Automation and the labor market: Evidence from technological change in Colombia, 2009–2017

Moreno Narvaez, Cristian Camilo

Universidad del Rosario

22 November 2022

Automation and the Labor Market: Evidence from Technological Change in Colombia, 2009-2017

Autor

Cristian Camilo Moreno Narvaez

Submitted as a requirement to opt for the degree of

Master in Economics

Advisors

Jorge Pérez Pérez

Paul Rodriguez Lesmes

Department of Economics

Universidad del Rosario

Bogotá - Colombia

2022

Automation and the Labor Market: Evidence from Technological Change in Colombia, 2009-2017*

Cristian Camilo Moreno Narvaez†

November 15, 2022

Abstract

This paper studies differences in the labor markets for occupations with different automation risk, and how actual automation may induce changes in wages and employment. Using data from Colombia between 2009 and 2017, we compute wage disparities by automation risk. We find that 62% of the occupied people in Colombia are at high-risk of automation. In the same way, we find that 71% of informal workers are at high-risk, while 56% of formal workers are. The wage return to education are highest in the less automatable occupations. We then look at the effects of actual education, measured by ICT investment. On wages and employment, automation increases employment, decreases wages and the wages gap by skill. Education acts as a protection mechanism against new automation technologies.

JEL Classification: J23, J24, J31, O33

Keywords: Labor demand, Automation, Employment, Wages, Technological change.

*I would like to acknowledge my supervisors, Jorge Perez and Paul Rodriguez for making this work possible. Their help and guidance enabled the project to be completed. I would also want to thank my family for their constant support and understanding while I was investigating and writing my thesis.

†Department of Economics, Universidad del Rosario. E-mail: cristian.moreno@urosario.edu.co

1 Introduction

Job automation is the development and inclusion of new technologies in firms' value chains, allowing capital to substitute for labor in a set of tasks (Acemoglu & Restrepo, 2020). Task automation brings about social, economic, and environmental changes (ONU, 2016). Research on job automation is essential nowadays because economies have to improve their labor policies to protect the workers' welfare in terms of wages and labor stability, while accounting for the current digital transformation. Martinez et al.(2020) expose that new technologies such as robots, artificial intelligence, and self-driving vehicles have a significant development in this transformation. They called it the fourth technological revolution. This revolution focuses the effort of innovation on introducing new technology able to carry out tasks performed by humans, rather than on the development of more productive advances in already existing machines (Barbieri et al., 2019). Therefore, it is relevant to study the effect of automation on aggregate employment, wages, and wage inequality to understand the impact of the current technological wave in developing countries like Colombia.[1]

We examine aggregate wages and employment in Colombia to answer the following research question: What has been the impact of automation on the Colombian labor market over the last decade, with the appearance of new automation technologies? To answer this, we divide our analysis in two parts. First, we estimate automation risk in Colombia, using the findings of Frey & Osborne (2017), to calculate the potential impact on aggregate employment and wages. We also relate the return on education, calculated by a Mincer equation, with the probability of automation by sector. Finally, we measure the Information and Communication Technologies (ICT) exposure[2] to estimate the automation effect on labor market outcomes by firm size and education level.

In the first part of the paper, we analyze automation risk in the Colombian labor market by combining automation probability data by Frey & Osborne (2017) with microdata from the Great Household Integrated Survey (GEIH) at the industry level in Colombia. We find that 62% of the occupied people are at high-risk of automation according to the tasks set performed by each worker. We find that education plays a role as a protection mechanism for the appearance of new automation technologies in the production chains. Among workers whose highest educational attainment is elementary school (1st°-5th°) , the share of high-risk workers is on average 75%. In contrast, middle and college-educated workers have a high-risk share of 63% and 41%, respectively.

An important aspect of the Colombian labor market is the high share of people in the informal sector. Our findings show that 71% of the people in the informal sector are at high risk of automation, whereas only 56% of people in the formal sector are at high-risk. We estimate return to education with a Mincer equation to analyze the relationship with the automation probability. Results show that education returns and automation risk have a stable negative relationship from 2009 to 2017.

[1] In developed countries, Autor (2019) highlights the growing proportion of workers earning low and high wages in the United States, as well as the decreasing share earning middle wages, as a result of international trade and technological changes.

[2] Hötte et al. (2021) expose another kind of automation which relates the technologies focus on cognitive tasks. They analyze the automation effect on the tax structure and use the interaction of two kinds of automation. The literature commonly uses robots to measure the automation of routine tasks as well as investment in ICT to measure the automation of cognitive tasks.

In the second part of the paper, we examine the relationship between a measure of automation (ICT exposure), employment, and wages by skill level. Using a two-way fixed effects model, we find that ICT exposure has a greater impact on employment and wages in the population in firms with more than 50 employees than on those of workers in smaller firms.

Our estimates show a positive effect for those working in organizations with more than 50 employees. A 10% increase in ICT exposure, in this group, leads to a rise in the employment of high-skill, low-skill, and total workers of 0.156%, 0.091%, and 0.154%, respectively. This means that the impact of workers with high skills leads to the total effect of aggregate employment in Colombia. In terms of wages, increasing ICT exposure by 10% results in a decrease of 0.276% and 0.0233% for high-skill and low-skill employees, respectively, and a decrease of 0.239% in wage premium. The results confirm that high-skill workers supply are more inelastic to wage changes than low-skill workers, resulting in a greater impact on the first kind of workers and a significant impact on the wage premium. The findings imply that ICT exposure generates employment in the Colombian economy while simultaneously decreasing aggregate wages and the wage premium.

This paper contributes to the literature on automation in two ways. First, it computes the automation risk in Colombia's labor markets, showing the potential impact across education levels and economic sectors. We do this by adapting automation risk probabilities from Frey & Osborne (2017), who calculated the automation risk that each occupation has in the United States[3]. Some other papers have calculated automation risks outside the United States. Arntz et al. (2016) estimate the automation risk of jobs for 21 OECD countries based on a task-based approach and find that 9% of jobs are automatable. According to Brambilla et al. (2021), the current automation process is unlikely to have a significant impact on Latin America's employment rate, with unskilled and semi-skilled workers carrying a substantial portion of the adjustment costs. We estimate the automation risk, by education level to understand its ideas as a possible protection mechanism from new technologies. Bustelo et al. (2019) and Bustelo et al. (2020) concentrate on the gender wage disparity in four Latin American nations. They use the task-based approach to consider that automation can displace specific tasks within an occupation rather than whole occupations. The main contribution is making the connection between gender and the "skills of the future" (STEM)[4] in order to understand how the gender gap relates to the automation risk. Cebreros et al. (2019) investigate the sorting of young employees in the Mexican economy by relating the automation risk to the informal sector and economic cycle. Our paper contrasts the risks associated with automation in the formal and informal sectors.

The second contribution of the research is to give evidence of how automation affects employment and wage growth by skill in Colombia. Acemoglu & Restrepo (2018b) propose a theoretical approach to automation and its different impacts on high and low-skill workers.[5] In the same way, the literature

[3]In the paper, the authors compute a Machine Learning model to obtain that the 42% of total US employment is at a high-risk to be automated in the next ten years. They use an occupation-based approach to relate the job automation.

[4]Science, Technology, Engineering, and Mathematics.

[5]The authors expose two different effects related to automation, productivity and displacement effect. Productivity effect generates a positive effect from automation, increasing the value-added and raising the labor demand of non-automated tasks. The displacement effect replaces labor from previously assigned tasks, changing production task content against labor and lowering labor share.

explores the polarization of the skill level of workers by modeling the supply and demand for skills and assuming two distinct skilled groups that do two different and imperfectly substitutable activities (see Autor & Dorn, 2013; Autor, 2019; Acemoglu & Autor, 2011). This research investigates the impact of automation on aggregate employment and wages, using ICT investment as a measure for automation. Previous studies have shown the relationship between ICT and automation. Dottori (2020) suggests an approach to estimate the effect of automation, taking into consideration the model's contemporary trade and ICT developments. In the same context, Taniguchi & Yamada (2019) estimate aggregate production extended to account for capital-skill complementary and skill-biased technological change. They found a significant influence of the observed expansion of ICT capital equipment on high-skill labor around the world. Some studies emphasize on the impact of robots as an automation on manual and routine jobs, whereas others focus on the impact of ICT investment as an automation of cognitive tasks. Benmelech & Zator (2022) investigate corporate investment in automation using data from Germany on business-level automation choices and cross-country statistics on robotization. In the same way, we follow Hötte et al. (2021), who estimate the impact of manual (robots) and cognitive (ICT) automation on taxation, to see the effect of technological change about cognitive tasks on aggregate employment and wages in Colombia.

This paper is divided into six sections. The second section introduces the information sources used and descriptive statistics. The third section shows stylized facts on the relationship between risk of automation and returns to education. The fourth section explains the empirical strategy. The fifth section explains the results of the central hypothesis about the automation effect, while the sixth exposes some mechanisms which may lead to automation protection. The last section concludes.

2 Data and Descriptive Statistics

We use two data sources on the likelihood of automation and labor market variables to calculate the automation risk, the main variable we use to analyze wage dynamics related to occupation automation and its potential impact on the labor market. For the probability of automation data, we employ the estimates proposed by Frey & Osborne (2017) for occupations in the United States. For wages and employment, we use the information on salaried employed individuals from Colombian labor Force Statistics (GEIH[6]) provided by the National Administrative Department of Statistics (DANE). Our initial findings combine the two data sources, allowing us to categorize Colombian workers based on the probability of automation technologies replacing their jobs. The second set of findings combines the GEIH and the Technological Development and Innovation survey (EDIT) to compute labor market dynamics and the exposure to investment in Information and Communications Technologies (ICT) for high and low-skill workers in order to model the impact of automation on employment and wage inequality.

Probability of automation data. We use detailed occupational data from Frey & Osborne (2017) who estimate automation risk by occupation in the United States. They use O*NET data to identify bottlenecks in task computerization and classify 702 occupations at a six-digit level using the 2010 Standard Occupational Classification (SOC-10) based on the likelihood that the tasks associated with an occupation can be automated. These estimates consider whether automation of a specific occupation is technologically feasible and whether technological and artificial intelligence advancements will allow for

[6]Great Household Integrated Survey, GEIH for its Spanish acronym.

automation of the occupation in the future. We use estimates based on the O*NET database because the tasks associated with a particular occupation are similar across countries.

Labor market data. We use the Great Household Integrated Survey (GEIH), a monthly cross-sectional survey that characterizes the primary outcomes of the Colombian Labor Market. The survey has representative information about wages, education level, occupation, sector of employment, and age, for 23 main cities and their metropolitan areas in Colombia.[7] We express wages in constant prices from 2018, using the consumer price index (IPC). We use the pool of individuals who answered the survey between 2009 to 2017, reported their schooling years, and were occupied as salaried. In this period, we identify 83 occupations and 61 employment sectors at a two-digits CIIU-4 level.

We use a series of crosswalks; figure 1 shows the crosswalks to match the 2010 Standard Occupational Classification (SOC-10) in the United States to the National Classification of Occupations (CNO-70) in Colombia. Appendix A provides details on the construction process. [8] With this match, we estimate the automation risk for each of the 82 occupations in the GEIH. Similarly, we investigate informality and its relationship with the automation risk. We define informal workers using DANE's definition.[9]

Information and Communications Technology (ICT) data. We use the survey of Technological Development and Innovation (EDIT) in the Manufacturing Industry, Services, and Trade provided by DANE. The EDIT database characterizes the innovation activities and technological development of Colombian firms. We use investment in information and communication technologies to generate a variable that is the sum of investments in equipment and machinery, internal research and development activities, acquisition of internal research and development, acquisition or use of intellectual property, technical assistance and consultancy, information and communication technologies, software development, and data analysis activities.[10] The database contains 16.891 firms with ten or more employees or a value-added higher than the measure established in the annual manufacturing survey (EAM) between 2009 and 2017. We conclude our analysis in 2017 to eliminate the shock from the Venezuelan exodus (Santamaria, 2019).

Descriptive statistics. Table 1 shows descriptive statistics about wages, education, and automation risk for 2,780,565 observations from 2009 to 2017, with an average of 231,731 observations per period. This sample represents 20,6 million workers per period on average, using the survey weights. Workers in Colombia have an average of 8.94 years of education and a monthly real wage of $1,015,396 (constant Colombian prices of 2018). The average age of employed people is around 39 years old in the sample. Panel A shows that the education level with the highest participation of workers in the total labor force is elementary school (27.59%), followed by middle school (26.82%), and college (25.41%). Panel B shows

[7]The main 23 cities are: Armenia, Barranquilla, Bogotá, Bucaramanga, Cali, Cartagena, Cúcuta, Florencia, Ibagué,Manizales, Medellín, Montería, Neiva, Pasto, Pereira, Popayán, Quibdó, Riohacha, Santa Marta, Sincelejo, Tunja,Valledupar, Villavicencio.

[8]In the following link is the repository of the process of crosswalk: https://github.com/ccamilocristian/Crosswalk_Frey_Osborne_GEIH

[9]DANE defines informality as the particular workers and the workers in companies, businesses, or firms with less than five employees, including the employer or partner; familiar workers without salary in companies with less than five employees, domestic workers with less than five employees, day laborers or laborers in enterprises with five or less workers, self-employed workers working in establishments of up to five people, employers or employers in companies of five workers or less, and exclude workers and government employees.

[10]We exclude the city Quibdó, due to the volatility of its ICT investment measure.

the automation risk for each category.[11] We estimate that 62.4% of workers in Colombia are at high-risk of automation, while 11.1% are at medium-risk and 26.4% are at low-risk. This contrasts with the 44% of employment at high-risk of automation for the United States, and the 65% for Mexico estimated by Frey & Osborne (2017) and Cebreros et al. (2019), respectively.

Panel C depicts the labor force share in Colombia by sector. Wholesale and retail have the highest economic share at 20.7%, followed by agriculture and manufacturing at 16.6% and 12.4%, respectively. These sectors are highlighted here because they are the most representative. The tasks set performed by workers in these sectors is primarily made up of routine and cognitive tasks, particularly in manufacturing. To analyze automation risk by education and firm size, we divided people into those who work in firms with less and more than 50 employees, as well as into people with high (college) and low-skill levels (middle school). Table 2 displays the descriptive statistics of the variables with a focus on the education level and automation, measured by the investment in ICT. We compute the variables as differences from 2009 to 2017. Workers' shares in this period are 76% and 24% for companies with less and more than 50 employees, respectively. Panel A shows a 1.02% decrease in ICT exposure for people with high skills working in firms with less than 50 employees and a 3.6% decrease for people working in firms with more than 50 employees. Panel B shows a decrease in ICT exposure for low-skill workers compared to high-skill people. Workers in firms with less than 50 employees have a 4.9% decrease in ICT exposure, while those in firms with more than 50 employees have a 7.5% decrease. These results leads the heterogeneous behavior of the ICT exposure across sectors and cities, which has high standard deviation due to an increase of workers and ICT investment.

3 Automation Risk

This section examines potential job automation risk in Colombia. By matching Colombian labor market data and job automation probabilities, we obtain an approximate measure of automation risk by occupation and economic sector.

3.1 Employment at Risk in Colombia

We begin with the estimates of Frey & Osborne (2017), identifying the workers most vulnerable to automation. Figure 2 depicts the share of workers per year in each risk category, demonstrating that worker shares in each risk category are stable over time. Employment concentrates mainly on occupations with a low and high-risk of automation. Medium-risk occupations, on the other hand, account for a relatively small proportion of total employment. In the Colombian labor market in 2017, 62% of employed people are classified as high-risk. At the same time, 26% are at low-risk, and 11% are at medium-risk. Based on our sample, the share at high-risk in Colombia is significantly higher than the percentage of total US employment at high-risk calculated by Frey & Osborne (47%).

Figure 3 disaggregates our estimates of automation risk by educational attainment. Workers without college education tend to be more exposed to automation. Among workers whose highest educational

[11] Following Frey & Osborne (2017), we categorize occupations as low automation risk (less than 40%), medium automation risk (40-70%), and high automation risk (>70%)

attainment is elementary school, the share of high-risk workers is on average 75%. In contrast with this education level, middle and college-educated workers have a share of 63% and 41%, respectively. We highlight that the low-risk category has the highest share in college, whereas in the rest of the education levels high-risk has the highest share. These results suggest that workers with higher education may perform non-routine tasks, such that accumulating human capital through investment in education could be a mechanism for workers to protect from the risk of automation.

Figure 4 shows the share of workers in each automation risk category by sector. All sectors have a relatively high participation of high-risk employment, and this share has not changed significantly in this decade. Agriculture, hunting, forestry, and manufacturing industries have a large concentration of employment in the high-risk category because the structure of their jobs includes a set of tasks sensitive to computerization (i.e., Farmers, carpenters, tailors, and dressmakers). However, some occupations are unlikely to be automated, at least soon, because automation will not affect jobs that involve intangible skills (e.g., creativity, flexibility, adaptability) or where human interaction processes produce the quality of the output that firms need (Autor, 2015).

Nevertheless, in the last decade, figure 5 shows a decrease in the workers with elementary and secondary education levels, while the participation is increasing in college and middle. We see what could be a displacement of jobs across education levels and sectors to perform new non-automated tasks. This displacement is because workers protect against the automation risk by increasing their years of education. In sum, the Colombian labor market shows significant participation of high-risk employees, concentrating their education in elementary, secondary, and middle school. However, a considerable part of these workers is in sectors with sets of tasks difficult to automate, for example, construction, agriculture, and hunting jobs. Therefore, an analysis of the displacement effect is relevant to explain the evolution of the share of workers across industries and education levels.

Informality in the labor market. An essential aspect of the labor market composition in developing countries is informality and its relationship with small and medium-sized enterprises. Based on the GEIH, during 2009-2017 there were on average, 43% of informal workers (Figure 6). Cebreros et al. (2019) argue that the adoption of automation technologies is not profitable or feasible in the informal sector. Given that self-employment and small establishments are typical of the informal sector, automation may have a differential effect across formal and informal employment. Figure 7 reports the distribution of Colombian employment, distinguishing the formal and informal sectors across automation risk categories. In the formal sector, on average, 56% of the workers are at high risk of automation, a constant trend from 2009 to 2017. In contrast, the 71% of the people in the informal sector are at the same automation risk category. There is a decreasing trend of workers in the high automation risk category for the formal sector, where the participation passed from 57.5% in 2009 to 56.5% in 2017. The movement of people from low-skill to high-skill as a form of protection against automation can explain the re-composition of the share of employment at risk. Also, we highlight that informal employment is particularly prevalent in agriculture, fishing, mining, and retail, which concentrate employment in occupations with a high degree of automation risk. Considering the nature of the occupations in these sectors, we see informality as a mechanism that affects the degree of automation in Colombia and other developing countries.

Geographic risk of automation. Another relevant dimension to automation risk is its geographical differences. Departments with a high concentration of sectors like agriculture, mining, and fishing have a higher share of people at high risk of automation. Furthermore, departments with weak institutions will allow firms to hire people for routine tasks in industries such as manufacturing, mining, and agriculture. Figure 8 shows the average risk of automation for the 23 departments in the GEIH. Departments on the Pacific coast, such as Chocó and Nariño, show a near 70% of automation risk. On the Caribbean coast, states such as Magdalena and Guajira show a 66% risk. In the center of Colombia, departments show a 62% automation risk. Notice that Bogota, as a capital district, has the lowest risk of automation in the country, at nearly 54%. Another essential aspect to highlight is the economic incentives that the departments have to invest in new technology. Developing countries have to differentiate their investment policies by state to take into account the heterogeneous distribution of sectors. Thus, the level of development in the countries relates to the strength of the economic policies focused on technological innovation and politics to protect jobs from automation.

3.2 Wages and Automation Risk

Figure 9 depicts average real wages by automation risk category. The average real wage is $930.889, $873.429, and $1.584.82 in the medium, high, and low-risk categories, respectively. We see that the real wage in the high-risk category is similar to that in the medium-risk category. In the early period (2009), the wages of these two categories were close, but around the year 2016, the inequality increased. On the other hand, there is a prominent gap between the average real wage for the low-risk category and the others, although it decreases over time. Considering figure 2, the potential impact of automation in Colombia could be prominent in the high-risk category where labor participation is close to 60%, and the real wage is the lowest.

Now, we analyze wage inequality between workers in college (skilled) and middle school (unskilled) with the wage premium, a summary measure of the market's valuation of skills. For this purpose, we compute the real wages of both education levels, taking into account the risk categories, and finally measure the wage premium as the ratio of the wages. Figure 10 shows the evolution of real wages in these two categories. In the middle education level, inequality decreases significantly over time, while for the college education level it is stable. With this result, we can see in panel A how the return of education in middle school decreases for the workers with low automation risk category and increases for employees in high and medium risk. On the other hand, panel B shows that although the return of each risk category decreases, the gap has no evident change over time. For the middle education level, real wages for low automation risk category decreased in 13% from 2009 to 2017, whereas medium-risk increased in 13%. Real wages for high automation risk have no change over time. In college education level, low-risk workers reduced real wages by 12%, while medium and high-risk workers reduced real wages by 11% and 5%, respectively.

Figure 11 depicts the wage premium from middle and college education levels, showing that the low automation risk category has the highest level of inequality. For this risk category, the disparity has no change over time. In contrast, until 2015, inequality in the medium-risk category had decreased.This wage premium behavior means that for workers at medium-risk, the return of education in middle school increases more than the college degree, which describes the decrease in the gap.

We now want to adjust the real wages using a Mincer equation model. It generalizes the idea of

connecting the human capital theory with the microdata information to study wage inequalities (Rosen, 1992), quantifying the earnings in the labor market as a function of the accumulation through investment in education.[12] We measure the wage taking into account the benefits of investment in education in order to compute wage inequality between the automation risk categories in the Colombian microdata.
The Mincer equation can be expressed as follows:

$$ln(w_i) = \alpha + \rho_e Edu_i + \beta_0 Exp_i + \beta_1 Exp_i^2 + \epsilon_i \tag{1}$$

Where w_i represents wage for the person i, Edu_i represents years of education, Exp_i stands for experience, and ϵ_i is the error term. In this linear regression, many authors have interpreted ρ_e as the Internal Rate of Return to Education (IRR). [13] This approach presumes that education has the same effect regardless of educational level. To relax this assumption, we estimate a Mincer equation with splines to differentiate because those who complete a level of education receive greater retribution than those who do not.

Following Alvarez et al.(2017), we construct splines to disaggregate the effect of each educational level on the wage of a particular person. Therefore, our analysis follows a regression of the form:

$$ln(w_i) = \alpha + \rho_e Edu_i + \beta_0 Exp_i + \beta_1 Exp_i^2 + \beta_2 \phi_{sec} + \beta_3 \phi_{coll} + \beta_4 \phi_{post} + \epsilon_i \tag{2}$$

Where ϕ denotes the splines that correspond to secondary, college, and postgraduate education levels, we compare wage inequality across the high, medium, and low probability of automation using the residuals of the equation. Appendix B shows an explanation of the splines construction. Once we estimate the model, we relate each individual's residuals to the probability of automation and then aggregate by the risk category. The constant α represents the base wage without human capital accounting for experience, considering the relationship between education and experience. Figure 12 shows that before 2012, the workers at low-risk earned more than the others in the two categories. However, after this year, the medium-risk becomes the category where the average workers has the highest residual wage. The high-risk employees throughout the period earn less than the others.

As discussed, there is wage inequality between high-skill and low-skill workers, specifically those at high risk of automation. Cebreros et al. (2019) exposed that firms will have incentives to automate those jobs, resulting in the most significant net savings, defined as the net cost of adopting automation technologies. In the Colombian case, the firms will want to get the workers in low-risk automation but with high-skill because they have a low cost-productivity relationship, described by Acemoglu and Restrepo (2018b) as the productivity effect. However, the firms will displace the workers in this automation risk category because it is cheaper to adopt new automation technologies than to hire low-skilled workers, based on the structure of the tasks performed in each occupation. We will analyze the relationship between the return to education and the wage premium to see how education allows protection from job automation.

[12] Mincer (1974) models the natural logarithm of wages as a function of the years of education and the potential experience in the labor market (We use age as a proxy for experience).

[13] The IRR means the average percentage change of a person's wage if his education level increases in a unit, ceteris paribus. Based on Mincer's model framework, we can affirm that the IRR and experience are positive because as education levels rise, so does productivity and, consequently, salaries. Mincer also proposes the equation as a concave function because one year of experience raises wages slower than the previous year.

Educational Returns. It is possible that the relationship between earnings and automation risk comes from the interaction between schooling and the composition of the tasks set by occupation in each economic sector. The IRR (Internal Rate of Return to Education) differs across automation risk levels. Figure 13 depicts that sectors' IRR is negatively related to the automation risk. In 2009, passing from a high-risk category (0.7) to a low-risk category (0.4) increased the return to education by 3.3 percentage points, but in 2017 the increase was 4.05 percentage points. We test the hypothesis of the negative relationship and constant over 2009 to 2017 of the educational returns and probability of automation. We applied bootstrapping to see the statistical significance of the change. The result suggests the incentive of workers has a displacement effect. People with high-risk automation move to low-risk automation, improving their education level as a mechanism of protection from new automation technologies.

According to Alvarez et al.(2017), there is a significant difference between post-secondary and pre-university returns, not only in terms of levels but also of trends: pre-university educational returns have been steadily declining, whereas post-secondary educational returns appear to have stabilized. As we see in figure 5, there is a re-composition in the share of education levels, especially in college, elementary, and middle school.

Wage inequality. Acemoglu & Restrepo (2021) based their study on the effects of skill-biased technological change where the employment levels of high-skill and low-skill workers take an essential role in aggregate production. They argue that much of the change in the wage structure of the United States is being driven by the automation of tasks previously performed by certain types of workers in certain industries (For example, numerically controlled machinery or industrial robots could replace factory workers in manufacturing, while specialized software could substitute for clerical workers.).

In 2009, passing from a high-risk category to a low-risk category decreased wage inequality by 8.22 percentage points. In 2017, it increased by 8.12 percentage points. Figure 14 analyzes the wage premium in each risk category, showing a wage inequality increase in the low-risk category and the decrease in the high-risk category. Cebreros et al. (2019) show that median wages for an occupation are negatively related to the automation risk in Mexico, and suggest that occupations at high-risk of automation may actually experience lower worker displacement than less vulnerable occupations due to lower wages, which reduce the incentives to automate. Workers at high-risk from the early years displace workers with low probabilities of automation through increasing education levels, which leads to an increase in wages because there is a productivity improvement, changing the relationship between wage inequality and the probability of automation.

4 Measuring ICT Exposure in Labor Markets

In this section we implement an empirical approach based on Acemoglu & Restrepo (2018b), to estimate the impact of automation on wage premiums and employment.[14] There is no data on robots in Colombia; thus, our identification strategy uses information and communications technology data based on the research conducted by Hötte et al. (2021). They use investments in new machinery (ICT) as the measure of cognitive automation, another kind of technological change. Our unit of analysis is a city-sector

[14] Appendix C provides details about the model.

cell (Using CIIU Rev. 4 sectors).

The baseline model can be specified as follows:

$$\Delta y_{ct}^{si} = \alpha' \Delta X_{ct}^{si} + \beta \Delta ICTExposure_{ct}^{si} + \epsilon_t^{si} + \eta_c^{si} + u_{ct}^{si} \tag{3}$$

Where c denotes a city-CIIU4 sector cell; the starting year is 2009, and the ending year is 2017. The index s identifies high and low-skill employment groups, and i represents the population working in firms with less or more than 50 employees. The dependent variables y_{ct} are the wage premium, employment, and wage growth measured as log differences. We use long-term changes from 2009 to 2017 to reduce the impact of yearly measurement error. The vector X_{ct} includes a set of control variables to allow for different trends across cells depending on their socio-demographic structure. The variable are the share of the female population, the share of the population between 20 and 40 years, and the share of the population between 40 and 60 years. We also consider time and city-sector fixed effects in order to control observable and unobservable systematic differences between observed time units, as well as to control for average differences across city-sector cells that may affect the predictor variables, removing the effect of those time-invariant characteristics so that we can assess the net effect of the predictors on the outcome variable.

In the literature automation is measured using data on industrial robots from the International Federation of Robotics (IFR). The IFR (2020) defines the industrial robot as an automatically controlled, reprogrammable, multipurpose manipulator for industrial applications. Authors like Ge & Zhou (2020) and Aghion et al. (2020) measure robot exposure to estimate the effect of automation on labor market outcomes because robots can replace humans in executing specific routine tasks. For the purpose of this research, we use ICT capital intensity to measure another type of automation related to cognitive tasks. Investment in information and communication technologies can be adapted flexibly to a wide range of tasks, many of which do not have a clear alternative in the range of tasks performed by humans.

We consider ICT exposure as the measure of automation, calculating ICT investment exposure by city-ciiu4 sector cell:

$$\Delta ExposureICT_{ct} = \Sigma_j \frac{L_{csj}}{L_{c2008}} \Delta ICT_{jt} \tag{4}$$

Where ΔICT denotes the log difference in sector-level ICT investment, and Lcj denotes the number of employees in cell city-sector c and sector j (These sector are a more specific classification of CIIU4 sectors). We use initial employment at the city-CIIU4 level $Lc2008$ to normalize ICT investment across industries; Acemoglu & Restrepo (2017) note that this decision also contributes to mitigating mechanical correlation or mean reversion issues associated with changes in industry employment in anticipation of the subsequent introduction of new technology. Thus, ICT exposure is essentially a Bartik-style measure that combines industry-level variation in ICT expenses with baseline employment shares.

Estimation issues. The previous model has an endogeneity issue due to shocks in demand factors associated with current global economic dynamics. The estimation strategy considers that changes in ICT investment may be endogenous to demand shocks that affect employment and may be correlated with current trends affecting employment transformation or distribution across industries. In order to address this problem, our identification model takes into account fixed effects for time and state effects to control average differences across cities, removing the effect of those time-invariant characteristics. Other authors have alternatively considered an instrumental variable approach to address the endogeneity.

A potential application for Colombia would be to address this issue with a measure of ICT exposure for developed countries that have high commercial interaction with Colombia. Unfortunately, no data is available for the ICT exposure in more developed countries to implement this strategy. Another issue is that we only estimate automation that affects cognitive tasks using ICT exposure; however, the literature commonly uses robots to measure the automation of manual and routine tasks. Hötte et al. (2021) analyze the effect of automation on tax structure and use the interaction of both kinds of automation. Finally, to address the level of industrialization in the country, we separate the population into people working in firms with a different number of employees to consider the differential ICT intensity.

4.1 The Impact of Automation on the Colombian Labor Market

This section summarizes our key findings based on the previous model. We show that automation affects the wages and employment of high and low-skill workers differently depending on the size of the firms where they work, resulting in wage inequality, as we see in the analysis of automation risk and educational returns.

4.1.1 Employment

Table 3 shows the results of the employment growth in panel A, segmented by the number of employees in the firms and the skill level. We see a negative impact of automation on the people with high-skill working in enterprises with less than 50 employees and a positive effect on workers with low skills. Although these results are non-statistically significant, we see a negative effect of ICT exposure on employment growth (third column "Total"), which suggests that workers with high-skills are affected to a greater degree the employment than workers with low-skills. Continuing, we observe a positive effect for the total, high-skill, and low-skill workers in the category of people working in enterprises with more than 50 employees. When ICT exposure increases by 10%, the employment of high-skill, low-skill, and total workers increases by 0.156%, 0.091%, and 0.154%, respectively. From table 4, we see that ICT exposure explains the positive effect of the labor supply, which leads to the previous result in employment over 2009 to 2017. The table shows that the labor supply of low-skill people has a higher automation impact than that of high-skill people.

4.1.2 Wage

Table 3 provides the wage growth results in panel B, also divided by the number of employees in the firm and the skill level. We observe that ICT exposure has a positive effect on wage growth for workers with high and low-skills in enterprises with less than 50 employees. In contrast, ICT exposure has an effect on the wage premium, which means that this kind of automation reduces inequality in wages for workers with high and low-skills. In the case of people in firms with more than 50 employees, we estimate that ICT exposure has a negative effect on the wages of workers with high and low-skills. Arango et al. (2019) show that in the Colombian scenario, high-skill workers are more inelastic to wage changes than low-skill workers, resulting in a greater impact on the first kind of workers and a significant impact on the wage premium. In the effect of our study, we found that ICT exposure has a negative influence on wage premiums. Increasing ICT exposure for high-skill workers by 10% has a statistically significant effect on wage growth of -0.276%. The negative impact on wage growth for low-skilled employees is 0.023%. These results have a negative effect on the wage premium of 0.239%.

5 Discussion

In general, we observe that automation, as explained by ICT exposure, has different effects when firm size and skill level are considered. For workers in enterprises with less than 50 employees, the impact on employment growth is negative, whereas the effect is positive for workers in firms with more than 50 employees. On the other hand, ICT exposure has a positive effect on wage growth in workers with high and low-skills who work in companies with less than 50 employees, and a negative effect for people in firms with more than 50 employees. We observe a negative impact on the wage premium for the workers in both firm-sized groups.

5.1 Summary of the Effects of Automation

Summarizing, high and low-skill workers at firms with more than 50 employees see a decrease in wages and an increase in employment as a result of automation. People at companies with less than 50 employees, on the other hand, face wages decrease, with different effects on employment depending on skill level. ICT exposure increases employment for low-skill employees while decreasing employment for high-skill workers.

We focus on the population where the results are statistically significant, more than 50 employees. It is not clear whether the productivity effect has a greater impact than the displacement effect. The results show that ICT exposure has a positive effect on the labor supply of high and low-skill workers, which explains the positive impact on employment growth. Workers with sets of non-automatable and automatable tasks are attractive for firms to satisfy the needs of people to perform new sets of tasks that new technology can not.

Based on our findings, we explore two scenarios that could be possible channels for our results to emerge, even if not formally tested in the empirical analysis. First, ICT exposure could generate a productivity effect greater than the displacement effect for high-skill workers and a displacement effect greater than the productivity effect for low-skill workers. In this scenario, workers with low skills would have incentives to enhance their education level to improve their skills, which would serve as a tool to protect them against automation. Also, high-skill workers would become more productive. An alternative scenario would be where the displacement effect has a greater impact on high-skill workers acquiring new technologies. In this case, automation would displace high-skilled workers (i.e., replaced by technology) as their set of tasks resulted in being easily automatable, and low-skill workers would become more productive in non-routine tasks. Finally, it is worth mentioning that even if ITC exposure would affect both high- and low-skilled workers in the same direction (that is, either the displacement or the productivity effect would prevail for both types of workers), this would result in higher wage inequality as the wage of high-skill workers tend to be more inelastic than the wage of low-skill ones.

5.2 Mechanisms

The increasing use of new technology focusing on automation generates a discussion about whether there are some conditions in the economic sectors and countries that could generate unemployment protection from implementing this new technology.

The first mechanism to explore is the shortage of capital that faces most developing countries due to automation equipment's being more expensive than other equipment, which means firms have a disadvantage over others in whether they have enough capital to acquire new technology. OIT (1966) explains that capital can be treated as a scarce commodity, where the pressure to use capital-saving technology is complementary to the effort to provide new job opportunities for the workforce. In the same way, enterprises' decisions depend on the different costs and benefits and the expected cost factor to generate an inclusion of automation in their value chains. One of the principal costs in the firms' decision is the interest charge on borrowed funds. In this case, monetary policy plays a principal role in incentivizing firms to acquire new technology where a low-interest charge leads to relatively lower prices for acquiring capital for new machinery.

Another important mechanism is the industrialization level of the countries, which relates to the facility to connect the regions and transport the new technology. This implementation of new technology is an issue in countries with low institutionality and high internal conflict because the firms have no economic incentives to acquire technology, also due to inadequate communication lines, which implies a high cost to implement, and they move more to industrialized zones such as the capital of the states.

The firms that used to generate guides and courses for their employees focus on improving the skill sets of their workers in order to generate high productivity due to the introduction of new technology. This extensive assistance in firms to keep employment creates protection for workers with low skills through automation thanks to moving to another set of tasks or upgrading the same task in each occupation.

Finally, the informal sector, characterized by self-employment and employment in micro establishments, is typically more credit constrained and less technologically advanced than formal sector establishments. Thus, the formal sector used to generate more impact for the acquisition of automation technology than the informal sector. Cebreros et al. (2019) show evidence that the informal sector has either no economic incentive or no possibility to adopt automation technologies and highlight the necessity of incorporating the technological possibilities and economic incentives for the adoption of automation technologies into any analysis that seeks to quantify the magnitude of the threat posed by automation.

6 Conclusion

This study examines how automation affects wages, employment, and wage premium. First, we present stylized facts from the relationship between automation risk and the Colombian labor markets. This relationship demonstrates that education serves as a protection mechanism for workers against the arrival of new automation technologies into production chains. In particular, we find that 62% of the occupied people in Colombia are at high-risk of automation due to the tasks set performed by each worker. The people with college have 41% in high-risk, and the riskiest category of education is elementary (1st°-5th°) with 75%. On the other hand, we find that 71% of informal workers are at high-risk, while 56% of formal workers are.

In order to analyze the relationship between real wages and probability of automation, we compute the Mincer equation to adjust wages, and obtain return to education. We find a negative relationship between education returns and probability of automation. In the same way, the study shows the wage inequality between high and low-risk of automation workers. Results suggest the incentive of workers to move from high-risk of automation occupations to low-risk improving the education level. These findings validate the analysis from the stylized facts about the education as a protection mechanism for workers against automation.

The research, on the other hand, investigates the relationship between automation, as measured by ICT exposure, and labor market outcomes such as wages and employment. We find that automation has different effects depending on skill level and firm size. According to the findings, ICT exposure positively affects employment growth for workers in organizations with more than 50 employees and negatively impacts workers in enterprises with less than 50 employees. Workers with sets of non-automatable and automatable tasks are attractive for large firms because they perform tasks that new technologies can not. Wage growth is negative for workers in firms with more than 50 employees and positive for those in companies with less than 50 employees. Low-skill workers improve the skills as protection for automation and become more productive in non-automatable tasks. On the other hand, high-skill workers are more productive and are displaced because they have sets of tasks highly automatable.

The estimation issues in the present study are presented as a potential line of research. Future research can use ICT exposure from developed countries with high commercial interaction with Colombia as an instrument. Another potential line of research is to estimate automation of manual and routine tasks with data about robots by sectors in Colombia. Finally, we propose to differentiate the automation impact on labor market outcomes by informal sector. This proposal will address the mechanism in developing countries where the informal sector has a greater share than the formal sector.

7 References

Acemoglu, D. & Autor D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. Handbook of Labor Economics, 1043-1171.

Acemoglu, D., & Restrepo, P. (2018a). Modeling Automation. AEA Papers and Proceedings, 108: 48–53.

Acemoglu, D., & Restrepo, P. (2018b). Low-Skill and High-Skill Automation. Journal of Human Capital, University of Chicago Press, vol. 12(2), pages 204-232.

Acemoglu, D., & Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. Journal of Economic Perspectives 33 (2): 3–30.

Acemoglu, D. & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. NBER working paper 23285.

Acemoglu, D. & Restrepo, P. (2021). Tasks, Automation, and the Rise in US Wage Inequality. NBER

Working Papers 28920.

Aghion, P. & Antonin, C. & Bunel, S. & Jaravel, X. (2020). What are the Labor and Product Market Effects of Automation? New Evidence from France. CEPR Discussion Paper No. DP14443.

Aksoy, C., Ozcan, B. & Philipp, J. (2020). Robots and the Gender Pay Gap in Europe. IZA Discussion Papers 13482, Institute of Labor Economics.

Alvarez, O; Gaviria, A.; Jiménez, M. & Tenjo, J.(2017). Evolution of Returns to Education in Colombia (1976-2014). Coyuntura Económica, Fedesarrollo, vol. 47(1 y 2), pages 15-48.

Arango, L., Castellani, F. & Obando, N.(2019). Heterogeneous Labour Demand in the Colombian manufacturing sector. Journal Labour Market Res 53, 1.

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, Paris.

Autor, D. (2015). Why are there still so many Jobs? The History and Future of Workplace automation. Journal of Economic Perspectives, 29(3), 3–30.

Autor, D (2019), Work of the Past, Work of the Future. NBER working paper 25588.

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

Benzell, S. & Kotlikoff, L. & Lagarda, G. & Sachs, J. (2015). Robots are Us: Some Economics of Human Replacement. NBER Working Paper No. w20941.

Barbieri, L., Mussida , C., Piva, M. & Vivarelli, M. (2019). Testing the Employment Impact of Automation, Robots and AI: A Survey and some methodological issues. Quaderni del Dipartimento di Politica Economica, Universita Cattolica del Sacro Cuore.

Benmelech, E. & Zator, M.(2022). Robots and Firm Investment. NBER Working Paper No. w29676.

Brambilla, I., César A., Falcone G., Gasparini, L. & C. Lombardo (2021). The Risk ofAutomation in Latin America. Documentos de Trabajo del CEDLAS Nº 281, CEDLAS-Universidad Nacional de La Plata.

Bustelo, M., Egana-delSol, P., Ripani, L., Nicolas, S., & Viollaz, M. (2020). Automation in Latin America: Are Women at Higher Risk of Losing Their Jobs? Inter-American Development Bank.

Bustelo, M., Flabbi, L., & Viollaz, M. (2019). The Gender Labor Market Gap in the Digital Economy. IDB Working Paper Series 1056.

Cebreros, A., Heffne, A., Livas, R. and Puggioni, D. (2019). Automation Technologies and Employment At Risk: The Case Of Mexico. Banco de México.

Crowley, F.& Doran, J. (2019). Automation and Irish Towns: Who's Most at Risk?, SRERC Working Paper Series, No. SRERCWP2019-1, University College Cork, Spatial and Regional Economic Research Centre.

Daisuke, A., Daiji, K. & Yukiko,S. (2020). Robots and Employment: Evidence from Japan, 1978-2017.

Discussion papers 20051, Research Institute of Economy, Trade and Industry (RIETI).

Dottori, D. (2020). Robots and Employment: Evidence from Italy. Questioni di Economia e Finanza (Occasional Papers) 572, Bank of Italy, Economic Research and International Relations Area.

Faber, M. (2018). Robots and Reshoring: Evidence from Mexican Local Labor Markets. Working papers 2018/27, Faculty of Business and Economics - University of Basel.

Frey, C. & Osborne, M. (2017). The Future of Employment: How Susceptible are Jobs to Computerisation? Technological Forecasting and Social Change, 114, 254–280.

Hirvonen, J. & Stenhammar, A. & Tuhkuri, J.(2022). New Evidence on the Effect of Technology on Employment and Skill Demand. ETLA Brief 108, The Research Institute of the Finnish Economy.

Hötte, K. & Theodorakopoulos, A. & Koutroumpis, P. (2021). Does Automation Erode Governments' Tax Basis?. An Empirical Assessment of Tax Revenues in Europe. Papers 2103.04111, arXiv.org.

Ge, S. & Zhou, Y. (2020). Robots, Computers, and the Gender Wage Gap. Journal of Economic Behavior & Organization, Elsevier, vol. 178(C), pages 194-222.

IFR (2020). International Foundation of Robotics. Proprietary Data on Industrial Robot Deliveries and Stocks.

Kugler, A., Kugler, M., Ripani, L. & Rodrigo, R. (2020). U.S. Robots and their Impacts in the Tropics: Evidence from Colombian Labor Markets. NBER Working Papers 28034, National Bureau of Economic Research.

Lerch, B. (2021). Robots and Nonparticipation in the US: Where Have All the Workers Gone?. IdEP Economic Papers 2003, USI Università della Svizzera italiana.

Martinez, R., Palma, A. & Velásquez, A. (2020). Revolución Tecnológica e Inclusión Social: Reflexiones sobre Desafíos y Oportunidades para la PolíticaSsocial en América Latina. Serie Políticas Sociales, N° 233, Comisión Económica para América Latina y el Caribe (CEPAL).

Nedelkoska, L. & G. Quintini (2018). Automation, Skills Use and Training. OECD Social, Employment and Migration Working Papers, No. 202, OECD Publishing, Paris.

Micco, A. (2019). Automation, Labor Markets, and Trade. Working Papers wp486, University of Chile, Department of Economics.

Mincer, J. (1974). Schooling, Experience and Earnings. Columbia University Press - NBER.

OECD. (1996). Job Creation and Loss: Analysis, Policy and Data Development. París.

OIT. (1966). Automation in Developing Countries. Cepal.

ONU (Organización de las Naciones Unidas). (2016). Global Sustainable Development Report. Publicación de las Naciones Unidas.

Rosen, S. (1992). Distinguished Fellow: Mincering Labour Economics. Journal of Economic Perspectives, núm. 6, vol. 2, pp. 157-170.

Santamaria, J. (2019). Venezuelan Exodus: The Effect of Mass Migration on Labor Market Outcomes.

University of Minnesota.

Taniguchi, H; Yamada, K. (2022). ICT Capital–Skill Complementarity and Wage Inequality: Evidence from OECD Countries. Papers 1904.09857, arXiv.org

Zemtsov S., Barinova V. & Semenova R. (2019). The Risks of Digitalization and the Adaptation of Regional Labor Markets in Russia. Foresight and STI Governance, vol. 13, N° 2, 84–96.

Appendices

A Match between Frey & Osborne's results and GEIH

In order to obtain the risk of automation per occupation in Colombia, which is the principal variable to analyze wage dynamics related to the automation of occupations and its potential impact on the labor market, we combine the probability of automation and the labor market data.

Frey & Osborne (2017) estimate a probability for each occupation found in the O*NET database by the 2010 Standard Occupation Classification (SOC-10) at a six-digit level in the United States. Following these results, we match the probability of automation in Colombia and the United States by the occupation classification with the crosswalks SOC-10 versus the National Classification of Occupations (CNO-70). The standard code in Colombia is the National Classification of Occupations[15] which is at a two-digit level and follows the International Standard Classification of Occupations 1988 (ISCO-88).

Methodology:

1. In conjunction with the Faculty of Economics of the University of Warsaw, Wojciech Hardy of the Institute for Structural Research developed a correlation for the SOC-10 at six digits with the ISCO-08 at four digits (International Standard Classification of Occupations)[16].
2. On the other hand, the International Labour Organization[17] created a four-digit equivalent of ISCO-08 and ISCO-88.Based on this correspondence, it was crossed with the base of paragraph number 1, yielding an input base of the form SOC-10 and ISCO-88.
3. Transcribe into a database the correlative made by DANE from the ISCO-88 code of four digits to the CNO-70 code of four digits[18].
4. Group the CNO-70 code into two digits and relate the corresponding ISCO-88 code to four digits. At this point, the grouping is cleaned, with duplicates and ISCO-88 occupations reduced to four digits that do not correspond to the CNO-70 subcategory to the two digits used by DANE in the GEIH.
5. We obtain a correlative of the CNO-70 code to two digits with the SOC-10 code to six digits by crossing the database of subsection number two with the ISCO-88 key with the database of the previous paragraph.
6. Cross the basis of automation risk probabilities in the paper by Frey & Osborne (2013) with the base of the previous paragraph by the SOC-10 6-digit key. Then, group at the level of CNO-70 and leave the average probabilities related to the SOC-10 codes in the correlative of point 5.

[15] The Servicio Nacional de Aprendizaje (SENA) is the Colombian institution that has the objective of updating codes

[16] Hardy, W. (2016). IBS. Obtained from Occupation classifications crosswalks – from O*NET-SOC to ISCO: https://ibs.org.pl/en/resources/occupation-classifications-crosswalks-from-onet-soc-to-isco/

[17] OIT. (2008). Estructura de la CIUO-08 y concordancias previas con la CIUO-88. Obtained from OIT: https://www.ilo.org/public/spanish/bureau/stat/isco/isco08/index.htm

[18] DANE (2005). Clasificación Internacional Uniforme de Ocupaciones Adaptada para Colombia CIUO-88 A. C. Bogotá D.C.: DANE

7. Cross the GEIH of Colombia input base for the reference years (the key is the variable OFICIO) and the database of point 6 at the level of CNO-70 (the key is the CNO-70) to calculate the probability of automation risk for the Colombian case data from Frey & Osborne.

The crosswalks construction has some loss of information. Beginning with Wojciech Hardy's crosswalk of SOC-10 to ISCO-08, there is no loss of information when we combine this correlative with the ISCO-08 to ISCO-88. However, in the merger of ISCO-88 and CNO-70, we lost 314 (37%) occupations. Finally, when we merge the correlative of SOC-10 to CNO-70 with the results of Frey and Osborne, 106 occupations (15%) are not related to Colombian occupations. Some occupations are auditors, lodging managers, foresters, and telemarketers.

After matching the occupations with the use of the crosswalks, we lose 1% of the representativity of the occupations in the GEIH, which have no automation probability related. Even without these occupations, the results are still significant for the analysis. The occupations lost in the matching are Chiefs of bonded personnel, Petty officers, supervisors and foremen, and Merchants and Sellers not qualified under other epigraphs.

B Splines

The splines allow estimating different effects of different education levels on each worker's wage. We use linear splines to capture these secondary, college, and post-college effects. The interpretation of the coefficient of the segments is as a market premium associated with having a specific education level.

The following are the steps to construct the linear splines:

1. Generating the variables dummy for the three education levels:
   $D_{sec} = 1$ if year of education $>= 5$
   $D_{Coll} = 1$ if year of education $>= 11$
   $D_{post} = 1$ if year of education $>= 16$
2. Interacting the dummies with the years of education:
   $\phi_{sec} = D_{sec} * (education - 5)$
   $\phi_{coll} = D_{Coll} * (education - 11)$
   $\phi_{post} = D_{post} * (education - 16)$

C Low-Skill and High-Skill Automation

The empirical approach, as described by Acemoglu & Restrepo (2018b), estimates the impact of robot exposure on wage premiums and labor employment. In the model, low-skill, high-skill labor or capital can perform a continuum of tasks, and they compete with each other assuming that developments in artificial intelligence allow capital to compete against high-skill labor in complex tasks or low-skill in routine and manual jobs.

The model starts with a static economy with a given supply of capital and inelastically supplied low-skill and high-skill labor. It considers an economy with a unique final good Y, produced by combining a continuum 1 of tasks $y(i)$ with an elasticity of substitution j ϵ $(0, \infty)$, uses the tasks as a unit of labor activity to produce and workers apply their provision of skills for tasks in exchange for a salary:

$$Y = [\int_0^1 y(i)^{\frac{(\sigma-1)}{\sigma}} di]^{\frac{\sigma}{(\sigma-1)}} \tag{5}$$

The final good is produced competitively, consumer utility is defined over the unique final good, and we normalize its price to 1. Final-good producers can produce each task with machines (capital) or labor, and there are two types of labor, high- and low-skill. All tasks can be produced by both types of labor, though they have different productivities in each task. The functions $\gamma_H(i)$, $\gamma_L(i)$, and $\frac{\gamma_H(i)}{\gamma_L(i)}$ are continuous and strictly increasing.

The model, therefore, abandons the supermodular comparative-advantage structure across all factors and tasks by assuming that there exists J ϵ (0, 1) and i ϵ (0, 1) such that, when automated, tasks i $<$ J can be produced with capital with productivity 1, while tasks i $\geq$ J can be produced with capital with productivity $\gamma_K \geq 1$. (Model refers to tasks i $<$ J as "simple" tasks and to i $\geq$ J as "complex" tasks.)

The model characterizes the implications of low-skill automation, which corresponds to an increase in the tasks set that capital can perform at the bottom of the distribution, and high-skill automation, which corresponds to an increase in the set of tasks that capital can perform toward the top of the distribution. It demonstrates that both types of automation have two distinct effects: a displacement effect and a productivity effect. The displacement effect negatively affects the labor market wealth of the directly affected factor by taking away tasks, whereas the productivity effect tends to increase the wages of all factors.

When the displacement effect dominates, factors affected by automation experience a decline in their wages, most interestingly, the displacement caused by automation also creates wave effects. High-skill automation displaces high-skill labor, which may compete with low-skill labor in other tasks and displace this latter group. For instance, high-skill automation can reduce the real wages of both low-skill and high-skill labor. However, the displacement effect on the directly affected group is always greater, so low-skill automation increases the inequality between high-skill and low-skill labor, whereas high-skill automation has the opposite effect.

Figure 15 depicts the relationship among the continuous set of tasks, Where A and C are the technologically automatable tasks, B are performed by low-skill labor, and D is performed by high-skill labor. The technologically feasible combinations of factors to produce different tasks are given by:

$$y(i) = \begin{cases} \gamma_H(i)h(i) + \gamma_L l(i) + k(i) & if \quad i \;\epsilon\; [0, I_L] \\ \gamma_H(i)h(i) + \gamma_L l(i) & if \quad i \;\epsilon\; [I_L, J] \\ \gamma_H(i)h(i) + \gamma_L l(i) + \gamma_K k(i) & if \quad i \;\epsilon\; [J, I_H] \\ \gamma_H(i)h(i) + \gamma_L l(i) & if \quad i \;\epsilon\; [I_H, 1] \end{cases} \tag{6}$$

$h(i)$, $l(i)$, and $k(i)$ denote the total quantities of high-skill labor, low-skill labor, and capital utilized in the production of task i, respectively, and Γ is the effective share. In "short-run equilibrium" is defined by factor prices—wages and a capital rental rate—of high-skill labor, low-skill labor, and capital, W_H, W_L, and R, respectively—such that final-good producers minimize costs and clear the markets of the three factors. Differentiating the equilibrium equation, the model obtains:

$$\begin{cases} \sigma \frac{dW_H}{W_H} = \frac{d\Gamma_H}{d\Gamma_H} + \frac{dY}{Y}, \\ \sigma \frac{dW_L}{W_L} = \frac{d\Gamma_L}{d\Gamma_L} + \frac{dY}{Y}, \\ \sigma \frac{dR}{R} = \frac{d\Gamma_K}{d\Gamma_R} + \frac{dY}{Y}. \end{cases} \tag{7}$$

The first part of the differentiation is the displacement effect which is a negative effect, and the second part is the productivity effect which is a positive effect. The ratio of high-skill to low-skill wages refers to wage inequality $\omega{=}\frac{W_H}{W_L}$.

Low-skill automation increases wage inequality.

$$\frac{dw}{wdI_L} = \begin{cases} \frac{\gamma_L(I_L)^{\sigma-1}}{\Gamma_L} > 0 & if\ M \geq\ J, \\ \frac{\gamma_L(I_L)^{\sigma-1}}{\Gamma_L} \frac{\sigma\epsilon}{\sigma\epsilon+(\gamma_H(M)^{\sigma-1}/\Gamma_H)+(\gamma_H(L)^{\sigma-1}/\Gamma_L)} > 0 & if\ M <\ J, \end{cases} \tag{8}$$

High-skill automation decrease wage inequality.

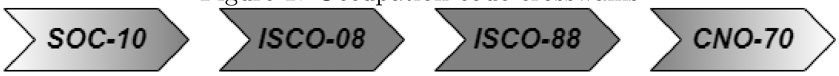
$$\frac{dw}{wdI_H} = \begin{cases} \frac{\gamma_L(I_H)^{\sigma-1}}{\Gamma_H} < 0 & if\ M >\ J, \\ \frac{\gamma_L(I_H)^{\sigma-1}}{\Gamma_H} \frac{\sigma\epsilon}{\sigma\epsilon+(\gamma_H(M)^{\sigma-1}/\Gamma_H)+(\gamma_H(L)^{\sigma-1}/\Gamma_L)} < 0 & if\ M \leq\ J, \end{cases} \tag{9}$$

Intuitively, when tasks reallocate from a factor, the displacement effect matters because the price of that factor falls dramatically; this is because such displacement forces more of that factor to work in the remaining tasks, resulting in a downward sloping demand for these tasks. The productivity effect is caused by the fact that automation involves substituting cheaper capital for labor (and we know that capital has to be cheaper, otherwise, companies would not have used capital instead of labor). Such substitution helps boost the economy's productivity and output. Because tasks are q-complements in the final good's production, an increase in output raises demand for all tasks and thus the price of all factors.

Low-skill automation, for example, reduces the share of tasks performed by low-skill labor, while high-skill automation reduces the share of tasks performed by high-skill labor. Combining the displacement and productivity effects generates the direct impact of automation on wages. In general, because these two effects are complete opposites, we cannot determine the impact of automation on all factor prices with certainty. However, the authors emphasize the gap between the effective cost of production by capital and labor inputs to characterize when one effect dominates; because the capital price (rental rate) rises when capital is scarce, this leads to a comparison in terms of the level of capital stock in the economy.

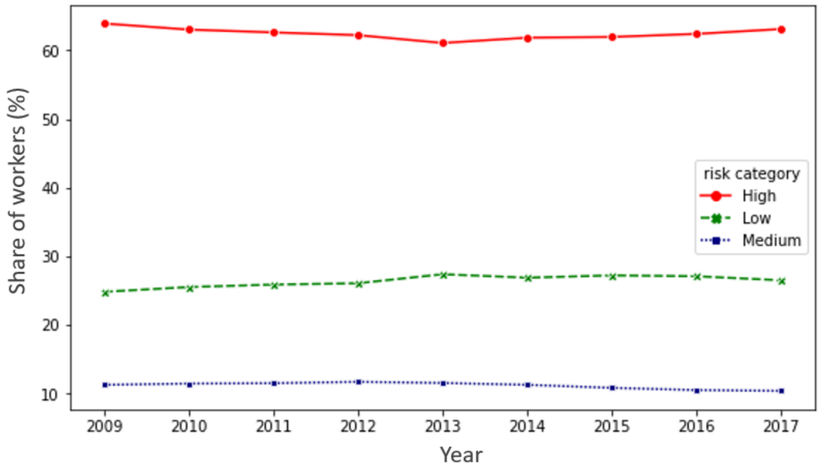
D Figures and Tables

Figure 1: Occupation code crosswalks



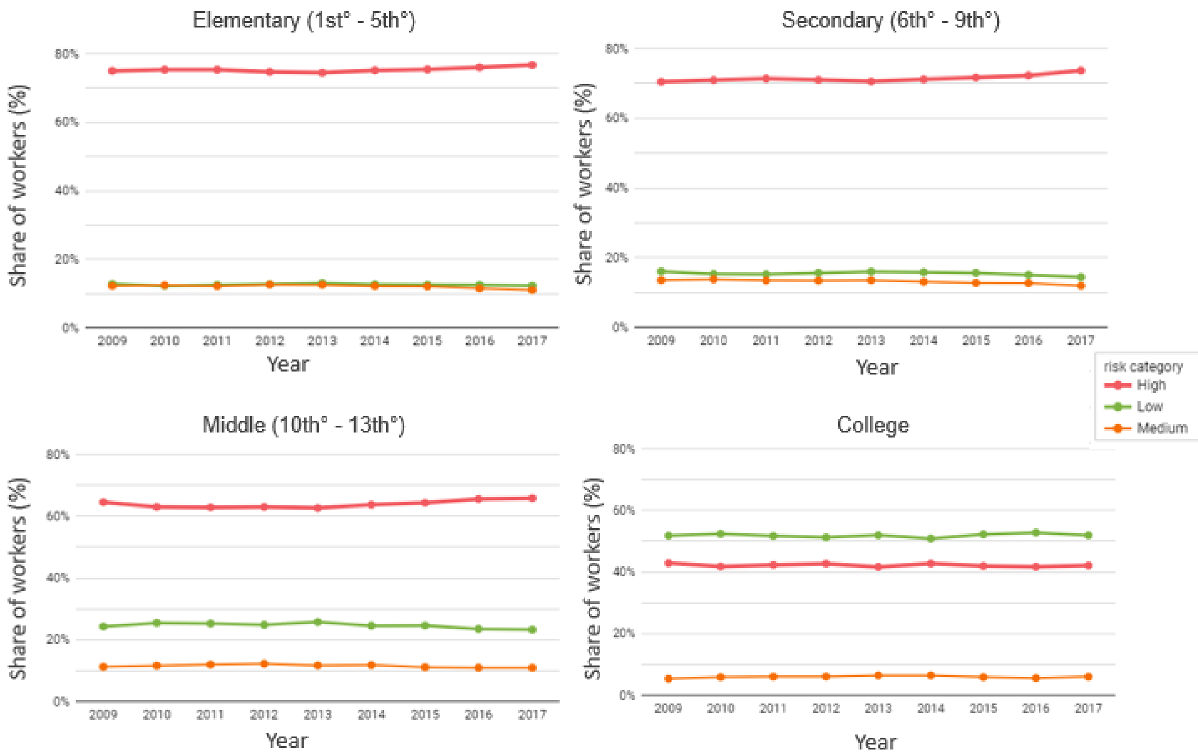
Note: Author's elaboration.

Figure 2: Share of workers by risk category



Note: The figure displays the share of workers by automation risk category for 2009-2017. Risk categories consider the results of Frey & Osborne (2017) to compute the probability of automation by occupations. The occupations are classified as low-risk (lower than 40%), medium-risk (40-70%), and high-risk (>70%).
Source: Author's calculations based on the Great Household Integrated Survey (GEIH) and the crosswalk by occupation with Frey & Osborne's results.

Figure 3: Share of workers in each risk category by education level

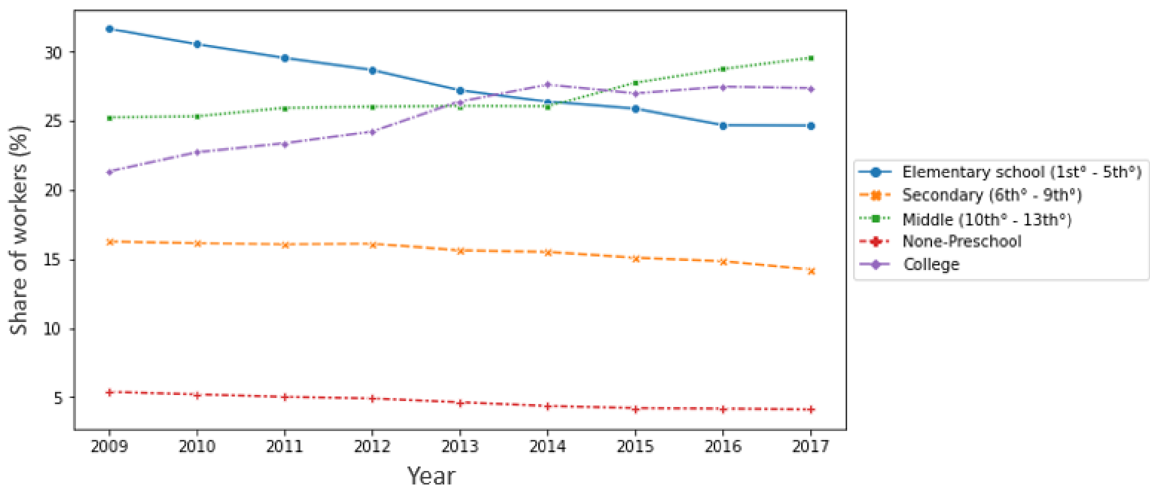


Note: The figure displays the share of workers in each automation risk category by education level. Source: Author's calculations based on the Great Household Integrated Survey (GEIH) and the crosswalk by occupation with Frey & Osborne's results.

Figure 4: Share of workers in each risk category by economic sector

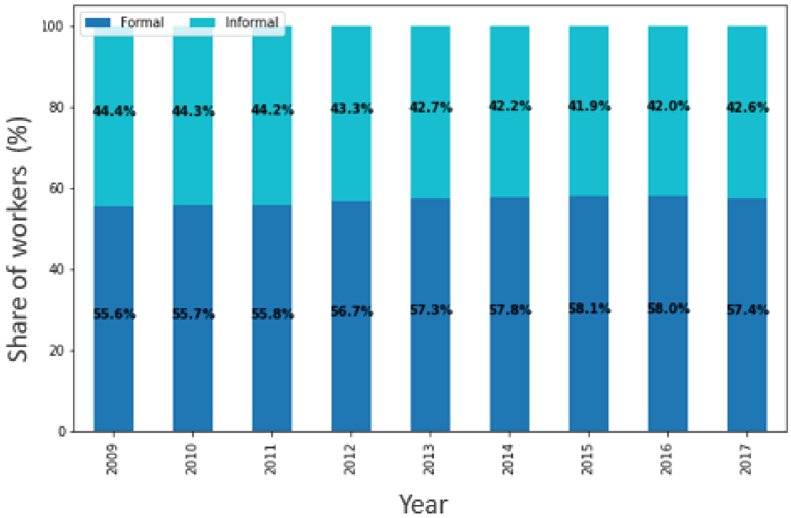
Note: The left panel shows the distribution of automation risk by sector in 2009, and the right shows the distribution in 2017. The classification of sectors (CIIU) is at a two-digit level. Source: Author's calculations based on the Great Household Integrated Survey (GEIH) and the crosswalk by occupation with Frey & Osborne's results. Agriculture, hunting, forestry, and manufacturing industries have a large concentration of employment in the high-risk category because the structure of their jobs includes a set of tasks sensitive to computerization (i.e., Farmers, carpenters, tailors, and dressmakers).

Figure 5: Share of workers by education level



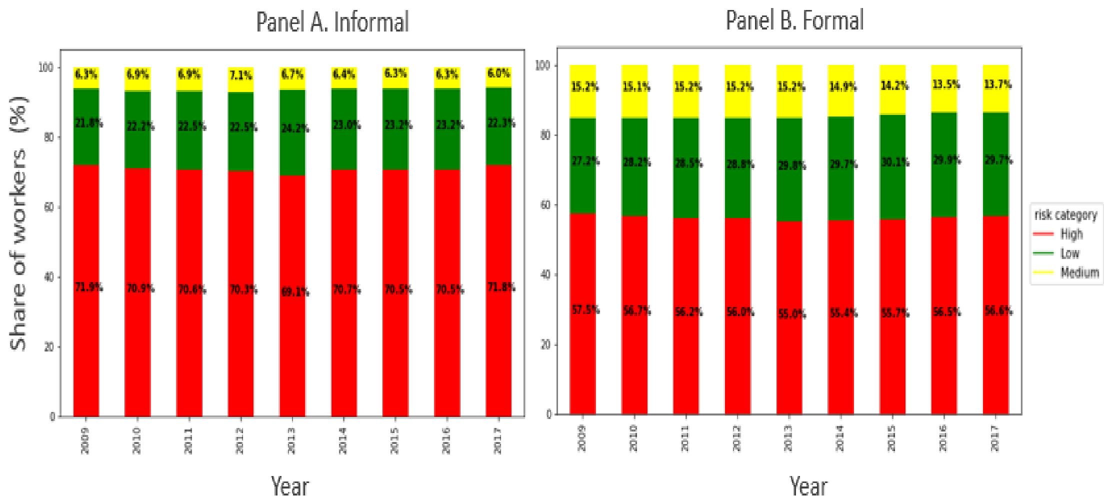
Note: The figure shows the percentage participation of workers by education level. Source: Author's calculations based on the Great Household Integrated Survey (GEIH).

Figure 6: Distribution of Formality in Colombia



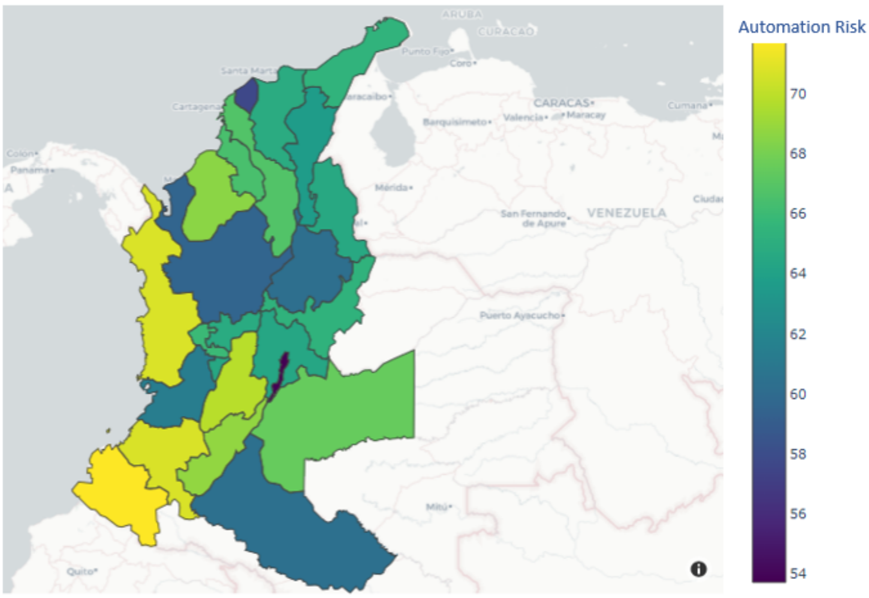
Note: The figure depicts Colombia's informal and formal sector distribution between 2009 and 2017. The formal sector has significant participation; however, there is also high participation of the informal sector in the Colombian labor market. Source: Author's calculations based on the Great Household Integrated Survey (GEIH).

Figure 7: Risk category of automation by Formality



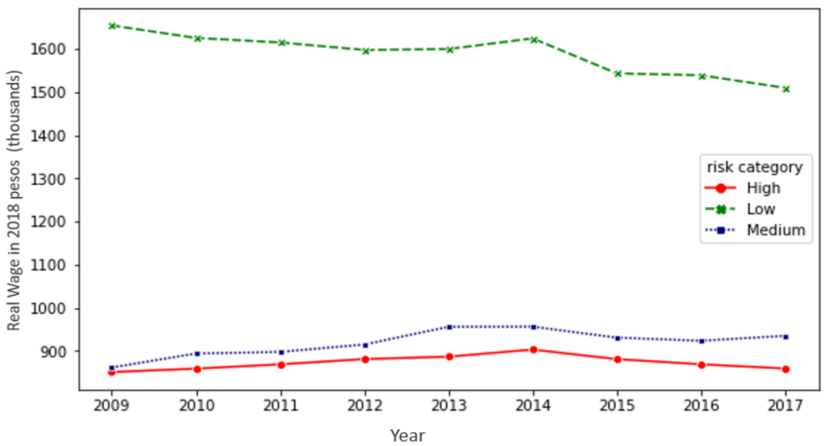
Note: The figure shows the participation of each risk category in the formal and informal sectors. We highlight that the high-risk category in both sectors is the most relevant. Source: Author's calculations based on the Great Household Integrated Survey (GEIH) and the crosswalk by occupation with Frey & Osborne's results.

Figure 8: Risk category of automation by Department (%)



Note: The figure shows the mean risk of automation, considering the distribution of occupations by sectors in each department reported in the GEIH (23 departments). The colors indicate the automation risk level, and the distribution has the information for 2017. Source: Author's calculations based on the Great Household Integrated Survey (GEIH) and the crosswalk by occupation with Frey & Osborne's results.

Figure 9: Real wage in 2018 Colombian pesos per risk category

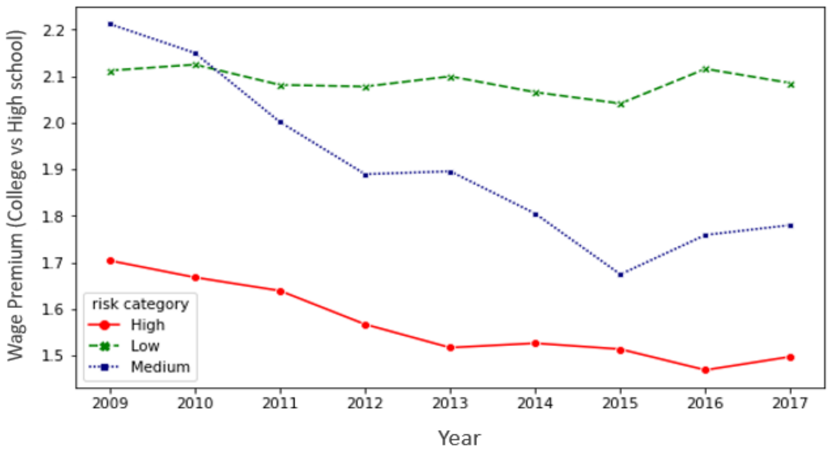


Note: Real wages in 2018 Colombian prices by automation risk category. Source: Author's calculations based on the Great Household Integrated Survey (GEIH) and the crosswalk by occupation with Frey & Osborne's results.

Figure 10: Real wage in 2018 Colombian pesos of Middle and College education level by risk category.

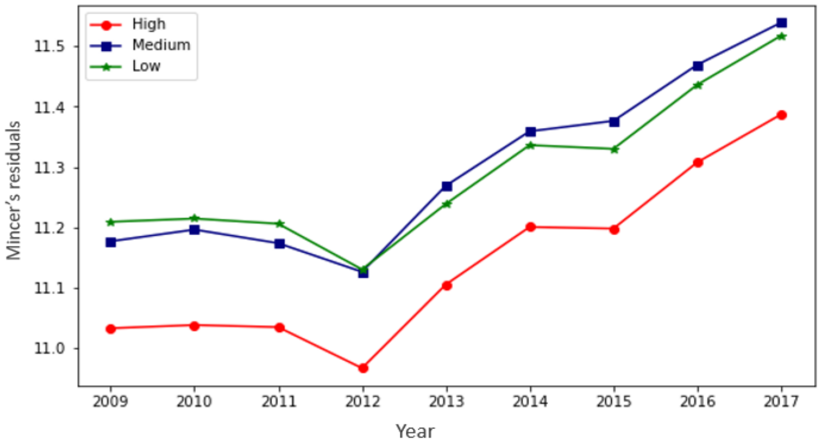
Note: Evolution of the real wage in 2018 Colombian prices by risk category and education level. Middle (Panel A) and College (Panel B). The figure shows the inequality in the real wage from both education levels, specifically in the low risk of automation category. Source: Author's calculations based on the Great Household Integrated Survey (GEIH) and the crosswalk by occupation with Frey & Osborne's results.

Figure 11: Wage premium (College/Middle school) by automation risk category



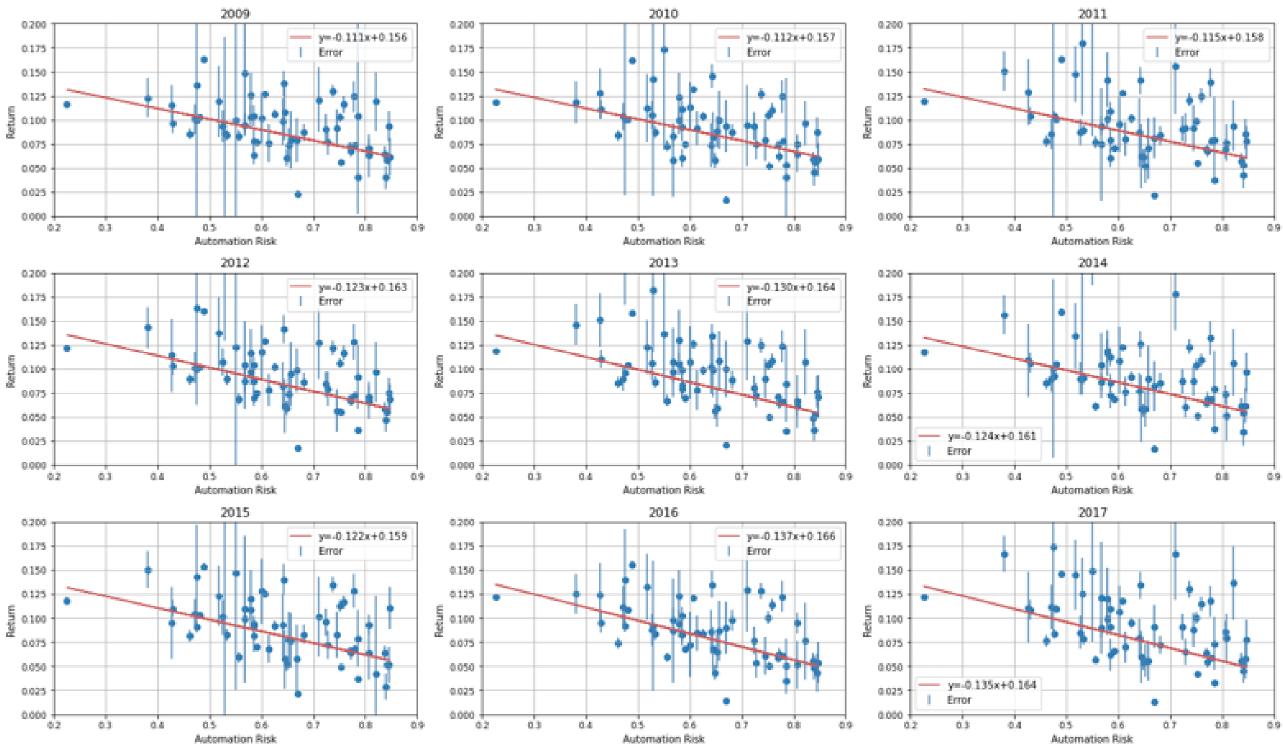
Note: In order to see the wage gap across risk categories of automation, we compute the wage premium between the wage of workers in College and High school education levels. The figure shows the persistent inequality in the low-risk category; however, the medium-risk category is important because it considerably decreases the wage premium. Source: Author's calculations based on the Great Household Integrated Survey (GEIH) and the crosswalk by occupation with Frey & Osborne's results.

Figure 12: Mincer's residuals by category



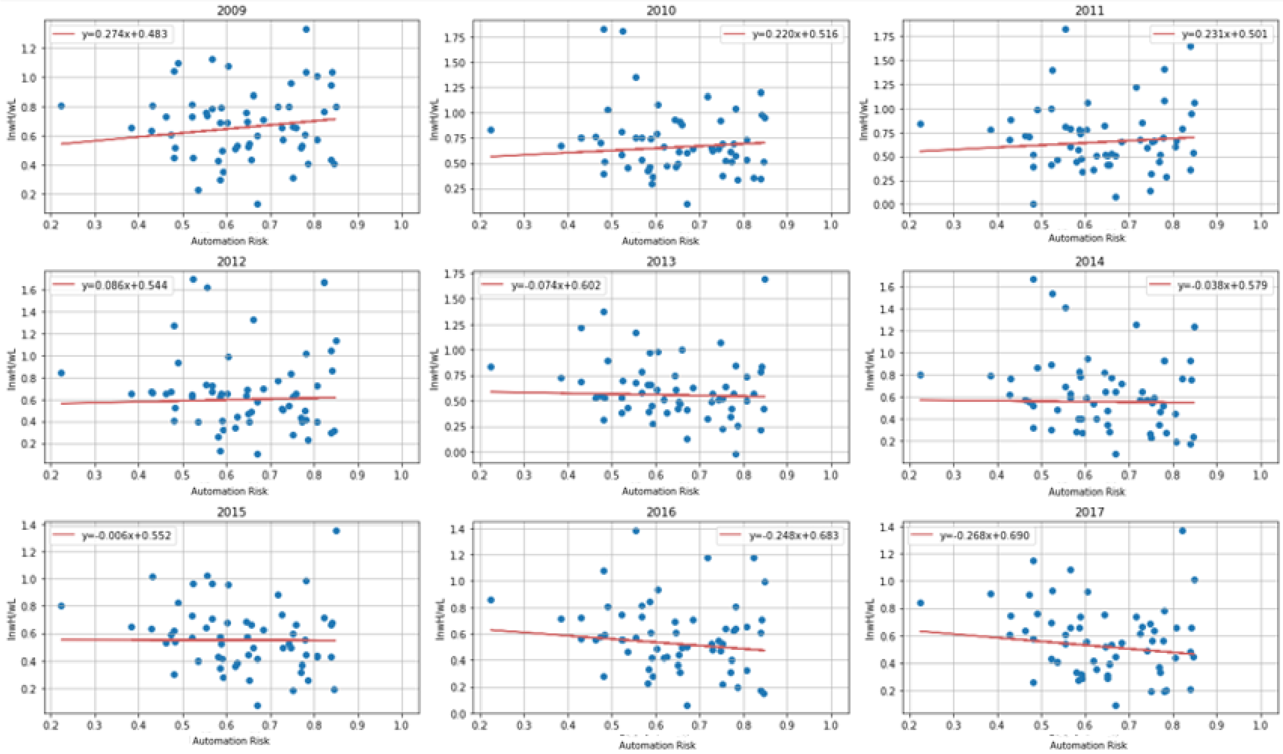
Note: We compute a Mincer equation to adjust the real wages by observable, considering educational splines. We relate the residuals with the probability of automation to obtain the real wage adjusted to each risk category. Source: Author's calculations based on the results of the Mincer equation.

Figure 13: Relationship between Educational Returns and Automation Risk



Note: The relationship between educational returns and automation risk is negative across the reference period. The observations are the economic sectors in Colombia, the vertical lines are the standard error of each sector. It shows the linear regression between return and automation risk. We test the hypnotize of the negative relationship and constant over 2009 to 2017 of the educational returns and probability of automation. We applied bootstrap to see statistical significance of the change. Source: Author's calculations based on the returns of the Mincer equation and the probability of automation computed by sectors.

Figure 14: Relationship between Wage premium (College/Middle school) and Automation Risk



Note: The relationship between wage premium (College/Middle school) and automation risk, which begin with a positive correlation in the last years, takes a negative correlation across the reference period. The observations are the economic sectors in Colombia. It shows the linear regression between wage inequality and automation risk. Source: Author's calculations based on the returns of the Mincer equation and the probability of automation computed by sectors.

Figure 15: Relationship among tasks

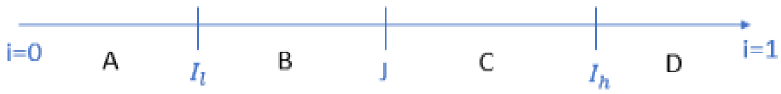


Table 1: Descriptive statistics

Variable	Mean	Std. Deviation
Years of education	8.94	4.81
Age	38.89	13.84
Automation Prob.	65.93%	30.77%
Informal	43.01%	49.50%
Real Wage in 2018 pesos	$1,015,396	$1,831,960
Panel A: Education level (Share %)		
College	25.41	
Middle (10o - 13o)	26.82	
Secondary (6o - 9o)	15.52	
Elementary school (1o - 5o)	27.59	
None-Preschool	4.62	
Panel B: Automation Risk (Share %)		
High	62.42	
Medium	11.16	
Low	26.41	
Panel C: Sectors (Share %)		
Wholesale & Retail	20.7	
Agriculture	16.69	
Manufacturing	12.4	
Transportation	8.25	
Hotels & Restaurants	6.02	
Construction	5.79	
Education	4.19	
Mining	1.02	
Observations	2,780,565	

Note: Author's calculations based on Great Household Integrated Survey (GEIH) from 2009 to 2017. The variables related to education level and automation risk. Statistics use the survey weights by household. Informality takes into account the definition of DANE. Automation risk made with a series of crosswalks by occupation, based on the results of Frey & Osborne (2017). Automation categories are divided as low-risk (less than 40%), medium-risk (40-70%), and high-risk (>70%)

Table 2: Descriptive statistics by skill and firm size

Panel A. High Skill				
	Less than 50 employees		More than 50 employees	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
ICT Exposure	-0.0102	0.863	-0.036	1.156
Employment	0.0586	0.292	0.034	0.217
Wage	0.1594	1.364	0.104	0.368
Female share	0.0009	0.1	0.001	0.082
Age 20-40	0.0018	0.097	0.002	0.086
Age 40-60	-0.0028	0.073	-0.003	0.081
Obs	1594		1612	
Panel B. Low Skill				
	Less than 50 employees		More than 50 employees	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
ICT Exposure	-0.049	1.079	-0.07	1.245
Employment	0.035	0.324	0.021	0.177
Wage	0.082	0.311	0.082	0.269
Female share	0.001	0.11	-0.0002	0.059
Age 20-40	-0.007	0.125	-0.004	0.067
Age 40-60	0.006	0.12	0.004	0.064
Obs	1492		1584	
Panel C. Total Employment				
	Less than 50 employees		More than 50 employees	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
ICT Exposure	-0.02	0.456	-0.006	0.461
Employment	0.043	0.05	0.007	0.111
Wage	0.077	1.925	0.02	0.289
Female share	0.001	0.006	0.001	0.033
Age 20-40	-0.0007	0.005	0.001	0.035
Age 40-60	-0.0008	0.004	0.0003	0.036
Obs	1594		1638	

Note: Author's calculations based on Great Household Integrated Survey (GEIH). The variables related to education level and automation risk. The growth of the variables are measured as (100×) log difference. All socio-demographic, exposure, employment, and wage growth are computed from 2009 to 2017. The panels divides the population into high-skill (with college) or low-skill workers (with middle school).

Table 3: Effect of automation on employment and wage growth

Panel A. Employment growth						
	Less than 50 employees			More than 50 employees		
Variable	High Skill	Low Skill	Total	High Skill	Low Skill	Total
$\Delta ICT Exposure$	-0.00486	0.00537	-0.00263	0.0156***	0.00914**	0.0154**
	(0.00938)	(0.0131)	(0.0133)	(0.00542)	(0.00440)	(0.00625)
R^2	0.308	0.0573	0.169	0.0540	0.0449	0.00419
Observations	1594	1492	1594	1612	1584	1638
Control	✓	✓	✓	✓	✓	✓
Time-State effect	✓	✓	✓	✓	✓	✓

Panel B. Wage growth						
	Less than 50 employees			More than 50 employees		
Variable	High Skill	Low Skill	Total (Wage Premium)	High Skill	Low Skill	Total (Wage Premium)
$\Delta ICT Exposure$	0.0119	0.00516	-0.00329	-0.0276**	-0.00236	-0.0239
	(0.0118)	(0.0229)	(0.0157)	(0.0112)	(0.00322)	(0.0238)
R^2	0.0125	0.673	0.0136	0.740	0.773	0.0263
Observations	1594	1492	1594	1612	1584	1638
Control	✓	✓	✓	✓	✓	✓
Time-State effect	✓	✓	✓	✓	✓	✓

Note: Author's calculations.All regressions include a constant, and time fixed effect. P-values wild cluster bootstrap are obtained by implementing wild cluster bootstrap with 1000 repetitions by city-ciiu4 cell: (the STATA command used is cgmwildboot) Change between 2009 to 2017.*p<0.1; **p<0.05; ***p<0.01

Table 4: Effect of automation on labor supply growth

Variable	High Skill Labor Supply	Low Skill Labor Supply
$\Delta ICTExposure$	0.00826	0.0125**
	(0.00799)	(0.00573)
R^2	0.241	0.0401
Observations	1638	1638
Control	✓	✓
Time-State effect	✓	✓

Note: Author's calculations.All regressions include a constant, and time fixed effect. P-values wild cluster bootstrap are obtained by implementing wild cluster bootstrap with 1000 repetitions by city-ciiu4 cell: (the STATA command used is cgmwildboot) Change between 2009 to 2017.*p<0.1; **p<0.05; ***p<0.01