Mayo 2023

WP2-2023-002

N° de serie

DOCUMENTO DE TRABAJO

Agglomeration and Skills: Explaining Regional Wage Disparities in Colombia

Autores

Diana Londoño

Andrés García

COLOMBIA CIENTÍFICA
Conocimiento Global para el Desarrollo

ALIANZA EFI
Economía Formal e Inclusiva

Agglomeration and Skills: Explaining Regional Wage Disparities in Colombia[1]

Diana Londoño Aguirre[2], Andrés García Suaza[3]

Abstract

This article aims to analyze the differences in regional labor markets in Colombia based on new theories that integrate the analysis of occupational characteristics of the workforce, urban agglomeration, and specialization patterns. In this way, it contributes to the Colombian literature since i) it adapts the classification of occupations, an important input to understanding the occupational dynamics; ii) it quantifies the wage premium differentials, incorporating skills, and iii) it decomposes the differences through the Oaxaca-Blinder model, and iv) it analyzes the differences at various points of the wage distribution based on a RIF influence function. The results allow us to conclude that there is a premium in labor income differentiated in skills for the same occupation between cities in Colombia. Likewise, evidence of a significant effect in the sectoral, geographic specialization leads to greater differences in labor income. Finally, this analysis is a key input for public policy formulation that reduces adjustments and structural skills gaps between different cities in Colombia.

Keywords: agglomeration, salary differences, skills bonus, occupational heterogeneity.

JEL Codes: J24, J31, O18

Resumen:

Este articulo tiene como objetivo analizar las diferencias en los mercados laborales regionales en Colombia a partir de nuevas teorías que integran el análisis de características ocupacionales de la fuerza laboral, la aglomeración urbana y los patrones de especialización. De esta forma, se contribuye a la literatura colombiana puesto que i)Realiza una adaptación de la clasificación de ocupaciones, insumo importante para entender las dinámicas ocupacionales; ii)Cuantifica los diferenciales de prima salarial, incorporando las habilidades, iii) Descompone las diferencias a través del modelo de Oaxaca-Blinder y iv)Analiza

---

[1] Authors thank Colombia Scientific- Alianza EFI Research Program, with code 60185 and contract #FP44842-220-2018 funded by The World Bank through the call Scientific Ecosystems, managed by the Colombian Ministry of Science, Technology, and Innovation. The authors also thank the Inter-American Development Bank for financial support under the contract ID OC#7000001919.

[2] Researcher of Alianza EFI and Observatorio Laboral de la Universidad del Rosario, LABOUR. Correo electrónico: dianai.londono@urosario.edu.co

[3] Profesor, School of Economics, Universidad del Rosario; Researcher of Alianza EFI and Observatorio Laboral de la Universidad del Rosario, LABOUR. Correo electrónico: andres.garcia@urosario.edu.co

las diferencias en varios puntos de la distribución de los salarios a partir de una función de influencia RIF. Los resultados permiten concluir que para Colombia existe una prima en el ingreso laboral diferenciada en habilidades para una misma ocupación entre ciudades. Asimismo, se evidencia un efecto significativo en la especialización geográfica sectorial que lleva a mayores diferencias en los ingresos laborales. Finalmente, este análisis es un insumo clave para la formulación de política pública que conlleve la disminución de los ajustes y brechas estructurales de habilidades entre diferentes ciudades de Colombia.

Palabras claves: aglomeración, diferencias salariales, prima de habilidades, heterogeneidad ocupacional.
Clasificación JEL: J24, J31, O18

1. Introduction

The relationship between the size of cities and labor productivity is a topic that has been extensively studied in the literature (Marshall, 1890; Jacobs, 2016; Glaeser, 2008). The main determinant of this relationship is agglomeration economies in large cities, making it possible to boost productivity. Duranton and Puga (2004) describe three main channels: exchanging, pairing, and learning. These channels refer to the fact that the agglomeration or concentration of companies and workers in a specific place provides mutualization costs and knowledge externalities. Labor markets benefit from agglomerations through improvements in job opportunities, the quality of the match between labor supply and demand, and through the promotion of economies of scale and specialization (cluster activities).

There is also a significant thread in the literature that provides evidence of the positive effect of agglomeration on wages and its interaction with the level of specialization in large cities (Wheeler, 2001; Glaeser and Mare, 2001; Wheaton and Lewis, 2002; Combes et al. 2008, Matano and Naticchioni 2011, Baum-Snow and Pavan 2012). However, the effect of agglomeration on wages is heterogeneous among workers and depends on their characteristics. Specifically, workers with higher levels of qualification or skills are the ones who benefit most from agglomeration [4] (Combes et al., 2008; Matano and Naticchioni, 2011).

The differential benefits of agglomeration for workers of different skill levels are consistent with Duranton and Puga (2004), as workers in agglomerated cities can accumulate human capital more quickly (Rauch, 1993; Glaeser, 1999; and Glaeser and Henderson, 2017), together with the possibility of developing and learning interpersonal and cognitive skills from the interaction with their counterparts, taking advantage of a more specialized market scenario (Baum-Snow and Pavan, 2012; and Andersson et al.,

[4] There is also evidence of opposite results, as described in Graham and Melo (2009) and Rinz (2018).

2013). The interaction between agglomeration economies and the skill composition of the labor force are important sources for explaining regional wage differences.

Likewise, an extensive body of empirical literature incorporates agglomeration and worker skill as regional sources of wage disparity.[5] For example, Combes et al. (2008) study spatial disparities in France and quantify the explanatory power of the skill composition of the labor force, local endowment (e.g., geography, local institutions, etc.), and the interaction between workers and firms. The results suggest that the skills composition explains an important part of the regional salary differences, which is spatial classification evidence.[6]

For developing countries, Pan et al. (2016) and Neves et al. (2017) find a significant effect of agglomeration that varies between Chinese and Brazilian workers with different skill levels. In China, Pan et al. (2016) find that the effects of agglomeration are smaller compared to developed economies, while Neves et al. (2017) find a significant and positive correlation between the premium of cognitive and social skills and the size of the cities. Both the skill composition and the skill premium are crucial to understanding regional wage differentials.

This article studies the differences in regional labor markets in Colombia and how these are related to the levels of agglomeration of cities and the composition of skills. With this in mind, a methodology is proposed to measure the skills in occupations in Colombia based on the Occupational Information Network [O*NET] and following the previous contributions of Autor et al. (2003), Autor & Dorn (2013), and Autor (2013). Besides the above, decomposition methods are also implemented to study the role of agglomeration and workers with high levels of qualification or skills on the regional wage differential.

Measuring skills is not a trivial task. Pioneering studies used the educational level as an approximation of ability. However, this does not consider important dimensions associated with workers' abilities to perform specific tasks. An adequate measure of skills[7] requires identifying important dimensions, such as cognitive, social, and motor skills, not just staying with educational achievement. In the recent literature, where this perspective of horizontal differentiation of skills is refined, the difficulty of measuring the task content of a job lies in the fact that the characteristics of the occupations, which change over time, often until reaching generate new labor categories, added to the fact that these patterns can vary between countries (Autor & Handel, 2013; Spitz-Oener, 2006) (Arntz et al., 2016).

---

[5] Some seminal works have explored the relationship between the regional wage differential and different factors such as migration, living costs, and the labor force composition (Krumm, 1983; Newman, 1982).
[6] This is the result of different circumstances, such as the fact that large cities offer a wide range of services that attract workers and promote the accumulation of human capital.
[7] A skill is understood as the entire combination of skills, experience, and knowledge on a specific subject. At the same time, a task is related to a specific action needed to perform a job and requires specific skills.

The pioneering work of Autor et al. (2003) proposes a classification of jobs in terms of the composition of skills required. It provides evidence of the relationship between the task content of jobs and technological progress, trade opportunities, and the skills premium. A classification based on O*NET allows job task content to be identified in various categories, including analytical, manual, and cognitive skills and subcategories based on whether they are routine or non-routine. Because of this structure, these contributions have been a crucial input for the growing literature dedicated to research on how skills impact labor market outcomes, such as the skills mismatch and the skills premium.

Other important contributions are those made by Handel & Autor (2013), who examine links between the endowment of human capital, occupations, skills, and wages. Autor & Dorn (2009) investigate the impact of technological progress on demand for skills under the concept of wage polarization. These contributions show that jobs with a high routine and manual content are at risk of being replaced by automation or machines. In contrast, jobs with high cognitive and non-routine content complement technical and technological progress and, therefore, will gain share over time. (Bachmann (2018), Cortes (2018), Danieli (2017), Goos et al. (2011), Goos & Manning (2007), Firpo et al. (2011, 2018), Autor & Dorn (2009), Bhalotra & Fernández (2018)) (see Autor et al., 2006, 2008; Autor & Dorn, 2013; Bhalotra & Fernandez, 2018; Bizopoulou, 2017; Firpo et al., 2011, 2018).

The previous paragraphs show that the literature on the relationship between agglomeration and wages has grown in recent years. Following this line, Matano and Naticchioni (2011) propose an analysis using a database at the employer-employee level for Italy. They find evidence of the complementary effect of agglomeration and skill composition[8] and register evidence in favor of a specialization premium, which shows that workers in more specialized cities (concentration of workers in a particular sector) receive higher wages (see also Wheaton and Lewis, 2002). The authors argue the results are explained because non-routine skills (e.g., analytical, social, and cognitive skills) have a higher premium (Glaeser and Maré 2001; Wheeler 2006; Yankow 2006; Gould, 2007; Melo et al., 2009; Bacolod et al., 2009; Puga, 2010; and Florida et al., 2012; Andersson et al., 2013; and Neves et al., 2017). Complementarily, Wheeler (2001, 2004, and 2007) shows evidence that the impact of agglomeration is also on wage inequality.

For the development of this article, information from the Great Integrated Household Survey (GEIH, for its Spanish acronym) is used, a survey that collects data every month throughout the Colombian territory on the workforce for the period 2019. This analysis is initially presented for the 13 main cities and their metropolitan areas since these have significant differences in terms of population, the performance of their labor markets, and level of sectoral specialization. According to the official statistics of the National Administrative Department of Statistics (DANE, for its Spanish acronym), Bogotá Capital

[8] There is also additional mixed evidence with negative results presented in the study by Rinz (2018) and no effect in Melo (2009).

District has over eight million inhabitants. In contrast, Pasto has about three hundred thousand. Similarly, for 2019, Barranquilla's unemployment was 7.8%, while Cúcuta's was 15.8%.

Duranton (2016) estimated that the wage elasticity of the agglomeration was 5%, which implies the existence of a wage premium for city size in line with what was postulated by Baum-Snow and Pavan (2012). Previous works on regional wage differentials have focused on explaining these dynamics from the perspective of segmented markets (Nupia, 1997, Galvis, 2002; Mesa et al., 2008). Given this, it is proposed to explore the sources of this wage differential by studying the specific role of the skills composition and the skills premium.

One of the main challenges of this task is to adapt the skills measures of the O*NET scheme to the Colombian case and the GEIH as the chosen source of information. Based on the Autor & Dorn (2013) classification, matching categories were constructed between the National Classification of Occupations (CNO70)[9] and the O*NET, assigning levels to three types of skills for each worker, specifically: routine, manual and cognitive. This information was used to estimate salary equations by city. Specialization indices were included at the local and sectoral level in the style of the proposals by Combes (2000), Mion & Naticchioni (2009), Matano & Naticchioni (2012), and Glaeser & Henderson (2017). These indices represent the relative share of each sector concerning total employment in a particular city, comparing this proportion with the same measure for the entire country.

Finally, it was studied that these three factors explain regional wage differentials. In particular, the Oaxaca-Blinder decomposition (Oaxaca, Blinder, Fortin) was used, using Bogotá as a reference point and comparing it with the rest of the cities. Finally, an additional analysis is carried out at various points of the wage distribution from a Recentered Influence Function [RIF]. The results provide evidence of the effect of agglomeration on work and its interaction with the skills premium for the case of Colombia. The relative importance of the skills premium and the effect of specialization are shown as explanatory variables of regional wages. Differences in the wage premium for specific skills are considerable in magnitude and remain significant after controlling for differences in returns to education across cities.

The rest of this article is organized as follows; Following this first introductory section, the second section presents the methodology and explains the adaptation of the classification of occupations used to measure the composition of skills in Colombia. The third section describes the relationship between skill composition and agglomeration patterns. The fourth section contains the results for the decomposition of the wage

[9] This classification is used in the GEIH as a referent categorizer for the variable job performed by the employed person.

differential. Finally, the fifth section presents the conclusions drawn from the results of the exercises.

2. Measurement of skills content for occupations in Colombia

We will call skill content the level of intensity with which an occupation requires a skill. Skill is generally defined as applying knowledge about a subject to perform a task. While the task is related to the specific actions needed to perform a job, it requires specific skills.

For example, driving involves skills such as reacting in a short time or memory. Therefore, an occupation has different skill levels measured through a series of tasks or set of actions that a worker performs in the job. These tasks are of multiple nature. For example, some are repetitive or routine, and others are related to analytical knowledge and reasoning, such as reading or calculating equations. Measuring the level of each dimension attached to each occupation requires detailed information about the tasks and their intensity.

Colombia has an occupational classification[10] that provides a framework for labor market analysis and is the basis for the data collection on this subject, including household surveys.[11] However, as in many other countries, data on the intensity required for these occupations to perform specific tasks is not available. Therefore, it is necessary to associate the National Occupational Classification of Colombia (CNO, for its Spanish acronym) [12] with the International Standard Classification [ISCO], which allows measuring the content of tasks by associating it with databases that measure the intensity of skills for other countries such as the USA.

To study the change in skill requirements in the US labor market and test the polarization hypothesis, Autor et al. (2003) and Autor & Dorn (2013) provide a methodology to measure five skill categories: non-routine analytical (Analytic), non-routine interactive (Interactive), non-routine manual (NR. Manual), routine cognitive (Cognitive) and manual routine (R. Manual).

Annex 1. (see Annex) provides definitions and examples of each category of competencies. Based on the occupational classification system registered in the Survey of the American Community [COC] and the Dictionary of Occupational Titles [DOT], the documents propose a method to construct a set of skills scores consistent with the O*NET, a modern version of DOT. In particular, DOT makes it possible to identify the intensity of a task,

[10] Organized and aggregated system of data related to occupation, understood as a set of jobs whose main tasks and functions are characterized by a high degree of similarity that guarantees coherence between the collection, processing, analysis, and dissemination. Constituted in an instrument of harmonization and statistical infrastructure.
[11] For this exercise, where the GEIH of the DANE is used, the classification of occupations used in the GEIH corresponds to the National Classification of Occupations (CNO70).
[12] In Colombia, there are two occupational classifications: i) The National Classification of Occupations (CNO, for its Spanish acronym) and ii) The International Standard Classification of Occupations Adapted to Colombia –[ISCO-A.C.]. Both classifications since their generation have several updates.

such as performing calculations, repetitive movements, or interacting with customers. Also, it allows matching the occupation with the categories collected by the census and household surveys. For example, Autor et al. (2003) aggregate the occupations in 450 categories available in the COC and select a set of variables to measure each skill. These are coded on the 1-10 scale, where high values refer to more intensive use of the task, and skill measures are estimated as an index that aggregates task intensity using principal components analysis.

The main challenge is establishing a bridge between the DOT-COC and the CNO70 to assign skills measures to Colombian workers. This requires the use of additional occupational classifications and their corresponding correlative tables. In particular, the Standard Occupational Classification System [SOC], the International Standard Classification of Occupations [ISCO], which is an international reference, and two versions adapted for Colombia (ISCO88 and ISCO08) are used. The process matches the DOT-COC occupations (485 occupations) with the SOC, which maps directly to the international ISCO (436 occupations). Likewise, the ISCO has the corresponding versions adapted to Colombia (449 occupations) that serve as a reference for the CNO70 (285 occupations). If a correlative table merges various occupations from one classification to another, the measures are collapsed using an average.

The adaptation for Colombia of these measures is made for data from the GEIH[13] for the 13 main urban areas[14] for 2019.[15] The adaptation variable is the job variable, which allows, at an aggregate level, 83 categories of the CNO70 and 43 categories of the ISCO 08 A.C.[16] to relate what the employed person does in his job worked. Table 1 and

Annex 2 present the result of the adaptation and imputation of the skill intensity indicator under the ISCO 08 A.C classification.[17]

[13] This survey contains information on socioeconomic conditions, with a particular emphasis on employment, which focuses on the supply component of the labor market.

[14] The DANE defines an Urban Area (Metropolitan Area, AM, for its Spanish acronym) as an area of influence that includes surrounding municipalities, which with the city make up a single non-discontinuous urban fabric and have been legally recognized. The urban areas are: Medellín AM-MED-, Barranquilla AM-BAR-, Bogotá AM-BOG-, Manizales AM-MAN-, Ibagué AM-IBA-, Montería AM-MON-, Cartagena AM-CAR-, Villavicencio AM-VIL-, Pasto AM-PAS-, Cúcuta AM-CUC-, Pereira AM-PER-, Bucaramanga AM-BUC-, Cali AM-CAL-.

[15] The reference period was chosen based on data availability, in this case, the last year with available information.

[16] ISCO 08 A.C is a classification adapted by DANE from the international ISCO of the International Labour Organization [ILO], which is characterized by having a hierarchical structure based on four categories: large group, principal subgroups, subgroups, and primary groups. In addition to the structure, the classification is built under two relevant criteria: the level of skills and the specialization of skills.

[17] The results are presented using ISCO 08 A.C at the subgroup level due to international comparability and the classification structure based on skill levels and skill specialization.

When analyzing the results of the adaptation at the level of Large Groups and Principal Subgroups,[18] consistency is observed in the skill intensity index associated with the characteristics (nature of work performed, level of formal education, areas of knowledge, tools, and machines used, materials with which one works, and types of goods and services produced) typical of the occupations of ISCO 08 A.C.

An example of this is highlighted in red in Table 1, where for the large group of directors and managers occupations, there is a higher skill intensity index for non-routine analytics (5.27) and interactive (interpersonal) non-routine (7.70), where skills associated with leadership, management, control, planning, responsibility, among others, are required. Compared to a lower routine manual index (2.80) involving skills associated with strength, endurance, flexibility, balance, and coordination.

Another example is that of the group of Officials, Operatives, Craftworkers, and Related Trades, who have higher routine cognitive skills (7.67) and routine manual skills (4.48) compared with non-routine interactive skills (0.90). These busy people have greater abilities associated with adaptability to situations that require the precise achievement of established limits, tolerances, or standards and skills to move their fingers and manipulate small objects with their fingers, quickly or precisely. Skills that influence strength, endurance, flexibility, balance, and coordination.

Table 1. Skill intensity index for occupations classified under ISCO 08 A.C at the Large Group level.

Code	**Large Group**	**Non-Routine Analytical Index**	**Non-Routine Interactive Index**	**Routine Cognitive Index**	**Routine Manual Index**	**Non-Routine Manual Index**
0	Military forces	2,13	1,11	1,07	2,53	3,28
1	Directors and Managers	5,27	7,70	0,70	2,80	0,39
2	Professionals, Scientists, and Intellectuals	6,01	5,53	2,79	3,70	0,67
3	Intermediate Level Technicians and Professionals	4,46	3,09	4,08	3,77	1,05
4	Administrative Support Staff	3,04	1,19	5,00	3,95	0,21
5	Service Workers and Sellers of Shops Markets	2,79	1,48	1,33	3,21	1,44

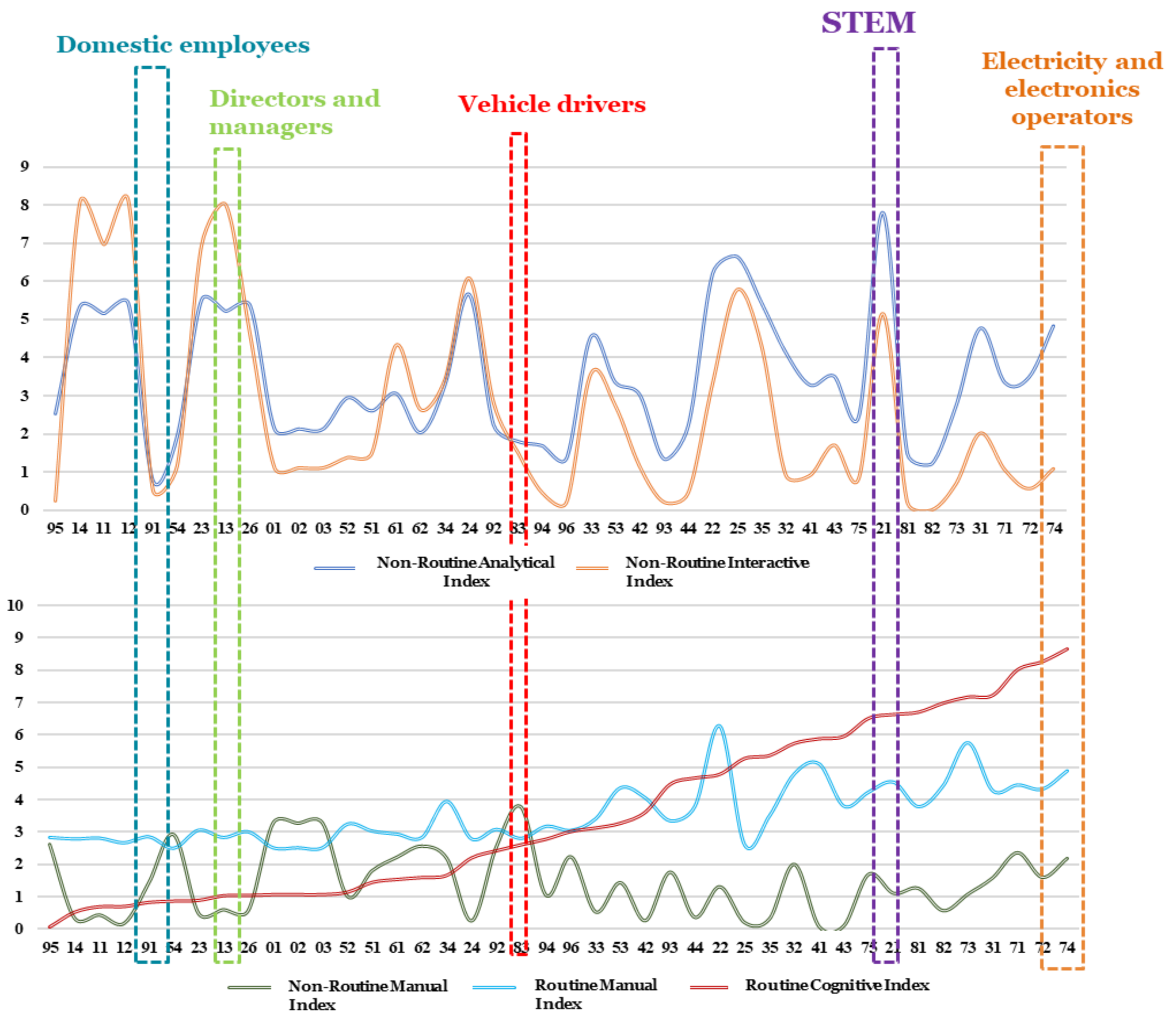
[18] To obtain the value of the skills index at the Large Group level, each of the indices obtained at the level of principal subgroups was averaged.

Code	Large Group	Non-Routine Analytical Index	Non-Routine Interactive Index	Routine Cognitive Index	Routine Manual Index	Non-Routine Manual Index
6	Farmers and Skilled Agricultural, Forestry, and Fisheries Workers	2,57	3,52	1,56	2,90	2,40
7	Officials, Operatives, Craftworkers, and Related Trades	3,31	0,90	7,67	4,48	1,97
8	Plant and Machine Operators and Assemblers	1,73	1,20	3,60	3,06	3,16
9	Elementary Occupations	1,17	0,56	2,03	3,02	1,63
Source: Authors' work						

Finally,

Figure **1** presents some additional examples for the skill intensity indices for various principal subgroups, such as i) Domestic workers with a higher index for routine manual skills, ii) Directors and managers with a higher index for cognitive and interactive skills non-routine, iii) drivers with a higher rate of non-routine manual skills and iv) STEM or science and engineering occupations with a higher rate of non-routine cognitive skills.

Figure 1. Skill intensity index for occupations classified under ISCO 08 A.C at the Principal Subgroup level



Cod	Name
01	Officers of the armed forces
02	Non-commissioned officers of the armed forces
03	Other members of the armed forces
11	Executive directors, managerial personnel in administration and legislatur
12	Administrative and commercial directors
13	Directors and managers in production and service sectors
14	Hotel, restaurant, trade, and other service managers
21	Professionals in sciences and engineering
22	Healthcare professionals
23	Education professionals
24	Business and administration professionals
25	Information and communications technology professionals
26	Legal, social, and cultural professionals
31	Technicians and mid-level professionals in sciences and engineering
32	Mid-level healthcare technicians and professionals
33	Mid-level finance and administration technicians and professionals
34	Mid-level technicians in legal, social, cultural, and related services
35	Information and communications technology technicians
41	Office workers
42	Customer service representatives
43	Accounting assistants and material recording clerks
44	Other administrative support personnel
51	Personal service workers
52	Salespersons
53	Personal care workers
54	Security service personnel
61	Skilled farmers and agricultural workers
62	Skilled forestry workers, fishermen, and hunters in market-oriented activit
63	Agricultural, fishing, hunting, and subsistence gathering workers
71	Construction officers and operators (excluding electricians)
72	Metallurgy officers and operators; mechanics and machine repairers
73	Craft and graphic arts workers
74	Electricity and electronics officers and operators
75	Food processing officers, textile workers, carpenters, and related workers
81	Fixed facilities and machine operators
82	Assemblers
83	Vehicle drivers and heavy mobile equipment operators
91	Domestic and cleaning personnel
92	Agricultural, fishing, and forestry laborers and unskilled workers
93	Mining, construction, manufacturing, and transportation laborers and uns
94	Food preparation assistants
95	Street vendors of services and related occupations
96	Waste collectors and other elementary occupations

Source: Authors' work

2.1 Empirical Strategy

Considering that skill directly affects wage determination due to its relationship with worker productivity and relative scarcity in the market, as previously evidenced, this section describes the econometric strategy used to study how variations in skill distribution affect wage levels and inequality at the regional level. The Oaxaca-Blinder decomposition method is employed, which allows for an accounting exercise to capture the portion of the wage difference that is attributable to a specific factor, in this case, skills.

This procedure, known in the literature as the Oaxaca-Blinder decomposition (Blinder, 1973; Oaxaca, 1973), divides the wage differential between two groups into a part that is "explained" by differences in productivity-related characteristics between the groups, such as education or work experience, and a residual part that such differences in wage determinants cannot explain. This "unexplained" part is often used to measure discrimination but also includes the effects of group differences in unobserved predictors. In other words, the Oaxaca-Blinder procedure provides a way to decompose wage changes or differences into a wage structure effect and a composition effect. Furthermore, it allows for further disaggregation of these two components so that the contribution of each covariate can be assessed.

This method has been used by a considerable amount of literature, such as Bhalotra and Fernández (2018), Firpo et al. (2011-2018), Matano and Naticchioni (2018), and Bacolod (2016), among others, who aim to analyze wage differentials between subgroups through mean comparisons. In all these cases, the key question of economic interest is which factors explain changes (or differences) in the distributions.

Compared to other decomposition techniques, this method has several advantages due to its simplicity of implementation and efficiency (Busso et al., 2014; Firpo, 2007). It imposes fewer assumptions on the data and provides more informative results since it does not impose a functional form on the earnings equation and allows for the evaluation of differences across the entire distribution. However, the main difficulty lies in the analysis itself, as it only allows for comparisons between groups (pairs) of cities and does not establish the statistical significance of each component of the wage differential. Annex 4 provides a more detailed theoretical description of the Oaxaca-Blinder decomposition (1973).

Equations to be Estimated

In line with those mentioned above, a two-stage model is estimated. In the first stage, the importance of these skills in wage determination is examined by analyzing differences between cities. For this purpose, the skill premium equation is estimated using a linear specification of the Mincer type for worker i in city j during period t.

$$LnW_{ijt} = \beta_{jt}X_{ijt} + \lambda_{jt}S_{ijt} + \alpha_{jt}Z_{ijt} + \varepsilon_{ijt} \qquad (1)$$

Where LnW_{ijt} represents the natural logarithm of hourly wages, X_{ijt} is a set of socioeconomic characteristics such as age, gender, years of schooling, type of worker (wage employee or self-employed), type of employment (formal or informal), S_{ijt} refers to the skills associated with the worker's occupation, Z_{ijt} is an index of geographic specialization[19], and ε_{ijt} represents unobserved factors. β, λ, and α are parameters to be estimated.

Once the Mincerian regressions are estimated in the first stage, a decomposition is performed to isolate the effect of skills on changes in average wage levels and wage inequality. For this purpose, a comparison is made between Bogotá (City 0), the main city (with higher labor incomes and a larger number of participants in the labor market), and the rest of each city (City 1).

This decomposition model can be written as follows:

$$\begin{aligned} E(LnW \mid c = 1) - E(LnW \mid c = 0) = {} & \beta_{t1}[E(X_t \mid c = 1) - E(X_t \mid c \\ & = 0)] + \lambda_{t1}[E(S_t \mid c = 1) - E(S_t \mid c \\ & = 0)] + \alpha_{t1}[E(Z_t \mid c = 1) - E(Z_t \mid c = 0)] + E(X_t \mid c \\ & = 0)(\beta_{t1} - \beta_{t0}) + E(S_t \mid c = 0)(\lambda_{t1} - \lambda_{t0}) + E(Z_t \mid c \\ & = 0)(\alpha_{t1} - \alpha_{t0}) \end{aligned} \qquad (2)$$

$$= A + B + C + D + E + F$$

A + B + C is the composition effect (EC) or the portion of the wage differential associated with differences in population characteristics. At the same time, D + E + F refers to the structure effect (ES) or the part of the wage differential due to prices of characteristics, skill, and geographic specialization. In turn, B + E measures the portion of the wage differential explained by skills.

[19] Este índice de especialización del sector a nivel local se calcula siguiendo a autores como Mion & Naticchioni (2009) y Matano & Naticchioni (2012). Quienes estiman este indicador como la relación entre la participación del empleo sectorial en el empleo total en cualquier ciudad sobre la participación correspondiente al empleo sectorial a nivel nacional y el empleo total a nivel nacional.

Finally, an additional decomposition exercise is conducted to study the reasons behind differences in other points of wage distribution or wage inequality. For this purpose, the methodology proposed by Firpo, Fortin, and Lemieux (2009) is employed, which suggests the use of recentered influence functions (RIF). This function-based method has a similar logic to the Oaxaca-Blinder decomposition (1973) as it allows for the decomposition of each factor and evaluation of results at the mean and other points of the distribution. In this case, we will evaluate the results at the 25th percentile, 75th percentile, and Gini coefficient. The equation to be estimated is as follows:

$$\boldsymbol{RIF_{ij}} = \boldsymbol{\beta_j X_{ij}} + \boldsymbol{\lambda_j S_{ij}} + \alpha_j \boldsymbol{Z_{ij}} + \boldsymbol{\varepsilon_{ij}} \quad (3)$$

3. Regional wage disparities, agglomeration, and skills

3.1. Regional disparities in regional labor markets

The Colombian labor market at the regional level is quite heterogeneous. The difference between the cities, especially the main cities, in terms of performance and the level of sector specialization has driven the growth in the wage gap. To carry out this analysis, the GEIH information for 2019 is used for the employed population,[20] defined as salaried and self-employed. The total sample[21] used corresponds to 137,281 observations, equivalent to 8,809,852 workers belonging to 83,822 households surveyed. Salary calculations are made based on hourly labor income, thereby analyzing the actual price of an hour of work for the different urban areas.

Figure 2 and **¡Error! No se encuentra el origen de la referencia.** summarize the main labor market indicators for the 13 main urban areas. In these, significant differences can be observed regarding the size of the population, the employed population, the informal population, and the hourly wage, among other characteristics between the cities.

To illustrate the above, one of the urban areas with the highest employment rate (61.5%), the highest labor income per hour (7,235), and the lowest informality rate (40.6%) is Bogotá. If this labor market is contrasted with Medellín, Cali, Bucaramanga, and Cúcuta, Medellín, for example, is a city with a labor market like that of Bogotá; in terms of hourly labor income (6,555) and the informality rate

[20] According to DANE, an Employed person, is a person of working age who, during the reference period: 1. worked at least one paid hour in the reference week; 2. did not work during the reference week but had a job; 3. unpaid family workers who worked in the reference week for at least 1 hour. These employed persons are categorized according to the economic remuneration in consideration for the work carried out. These may be wage earners, self-employed, domestic employees, and/or other family employees or day laborers.

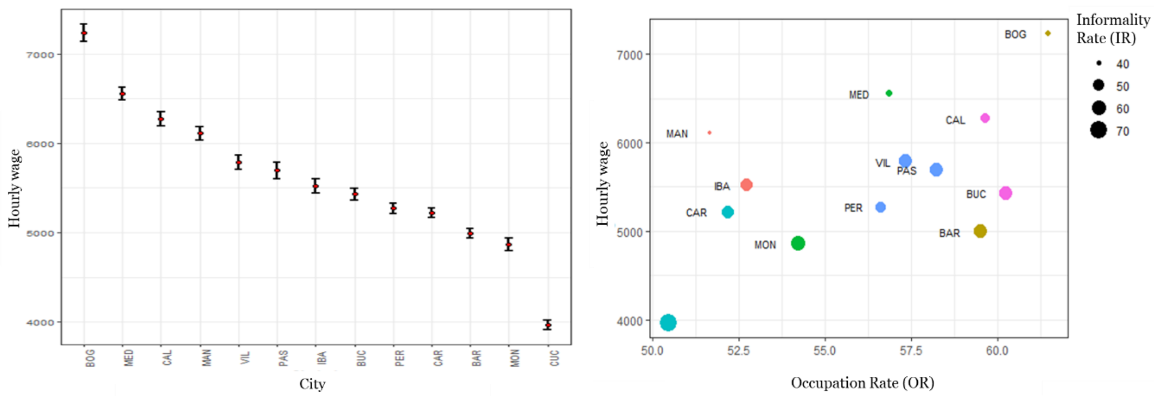
[21] The total sample used responds to a statistical standardization process carried out on the hourly labor income variable.

(41.6%), however, it has a much lower employment rate (56.8%). Bucaramanga is a city that has one of the highest occupancy rates among the regions (60.2%). Still, it has the lowest labor income per hour (5,431), 24.9% lower than labor income per hour than Bogotá, with 14.7 pp more in its informality rate (55.3%).

Figure 2. Labor market indicators for the 13 main cities and their metropolitan areas

A. Hourly wage

B. Occupation Rate (OR) and Informality Rate (IR)



Note: BAR: Barranquilla AM, BOG: Bogotá D.C., BUC: Bucaramanga AM, CAL: Cali AM, CAR: Cartagena, CUC: Cúcuta AM, IBA: Ibagué, MAN: Manizales AM, MED: Medellín AM, MON: Montería, PAS: Pasto, PER: Pereira AM, VIL: Villavicencio
Source: Authors' work

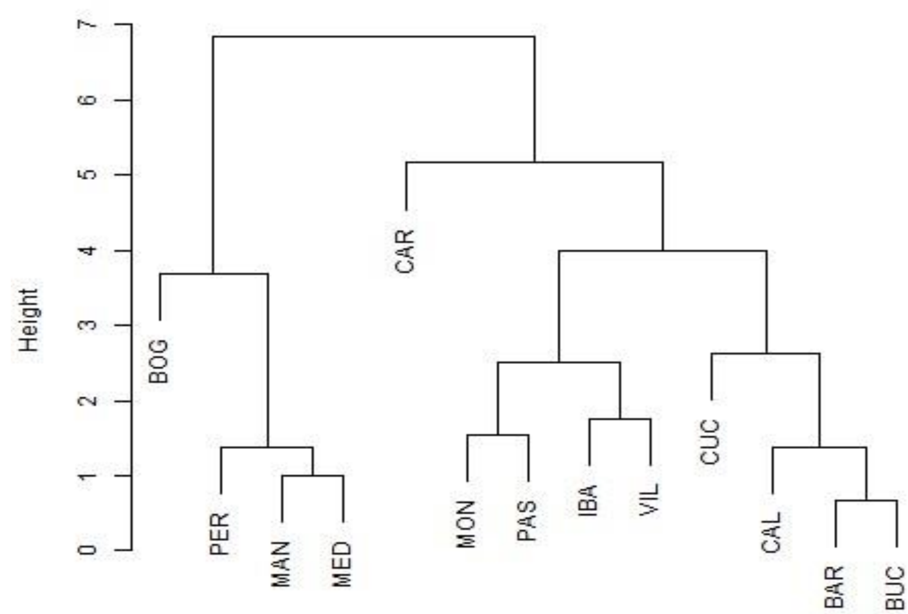
Likewise, Cali is a city that could be said to have a high employment rate (59.6%), a low level of informality (45.7%), and an average level of labor income per hour (6,273); however, in all its indicators, it presents strong differences to Bogotá. Finally, there is the city of Cúcuta, which during 2019 presented the worst labor market indicators, which led to a significant gap with the other regional labor markets, not only with Bogotá.

Observing the hourly wage in To illustrate the above, one of the urban areas with the highest employment rate (61.5%), the highest labor income per hour (7,235), and the lowest informality rate (40.6%) is Bogotá. If this labor market is contrasted with Medellín, Cali, Bucaramanga, and Cúcuta, Medellín, for example, is a city with a labor market like that of Bogotá; in terms of hourly labor income (6,555) and the informality rate (41.6%), however, it has a much lower employment rate (56.8%). Bucaramanga is a city that has one of the highest occupancy rates among the regions (60.2%). Still, it has the lowest labor income per hour (5,431), 24.9% lower than labor income per hour than Bogotá, with 14.7 pp more in its informality rate (55.3%).

Figure 2, panel A, there is a difference in favor of Bogotá throughout the entire distribution. This is greater at the 75th percentile, followed by the difference at the 25th percentile. This could show that, for an analysis of wage differences per hour, it is not enough to characterize the differences in the mean since their magnitude varies throughout the distribution. In addition, if the greatest differences are observed in the 75th percentile, it would be necessary to validate whether the greatest difference is occurring in workers with occupations with a higher qualification.

As a complement to the previous analysis, Figure 3 presents a dendrogram that allows analyzing the similarity of the labor markets based on the skill intensity indices. In observing this, cities such as MED and MAN have an important level of similarity in their labor markets, followed by PER and then BOG. These cities have the highest hourly wage allocation and are similar in their labor markets. This could show that skills demand patterns can also determine the wage allocation of cities. The right side of the figure presents more heterogeneous markets. However, certain proximity or similarities can be seen in some, such as MON and PAS or BAR and BUC. Finally, CAR is the most different city in terms of skills demand from all of them. The high unemployment rates could explain the high rates of informality and higher youth unemployment that this city faces.

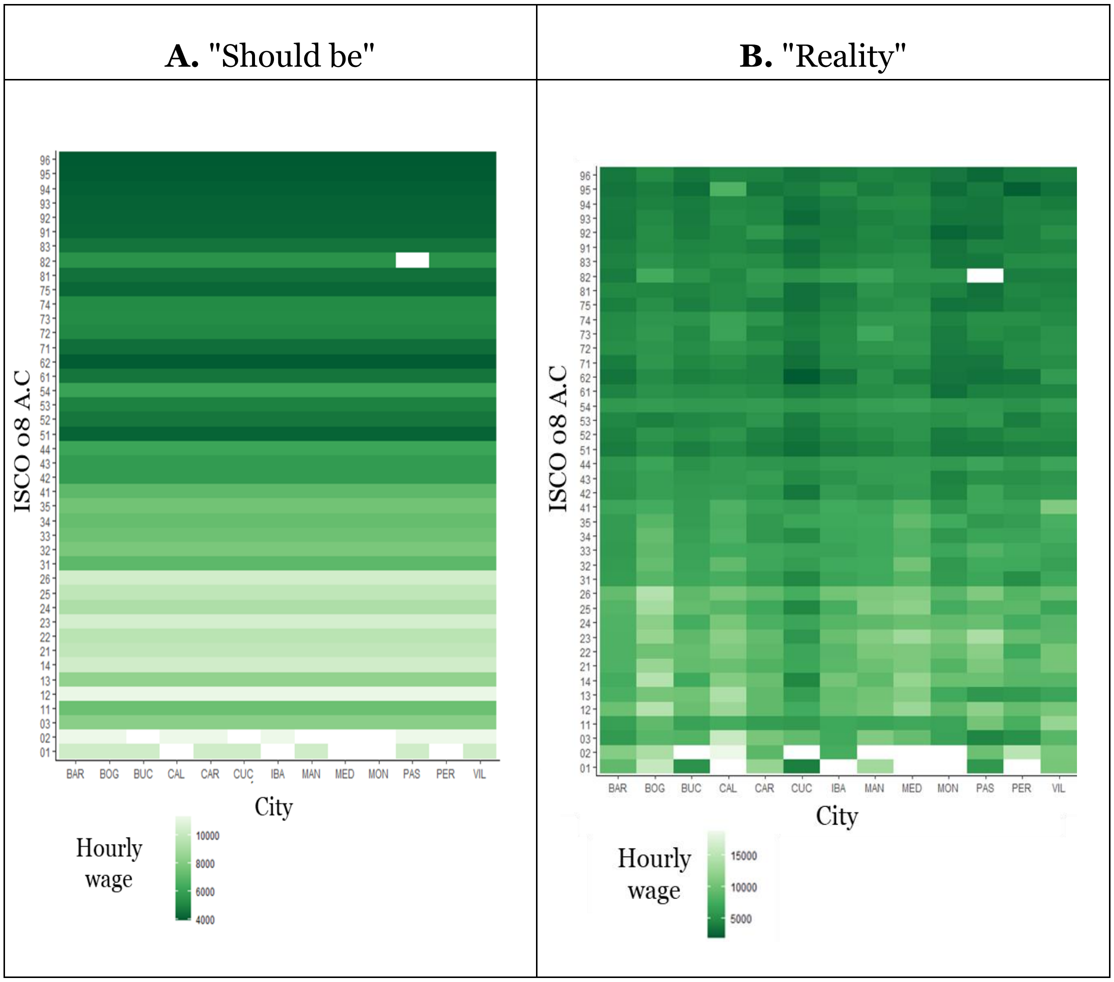
Figure 3. Regional comparison of labor markets based on skill intensity indices.



Source: authors' work using GEIH 2019.

Figure 4 presents a color intensity diagram that represents, in panel A, the "should be"; There is no difference between the cities in the hourly wage paid for the same occupation (in this case principal subgroup of occupations).[22] Panel B presents the "reality" of what is happening at the regional level in Colombia. This shows the hourly wage paid by the local labor market for each occupation in each city. As the intensity of the color increases, the average value paid for that occupation in that city decreases.[23] In addition, the greater intensity of the color within the occupation also shows the lower-paid value of that occupation in the entire Colombian labor market.

Figure 4. Hourly wage by occupation for the 13 urban areas



Note: the 2-digit ISCO 08 codes used are presented in Figure 1.
Source: Authors' work

An example of this is the occupations belonging to group 22 of Health Professionals, where you can find occupations such as doctors, dentists, professional nurses, etc. In cities like BOG, MED, and MAN, they have wages between 10,000 and 11,000

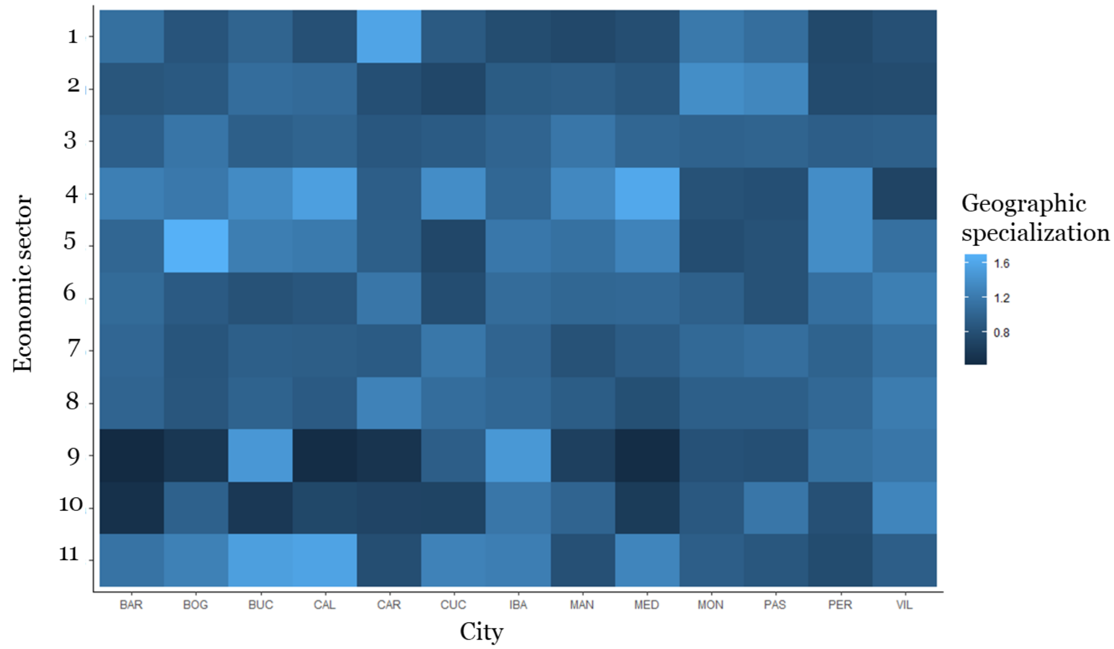
[22] The data associated in the illustration corresponds to the average salary of the occupation at the national level.
[23] The entry boxes associated with occupations found in the illustration are because there is no salary information for said occupation in some cities.

pesos per hour. Another of the occupational groups with the highest rates of remuneration is 31 mid-level technicians and professionals in science and engineering, where there are occupations such as electronics technicians, telecommunications technicians, systems technicians, etc., which are occupations that, because of technological impacts, have become occupations with a great impact and therefore the occupation for which the Colombian labor market presents the highest rate of hourly labor remuneration in the city of BOG with a value of 8,723 pesos, followed by cities such as MED, MAN and CAL with values of 8,658; 7445 and 7823.

As stated above, exploring the role of agglomeration economies applies to assessing the importance of urban economies and local sectoral specialization (Combes, 2000; Combes et al., 2008; Mion & Naticchioni, 2009; Glaeser & Henderson, 2017). Because agglomeration works as a major factor leading to regional disparities, especially in the labor market (Krugman, 1991; Fujita et al., 1999), salary levels and employment opportunities depend on the regional concentration of each market (Duranton, 2016). Following Matano & Naticchioni (2012), we constructed a specialization index for the sector at the local level based on the share of sectoral employment in total employment in any city over the share of corresponding sectoral employment at the national level.

Figure 5 presents the results of the local sectoral specialization index. In this, it is possible to observe not only by cities which is the sector in which the labor market is concentrated; but it also allows showing by sectors, which are the cities that develop activities concentrated in specific sectors to a greater extent. An example of this is the Manufacturing industry, a sector for which most cities (except VIL) present very high specialization rates. In the same way, when observing by cities, it is clear how Medellín is strongly specialized in Industry; Bogotá in financial intermediation; Bucaramanga and Ibague in agriculture and mines and Cartagena in transportation and telecommunications.

Figure 5. Specialization index by economic sectors for the 13 urban areas



Note: The economic sector: 1. Transport, 2. Health, 3. Other services, 4. Industry, 5. Financial, 6. Construction, 7. Trade, 8. Accommodation and meals, 9. Agriculture and mine, 10. Public administration, 11. Activities Households.
Source: authors' work using GEIH 2019

Finally, after describing and analyzing some of the regional differences observed from the data; It can be concluded in this section that there are significant variations in terms of the characteristics of the workers, the skills, and the sectoral specialization between the cities studied; which is why it is justified to delve into the analysis, considering the intensity of skills and the index of sectoral geographic specialization.

3.2. The regional hourly wage premium

Based on the estimation of equation 1 to evaluate skills in wage premiums, the results presented in Figure 6 and Annex 5 are obtained. It can be observed how educational attainment, age, gender, formality, and employment type generally affect wage premiums in each region. In the case of age, it is significant for all urban areas and, on average, increases hourly labor income by 3.0%. Regarding educational attainment, the increase in average hourly income ranges from 3.7% to 5.0%.

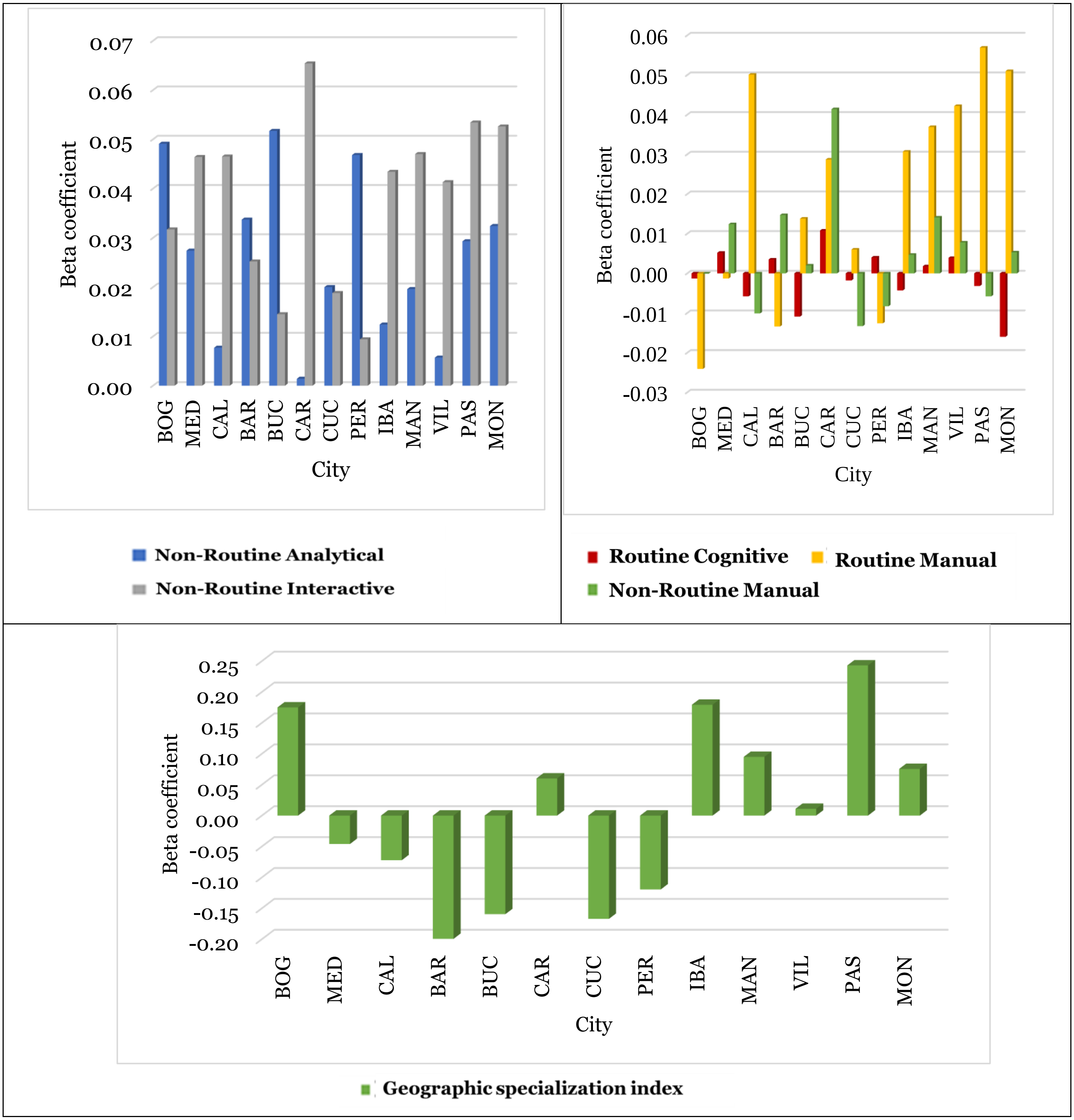
On the other hand, being self-employed or female reduces hourly labor income, with a more significant impact ranging from 11.0% to 20.0% compared to wage workers in the areas of Monteria, Pereira, Cucuta, and Cartagena, and from 12.0% to 18.0% for women in Barranquilla, Bucaramanga, Cartagena, Ibague, Cucuta, Pasto, and Monteria. Finally, formality significantly increases hourly labor income compared to informality, with more significant increases in cities such as Medellin (33.0%), Barranquilla (37.9%), Pasto (39.8%), and Cali (34.9%). In contrast, the increase ranged between 25.0% and 30.0% for the rest of the cities.

As well as high heterogeneity was found in wage premiums associated with sectoral geographic specialization and skills. It can be observed that non-routine analytical and interactive (interpersonal) skills positively affect wage premiums for all cities. In the case of cities like IBA, PAS, CAL, MON, VIL, and MAN, these skills are particularly important and substantially increase hourly labor income compared to Bogotá. For non-routine manual skills, CAR gives more importance to this type of skill, which may be closely related to its specialization index in the transportation sector (see Figure 6). Lastly, geographic specialization is not always beneficial in all cities. For MED, CAL, BAR, BUC, CUC, and PER, this index reduces hourly labor income, indicating that sectoral specialization can lead to an abundance of resources in each region but also increase the cost of those resources.

The main results of this estimation for each of the studied urban areas are described below.

Bogotá: Non-routine analytical skills have the highest increase in wage premiums in this region, around 4.9%, followed by interactive (interpersonal) skills at 3.2%. Routine manual skills decrease hourly labor income by 2.7%. The specialization index is significant and increases hourly labor income by 17.5%.

Figure 1. Premium salary skills and specialization



Source: authors' work using GEIH 2019.

Medellín: Interactive (interpersonal) skills have the highest increase in wage premiums in this region, around 4.6%, followed by analytical skills at 2.7% and non-routine manual skills at 1.2%. The specialization index is significant and negative, reducing hourly labor income by 4.7%.

Cali: Routine manual skills have the highest increase in wage premiums in this region, around 5%, followed by non-routine interactive (interpersonal) skills at 4.6%. The specialization index is significant and negative, reducing hourly labor income by 7.2%.

Barranquilla: Non-routine analytical skills have the highest increase in wage premiums in this region, around 3.3%, followed by interactive (interpersonal) skills at 2.5% and routine manual skills at 1.4%. The specialization index is significant and negative, reducing hourly labor income by 19.9%.

Bucaramanga: Non-routine analytical skills increase hourly labor income by 5.1%, followed by interactive (interpersonal) skills at 1.4%. Routine cognitive skills decrease hourly labor income by 1.1%. The specialization index is significant and negative, reducing hourly labor income by 15.9%.

Cartagena: For this region, non-routine interactive (interpersonal) and non-routine manual skills have the highest increase in hourly labor income at 6.5% and 4.1%, respectively. Routine cognitive and routine manual skills also increase hourly labor income by 1% and 2.8%. The specialization index is significant and positive, increasing hourly labor income by 6%.

Cúcuta: Non-routine analytical skills and interactive (interpersonal) skills have the highest increase in wage premiums in this region, around 2% and %, respectively. Non-routine manual skills, on the other hand, decrease hourly labor income by 1.3%. The specialization index is significant and negative, reducing hourly labor income by 16.7%.

Pereira: Non-routine analytical skills have the highest increase in wage premiums in this region, around 4.6%. The specialization index is significant and negative, reducing hourly labor income by 11.9%.

Ibagué: Non-routine interactive (interpersonal) skills have the highest increase in wage premiums in this region, around 4.3%, followed by routine manual skills at 0.5%. The specialization index is significant and positive, meaning it increases hourly labor income by 17.9%.

Manizales: Non-routine interactive (interpersonal) skills have the highest increase in wage premiums in this region, around 4.6%, followed by routine manual skills at 3.6%, non-routine analytical skills at 1.9%, and non-routine manual skills at 1.4%. The specialization index is significant and positive, increasing hourly labor income by 9.5%.

Villavicencio: Routine manual skills have the highest increase in wage premiums in this region, around 4.2%, followed by non-routine interactive (interpersonal) skills at 4.1%. The specialization index is not significant.

Pasto: Routine manual skills have the highest increase in wage premiums in this region, around 5.6%, followed by non-routine analytical skills and non-routine interactive (interpersonal) skills at 2.9% and 5.3%, respectively. The specialization index is significant and negative, reducing hourly labor income by approximately 24.3%.

Montería: Non-routine interactive (interpersonal) skills have the highest increase in wage premiums in this region, around 5.2%. Similarly, non-routine analytical and routine manual skills also increase hourly labor income by 3.2% and 5.1%, respectively. Routine cognitive skills decrease hourly labor income by approximately 1.6% in this region. The specialization index is significant and positive, meaning it increases hourly labor income by 7.5%.

Finally, in terms of agglomeration, an exercise was conducted relating wage premiums to population density (see Annex 6). This exercise shows that more agglomerated markets tend to have higher turnover, and non-routine analytical skills have a much higher wage premium, consistent with findings in the literature. On the other hand, interactive (interpersonal) skills have a higher premium in less agglomerated cities or smaller labor markets.

4. Decomposition of regional wage disparities

The results obtained from the decomposition exercise proposed by Oaxaca & Blinder (1973) and the additional decompositions using RIF for the P25, P75, and GINI, as presented in equations 2 and 3, can be observed in Figure 7 and Annexes 7 and 8. Figure 7 illustrates the decomposition exercise for the mean, where the specific effect of skills on the change in average level and inequality of labor income is isolated. It can be observed that the structural differences favor Bogotá and explain 25.8% of the income difference between Bogotá and the rest of the cities. However, this percentage varies significantly among city groups, identifying that the largest differences are for cities like Barranquilla (34.6%), Cúcuta (53.2%), Montería (37.8%), and Pasto (27%). It is worth noting that this differential is primarily due to the differences found in the unexplained part (Effect of structure-ES).

On the other hand, when decomposing these differences into the explained part (Effect of composition-EC) and unexplained part (Effect of structure-ES) for each of the components of worker characteristics, skills, and sectoral geographic specialization, it is found that the greater weight lies in the unexplained part (ES). The explained part (EC), in most cases, favors the city of Bogotá, where worker characteristics (age, gender, years of schooling, formality, etc.) are the ones that explain the most. In contrast, the specialization index and skills have a small contribution. In the case of skills, non-routine analytical skills are the ones that generate the differential and are generally more valued in Bogotá compared to the rest of the cities (see Annex 7 and Figure 7).

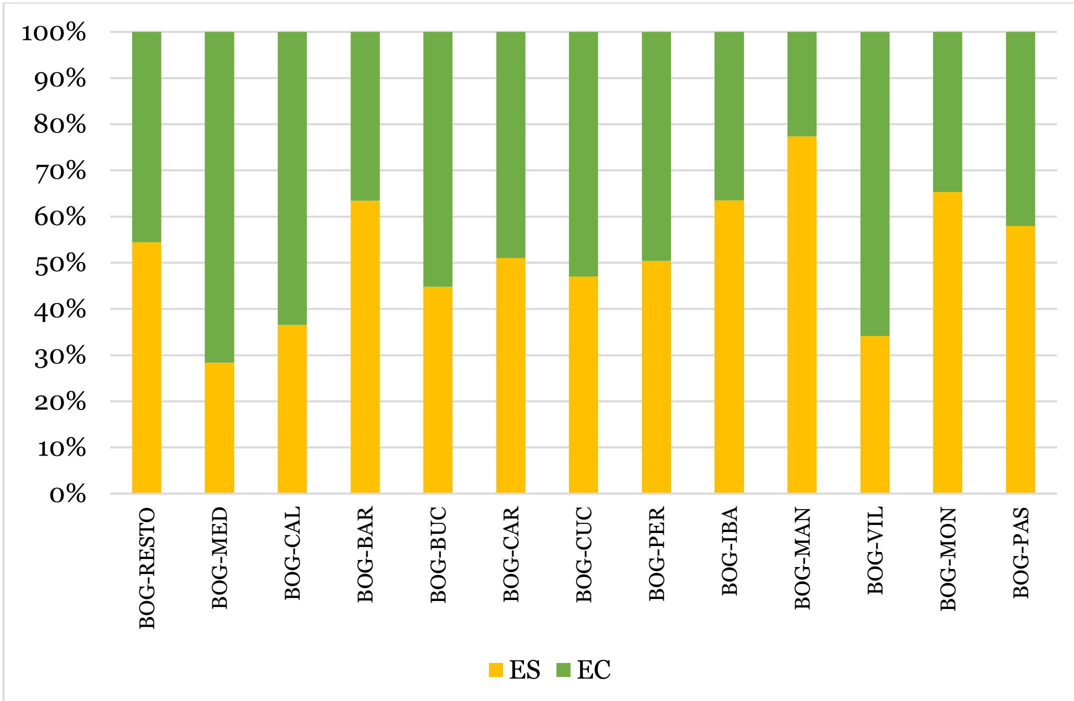
When analyzing the unexplained part (ES), a significant portion of the differential in favor of Bogotá was found. The majority of this can be attributed to the index of sectorial geographic specialization and the rest of the workers' labor characteristics. However, it is worth noting an interesting result regarding the skills component, where cities such as MED, CAL, BAR, BUC, CUC, PER, and VIL place higher value on skills compared to Bogotá. This effect may be due to factors of scarcity of routine manual skills in those regions.

In summary, for the EC, the differential in required skills has a greater impact on labor income than the level of sectorial geographic specialization. Conversely, in the ES, it is primarily determined by the level of sectorial geographic specialization. The EC and ES demonstrate how the gap between the rest of the cities and BOG is positive, meaning that in BOG, the premium on labor income is higher due to labor supply characteristics.

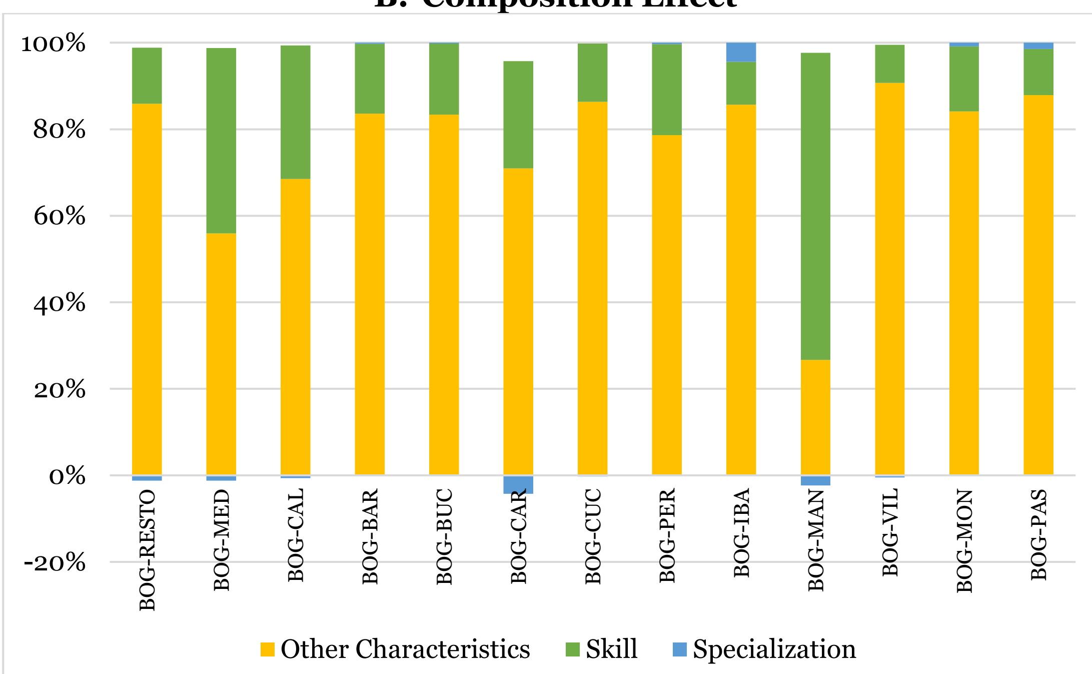
Analyzing the results of the quantile regression RIF (see Annex 8), we can see, for example, that in the case of the GINI coefficient, the EC factor widens the income gap between BOG and MED. In the case of the P75, a worker in BOG has higher labor income compared to a worker in MED, but in the latter city, the premium paid by the labor market narrows the gap to some extent. When decomposing the EC and ES, it is observed that for the EC in the GINI, P25, and P75, sectorial geographic specialization does not play an important role or is not significant in the gap, whereas skills composition of the population widens the gap. Finally, for the ES, the premium of specialization is indeed a relevant factor.

Figure 2. Oaxaca-Blinder Methodology Decomposition (1973)

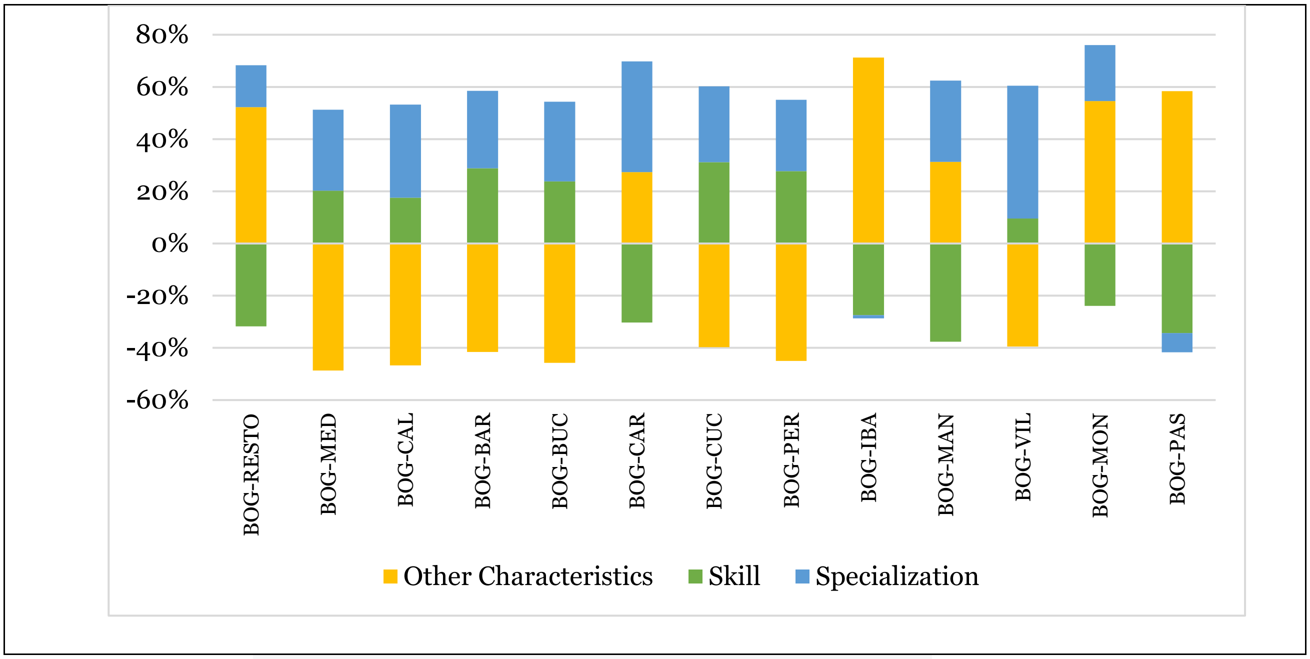
A. Aggregate decomposition



B. Composition Effect



C. Structure Effect



Note: EC = Composition Effect and ES = Structure Effect

Source: authors' work using GEIH 2019.

5. Conclusions

This article aimed to advance academically in studying regional salary disparities in Colombia, incorporating occupational characteristics of the labor force, urban agglomeration, and specialization patterns. It was found that there is a different salary premium between the urban areas studied for the different skills of the same occupation. A significant effect was found in the sectorial geographic specialization, generating differentials in the salaries of the different regional labor markets.

In previous studies found for Colombia, analyses of regional gaps were evidenced that do not consider factors such as skills and specialization. Therefore, this work becomes a pioneering work on this subject, advancing toward future research, where an in-depth analysis of polarization and skill imbalances (mismatch) are incorporated. In addition, based on this study, it is possible to propose public policy instruments that effectively reduce wage disparities.

However, before designing effective instruments, knowing the determinants of these differences is essential. For this, the work used an Oaxaca-Blinder type decomposition methodology and subsequent analysis of the differences in various points of the labor income distribution based on a RIF influence function. Here, the decomposition made it possible to analyze salary disparities in the part attributed to differences in the characteristics of the labor force, tasks-skills, and specialization and the part attributed to differences in the returns to these factors.

When conducting the decomposition exercise proposed for Colombia, it was possible to determine that the differences observed in labor income of the other cities with Bogotá are mainly explained by how the characteristics of the workers are remunerated in the first place, according to the income structure in each city. Still, getting a small share of skills and sectoral geographic specialization is also possible. In line with the above, the results show that if you want to formulate more efficient public policy instruments to reduce salary disparities between cities, it is not enough to increase workers' education years. Still, also, skills should be evaluated, and local sectoral specialization should be promoted. It is established that it is possible to characterize the differences between the labor income of the cities using the decomposition methods of the Oaxaca-Blinder type and RIF quantile influence functions.

There are different interrelated factors on which public policy can act to mitigate the negative effects of this process and generate and take advantage of its opportunities. For example, a skills imbalance in the workforce is partly because of individuals making educational and employment decisions with incomplete information about the economic returns of their choice. In addition, salaries do not always reflect the market value of skills, and even when they do, adjusting the supply of skills is time-consuming and costly. Therefore, the relevance of higher education, the strengthening of job training programs, and the flexibility of labor relations and their relationship with the sustainability of the social security system make up institutional pillars for this purpose.

Finally, in developing countries, the challenge is greater considering that informal employment relationships persist, characterized by low productivity levels and because they are the employment alternative for population groups that are more vulnerable to unemployment. In this way, skills gaps are a factor that can significantly increase informality and the quality of employment of workers. The profound discussions about labor relations on collaborative economy platforms are proof of this.

References

Acemoglu, D. (2010). Skill, Tasks and Technologies: Implications for Employment and Earnings. https://doi.org/10.1370/afm.75.

Acemoglu, D., and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. Handbook of Labor Economics (Vol. 4). Elsevier Inc. https://doi.org/10.1016/S0169-7218 (11)02410-5

Alvarez, A., & Hosftter, M. (2012). 50 Years of Job Vacancies in Colombia: The Case of Bogotá, 1960-2010. Borradores de Economía.

Andersson, M., Klaesson, J., & P. Larsson (2013). Sources of persistence in regional start-up rates—Evidence from Sweden. Journal of Economic Geography, 13(3), 529-551.

Arango, L. E. (2013). Puestos de trabajo vacantes según anuncios de la prensa de las siete principales ciudades de Colombia. Borradores de Economía.

Arntz, M., Gregory, T., & Zierahn, U. (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. OECD Social, Employment and Migration Working Papers, No. 189, OECD Publishing.

Autor (2013). The "task approach" to labor markets: An overview. Journal of Labor Economics, 31(S1), S589-S610.

Autor, D. and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market, 103(5), 1553–1597.

Autor, D. and Handel, M. (2013). Putting Tasks to the Test: Human Capital, Job Tasks, and Wages. Journal of Labor Economics, 31(S1), S59–S96. https://doi.org/10.1126/science.249.4967.345-a

Autor, D., & Dorn, D. (2009). This job is 'getting old': Measuring changes in job opportunities using occupational age structure. American Economic Review, 99(2), 45-51.

Autor, D., Lawrence, K. and Melissa, K. (2006). "The Polarization of the U.S. Labor Market." American Economic Review 96 (2): 189–94.

Autor, D., Lawrence, K. and Melissa, K. (2008). "Trends in U.S. Wage Inequality: Revising the Revisionists." Review of Economics and Statistics 90 (2): 300–323.

Autor, D., Levy, F. and Murnane, R. (2003). "The Skill Content of Recent Technological.

Bachmann, R. (2018). Job mobility in Europe, Japan and the U.S. A comparative study of wage dynamics. De Gruyter.

Bacolod, M., Blum, B. S., & Strange, W. C. (2009). Skills in the city. Journal of Urban Economics, 65(2), 136-153.

Bárcena-Martín, E., y Silber, J. (2018). On the Decomposition of the Esteban and Ray Index by Income Sources. Econometrics, 6(2), 17.

Baum-Snow, N., & Pavan, R. (2012). Understanding urban density: A model of the commute mode choice and the density of employment and housing. Journal of Urban Economics, 72(2-3), 276-291.

Bertola, G., & Ichino, A. (1995). Wage Inequality and Unemployment: United States vs. Europe. NBER Macroeconomics Annual, 10, 13-54.

Bhalotra, S and Fernandez, M. (2018). The Distribution of the Gender Wage. IZA (11640).

Bizopoulou, A. (2017). Task Profiles and Gender Wage-Gaps within Occupations. Working Paper, (June), 1–22.

Blinder, A. S. 1973. Wage Discrimination: Reduced Form and Structural Estimates. The Journal of Human Resources 8: 436–455.

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

Clark, G., & Ballard, K. (1981). The Demand and Supply of Labor and Interstate Relative Wages: An Empirical Analysis. Economic Geography, 57(2), 95-112.

Clarke, S. (2002). Market and Institutional Determinants of Wage Differentiation in Russia. Industrial and Labor Relations Review, 55(4), 628-648.

Combes, P. P. (2000). Overlapping jurisdictions and demand externalities: Towards a fiscal federalism with independent cities. European Economic Review, 44(1), 91-112.

Combes, P. P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2008). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. Econometrica, 76(5), 1447-1476.

Cortes, G. (2016). Where Have the Middle-Wage Workers Gone? A Study of Polarization Using Panel Data. Journal of Labor Economics, 34(1), 63–105. https://doi.org/10.1086/682289

Cortes, G. M. (2018). Where have the middle-wage workers gone? A study of polarization using panel data. Journal of Labor Economics, 36(S1), S89-S123.

Danieli, G. (2017). Crime spillovers and sentencing on the border: Evidence from state court data. Journal of Labor Economics, 35(4), 1045-1076.

Deming, D y Kah, L. (2017). Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. National Bureau of Economic Research.

Djumalieva, J y Sleeman, C. (2018). An Open and Data-driven Taxonomy of Skills Extracted from Online Job Adverts. Economic Statistics Centre of Excellence –ESCoE-.

Djumalieva, J; Lima, A and Sleeman, C. (2018). Classifying Occupations According to Their Skill Requirements in Job Advertisements. Economic Statistics Centre of Excellence –ESCoE-.

Duranton, G. (2016b). Determinants of city growth in Colombia. Papers in Regional Science, 95(1), 101–131. https://doi.org/10.1111/pirs.12225

Duranton, G. and Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. https://doi.org/10.1016/S0169-7218 (04)07048-0

Firpo, S., Fortin, N. and Lemieux, T. (2007). Unconditional quantile regression. NBER Technical Working Paper Series, (339).

Firpo, S., Fortin, N. M., & Lemieux, T. (2011). Occupational tasks and changes in the wage structure. Journal of Labor Economics, 29(3), 549-591.

Firpo, S., Fortin, N. M., & Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. The Review of Economics and Statistics, 100(4), 553-567.

Florida, R., Mellander, C., & Stolarick, K. (2012). Human capital in cities and suburbs. Journal of Economic Geography, 12(5), 843-872.

Fortin, N., Firpo, S., and Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. Econometrica, 64(5):1001–44.

Fredriksson, P. (1999). The Dynamics of Regional Labor Markets and Active Labor Market Policy: Swedish Evidence. Oxford Economic Papers, 51(4), 623-648. Retrieved from http://www.jstor.org/stable/3488722

Fujita, M., Krugman, P., & Venables, A. J. (1999). The spatial economy: Cities, regions, and international trade. MIT Press.

Galvis, L. Á. (2002). Pobreza, desigualdad y crecimiento en Colombia. Documentos CEDE, (2002-02), Universidad de los Andes.

Gil, Y. O. (2007). Canales de Búsqueda de empleo y duración de desempleo en el mercado laboral colombiano. Universidad del Valle.

Glaeser, E. L. (1999). Learning in cities. Journal of Urban Economics, 46(2), 254-277.

Glaeser, E. L. (2008). Cities, Agglomeration, and Spatial Equilibrium. Oxford University Press.

Glaeser, E. L., & Henderson, V. (2017). Urban economics and urban policy: Challenging conventional policy wisdom. Journal of Economic Perspectives, 31(3).

Glaeser, E. L., & Mare, D. C. (2001). Cities and skills. Journal of Labor Economics, 19(2), 316-342.

Goos, M., & Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in Britain. The Review of Economics and Statistics, 89(1), 118-133.

Goos, M., Manning, A., & Salomons, A. (2011). Explaining job polarization: Routine-biased technological change and offshoring. American Economic Review, 104(8), 2509-2526.

Gould, E. D. (2007). Returns to skills and the speed of reforms: Evidence from Central and Eastern Europe, China, and Russia. Journal of Comparative Economics, 35(2), 351-370.

Graham, D. J., & Melo, P. C. (2009). Agglomeration economies and labour productivity: Evidence from longitudinal worker data for GB's travel-to-work areas. Journal of Urban Economics, 65(1), 48-63.

Greig, M., Glancey, K., & Wilson, P. (2008). A review of Labour Shortages, Skill Shortages and skill Gaps. The home Office Migration Advisory Committee.

Grinis, I. (2017). The STEM Requirements of 'Non-STEM' Jobs: Evidence from UK Online Vacancy Postings and Implications for Skills & Knowledge Shortages. SSRN Scholarly Paper ID 2864225, Social Science Research Network

Handel, B. R., & Autor, D. H. (2013). The case against cutting class. Journal of Economic Perspectives, 27(3), 207-230.

Holt, R., Sawichi, S., & Sloan, J. (2010). A theoretical Review of Skill Shortages and Skill Needs. UK Commission for employment and skill.

Ikudo, A, Lane, J; Staudt, J and Weinberg, B. (2018). Occupational Classifications: A Machine Learning Approach. IZA DP No. 11738.

Juhn, C., Murphy, K. M., and Pierce, B. (1993). Wage Inequality and the Rise in Returns to Skill. Journal of Political Economy, 101(3):410–42.

Krugman, P. (1991). Increasing returns and economic geography. Journal of Political Economy, 99(3), 483-499.

Krumm, R. J. (1983). Regional Wage Differentials, Fluctuations in Labor Demand, and Migration. International Regional Science Review, 8(1), 23–45.

Lima, A and Bakhshi, H. (2018). Classifying occupations using web-based job advertisements: an application to STEM and creative occupations. Economic Statistics Centre of Excellence –ESCoE-

Londoño, E. (2013). Retornos a la Educación y Promedio Bayesiano de Modelos a partir de un pseudo panel: aplicación en Colombia. Documento de trabajo. Mimeo, Universidad del Rosario.

Manpower. (2008). La integración al Mercado Laboral del Talento Latinoamericano. Aprender-Emprender.

Marinescu, I and Wolthoff, R. (2016). Opening the black box of the matching function: The Power of words. National Bureau of Economic Research.

Matano, A., & Naticchioni, P. (2011). The impact of immigration on the wage distribution in Italy. Labour Economics, 18(6), 737-747.

Matano, A., & Naticchioni, P. (2012). Regional unemployment and productivity in Italian regions. Papers in Regional Science, 91(4), 897-913.

Mavromaras, K., Healy, J., Richardson, S., Slonane, P., Wei, Z., & Zhu, R. (2013). Final Report to the Australia Workforce and Productivity Agency (AWPA) by the National Institute of Labour Studies (NILS). National Institute of Labour Studies Flinders University, Australia.

Melo, P. C., Graham, D. J., & Brage-Ardao, R. (2009). The role of geography in unemployment dispersion: Evidence from spatial panel models. Regional Science and Urban Economics, 39(3), 255-267.

Mesa, J. A., González, A. R., & González, J. S. (2008). The relationship between road infrastructure and regional development in Spain. Journal of Infrastructure Systems, 14(2), 156-165.

Mihaylov, E. (2015). Returns to routine and non-routine job tasks: Evidence from Germany, (i), 1–20.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. Journal of Political Economy, 66(4):281–302.

Mion, G., & Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. Journal of Urban Economics, 65(2), 194-209.

Neves, P. C., Afonso, Ó., & Silva, S. T. (2017). Regional human capital and the shift in location of Portuguese manufacturing industries. The Annals of Regional Science, 59(1), 131-161.

Newman, R. (1982). Dynamic Patterns in Regional Wage Differentials. Southern Economic Journal, 49(1), 246-254.

Nupia, O. G. (1997). Cambio estructural y especialización industrial en Colombia. Lecturas de Economía, (47), 9-56.
Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. International Economic Review, 14(3):693–709.
Oaxaca, R., & Blinder, A. (1973). Interpreting the decomposition of the gender earnings gap. Journal of Human Resources, 8(3), 436-455.
Pan, W., Wu, Y., & Bai, X. (2016). Urban land expansion and structural change in the Yangtze River Delta, China. Journal of Urban Economics, 92, 61-75.
Puga, D. (2010). The magnitude and causes of agglomeration economies. Journal of Regional Science, 50(1), 203-219.
Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: Evidence from the cities. Journal of Urban Economics, 34(3), 380-400.
Rinz, K. (2018). The role of residential mobility in reproducing educational inequality. American Journal of Sociology, 123(5), 1422-1469.
Rothwell, J. (2014). Still searching: Job Vacancies and STEM skill. Metropolitan Policy Program.
Spitz-Oener, A. (2006). Technical Change, Job Tasks and Rising Educational Demands: Looking outside the Wage Structure, 24(2).
Van Sickle, J. (1954). Regional Economic Adjustments: The Role of Geographical Wage Differentials. The American Economic Review, 44(2), 381-392. Retrieved from http://www.jstor.org/stable/1818350
Wheaton, W. C., & Lewis, M. (2002). Urban decline and the future of American cities. Brookings-Wharton Papers on Urban Affairs, 3, 1-64.
Wheeler, C. H. (2001). Urban growth externalities and neighborhood incentives: Another cause of urban sprawl? Journal of Urban Economics, 50(3), 385-407.
Wheeler, C. H. (2004). The environmental implications of urban sprawl: A literature review and assessment. Journal of Planning Literature, 18(2), 229-242.
Wheeler, C. H. (2006). Urban development and land-use change: What drives the process in Australian cities? Urban Studies, 43(11), 1999-2014.
Wheeler, C. H. (2007). Recycling land for urban development: Using previously developed land to accommodate growth. Growth and Change, 38(2), 179-203.
Wheeler, J. O. (2001). Urban development and the natural environment: Pattern and process. Routledge.
Yankow, J. J. (2006). Moving to the sticks: Returns to the rural life for the educated. Journal of Urban Economics, 59(2), 193-213.

Annex

Annex 1. Task Categories

Types of Tasks	Measured by…	Examples skills/activities
Non-routine analytical tasks	General educational development, mathematics.	**Lowest level:** adds and subtracts 2-digit numbers; performs operations with units such as cup, pint, and quart. **Medium level:** calculates discount, interest, profit, and loss; Inspects flat glass and compiles sample-based defect data to determine variances from acceptable quality limits. **Highest Level:** Performs and supervises aerodynamic and thermodynamic systems analyses to determine design adequacy for aircraft and missiles, etc.
Non-routine interactive (interpersonal) tasks	Adaptability to accept responsibility for directing, controlling, or planning an activity.	Plans and designs private residences, office buildings, factories, and other structures; Applies accounting principles to install and maintain the operation of the general accounting system; Carries out legal actions in court proceedings; Collects and analyzes evidence, reviews relevant decisions; etc.
Routine cognitive tasks	Adaptability to situations that require the precise achievement of established limits, tolerances, or standards	Operates a billing machine to transcribe data from office records; Calculates degrees, minutes, and seconds of latitude and longitude using standard navigation aids; Measures bottle dimensions using gauges and micrometers to verify bottle manufacturing setup meets manufacturing specifications; Prepares and verifies voter lists from official registration records; etc.
Routine manual tasks	Ability to move the fingers and manipulate small objects with the fingers, quickly or precisely. Skills that influence strength, endurance, flexibility, balance, and coordination.	Mixes and bakes ingredients according to recipes; Hangs fasteners and decorative trim for items; Operates the tabulation machine that processes the data from the tabulation cards into printed records; Packs produce such as bulbs, fruits, nuts, eggs, and vegetables for storage or shipping; etc.
Non-routine manual tasks	Ability to move the hand and foot in coordination with each	**Lowest level:** Has machine that crimps eyelets; **Medium level:** Takes care of the beef cattle on the ranch; Drives the bus to transport passengers; Pick plums from

	other according to visual stimuli.	ornamental and shade trees. **Highest Level:** Performs gymnastic feats of skill and balance.
Source: Autor et al. (2003)		

Annex 2. Skill intensity index for occupations classified under ISCO 08 A.C at the Principal Subgroup level

Code	Principal Subgroup	Non-Routine Analytical Index	Non-Routine Interactive Index	Routine Cognitive Index	Routine Manual Index	Non-Routine Manual Index
01	Military Forces Officers	2.13	1.11	1.07	2.53	3.28
02	Military Forces Sub-Officers	2.13	1.11	1.07	2.53	3.28
03	Other members of the military	2.13	1.11	1.07	2.53	3.28
11	Executive directors, senior administrative staff, and legislators	5.16	6.97	0.69	2.83	0.44
12	Administrative and commercial directors	5.42	8.11	0.70	2.68	0.18
13	Directors and managers in the production and service sectors	5.22	8.01	1.03	2.85	0.60
14	Managers of hotels, restaurants, shops, and other services	5.33	8.06	0.52	2.80	0.35
21	Science and engineering professionals	7.80	5.15	6.62	4.55	1.11
22	Healthcare professionals	6.21	3.33	4.78	6.27	1.31
23	Education professionals	5.51	6.94	0.89	3.07	0.49
24	Business and management professionals	5.66	6.09	2.20	2.80	0.26
25	Information and communications technology professionals	6.65	5.78	5.26	2.62	0.21
26	Professionals in law, social, and cultural sciences	5.36	4.60	1.04	3.02	0.57
31	Technicians and professionals at the middle level of science and engineering	4.77	2.02	7.21	4.28	1.59

Code	Principal Subgroup	Non-Routine Analytical Index	Non-Routine Interactive Index	Routine Cognitive Index	Routine Manual Index	Non-Routine Manual Index
32	Mid-level health technicians and professionals	4.14	0.93	5.73	4.80	2.01
33	Mid-level technicians and professionals in finance and administration	4.55	3.58	3.12	3.42	0.53
34	Mid-level technicians and professionals in legal, social, cultural, and related services	3.28	3.42	1.66	3.95	2.20
35	Technicians in information and communications technology	5.45	4.34	5.36	3.51	0.32
41	Office workers	3.28	0.91	5.87	5.11	0.08
42	Employees dealing directly with the public	3.03	1.15	3.61	4.06	0.26
43	Accounting assistants and those in charge of registering materials	3.51	1.71	5.95	3.81	0.11
44	Other administrative support staff	2.22	0.46	4.67	3.81	0.37
51	Personal service workers	2.61	1.50	1.45	3.04	1.81
52	Sellers	2.95	1.39	1.14	3.26	1.02
53	Personal care workers	3.35	2.75	3.27	4.37	1.44
54	Protection services staff	1.93	1.15	0.87	2.52	2.93
61	Farmers and skilled farm workers	3.06	4.33	1.54	2.95	2.23
62	Skilled forest workers, fishers, and hunters in market-oriented activities	2.04	2.64	1.60	2.84	2.58
71	Construction officials and workers (excluding electricians)	3.35	1.06	8.00	4.46	2.37
72	Officials and operators of the metallurgy; mechanics and repairers of machines and related	3.49	0.57	8.25	4.33	1.61
73	Artisans and operators of the graphic arts and related	2.74	0.70	7.16	5.75	1.09
74	Officials and operators of electricity and electronics	4.83	1.09	8.64	4.89	2.19

Code	Principal Subgroup	Non-Routine Analytical Index	Non-Routine Interactive Index	Routine Cognitive Index	Routine Manual Index	Non-Routine Manual Index
75	Officials and operators of food processing, clothing, cabinetmakers, and related	2.45	0.83	6.51	4.24	1.70
81	Operators of fixed installations and machines	1.49	0.20	6.69	3.79	1.27
82	Assemblers	1.21	0.01	6.97	4.43	0.57
83	Vehicle drivers and mobile heavy equipment operators	1.81	1.52	2.61	2.82	3.78
91	Domestic and cleaning staff	0.79	0.55	0.83	2.87	1.44
92	Agricultural, fishing, and forestry workers and laborers	2.24	2.82	2.42	3.09	2.56
93	Workers and laborers in mining, construction, manufacturing, and transportation	1.35	0.21	4.47	3.36	1.77
94	Food preparation helpers	1.69	0.46	2.78	3.19	1.05
95	Street vendors of services and related	2.53	0.25	0.07	2.85	2.63
96	Garbage collectors and other elementary occupations	1.36	0.22	3.01	3.06	2.25
Source: Authors' work						

Annex 3. Labor market indicators for the 13 urban areas

Urban Area	Total Population	PET	PEA	Employed	Unemployed	Formal	Informal	TGP	TO	TD	TI	Hourly Wage
BOG	8264020	6810542	4698377	4186301	512076	2485410	1700892	68.99	61.47	10.90	40.63	7235
MED	3773109	3193306	2067926	1815349	252577	1060022	755327	64.76	56.85	12.21	41.61	6555
MAN	424653	359525	210891	185671	25219	111902	73770	58.66	51.64	11.96	39.73	6111
CAL	2547162	2105008	1434013	1255227	178786	682108	573119	68.12	59.63	12.47	45.66	6273
VIL	503406	401300	264466	230094	34372	101877	128217	65.90	57.34	13.00	55.72	5788
BUC	1095988	915856	614192	551611	62581	246628	304983	67.06	60.23	10.19	55.29	5431
PER	636463	528784	328148	299289	28859	157358	141931	62.06	56.60	8.79	47.42	5274
CAR	1005959	816273	457133	425816	31317	200531	225285	56.00	52.17	6.85	52.91	5222
IBA	543555	443218	279232	233666	45566	111412	122254	63.00	52.72	16.32	52.32	5525
PAS	386455	323386	210644	188328	22316	82892	105436	65.14	58.24	10.59	55.99	5696
MON	362747	290592	180760	157527	23233	65413	92114	62.20	54.21	12.85	58.47	4867
BAR	1914925	1545359	997803	919522	78281	403911	515611	64.57	59.50	7.85	56.07	4995
CUC	847608	678735	406773	342524	64249	98766	243758	59.93	50.47	15.79	71.17	3964

Source: Authors' work using GEIH 2019

Annex 4.The Oaxaca-Blinder decomposition method (1973).

Consider two groups of cities, A and B (e.g., Bogotá, Medellín...), an outcome variable Y (log hourly wages), and a set of predictors (e.g., education, work experience, occupations, and/or skills). The question is how much the difference in the average outcome of hourly wages between the groups of cities can be attributed to the predictors. This can be represented as follows:

$$\mathrm{R} = \mathrm{E}(Y_A) - E(Y_B) \tag{4}$$

Where E(Y) denotes the expected value of the outcome variable, which is explained by the differences in the predictor variables. The above result is based on a linear model with the following characteristics:

$$Y_\iota = X_\iota'\beta_\iota + \varepsilon_\iota, \qquad E(\varepsilon_\iota) = 0, \qquad \iota \in \{\mathrm{A}, \mathrm{B}\} \tag{5}$$

Where X is a vector containing the predictors and the constant, β contains the slope and intercept parameters, and ε is the error, the difference in the mean can be expressed as the difference in the linear prediction of the means for the specific group of regressors. That is:

$$\mathrm{R} = \mathrm{E}(Y_A) - E(Y_B) = \mathrm{E}(X_A)'\beta_A - E(X_B)'\beta_B \tag{6}$$

Where:

$$E(Y_\iota) = E(X_\iota'\beta_\iota + \varepsilon_\iota) = E(X_\iota'\beta_\iota) + E(\varepsilon_\iota) = E(X_\iota')\beta_\iota \qquad (7)$$

$$\text{With } E(\beta_\iota) = \beta_\iota \text{ y } E(\varepsilon_\iota) = 0$$

To identify the contribution of group differences in predictors to the overall difference in the outcome, equation (4) can be rewritten as a decomposition divided into three parts:

$$\mathrm{R} = \underbrace{[\mathrm{E}(X_A) - E(X_B)]'\beta_B}_{\mathbf{Z}} + \underbrace{\mathrm{E}(X_B)'(\beta_A - \beta_B)}_{\mathbf{F}} + \underbrace{[\mathrm{E}(X_A) - E(X_B)]'(\beta_A - \beta_B)}_{\mathbf{C}} \qquad (8)$$

The first term (Z) represents the part of the differential that is due to group differences in predictors ("endowment effect"). The second component (F) measures the contribution of differences in coefficients (including differences in the intercept). The third term (C) is an interaction term that takes into account the fact that differences in endowments and coefficients exist simultaneously between the two groups.

Decomposition (6) is formulated from the perspective of Group B. In other words, group differences in predictors are weighted by the coefficients of Group B to determine the endowment effect (Z). In other words, the Z component measures the expected change in the average outcome of Group B if Group B had the predictor levels of Group A. Similarly, for the second component (F), differences in coefficients are weighted by the predictor levels of Group B. In other words, the second component measures the expected change in the average outcome of Group B if Group B had the coefficients of Group A. Similarly, the differential can be expressed analogously from the perspective of Group A, as presented in equation (7).

$$\mathrm{R} = \underbrace{[\mathrm{E}(X_A) - E(X_B)]'\beta_A}_{\mathbf{J}} + \underbrace{\mathrm{E}(X_A)'(\beta_A - \beta_B) + [\mathrm{E}(X_A) - E(X_B)]'(\beta_A - \beta_B)}_{\mathbf{K}} \qquad (9)$$

Now, the "endowment effect" corresponds to the expected change in the average outcome of Group A if Group A had the predictor levels of Group B. The "coefficient effect" quantifies the expected change in the average outcome of Group A if Group A had the coefficients of Group B.

An alternative decomposition that is prevalent in the literature on discrimination stems from the concept that there exists a vector of non-discriminatory coefficients that should be used to determine the contribution of differences in predictors. Let's denote such vectors as the non-discriminatory coefficient vectors. The difference in outcomes can be written as follows:

$$\mathrm{R} = [\mathrm{E}(X_A) - E(X_B)]'\beta^* + [\mathrm{E}(X_A)'(\beta_A - \beta^*) - E(X_B)'(\beta^* - \beta_B)] \quad (10)$$

Where the first component (J) is the part of the differential in outcomes that is "explained" by the group differences in predictors ("quantity effect"), and the second term is the "unexplained" part. The latter is often attributed to discrimination, but in this case, it captures all potential effects of differences in unobserved variables.

The "unexplained" part in (8) can be further decomposed. Let $\beta_A = \beta^* + \alpha_A$ and $\beta_B = \beta^* + \alpha_B$, with α_A and $+ \alpha_B$ as a set of potential unexplained effect parameter vectors. K can then be expressed as follows:

$$\mathrm{K} = \mathrm{E}(X_A)'\alpha_A - E(X_B)'\alpha_B \quad (11)$$

In other words, the unexplained component of the differential can be subdivided into a part U_A that measures the potential unexplained effects in favor of Group A, and a part U_B that quantifies the potential unexplained effects against Group B.

$$U_A = \mathrm{E}(X_A)'\alpha_A \quad (12)$$

$$U_B = \mathrm{E}(X_B)'\alpha_B \quad (13)$$

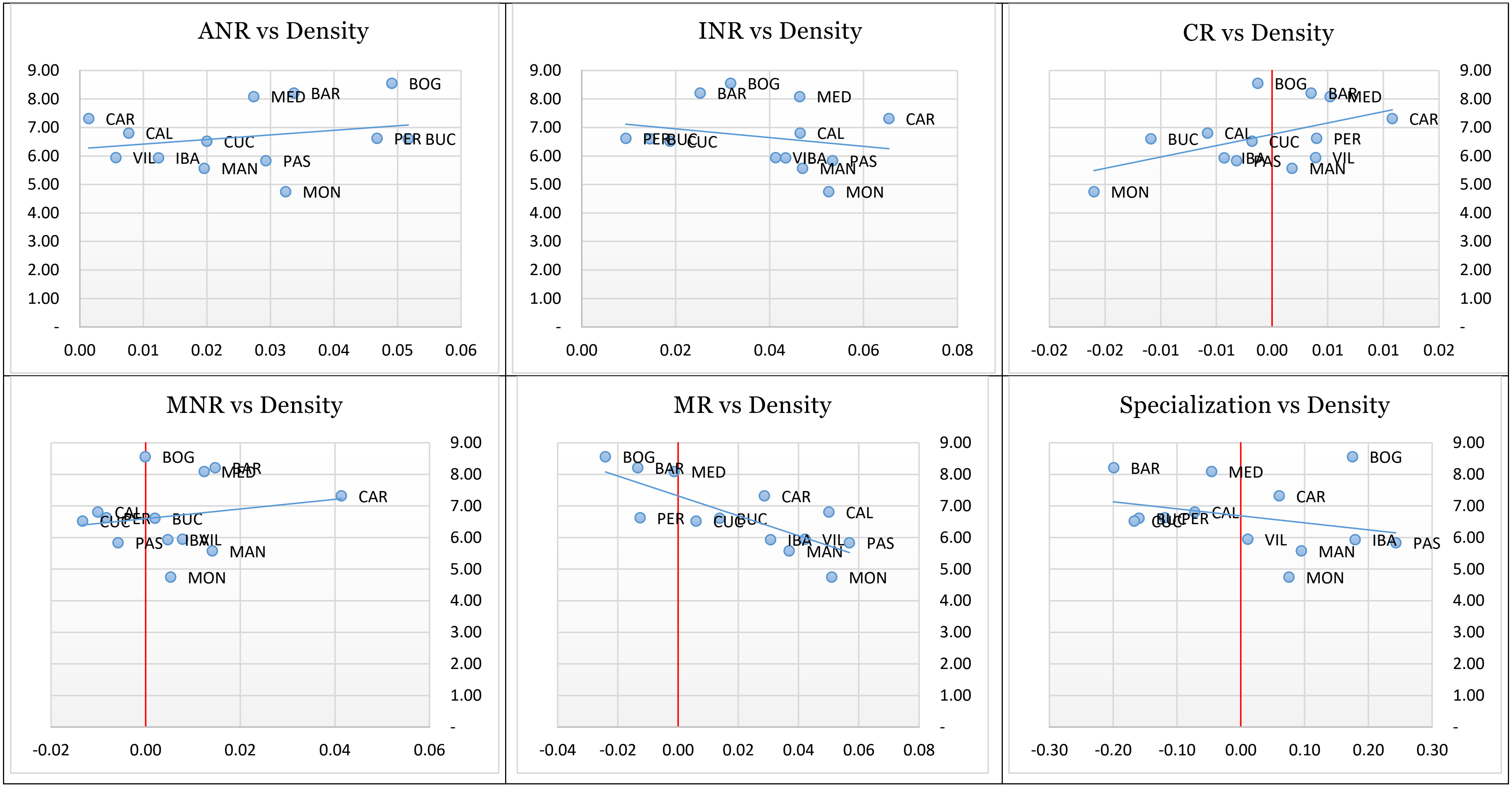
Annex 5. Premium wage

Variables	POOL	BOG	MED	CAL	BAR	BUC	CAR	CUC	PER	IBA	MAN	VIL	PAS	MON
	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour	Log Wage hour
Age	0.031***	0.0305***	0.0298***	0.0316***	0.0291***	0.0346***	0.0256***	0.0321***	0.0307***	0.0321***	0.0265***	0.0342***	0.0343***	0.0331***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
Age^2	-0.000***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0004***	-0.0002***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0004***	-0.0003***	-0.0003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Self employmente	-0.101***	-0.0913***	-0.0652***	-0.0504***	-0.0503***	-0.0946***	-0.1553***	-0.2051***	-0.1104***	-0.0983***	-0.0211	-0.0164	-0.1204***	-0.1306***
	(0.003)	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.013)	(0.017)	(0.014)	(0.017)	(0.014)	(0.015)	(0.014)	(0.016)
Women	-0.133***	-0.0835***	-0.0899***	-0.1003***	-0.1685***	-0.1345***	-0.1701***	-0.1421***	-0.1059***	-0.1210***	-0.0949***	-0.1127***	-0.1405***	-0.1836***
	(0.003)	(0.011)	(0.010)	(0.012)	(0.012)	(0.011)	(0.010)	(0.012)	(0.011)	(0.014)	(0.011)	(0.014)	(0.012)	(0.013)
Year of Education	0.050***	0.0593***	0.0551***	0.0573***	0.0465***	0.0377***	0.0380***	0.0312***	0.0450***	0.0473***	0.0483***	0.0522***	0.0536***	0.0465***
	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Formal	0.357***	0.2880***	0.3328***	0.3493***	0.3799***	0.3482***	0.3410***	0.3918***	0.2366***	0.3644***	0.2497***	0.3774***	0.3981***	0.4133***
	(0.003)	(0.013)	(0.012)	(0.015)	(0.014)	(0.013)	(0.014)	(0.018)	(0.012)	(0.015)	(0.013)	(0.014)	(0.014)	(0.016)
Non-Routine Analytical	0.025***	0.0491***	0.0274***	0.0077	0.0337***	0.0517***	0.0014	0.0200**	0.0468***	0.0124	0.0196**	0.0057	0.0293***	0.0324***
	(0.002)	(0.008)	(0.007)	(0.010)	(0.008)	(0.009)	(0.008)	(0.010)	(0.008)	(0.010)	(0.008)	(0.010)	(0.010)	(0.010)
Non-Routine Interactive	0.044***	0.0317***	0.0464***	0.0465***	0.0252***	0.0145**	0.0654***	0.0188**	0.0094	0.0434***	0.0470***	0.0413***	0.0534***	0.0526***
	(0.001)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.008)	(0.006)	(0.007)	(0.006)	(0.008)	(0.008)	(0.008)
Routine Cognitive	-0.004***	-0.0013	0.0052*	-0.0058	0.0035	-0.0109***	0.0108***	-0.0018	0.0040	-0.0043	0.0018	0.0039	-0.0032	-0.0160***
	(0.001)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Routine Manual	0.027***	-0.0241**	-0.0013	0.0501***	-0.0134	0.0138	0.0287**	0.0060	-0.0126	0.0307**	0.0369***	0.0422***	0.0569***	0.0510***
	(0.003)	(0.011)	(0.011)	(0.014)	(0.012)	(0.012)	(0.011)	(0.015)	(0.011)	(0.015)	(0.012)	(0.015)	(0.013)	(0.014)
Non-Routine Manual	0.004***	-0.0001	0.0124**	-0.0101	0.0147***	0.0020	0.0414***	-0.0133**	-0.0083	0.0047	0.0141**	0.0078	-0.0058	0.0053
	(0.001)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.007)	(0.006)	(0.007)	(0.005)	(0.007)	(0.007)	(0.006)
Geographic Specialization Index	0.115***	0.1750***	-0.0460***	-0.0722***	-0.1994***	-0.1595***	0.0601***	-0.1670***	-0.1195***	0.1793***	0.0950***	0.0110	0.2430***	0.0754**
	(0.005)	(0.026)	(0.015)	(0.022)	(0.037)	(0.028)	(0.018)	(0.028)	(0.026)	(0.056)	(0.025)	(0.033)	(0.042)	(0.037)
Constant	6.848***	6.7924***	6.9853***	6.8757***	7.1290***	7.1824***	7.0021***	7.2914***	7.2458***	6.7008***	6.8916***	6.7789***	6.2998***	6.5288***
	(0.014)	(0.060)	(0.054)	(0.067)	(0.071)	(0.064)	(0.058)	(0.075)	(0.062)	(0.092)	(0.059)	(0.078)	(0.067)	(0.062)
Area Fixed Effects	**SI**	**NO**	**NO**	**NO**	**NO**	**NO**	**NO**	**NO**	**NO**	**NO**	**NO**	**NO**	**NO**	**NO**
Observations	205,964	11,961	14,771	10,056	14,192	9,514	9,414	7,646	8,757	7,886	8,852	8,129	8,285	8,049
R-squared	0.420	0.418	0.375	0.341	0.298	0.357	0.454	0.378	0.339	0.380	0.392	0.354	0.514	0.472

Robust standard errors in parentheses

***** p<0.01, ** p<0.05, * p<0.1**

Annex 6. Wage premium vs. Agglomeration (Population Density)



Note: Population density is calculated as the total population of an area divided by its surface area in square kilometers (km2). ANR= Non-Routine Analytical, INR= Non-Routine Interactive, CR= Routine Cognitive, MR= Routine Manual, MNR= Non-Routine Manual, Geographic specialization index.
Source: Authors' work using GEIH 2019.

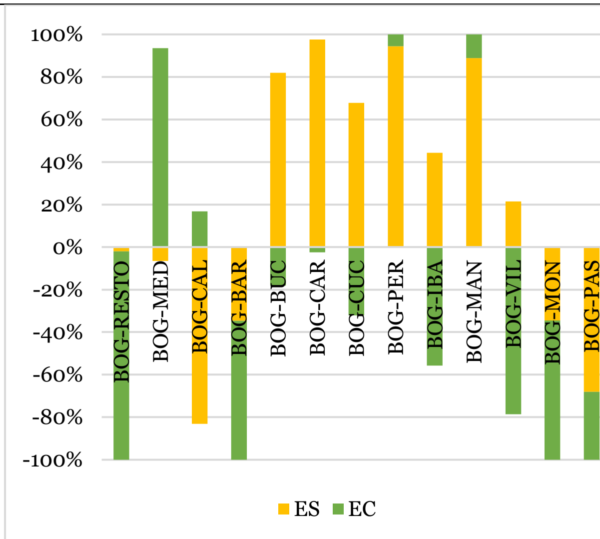
Annex 7. Oaxaca- Blinder (1973) descomposition

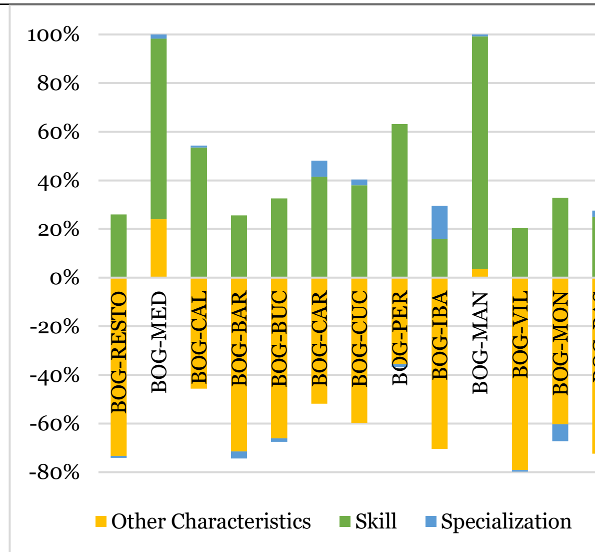
Área Urbana	BOG-RESTO	BOG-MED	BOG-CAL	BOG-BAR	BOG-BUC	BOG-CAR	BOG-CUC	BOG-PER	BOG-IBA	BOG-MAN	BOG-VIL	BOG-MON	BOG-PAS
Total Difference	0.258	0.066	0.129	0.346	0.217	0.222	0.532	0.221	0.238	0.085	0.205	0.378	0.270
Total Explained	0.118	0.047	0.082	0.127	0.120	0.109	0.282	0.110	0.087	0.019	0.135	0.131	0.114
Total Unexplained	0.141	0.019	0.047	0.220	0.097	0.113	0.250	0.111	0.151	0.066	0.070	0.247	0.157
Total Explained skills	0.016	0.021	0.026	0.020	0.020	0.030	0.038	0.023	0.009	0.014	0.012	0.020	0.012
Non-Routine Analytical	0.0046	0.0074	0.0097	0.0090	0.0126	0.0113	0.0154	0.0123	0.0038	0.0057	0.0038	0.0091	0.0046
Non-Routine Interactive	0.0113	0.0139	0.0154	0.0123	0.0073	0.0220	0.0228	0.0104	0.0054	0.0092	0.0081	0.0115	0.0072
Routine Cognitive	-0.0002	-0.0004	-0.0001	0.0000	-0.0001	0.0006	0.0000	0.0000	-0.0003	-0.0001	0.0008	-0.0012	-0.0003
Routine Manual	0.0004	0.0002	0.0001	0.0000	-0.0002	0.0000	-0.0003	0.0005	-0.0001	0.0001	-0.0001	0.0000	-0.0001
Non-Routine Manual	-0.0005	-0.0004	0.0005	-0.0009	-0.0001	-0.0043	0.0002	-0.0001	-0.0002	-0.0005	-0.0006	0.0003	0.0006
Total Explained Geographic Specialization	-0.0014	-0.0006	-0.0005	0.0003	0.0003	-0.0051	-0.0006	0.0004	0.0039	-0.0005	-0.0006	0.0011	0.0017
Total Explained Other Characteristics	0.104	0.027	0.057	0.106	0.100	0.084	0.245	0.086	0.075	0.005	0.124	0.110	0.100
Total Unexplained skills	-0.122	0.150	0.126	0.372	0.267	-0.087	0.380	0.308	-0.098	-0.100	0.032	-0.114	-0.323
Non-Routine Analytical	0.0878	0.0716	0.1327	0.0507	-0.0092	0.1509	0.0918	0.0070	0.1217	0.0972	0.1438	0.0548	0.0660
Non-Routine Interactive	-0.0276	-0.0319	-0.0310	0.0120	0.0364	-0.0697	0.0179	0.0445	-0.0256	-0.0329	-0.0214	-0.0443	-0.0475
Routine Cognitive	0.0047	-0.0221	0.0151	-0.0161	0.0323	-0.0398	0.0016	-0.0180	0.0100	-0.0104	-0.0171	0.0473	0.0065
Routine Manual	-0.1818	-0.0805	-0.2607	-0.0376	-0.1336	-0.1840	-0.1057	-0.0406	-0.1926	-0.2165	-0.2318	-0.2647	-0.2862
Non-Routine Manual	-0.0053	-0.0172	0.0139	-0.0212	-0.0029	-0.0656	0.0193	0.0120	-0.0066	-0.0197	-0.0109	-0.0085	0.0081
Total Unexplained Geographic Specialization	0.0617	0.2296	0.2562	0.3846	0.3437	0.1217	0.3551	0.3027	-0.0046	0.0826	0.1696	0.1020	-0.0695
Total Unexplained Other Characteristics	0.201	-0.360	-0.335	-0.537	-0.513	0.078	-0.485	-0.499	0.254	0.083	-0.132	0.258	0.549

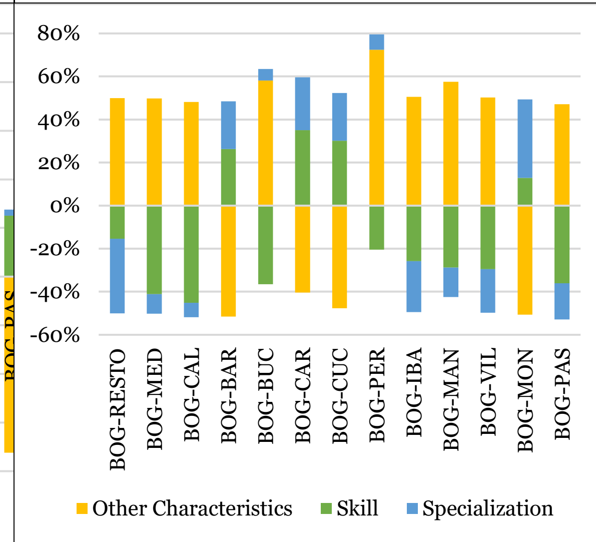
Source: Authors' work using GEIH 2019

Annex 8. Descomposition RIF

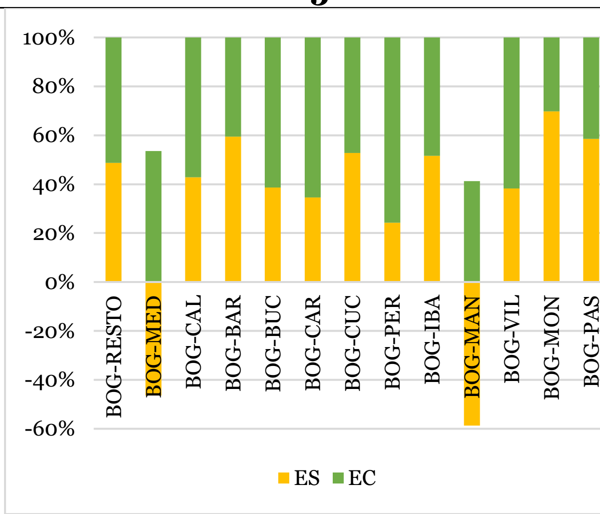
Aggregate descoposition	Composition effect	Structure effect
Gini	**Gini**	**Gini**

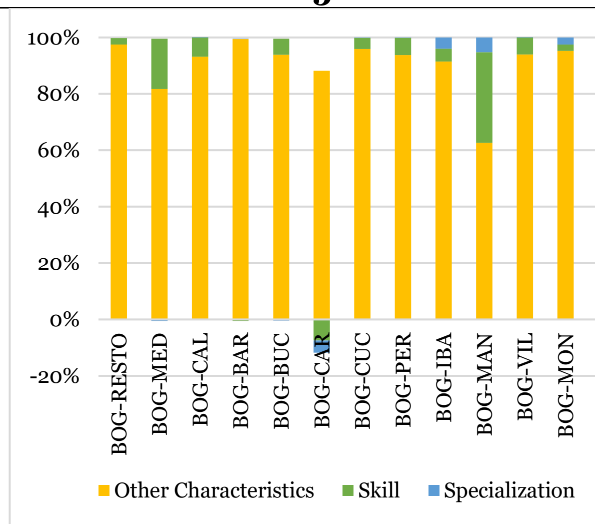


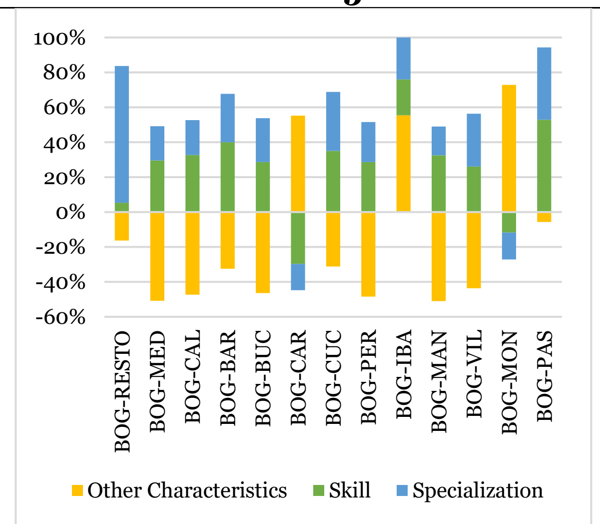




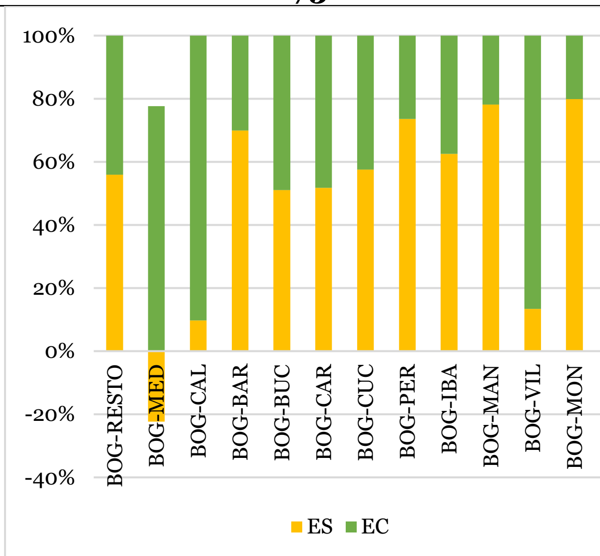
P25	**P25**	**P25**

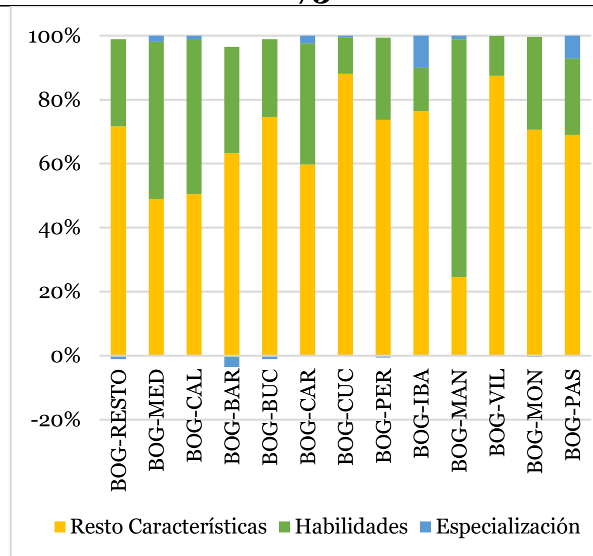


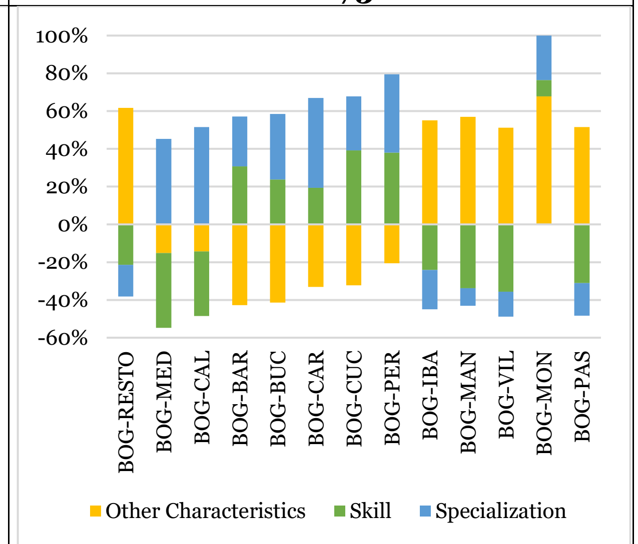




Aggregate descoposition	Composition effect	Structure effect
P75	P75	P75







Source: Authors' work using GEIH 2019

Robust standard errors in parentheses
***** p<0.01, ** p<0.05, * p<0.1**