In fact, in case that all industries would pay equally men and women, a gap of 1.4% in favor of women would be observed.

Table 6. Aggregate decomposition by occupation

	Salaried workers	Self-employed
Total gap	0.042 [0.033 , 0.051]	0.636 [0.623 , 0.656]
TPE	0.056 [0.045 , 0.065]	0.519 [0.478 , 0.561]
ESE	-0.014 [-0.019 , -0.007]	0.117 [0.084 , 0.157]

Source: Own calculations. 95% confidence interval is brackets estimated using 1000 bootstrapping simulations. Decompositions include the following controls: age, education, marital status, whether the individual is household head, number of working hours (in logs), firm size, informality status, and job qualification (job position as professional or director). TPE is the total pay gap and ESE is the employment share effect.

The results are remarkable different when studying the gender gap among self-employed. The pay effect is still the main driver; however the ESE is positive and has a significant contribution, 11.7 percentage points equivalent to 18.3% of the total gap. This might be related to gender sorting into industries for self-employed, which in this case is consistent

[14] This value differ from the previous analysis since the decomposition is performed on the logarithm of the earnings to facilitate the implementation and interpretation.

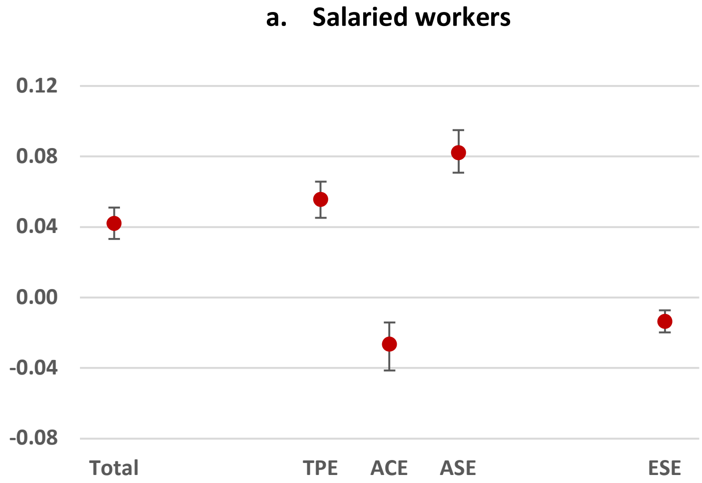
with the lower labor income in industries related to retail and services. These results remark the relevance of considering the sorting of the labor market to have a complete understanding of the gender gaps since it takes into account the idiosyncratic trends of salaried workers and self-employed given that female and male labor force are not uniformly allocated across industries.

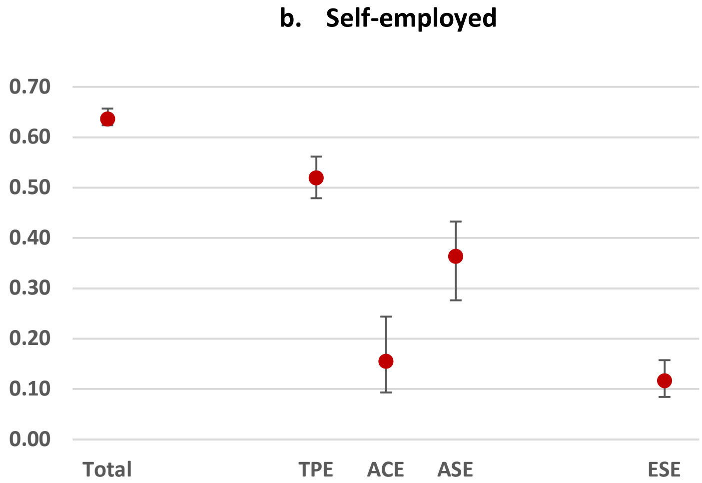
An important question rising here is at what extent gender differences in workers characteristics or gender differences in the returns account into the pay effect. To answer this, we perform the second step of the decomposition which enables to quantify these two margins. Figure 3 presents the point estimates and 95% confidence intervals for the ACE and ASE. For salaried workers, although the TPE is positive and the part of the gap that depends on the characteristics of the workers has a negative sign. In this scenario, if the characteristics were paid the same within industries, a pay gap of 4% (ESE+ACE) in favor of women would be observed. This is a significant result, as it highlights that the premium differential is the main driver of the gender gap for salaried worker. Noticeable, the gap prevails even when controlling for the number of hours worked, informality status and the qualification of the job.

The ASE is also the main component of TPE for the self-employed, but in contrast, the composition effect is positive. This result is aligned with Goldin (2014) who argues that women are self-selected into job where flexibility is a key factor of the labor supply decision. In sum, these decomposition exercises show that even if the participation of women and men in the industries were the same, and that men and women within each industry had the same characteristics, a gap of 8.2% and 36.38% (i.e., the ASE) would be observed between salaried workers and self-employed workers, respectively. Interestingly, the reminder, i.e., the sum between ACE and ESE, is negative for salaried workers, implying that these factors on the contrary reduce the gap. While in the case of the self-employed both factors expand the pay gap accounting by around 40% of the total.

One of the advantages of the proposed decomposition method is that it enables to quantify the contribution of each industry to the gender gap components. Figure 8 presents the percentage of this contribution which is measured as the corresponding elements of the summation in Equation (3). Two facts are evident. First, individual contributions at industry level are not equal, which implies that industries' features shape the composition and structure effect. Second, ASE is positive for all industries for both occupations, showing a systematic higher premia for men.

Figure 3. Two-step decomposition by occupation.





Source: Own calculations. 95% confidence interval is brackets estimated using 1000 bootstrapping simulations. Decompositions include the following controls: age, education, marital status, whether the individual is household head, number of working hours (in logs), firm size, informality status and job qualification (job position as professional or director). TPE is the total pay gap, ACE is the aggregated composition effect, ASE is the aggregated structure effect, and ESE is the employment share effect.

An expected result is that industries with the largest gender share difference and pay gender gaps exhibit stronger influence to the components. In particular, Building is the industry with the higher contribution to the ACE, which is in favor of women, but at the same time, everything else equal, generates the higher gap in term of characteristics' premia. Moreover, Public administration, which is a female-dominated industry, also provides a positive contribution to the ASE. Other industries with a remarkable importance are Manufacturing and Commerce. The mixed results across industries revealed by the ACE are also evidence of strong gender sorting patterns. That is, even controlling by industry, women and men are different in their demographic characteristics and job features.

Among the self-employed, the ACE and the ASE depend on industries such as Transportation, Building and Commerce. The first two correspond to male-dominated industries, while the latter is a sector with a similar gender employment share, but with the largest pay gap. It is important to highlight that Transportation and Building are also the industries that tend to close the gap through the ACE. Meanwhile, Transportation and Commerce account by 78.7% of the total ASE.

Finally, we explore the role heterogeneity across industries by performing industry-level decompositions. Figure 9 and Figure 10 in the Annexes present the composition and structure effects in a disaggregated manner for each characteristic considered in the decomposition. There are remarkable patterns when comparing salaried workers and self-employed. In terms of the CE, education plays a crucial role reducing the gender gap for both occupations. This is especially true for the Building and Transportation. In general, household head and marital status do not seem to be significant drivers of the pay gap. From the perspective of employment characteristics, factors such as working hours and job qualification are important forces leading to a positive gap in the case of self-employment. Regarding, working hours, the CE is positive in all industries, which is consistent with the fact that women choose jobs that allow them to balance work and personal life.

The analysis of the SE shows a higher contribution of education and working hours. This draws attention to two key issues. First, the labor income inequality observed in industries such as Building and Transportation is mediated by an asymmetric gender composition and the fact that an additional year of education has much higher returns for women in these industries. Secondly, there is additional evidence of the relevance of flexibility as a factor driving self-selection, since even when the return per marginal hour is higher for women, this group choose to work fewer hours. An additional interesting result is that the informality premium plays an important role among self-employed. In fact, this factor tends to increase the gender gap, i.e., an informal male self-employed earns much more than a female counterpart.

Previous findings show the extent to which industry adds to the gender gap. To get an overview of the results, we analyze the relationship between the level of feminization of the industries with the magnitude decomposition components (see Figure 11). It might be argued that the level of feminization is related to the observed behavior of the components. An expected result is that the higher the feminization, the lower the employment share effect. In contrast, ACE has a positive relationship with feminization while for ASE that is negative, although the latter is a weaker relationship. This shows that the most feminized industries, that coincide with labor intensive activities and possibly industries with lower productivity, are those contributing more to enlarge the gender gap.

Overall, our results suggest that the heterogeneity across industries is crucial to understand the sources of gender gaps. In fact, the CE is not the same for all industries, and so, differences in schooling or working hours between men and women vary across industries generates fluctuations in the pay gap. Using the above estimates, it is possible to perform counterfactual exercises to obtain additional evidence about the role of the industry heterogeneity in terms of the three components of the decomposition. In particular, we are able to assess how the gender gap would change in scenarios where variability in either gender employment shares, population composition within industries or premia, is eliminated. For this, the components of Equation (3) are estimated under three assumptions: i. the proportion of men and women is equal in all industries, ii. the CE of each industry is equivalent to the average of the labor market, and iii. the SE of each industry is equivalent to the average of the labor market.

We compare these scenarios to determine whether the variation between industries contributes to increase or decrease the gap[15]. The results in Table 7 indicate that if all industries have a 50-50 gender composition the gender gap would increase for salaried workers but reduce for self-employed. This implies that a higher participation of women in male-dominated industries does not necessarily contribute to reduce the pay gap. These mixed results are also found when the CE heterogeneity is dropped. On the one hand, CE heterogeneity tends to increase the gap between salaried workers. Without such variability, the gap would decrease by 0.9 percentage points. On the other hand, in contrast, the gap among the self-employed would increase by 14.2 percentage points. Finally, the heterogeneity in the SE provokes the smallest change with respect to the estimate gap for both occupations. Consequently, these finds support the idea that industry affect the gender gap mainly through the gender share and the employed composition. Besides, the strong contribution is observed for self-employed since employment shares' at the industry level contributes with 14.6 percentage points, 23% of the total gap.

Finally, a comparation regarding the results that would be obtained through an OB decomposition also provides an interesting discussion. Therefore, we present two versions of the OB decomposition (see Table 9 and Table 10), one including dummy variables at the industry level and other including the share of either male or female employment as controls. The latter is in the line of Fields & Wolf (1995) and Gannon et al. (2007). In relative terms, industry has lower participation in CE than in SE. The participation of industry variables in the CE shows quantitatively higher estimates in the case of salaried workers. Moreover, the direction of the effect coincides with that obtained in the two-step decomposition procedure, however, the OB is not considering the differences in the demographic characteristics of men and women between industries, which might tend to underestimate the contribution of the industry to the total gap.

[15] Of course, this exercise does not take into account general equilibrium effects, since it is assuming independence among the components, i.e., it is assumed that equalizing the premia across industries would not generate variations in gender composition or worker characteristics. However, these exercises can be informative about the sources of pay inequality between men and women.

Table 7. Gender gap and components heterogeneity between industries

	Salaried workers			
	Benchmark	**Equal shares**	**Equal ACE**	**Equal ASE**
Total gap	0.042	0.049	0.033	0.037
TPE	0.056	0.049	0.047	0.050
ACE	-0.026	-0.038	-0.035	-0.026
ASE	0.082	0.087	0.082	0.077
ESE	-0.014	0.000	-0.014	-0.014
	Self-employed			
	Benchmark	**Equal shares**	**Equal ACE**	**Equal ASE**
Total gap	0.636	0.490	0.778	0.612
TPE	0.519	0.490	0.661	0.494
ACE	0.155	0.215	0.297	0.155
ASE	0.364	0.275	0.364	0.339
ESE	0.117	0.000	0.117	0.117

Source: Own calculations. TPE is the total pay gap, ACE is the aggregated composition effect, ASE is the aggregated structure effect, and ESE is the employment share effect.

It is important to consider that the inclusion of industry-level dummies and controlling for employment share generate results that are not directly comparable. While the first allows us to derive results on the composition of the labor force and the aggregate industry premium, in the second case the ES can be associated with the premium of working in a sector with certain feminization level. Overall, our proposed decomposition allows for a more detailed disaggregation of these margins of the total gap.

5. CONCLUDING REMARKS

Reducing the gender pay gap is maybe one of the most urgent issues to improve de inclusiveness in the labor market. Policy makers have focused on designing policies reducing discrimination, and most recently on promoting programs that facilitates the access of women to better jobs. These policies have been motivated on the evidence of gender sorting into occupations and industries, another factor contributing to the gender pay gaps. We document the relevance of this sorting as a force driving the gender gap in Colombia. Particularly, we provide evidence that gender employment shares and heterogeneity in workers population and characteristics' returns affects the earning differences.

The presence of higher proportion of men or women in an industry has been an indicator of segregation. This factor is crucial to understand the sources of the gender gap. In fact, our estimates show that industries with a high participation of women are those with the highest gender pay gap, and this pattern is strong among self-employed. However, it is also obtained that for this group a more equal distribution of men and women across sectors would result in lower pay gap. What is behind is that men and women are taking into consideration preferences for occupations, flexibility and comparative advantages for job search decisions. This means that anti-discrimination regulations might be not enough to reduce gender gaps (see c.f. Morton et al. 2014), on the contrary, other elements are also determining the gap though the occupational choice decision.

In fact, as discussed by Clain (2000), Goldin (2004), and Eastough & Miller (2004), the higher gender pay gap among self-employed not only combines wage discrimination and premium gaps, but also relates to temporal flexibility and unpaid work. Even tax and subsidy's structure might play a role on determining gender equality (see c.f. Duval-Hernandez, 2021). In this way, promoting employment of women in male-dominated industries is a relevant strategy to reduce the job access gap, but in the light of our results it might not turn out in a reduction of the gender pay gaps. Therefore, a policy design must target to reducing segregation into high value added, which requires programs related to formation of aspiration and flexible arrangements (see Das & Kotikula, 2019, for further discussions).

REFERENCES

Allen, J., & Sanders, K. (2002). Gender gap in earnings at the industry level. *European Journal of Women's Studies*, *9*(2), 163-180.

Altonji, J. G., & Blank, R. M. (1999). Race and gender in the labor market. *Handbook of labor economics*, *3*, 3143-3259.

Atal, J. P., Ñopo, H., & Winder, N. (2010). Gender and Ethnic Wage gaps in Latin America at the turn of the Century. *Inter-American Development Bank*.

Babcock, L., & Laschever, S. (2009). *Women don't ask*. Princeton University Press.

Baker, M., & Cornelson, K. (2018). Gender-based occupational segregation and sex differences in sensory, motor, and spatial aptitudes. *Demography*, *55*(5), 1749-1775.

Blau, F. D., & Kahn, L. M. (1997). Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of labor Economics*, *15*(1, Part 1), 1-42.

Blau, F. D., Brummund, P., & Liu, A. Y. H. (2013). Trends in occupational segregation by gender 1970–2009: Adjusting for the impact of changes in the occupational coding system. *Demography*, *50*(2), 471-492.

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

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436-455.

Budig, M. J. (2006). Gender, self-employment, and earnings: The interlocking structures of family and professional status. *Gender & society*, *20*(6), 725-753.

Clain, S. H. (2000). Gender differences in full-time self-employment. *Journal of Economics and Business*, *52*(6), 499-513.

Chernozhukov, V., Fernández-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*(6), 2205-2268.

Cortes, P., & Pan, J. (2018). Occupation and gender. *The Oxford handbook of women and the economy*, 425-452.

Craig, L., Powell, A., & Cortis, N. (2012). Self-employment, work-family time and the gender division of labour. *Work, employment and society*, *26*(5), 716-734.

Das, S., & Kotikula, A. (2019). *Gender-based employment segregation: Understanding causes and policy interventions*. World Bank.

Duval-Hernandez, R., Fang, L., & Ngai, R. (2021). Taxes, Subsidies, and Gender Gaps in Hours and Wages.

Eastough, K., & Miller, P. W. (2004). The gender wage gap in paid-and self-employment in Australia. *Australian Economic Papers*, *43*(3), 257-276.

Fernandez, MP. (2006). Determinantes del diferencial salarial por género en Colombia, 1997-2003. *Revista Desarrollo y Sociedad*, (58), 165-208.

Fields, J., & Wolff, E. N. (1995). Interindustry wage differentials and the gender wage gap. *ILR Review*, *49*(1), 105-120.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. In *Handbook of labor economics* (Vol. 4, pp. 1-102). Elsevier.

Gallen, Y., Lesner, R. V., & Vejlin, R. (2019). The labor market gender gap in Denmark: Sorting out the past 30 years. *Labour Economics*, *56*, 58-67.

Gannon, B., Plasman, R., Ryex, F., & Tojerow, I. (2007). Inter-industry wage differentials and the gender wage gap: evidence from European countries. *Economic and Social Review*, *38*(1), 135.

Gibbons, R., & Katz, L. (1992). Does unmeasured ability explain inter-industry wage differentials?. *The Review of Economic Studies*, *59*(3), 515-535.

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

Haisken-DeNew, J. P., & Schmidt, C. M. (1997). Interindustry and interregion differentials: Mechanics and interpretation. *Review of economics and Statistics*, *79*(3), 516-521.

Hodson, R., & England, P. (1986). Industrial structure and sex differences in earnings. *Industrial Relations: A Journal of Economy and Society*, *25*(1), 16-32.

Horrace, W. C., & Oaxaca, R. L. (2001). Inter-industry wage differentials and the gender wage gap: An identification problem. *ILR Review*, *54*(3), 611-618.

Hundley, G. (2001). Why women earn less than men in self-employment. *Journal of labor research*, *22*(4), 817-829.

ILO (2018). *Global Wage Report 2018/19: What lies behind gender pay gaps*. International Labour Office.

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, *8*(4), 453-479.

Jones, F. L. (1983). On decomposing the wage gap: a critical comment on Blinder's method. *The Journal of human resources*, *18*(1), 126-130.

Krueger, A. B., & Summers, L. H. (1988). Efficiency wages and the inter-industry wage structure. *Econometrica: Journal of the Econometric Society*, 259-293.

Lechmann, D. S., & Schnabel, C. (2012). Why is there a gender earnings gap in self-employment? A decomposition analysis with German data. *IZA Journal of European Labor Studies*, *1*(1), 1-25.

Leuze, K., & Strauß, S. (2016). Why do occupations dominated by women pay less? How 'female-typical'work tasks and working-time arrangements affect the gender wage gap among higher education graduates. *Work, employment and society*, *30*(5), 802-820.

Leung, D. (2006). The male/female earnings gap and female self-employment. *The Journal of Socio-Economics*, *35*(5), 759-779.

Lin, E. S. (2007). On the standard errors of Oaxaca-type decompositions for inter-industry gender wage differentials. *Economics Bulletin*, *10*(6), 1-11.

Machado, J. A., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, *20*(4), 445-465.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour economics*, *12*(4), 577-590.

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

Morton, M., Klugman, J., Hanmer, L., & Singer, D. (2014). Gender at work: A companion to the world development report on jobs. *Washington, DC: World Bank*.

Nightingale, M. (2019). Looking beyond average earnings: Why are male and female part-time employees in the UK more likely to be low paid than their full-time counterparts?. *Work, Employment and Society*, *33*(1), 131-148.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693-709.

Oaxaca, R. L., & Ransom, M. R. (1999). Identification in detailed wage decompositions. *Review of Economics and Statistics*, *81*(1), 154-157.

O'Neill, J., & Polachek, S. (1993). Why the gender gap in wages narrowed in the 1980s. *Journal of Labor economics*, *11*(1, Part 1), 205-228.

Perry, G. (Ed.). (2007). *Informality: Exit and exclusion*. World Bank Publications.

Reilly, K. T., & Zanchi, L. (2003). Industry wage differentials: how many, big and significant are they?. *International Journal of Manpower*.

Rivero, R. Y., & Mesa, C. (1997). Ingresos laborales de hombres y mujeres en Colombia: 1976-1995. *Archivos de Maccroeconomía*.

Rothe, C. (2010). Nonparametric estimation of distributional policy effects. *Journal of Econometrics*, *155*(1), 56-70.

Sin, I., Stillman, S., & Fabling, R. (2020). What Drives the Gender Wage Gap? Examining the Roles of Sorting, Productivity Differences, Bargaining and Discrimination. *The Review of Economics and Statistics*, 1-44.

Tenjo, J., & Bernat, F. (2018). Diferencias por género en el mercado laboral colombiano: mitos y realidades. *Universidad Jorge Tadeo Lozano*.

Walker, J. R. (2009). Earnings, effort, and work flexibility of self-employed women and men: the case of St. Croix County, Wisconsin. *Journal of Labor Research*, *30*(3), 269-288.

WEF 2021. Global Gender Gap Report 2021

Yun, M. S. (2005). A simple solution to the identification problem in detailed wage decompositions. *Economic inquiry*, *43*(4), 766-772.

Yun, M. S. (2006). Revisiting Inter-industry wage differentials and the gender wage gap: an identification problem.

Annexes

Table 8. Socioeconomic characteristics of workers by occupation and industry

Variables	Manufacturing		Building		Commerce		Transport		Prof. serv.		Public adm		Other serv.	
	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men
Salaried workers														
%	38.0	62.0	11.2	88.8	47.6	52.4	27.1	72.9	47.7	52.3	62.0	38.0	51.2	48.8
Age														
Average	36.6	36.7	34.3	36.8	34.1	33.9	33.9	37.5	33.9	36.3	38.7	39.5	35.2	35.3
Education														
Average years	11.7	11.1	13.4	9.2	11.3	10.9	13.6	11.7	13.5	12.5	14.5	14.4	12.4	12.1
Marital status														
% with partner	47.2	59.6	49.4	62.8	46.5	51.8	43.9	60.0	44.6	54.8	51.5	60.6	43.3	50.7
Household head														
Yes	33.1	53.1	31.1	55.2	32.3	47.2	26.8	57.4	30.6	53.6	35.0	62.1	32.6	49.4
Self-employed														
%	49.5	50.5	2.3	97.7	50.7	49.3	6.7	93.3	61.0	39.0	65.6	34.4	52.8	47.2
Age														
Average	45.9	45.0	39.0	45.2	44.5	44.5	40.7	42.7	43.9	45.9	39.3	38.7	41.6	41.4
Education														
Average years	9.4	9.0	13.2	8.1	9.2	8.7	11.5	9.3	10.4	13.7	13.6	15.5	10.1	10.4
Marital status														
% with partner	55.7	64.1	35.2	63.6	56.1	62.9	50.8	63.8	44.9	56.7	51.5	50.1	49.7	47.9
Household head														
Yes	37.8	63.7	27.4	57.9	40.4	61.5	34.7	57.5	45.3	64.9	31.8	51.1	37.9	53.1

Source: GEIH 2019. Own calculations.

Table 9. OB decomposition for salaried workers

	OB + Industry dummies		OB + Industry shares	
	Effects	**%**	**Effects**	**%**
Total	0.042	100.0	0.042	100.0
CE	-0.038	-92.1	-0.043	-104.0
Industry	-0.004	-8.7	-0.008	-18.5
Others variables	-0.035	-83.4	-0.036	-85.5
SE	0.080	192.1	0.085	204.0
Industry	0.010	23.5	-0.057	-137.0
Others variables	0.070	168.6	0.142	340.9

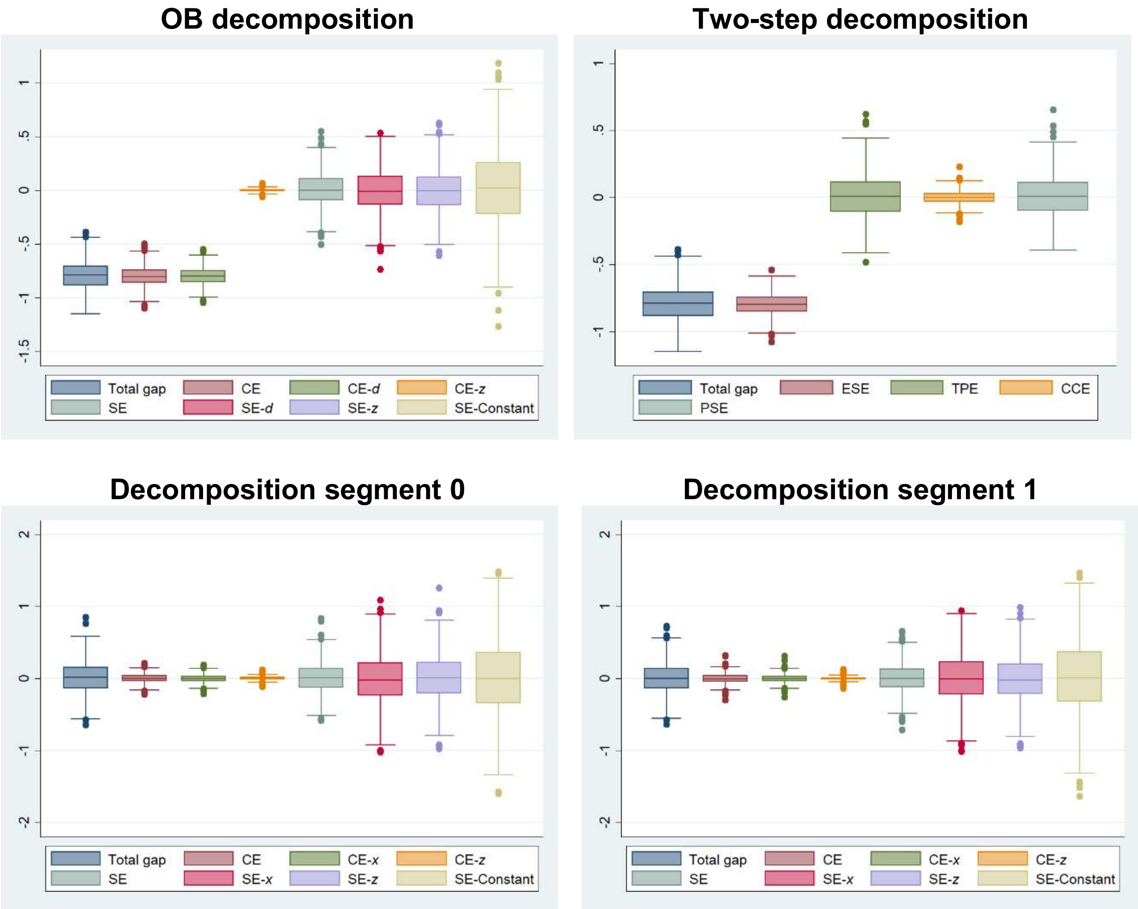
Source: Own calculations. Decompositions include the following controls: age, education, marital status, whether the individual is household head, number of working hours (in logs), firm size, informality status, and job qualification (job position as professional or director).

Table 10. OB decomposition for self-employed

	OB + Industry dummies		OB + Industry shares	
	Effects	**%**	**Effects**	**%**
Total	0.636	100.0	0.636	100.0
CE	0.263	41.4	0.302	47.4
Industry	-0.031	-4.9	0.004	0.7
Others variables	0.294	46.3	0.298	46.8
SE	0.373	58.6	0.335	52.6
Industry	-0.091	-14.3	-0.014	-2.1
Others variables	0.464	72.9	0.348	54.7

Source: Own calculations. Decompositions include the following controls: age, education, marital status, whether the individual is household head, number of working hours (in logs), firm size, informality status, and job qualification (job position as professional or director).

Figure 4. Simulation results scenario 1.

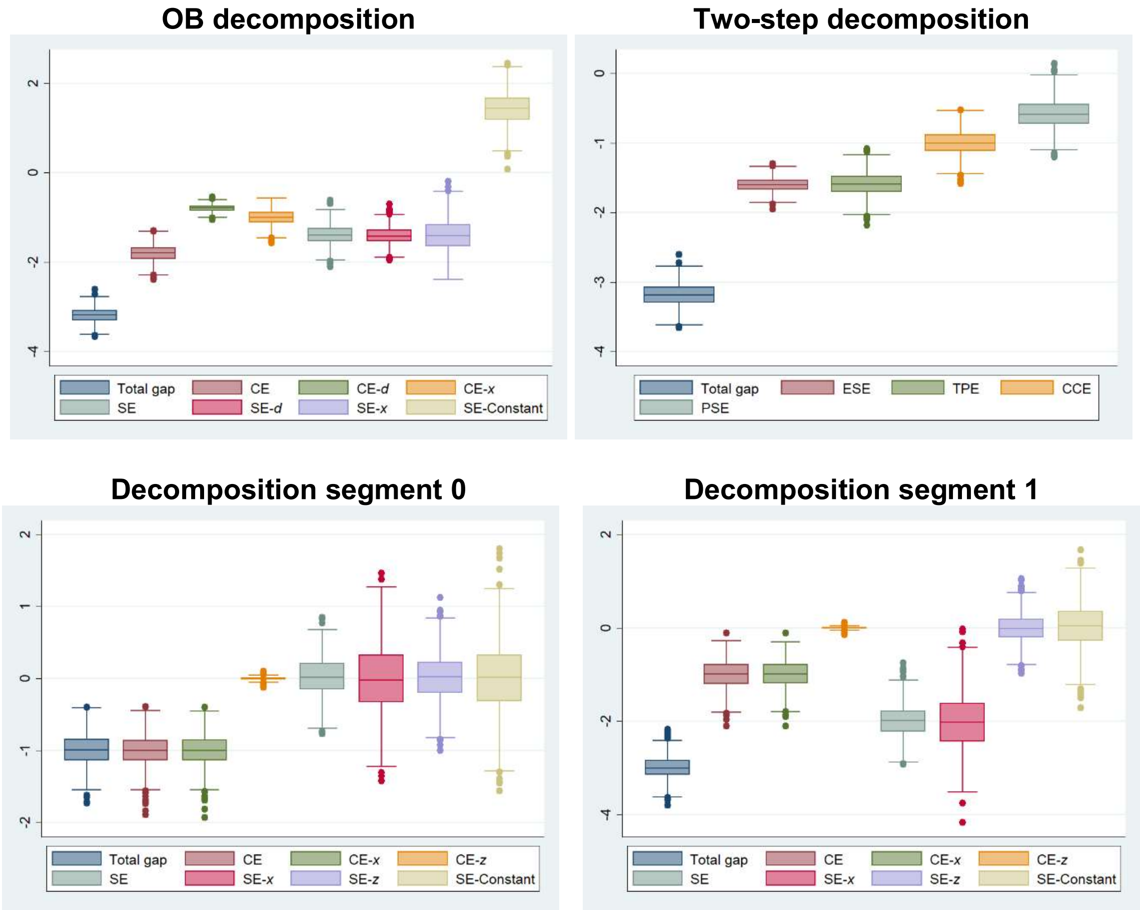


Source: Own calculations. This scenario considers difference in the share of individual by segment in each group.

Figure 5. Simulation results scenario 2.

Source: Own calculations. This scenario considers difference in the share of individual by segment in each group and a premium of the characteristics z in the segment 1.

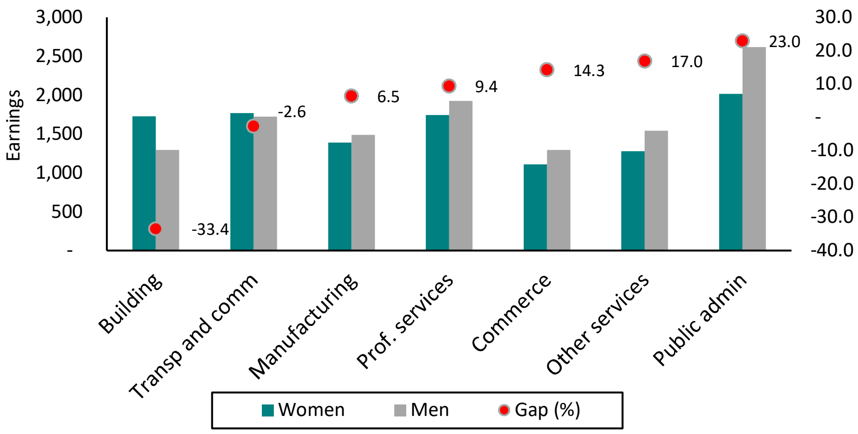
Figure 6. Simulation results scenario 3.



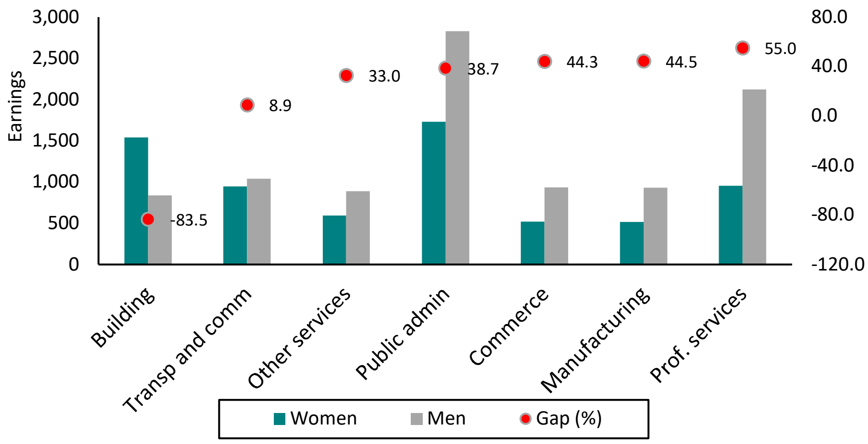
Source: Own calculations. This scenario considers difference in the share of individual by segment in each group, composition effect depending on x and a premium of this characteristic in the segment 1.

Figure 7. Pay gap by industry and occupation

a. Salaried workers

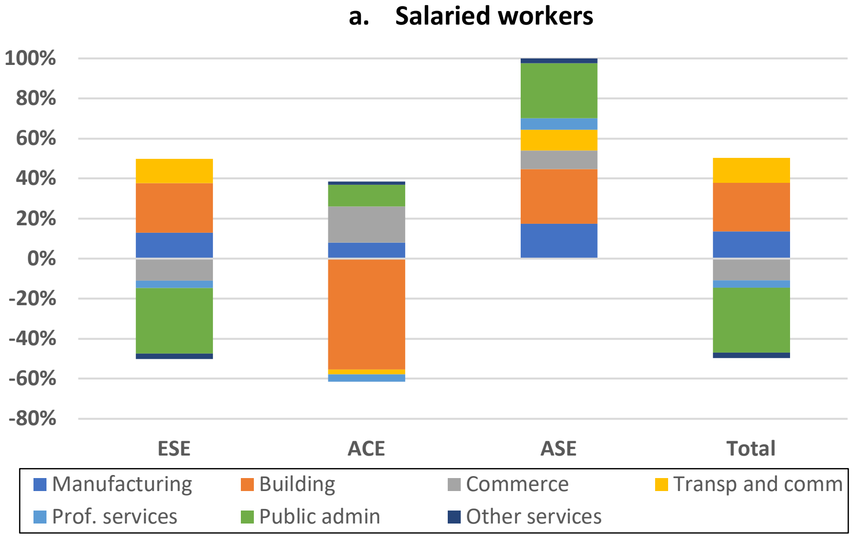


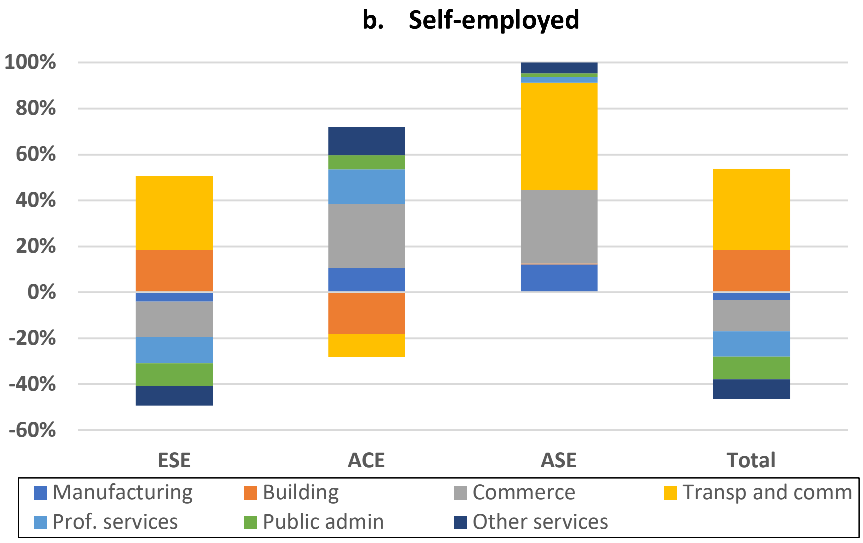
b. Self-employed



Source: GEIH 2019. Own calculations. Left axes measure the earnings in thousand Colombian pesos and right axes the gender gap in %.

Figure 8. Industry contribution to the decomposition components.

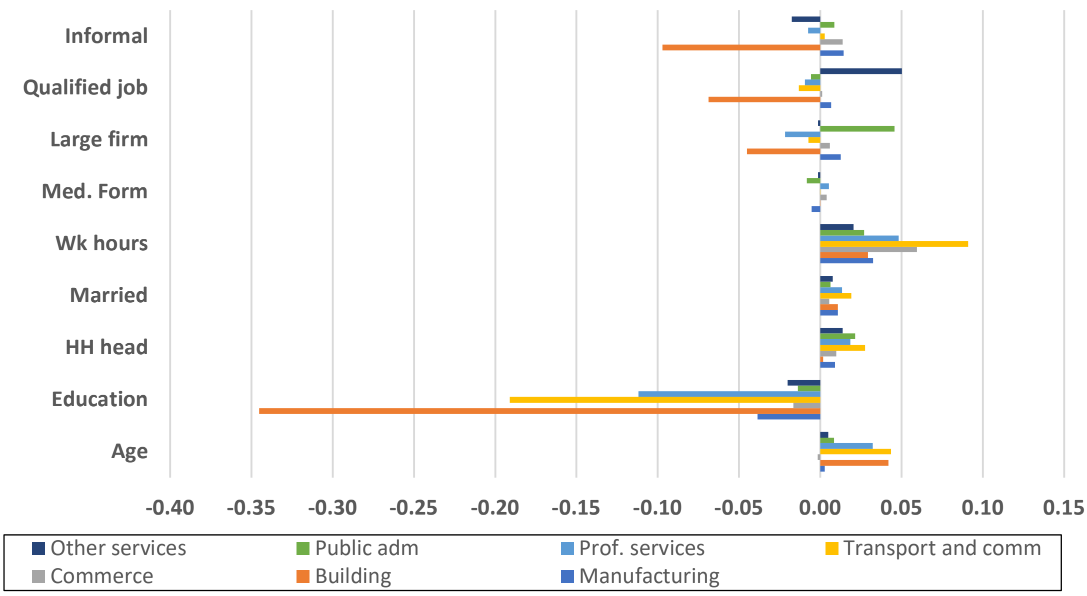




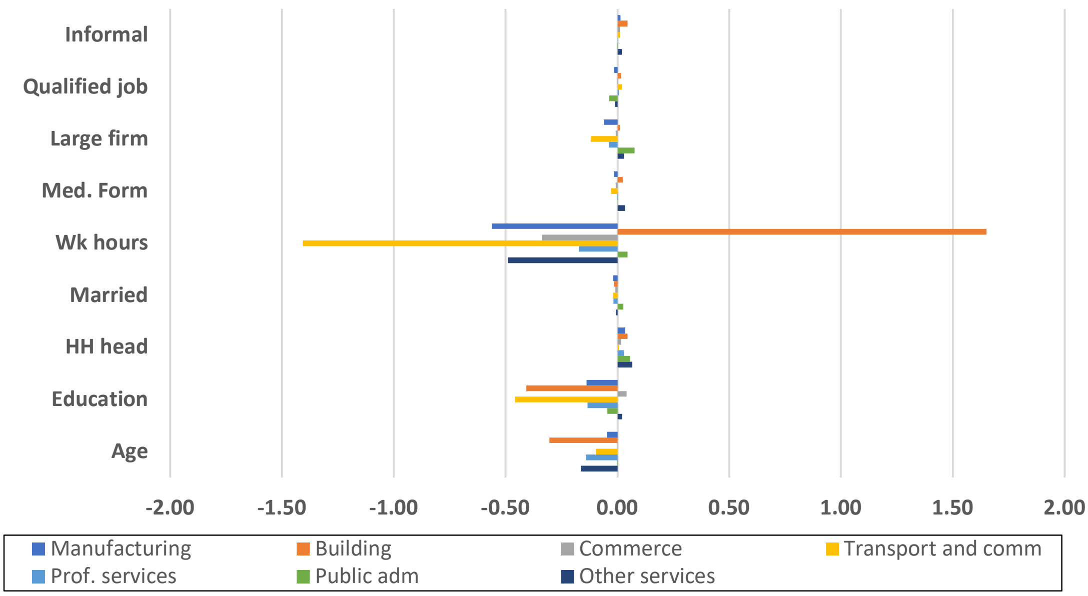
Source: Own calculations. Decompositions include the following controls: age, education, marital status, whether the individual is household head, number of working hours (in logs), firm size, informality status and job qualification (job position as professional or director). ACE is the aggregated composition effect, ASE is the aggregated structure effect, and ESE is the employment share effect.

Figure 9. Industry contribution to decomposition components for gender gap for salaried workers.

a. Composition effect



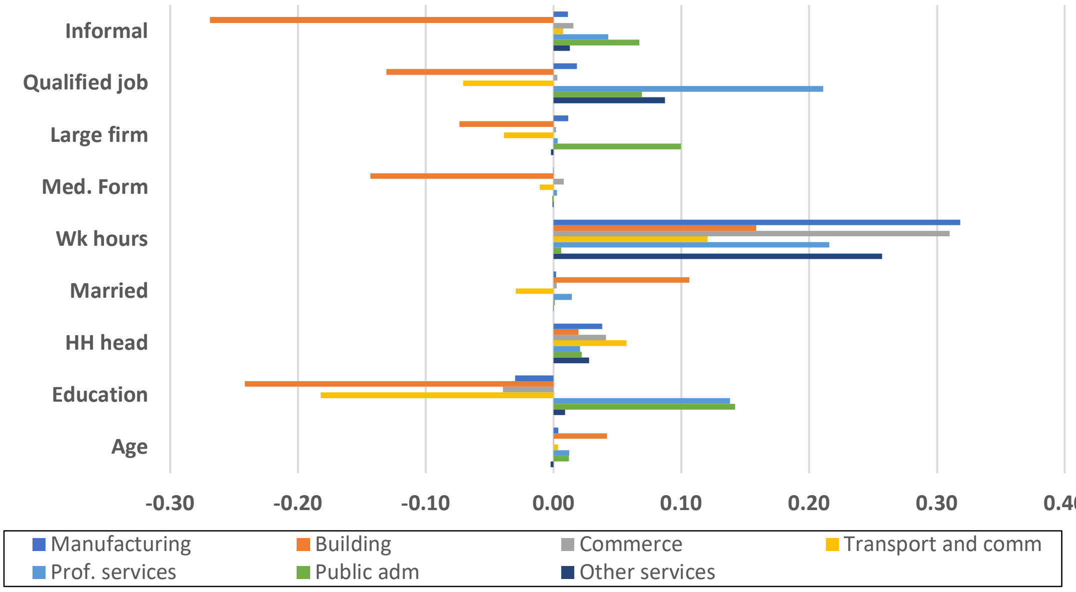
b. Structure effect



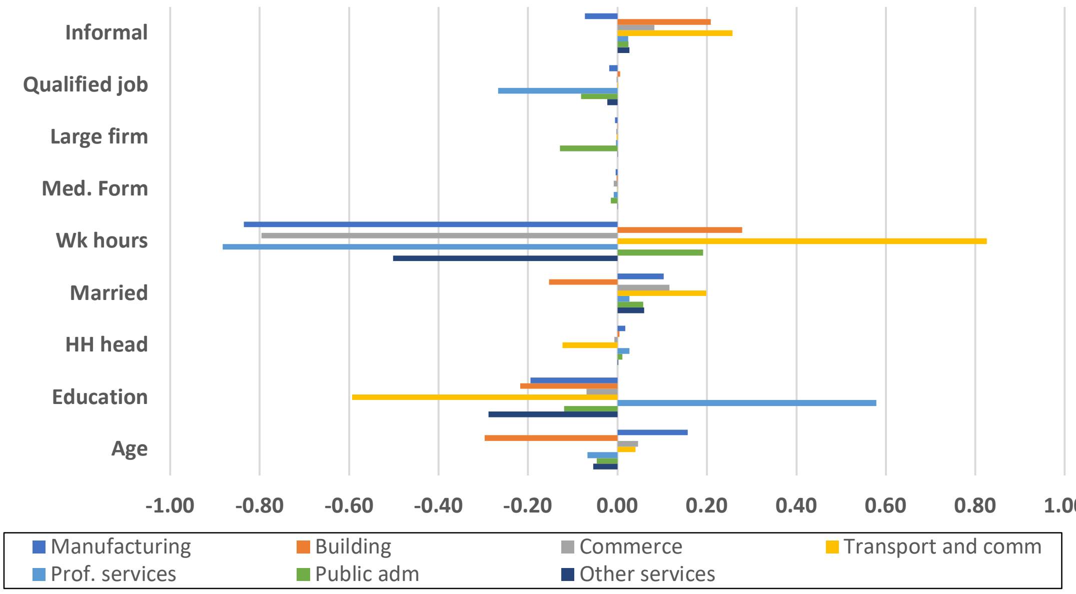
Source: Own calculations.

Figure 10. Industry contribution to decomposition components for gender gap for self-employed.

a. Composition effect

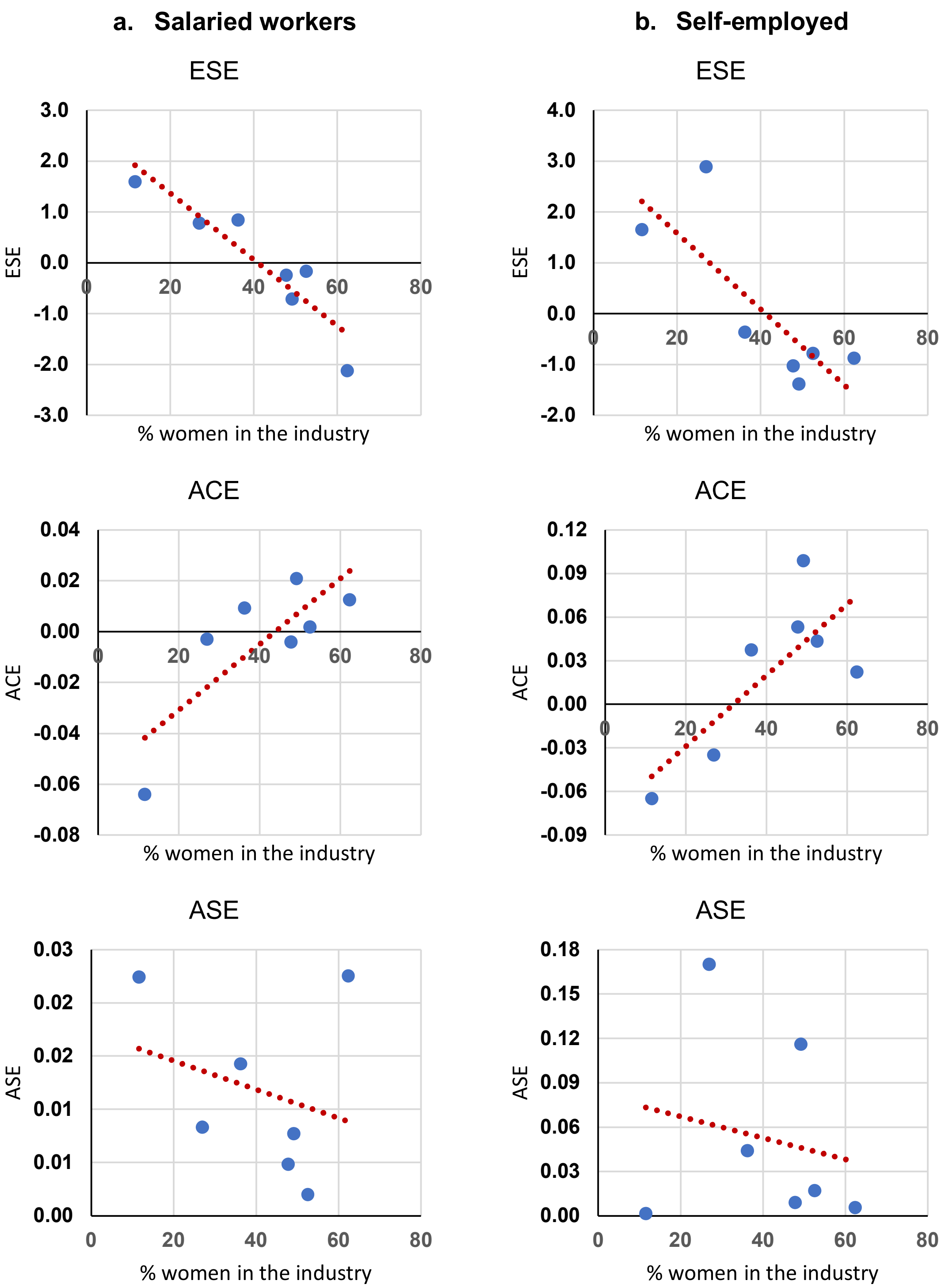


b. Structure effect



Source: Own calculations.

Figure 11. Industry feminization and decomposition components



Source: Own calculations.