

U-shaped wage curve and the Internet: *The Colombian case** *Internet y curva salarial en forma de U: El caso colombiano*

DIEGO A. MARTIN**

Abstract

While there is a broad consensus in the literature that there is a positive correlation between Internet usage and labor income in the richest countries, this link has not been proven in the developing world. This paper uses propensity score matching techniques and household survey data to estimate the effect of the Internet on wages in Colombia, a country that has experienced a relatively rapid diffusion of information and communications technology in recent years. The empirical results confirm that there is a positive and statistically significant relationship between Internet use and income in this country. Consistent with evidence gathered on developed countries in previous studies, the empirical results also suggest that workers in the middle of the skill distribution receive the lowest wage premium for using the Internet. However, contrary to most evidence from developed countries, low-skilled workers in Colombia enjoy the highest wage premium from Internet use, which illustrates the potential for new technologies to address inequality gaps between occupations.

Key words: Propensity score matching, skills, wage differentials, internet, income distribution.

JEL Classification: C14, J24, J31, L86, O15.

Resumen

Si bien existe un amplio consenso en la literatura acerca de la existencia de una correlación positiva entre el uso de Internet y el ingreso laboral en los países

* I thank the seminary of Universidad del Rosario and the anonymous peer reviewers for all the useful comments. I am also very grateful to Juan Vargas for his guidance in this document, which is my master's thesis for the Universidad del Rosario.

** CAF-Development Bank of Latin America, Research Department. E-mail: dmartin@caf.com

más ricos, este vínculo no se ha demostrado en el mundo en desarrollo. Este documento utiliza propensity score matching y datos de encuestas de hogares para estimar el efecto de Internet sobre los salarios en Colombia, un país que ha experimentado una difusión relativamente rápida de la tecnología de la información y la comunicación en los últimos años. Los resultados empíricos confirman que existe una relación positiva y estadísticamente significativa entre el uso de Internet y los ingresos en este país. De acuerdo con la evidencia reunida en países desarrollados en estudios previos, los resultados empíricos también sugieren que los trabajadores en el medio de la distribución de habilidades reciben la prima salarial más baja por usar Internet. Sin embargo, contrariamente a la evidencia de los países desarrollados, los trabajadores poco calificados en Colombia disfrutaban de la prima salarial más alta del uso de Internet, lo que ilustra el potencial de las nuevas tecnologías para abordar las brechas de desigualdad entre las ocupaciones.

Palabras clave: Propensity score matching, habilidades, diferencia salarial, internet, distribución salarial.

Clasificación JEL: C14, J24, J31, L86, O1.

1. INTRODUCTION

Do information and communications technologies (ICTs) impact productivity? This question has been extensively studied in the developed world since the introduction of the first computing technology in the 1950s. In the United States, for example, a large body of evidence demonstrates the positive effect of ICTs on economic dimensions such as wages, supply and demand of factors (Katz & Murphy, 1992; Krueger, 1993; Acemoglu, 1999). These patterns are also found in other developed countries, where ICTs have been found to contribute 0.2-0.9 percentage points per year to economic growth (Colecchia & Schreyer, 2002). The *canonical model* presents a framework to explain why ICTs impact economic variables. This model incorporates supply and demand for different skills, which are imperfect substitutes but produce similar goods. Technology can complement either high- or low-skilled workers, which in turn produces an increase in the demand for one group or the other (Katz & Murphy, 1992). The idea of a direct link between a specific type of worker and ICTs implies that technology has a heterogeneous impact on the labor market. According to this model, technology has simplified routine tasks and solved complicated problems, creating a specific wage distribution curve in which highly skilled workers are at the top and middle-skilled workers are at the bottom in the developed world (Autor *et al.*, 2003; Michaels *et al.*, 2014). Although this model is very useful for calculations and explanations, Acemoglu and Autor (2011) developed an improved version to demonstrate that in rich countries technology has depressed the real wages of low-skilled workers, created non-monotone

changes in the earnings distribution, and reduced the need to hire middle-skilled workers, among other effects.

The heterogeneous effect found in the wage distribution can also be observed in social relationships, including personal and political behavior, in which ICTs follow a model of “the rich become richer”, which means that technology exacerbates pre-existing inequalities, and the exact effects depend mainly on individual personalities. For example, extroverts using the Internet have better results in social involvement than introverts (Kraut *et al.*, 2002). However, other studies suggest the Internet is a tool that can help close the gap between the least and most advantaged populations (Steinfeld *et al.*, 2008). This technology can even build a bridge of communication between politicians and their electorates to reduce, at least to some degree, the asymmetry of information between government actions and what people need to improve their living conditions (Garcia-Murillo, 2013). The public sector can use ICTs to develop policies to reduce gaps along different dimensions, particularly in rural areas.

Do ICTs have a similar effect in developing countries, especially in Latin America? This relationship has been less frequently explored in this region for three main reasons. First, there is a lag in the absorption of ICTs in the developing world. Europe and North America, for example, have around 84 Internet users per 100 people, while South America has only 51 (World Bank, 2015). Second, there are no specialized databases that contain common information between countries, which is needed in order to produce standardized indicators to compare within the region. Third, estimating the impact of ICTs on different dimensions requires smart approaches because non-random access complicates the process of conducting an experiment. Nonetheless, the effect of ICTs in Latin America is a highly relevant question since they can be a powerful tool to improve living conditions and overcome poverty traps.

Colombia offers a unique opportunity to analyze how ICTs affect different dimensions of the labor market of a developing country that has experienced a recent growth in the number of Internet users because it has a database, which has never been used before, that makes it possible to present heterogeneous effects depending on workers' skill levels. It is also possible to test whether this country follows the prediction of the *canonical model*, and its expansion as carried out by Acemoglu and Autor (2011), or whether it is necessary to develop another model for developing countries.

The following three questions focus on the effect of the Internet in Colombia's wage distribution curve. First, is the increased efficiency in some work tasks generated by ICTs valued and compensated for in the labor market? A sophisticated improvement using ICTs, such as computers programming by themselves, is not necessary; whenever workers can either communicate easier or access information faster, they may receive a salary increase. Indeed, it is also important to analyze how this effect changes over time, since there is evidence that the wage premium has decreased in the United States from 2000 to 2001 (DiMaggio & Bonikowski, 2008). Second, given that people are more likely to use the Internet at work in productive activities than at home, is there a higher

wage premium for using ICTs in the workplace? Third, the Internet offers different kinds of applications for education, communication and entertainment, but can the labor market differentiate workers' efforts, at least imperfectly, in productive activities? For example, do employees using the Internet for education receive a higher reward than those who use it for entertainment purposes? Finally, there is considerable evidence from the developed world that technology has polarized the labor market into low-skilled workers with low salaries and high-skilled workers with high salaries (Acemoglu & Autor, 2011; Autor, 2015). To explore whether a developing country such as Colombia follows this pattern, I analyze the heterogeneous effects of ICTs on workers' income levels along three dimensions: places of access, activities on the web and workers' skills. This empirical support presents important ideas about how developing countries, especially those that are the most socioeconomically disadvantaged, can obtain the greatest benefits from using the Internet.

This paper uses Propensity score matching (PSM) techniques and non-parametric models to estimate the effects of the Internet on labor income. Although this model has some drawbacks, including the requirement to have individuals with the same probability of using the Internet in the treated and control groups (a technique known as a common support), the results must not change depending on the exclusion or inclusion of controls, and they must be robust to changing the matching method used (Abadie & Imbens, 2006). The paper presents several tests to demonstrate that these issues are overcome, and that it is possible to reliably estimate average treatment effects on the treated population.

The results address the hypotheses individually and demonstrate a positive and significant correlation between using the Internet and labor income. On average, Internet users earn 6.5% of one standard deviation (SD) more than non-users. Using this technology at work increases wages 11% more than using it at home, and users in productive activities earn 7.1% of one SD more than those who use the Internet for other applications. The wage premium depends on the tasks performed in a specific occupation, which proxies for skill level. Workers in the middle of the skill distribution in Colombia have benefited the least from using the Internet, which parallels trends in the developed world. Nonetheless, and contrary to the evidence gathered from the richest countries, the highest wage premium is found in the lowest tail of the distribution, where farmers and miners using the Internet gain 20.7% of one SD more than non-users in the same occupation. I find that the *canonical model* is applicable to the Colombian case, while Acemoglu and Autor's (2011) extension of the model partially applies in this situation given the fact that the salary distribution curve has a u-shaped form.

In the developing world, the relationship between ICTs, income, education and other variables has been explored extensively. To the best of my knowledge, this study is the first to use household surveys, non-parametric models and fixed effects of firm size, occupation and economic sector to analyze the Internet's impact on the wage distribution in Colombia. Therefore, it presents

new evidence and methodology related to Colombia, a benchmark for other developing countries, and provides corresponding technical support to justify investments in ICTs.

The remainder of the paper is structured as follows. Section 2 reviews the literature on the mechanism of the impact of new technologies and earnings. Section 3 describes the basic statistics of Internet use around the world as well as the descriptive statistics related to the hypotheses. It also presents the empirical strategy, which uses non-parametric models. Section 4 describes the results of tests of the matching model and addresses each of the hypotheses individually. Finally, Section 5 discusses the results and how the Colombian case is related to the international literature.

2. NEW TECHNOLOGIES AND EARNINGS

Many studies have explored the relationship between technology and wages, mainly in the developed world. Since the start of the computing era, there has been a rise in the demand for college graduates who can use this new technology, which creates a wage premium that favors technology users over the long term. This is also known as the *canonical model*, which has empirically explained the evolution of the skill premium generated by computers (Acemoglu & Autor, 2011). Computers can replace workers in some repetitive manual tasks, and help them carry out non-routine tasks such as communication and problem solving (Card & DiNardo, 2002; Autor *et al.*, 2003). The variations in wages are not uniform across skill levels, because the lower premium is at the middle of the distribution rather than the bottom, and the relationship between wages and skills has a u-shaped distribution. Indeed, the automation generated by technology reduces the demand for middle-skilled workers more than for lower-skilled workers (Autor & Dorn, 2013). This pattern is also found in 16 countries in Europe (Goos *et al.*, 2009) and the United Kingdom (Acemoglu, 1999; Goos & Manning, 2007).

The current labor market mainly requires the use of ICTs to access quality jobs because online skills help people perform their work more efficiently. Workers of relatively high ability are more likely to move to non-routine cognitive occupations, while people with low abilities tend to stay in routine jobs with a lower potential for salary growth (Cortes *et al.*, 2016). The Internet also makes it easier for workers to move to better jobs, because the demand for labor is concentrated in urban areas (McDonald & Crew Jr, 2006). In the United States, workers who use the the Internet earn 13.5% more than non-users. This wage premium is even higher in industries that are less intensive in technology, in which the only people who use Internet are those who can perform the most difficult tasks (Goss & Phillips, 2002). Using panel data, DiMaggio and Bonikowski (2008) also found a 20% increase in the hourly wages of workers for using Internet, not just computers. Bartel *et al.* (2007) found that companies encourage workers to undertake training in computers when they invest in ICTs.

Gust and Marquez (2004), for example, show that the more flexible regulation in information technology in the United States partly explains its higher productivity than some countries in the OECD, such as Germany, France, Italy and the United Kingdom. Paunov and Rollo (2016) emphasize that using the Internet also increases productivity in developing countries, and find that firms with the most sophisticated productivity experience the greatest gains from this technology.

The massive shift within the media to online channels has made the web a center of culture, education, politics and personal relationships, which has made the Internet as important in the labor market as in daily life. This technology can help reduce individuals' sense of isolation, social exclusion and decrease the impact of physical disabilities, which also affect productivity (Foley, 2004; Chigona *et al.*, 2009; Dobransky & Hargittai, 2016). Big changes in technology are not required in order to produce significant effects on the quality of life. Blanco and Vargas (2014), for instance, show that sending messages to vulnerable populations explaining their rights and where they can ask for them increases access to public aid and their welfare. Thus, this study explores the uses of technology in the developing world, which can inform policies and projects designed to increase the use of the Internet.

3. DATA AND EMPIRICAL STRATEGY

There is generally a positive correlation between Internet users and income around the world (Katz & Murphy, 1992; Colecchia & Schreyer, 2002; S.-Y. T. Lee *et al.*, 2005). Figure A-1 shows that an increase in the number of Internet users is positively correlated with the logarithm of per capita GDP. However, there is heterogeneity between regions. The developed world is at the top right of the figure with high levels of income and Internet users, while Latin America is in the middle of the figure with around 45 Internet users per 100 people and lower income levels. Whereas in 1994 only 0.10% of Colombians used the Internet, by 2016 this had increased to 58%, representing an average growth of 22% per year over the last two decades (World Bank, 2015). The United States, a benchmark in the developed world, surpassed Colombia's current Internet penetration rate in 2002, and in the last year it reached 76% (World Bank, 2015). However, the average growth in the United States over the last 20 years has only been 8% per year, a much lower rate than in Colombia (World Bank, 2015)¹.

To analyze the hypotheses in Colombia, this study uses the Great Integrated Household Survey (GEIH, acronym in Spanish) conducted yearly from 2009 to 2011. This is a nationally representative cross-section survey and it contains demographic and industry data at the industry level. The dependent variable is standardized monthly labor income, and the sample is restricted to workers aged

¹ The Colombian government invests 0.8% of its total budget in ICT projects; this sector ranks 15 out of 30 with the highest investment (MinHacienda, 2014).

18 to 65. This sub-sample is used in order to be able to compare the Colombian case with other countries with different levels of income, presented by DiMaggio and Bonikowski (2008) and Benavente *et al.*, (2011). The results are robust to using the full population. Only this wave included questions on whether respondents had used the Internet in the last 12 months and where they accessed the Internet². The survey also asked respondents whether they use the Internet for education or finance, which I used to analyze productive activities³. Finally, I used responses related to the frequency of Internet use in my robustness checks⁴.

Controlling for economic activities, occupation and firm size is one of the most important aspects of this study because this information proxies for productivity, which is an important improvement compared with other studies. These variables reduce possible omitted variable bias because they create a cell, as small as possible, in which two workers have the same ability, but one uses the Internet and the other does not. Economic activity is divided into 13 sectors using the Classification of All Economic Activities (ISIC, 2008): agriculture, mining, manufacturing, electricity, construction, transport, financial service, housing sector, public administration, education, health and domestic service. Occupations are classified into 10 categories using the Standard Occupational Classification System (SOCPC, 2010): professional specialty, executive and managerial, service, sales, machine operator, cleaners and laborers, professional products, transportation, farming and mining. Firm size comprises five categories depending on the number of workers. Each category of the variable is included as a fixed effect in all estimations. Therefore, these covariates indicate the productivity of each worker and compare Internet users and non-users in the same firm's economic activity (sector), in the activities that workers perform (occupation) and in the places they work (firm size).

Using this database, it is possible to build a general profile of Internet users. Table 1 presents the respondents' individual characteristics by location of access and productive activities on the web. It is clear that these aspects are not mutually exclusive, because an individual can use the Internet anywhere. On average, Internet users are women, younger, better educated, well paid, and have more experience with technology than non-users. Around 40% of people in both groups are homeowners, and thus enjoy a similar standard of living. Both groups have similar locations of access and activities in the Internet. These characteristics

² The options were accessing the Internet at home, at work, in educational institutions, in free public access centers, in paid access centers, in the house of another person.

³ The original question was worded "For which of the following services or activities, do you use the Internet?" The options are obtaining information, communication, electronic banking and other financial services, education and learning, transactions with government agencies, entertainment.

⁴ The original question was worded "How often do you use Internet?" The options are at least once a day; at least once a week, but not every day; at least once a month, but not every week and less than once a month.

TABLE 1
SOCIO-DEMOGRAPHIC STATISTICS BY PLACE OF INTERNET USE

	Non-users	Everywhere	At work	At home	Productive	
					non-users	users
Logarithm of income	13.276	13.632	13.839	13.836	13.588	13.699
Standardized income	-0.326	0.211	0.590	0.519	-0.152	0.301
<i>Panel A: Population</i>						
Schooling	11.812	13.797	14.545	14.513	13.614	14.468
Age	35.757	33.073	34.871	35.178	31.004	35.074
Women	0.479	0.525	0.562	0.524	0.481	0.603
<i>Panel B: Technology experience</i>						
Telephone at home	0.419	0.618	0.650	0.779	0.630	0.618
Satellite TV at home	0.653	0.808	0.837	0.907	0.814	0.799
PC at home	0.263	0.615	0.669	0.925	0.640	0.614
<i>Panel C: Proxy for wealth</i>						
Property owner	0.415	0.423	0.440	0.478	0.407	0.463

Notes: The average of each variable is presented in the table. The sample is restricted to employed workers aged 18 to 65. The standard deviation of the labor income is \$262 and the mean is \$295 for this sample. Internet users are workers accessing from any of the following places: at home, at work, in educational institutions, in free public access centers, in paid access centers, in the house of another person (relative, friend, neighbor).

Source: Colombian Great Integrated Household Survey (GEIH), 2009 to 2011.

show that it is possible to find similar individuals who use and do not use the Internet, which is an advantage of using a household survey.

Table 2 shows Internet users distributed by firm size, economic sector, occupation and type of worker. Whereas big companies have the highest number of Internet users (78% of workers), the number of productive users is smaller than non-productive users in all firm sizes. Small companies seem to have the poorest control over employees' activities on the web, because only 22% of workers use the Internet for education and financial activities in these companies.

Access to and activities on the web by occupation does not present a u-shaped distribution in Colombia. The financial services sector, which has the highest percentage of Internet users and average wage, has more than twice as many people using the Internet for education and financial services as the lower-income domestic service sector. Indeed, there are almost three white-collar workers for every two blue-collar workers using this technology. This does not represent evidence against a possible u-shape in the wage curve given the fact that the low supply of Internet users in a certain occupation can result in a wage premium depending on whether there is a large demand for workers with this skill in a particular sector.

The literature is divided over which way of estimating the relationship between new technologies and earnings is more effective: parametric models using

TABLE 2
INDUSTRIAL STATISTICS BY PLACE OF INTERNET USE

	Non-users	Everywhere	At work	At home	Productive	
					non-users	users
<i>Panel A: Firm size</i>						
101 and over employees	0.223	0.777	0.694	0.669	0.532	0.468
51-100 employees	0.278	0.722	0.625	0.583	0.574	0.426
11-50 employees	0.285	0.715	0.607	0.558	0.600	0.400
6-10 employees	0.342	0.658	0.489	0.460	0.654	0.346
2-5 employees	0.448	0.552	0.288	0.316	0.725	0.275
<i>Panel B: Economic sector</i>						
Financial service	0.132	0.868	0.840	0.799	0.409	0.591
Public administration	0.191	0.809	0.769	0.705	0.567	0.433
Education	0.210	0.790	0.705	0.711	0.445	0.555
Housing sector	0.255	0.745	0.676	0.608	0.564	0.436
Health	0.272	0.728	0.636	0.593	0.598	0.402
Electricity	0.311	0.689	0.610	0.552	0.611	0.389
Transport	0.321	0.679	0.576	0.503	0.647	0.353
Manufacturing	0.383	0.617	0.412	0.445	0.687	0.313
Commerce	0.397	0.603	0.392	0.393	0.698	0.302
Agriculture	0.422	0.578	0.394	0.382	0.708	0.292
Mining	0.429	0.571	0.408	0.481	0.663	0.337
Construction	0.441	0.559	0.360	0.353	0.718	0.282
Domestic Service	0.512	0.488	0.281	0.282	0.735	0.265
<i>Panel C: Occupation</i>						
Professional specialty	0.101	0.899	0.872	0.861	0.347	0.653
Administrative support	0.114	0.886	0.864	0.835	0.438	0.562
Executive and managerial	0.150	0.850	0.811	0.744	0.497	0.503
Sales	0.360	0.640	0.399	0.433	0.698	0.302
Service	0.371	0.629	0.437	0.431	0.707	0.293
Machine operator	0.426	0.574	0.230	0.374	0.756	0.244
Precision production	0.487	0.513	0.174	0.290	0.775	0.225
Transportation	0.488	0.512	0.176	0.313	0.786	0.214
Cleaners and gardeners	0.499	0.501	0.208	0.267	0.750	0.250
Farming and mining	0.589	0.411	0.144	0.175	0.802	0.198
White collar	0.118	0.882	0.853	0.824	0.415	0.585
Blue collar	0.400	0.600	0.378	0.392	0.708	0.292
<i>Panel D: Type of worker</i>						
Salaried worker	0.288	0.712	0.592	0.565	0.591	0.409

Notes: The average of each variable is presented in the table. The sample is restricted to employed workers aged 18 to 65. Economic sector is built using International Standard Industrial Classification of all Economic Activities (ISIC, 2008). Occupation is built using Standard Occupational Classification System (SOCPC, 2010). The standard deviation of the labor income is \$262 and the mean is \$295 for this sample. Internet users are workers accessing from any of the following places: at home, at work, in educational institutions, in free public access centers, in paid access centers, in the house of another person (relative, friend, neighbor).

Source: Colombian Great Integrated Household Survey (GEIH), 2009 to 2011.

ordinary least squares (OLS) or non-parametric models using PSM techniques. First, Krueger (1993) was the pioneer to find a wage premium for using ICTs. There is also evidence, using the same model, for countries such as the United States, United Kingdom and Australia (Goss & Phillips, 2002; Arabsheibani *et al.*, 2004; Chiswick & Miller, 2007). Second, Rosenbaum and Rubin (1983) use the PSM method to conduct causal inference, and many impact evaluation policies use this methodology (Heckman *et al.*, 1998; Heckman *et al.*, 1999). This approach is widely used to analyze household survey data related to ICTs and labor income (DiMaggio & Bonikowski, 2008; Navarro, 2010; Benavente *et al.*, 2011).

Parametric models estimate bias coefficients since they do not address two issues: omitted variables and the reflection problem. The associate coefficient for using the Internet may reflect the fact that some workers were more productive even before using the Internet (Entorf *et al.*, 1999). It is also possible that Internet connections are expensive, and only well-paid workers can afford access, which implies reverse causality. Indeed, DiNardo and Pischke (1997) doubt the empirical results in favor of using ICTs, particularly computers. Even though they find a similar wage premium for using computers in Germany and the United States in 1997, they argue that there is also a large wage premium for using calculators, telephones, pens and pencils. They claim that those results show the problem of selection bias, where workers' characteristics decide whether they have access to these tools. According to these authors, given the unlikelihood of devising an experiment in which Internet access is randomly assigned, the best way to estimate the effect of technology is to include education and fixed effects in the model. There is no evidence of a computer wage premium in countries such as Ecuador or Great Britain, where computer skills are not as important as math and reading skills (Oosterbeek & Ponce, 2011; Borghans & Ter Weel, 2004). S.-H. Lee and Kim (2004) only find a premium wage for Internet users in the United States for 1997; they find no statistically significant results for 1998 or 2000.

The first way to analyze the correlation between income and Internet use is to estimate equation 1 using the OLS method:

$$(1) \quad Y_{ifmt} = \alpha_0 + \theta Internet_{ifmt} + X_{ifmt} + \kappa_{imt} + \delta_j + \gamma_t + \mu_{ifmt}$$

where Y_{ifmt} is the monthly standardized labor income of individual i , working in a firm, economic activity and occupation f , in municipality m in year t . $Internet_{ifmt}$ is the treatment variable, equal to 1 if a worker uses the Internet, and 0 otherwise. X_{ifmt} are socio-demographic variables such as age, education, and experience with technology. κ_{imt} is a set of fixed effects that combines firms' size, economic sector and occupation⁵. These variables build a cell that compares Internet users and non-users controlling for workers' ability. This is a good proxy, because it shows where a person works and what she or he does there.

⁵ κ_{imt} can be also written as: $\kappa_{imt} = \kappa_{ismt} + \kappa_{iemt} + \kappa_{iomt}$, where κ_{ismt} is firm size fixed effect, κ_{iemt} is the economic sector fixed effect and κ_{iomt} is occupation fixed effect.

Meanwhile, δ_m is the municipality fixed effect, which controls for aggregate shocks at the municipal level such as geographic and weather conditions. It also controls for relevant variables related to characteristics of cities, such as the fact that big cities have a higher probability of accessing the Internet because they enjoy more services than small ones. γ_t is the year fixed effect, which controls for aggregate shock in time such as inflation or macroeconomic conditions of whole country. μ_{ijmt} is the error term.

How can we estimate the impact of ICTs on the labor market? It is unlikely to find a natural experiment in which ICT access is randomly assigned. A controlled experiment can even fail, since people assigned to not use the Internet can easily find different ways to access the web such as via smartphones. For example, the experimental evaluation of Colombia's Computer for Education program experienced problems with treatment group confidence, and randomly receiving computers at school was not found to impact students' test scores. This does not necessarily mean that computers are not useful for education; the authors argue that the results are explained by the failure to incorporate them into the educational process (Barrera-Osorio & Linden, 2009). In this context, PSM is a useful method of finding unbiased estimators when the treatment group is not randomly assigned. Intuitively, this non-parametric model allows us to find, for each Internet user (i.e., treated individual), an individual who is exactly the same, except they do not use the Internet. This model assumes that the treatment is determined exclusively by observable variables of the individuals, and that gaps in unobservable variables are closed at the same time as gaps in observable variables, which is called conditional independence (Angrist & Pischke, 2009). How does PSM work in comparison to an experiment? Dehejia and Wahba (2002) consider causal inference and sample selection bias in non-experimental cases in which only some individuals in the treatment and control groups are comparable, and building a sub-sample of individuals who share pre-treatment characteristics is difficult. They replicate the results of LaLonde (1986) involving an experimental evaluation of a training program in the United States using the PSM methodology. They conclude that both methods succeed at focusing on the small sub-set of treated and control individuals, which is the common support in matching models. Using these models, Navarro (2010) finds that the Internet has increased labor income in Honduras (30%), Brazil (29%), Chile (26%), Costa Rica (24%) and Mexico (18%). Paraguay is the only country in his sample that did not demonstrate a statistically significant effect.

The matching process requires two steps. The first step is estimating the propensity score using a probit model, which is a conditional probability of using the Internet following equation 2:

$$(2) \quad Internet_{ijmt} = \alpha_0 + X_{ijmt} + \kappa_{imt} + \delta_j + \gamma_t + \mu_{ijm}$$

where $Internet_{ijmt}$ is a dichotomy variable equal to 1 if a worker uses the Internet. The controls are demographic characteristics, firm size, economic sector, occupation, and fixed effects of the municipality and time, the same variables as in

equation 1. In the second step, using the propensity score, each treated individual is matched with a control individual, reducing the gaps in observable variables between Internet users and non-users. The PSM calculates an average effect of the treatment on the treated (θ_{ATT}), which means the estimated parameter is only for a sub-sample, in this case workers, instead of the full sample (Angrist & Pischke, 2009). Equation 3 shows the general coefficient estimated by PSM.

$$(3) \quad \theta_{ATT} = \sum_{i=1}^I \text{mean} \left[(Y_i | \text{Internet}_i = 1) - (Y_{c(i)} | \text{Internet}_i = 1) \right]$$

where θ_{ATT} is the difference in the outcomes between the most similar treated and control workers. I is the number of workers in the sample, Internet_i is the treatment, and equal to 1 when individual I uses the Internet. c_i is the set of control workers, who do not use the Internet but are very similar in all other characteristics. The simplest and most intuitive way to match workers is by using the nearest-neighbor distance, which compares a treated individual with the closest control individuals in propensity score, following the Euclidean distance. This method, however, can be powerless to close gaps in observable variables when there are many covariates and fixed effects, which risks being unable to compare the most similar individuals. The Mahalanobis distance, meanwhile, offers the advantage of closing gaps in most variables because the distance includes a matrix of variance and covariance of characteristics, which gives more information with which to match treatments and controls (Rubin, 1978).

This paper does not use instrumental variables for two main reasons. First, using household surveys provides the opportunity to control for all the necessary characteristics related to using the Internet. In addition, including fixed effects for municipality, sector and industry makes comparisons between two types of individuals, Internet users and non-users, as close as possible. Second, it is almost impossible to think of an instrument as relevant and exogenous if it is correlated with using ICTs but not income. This variable needs to vary across individuals, or at least within households, given the available information. Every type of exogenous variation related to geography, such as the slope of the municipality, is included in the model for the fixed effect of cities. This study uses the requirements suggested by Dehejia and Wahba (2002) and Abadie and Imbens (2016) to yield an accurate treatment effect in non-experimental settings.

4. RESULTS

The OLS model shows there is a positive correlation between Internet use and income in Colombia, controlling for socio-demographic characteristics, firm size, economic sector, occupation, and fixed effects of municipality and time (results reported in Table 3). This correlation is statistically significant under different specifications. Workers gain 6.9% of one standard deviation

TABLE 3
PARAMETRIC MODEL: ORDINARY LEAST SQUARE

	<i>Dep. Var.: Standardized labor income</i>				
	(1)	(2)	(3)	(4)	(5)
Internet use	0.169*** (0.005)	0.125*** (0.005)	0.116*** (0.004)	0.095*** (0.004)	0.069*** (0.005)
<i>Controls</i>					
Socio-demographic	✓	✓	✓	✓	✓
Firm size		✓	✓	✓	✓
Economic Sector			✓	✓	✓
Occupation				✓	✓
Municipality and time FE					✓
R-squared	0.367	0.386	0.392	0.419	0.429
No. of observations	468,166	468,166	465,331	462,014	462,014

Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. FE denotes municipality and time fixed effects. Internet users are workers accessing from any of the following places: at home, at work, in educational institutions, in free public access centers, in paid access centers, in the house of another person (relative, friend, neighbor).

more when they use this technology. Although this model includes controls for industry characteristics, a proxy for productivity, it estimates bias coefficients; it is therefore necessary to use PSM. The first step in implementing this methodology is to run a probit model using equation 2. Table 4 presents the marginal effects for different treatments: Internet users, place of access, and activities on the web. The control group is not using the Internet for the first two cases and is using this technology for entertainment for the last case. Whereas schooling and experience using devices increase the probability of using the Internet in all treatments, age reduces the probability. In the labor market, working in a big firm, in financial services or as an activity demanding at least some years in college, and professional activities, increases the probability of using the Internet on both places of access and in productive activities.

One of the main drawbacks associated with using matching models is that they must overcome three main tests to find unbiased estimators (Smith & Todd, 2005; Arceneaux *et al.*, 2006; Porto, 2016). The first requirement is a common support in the probability of using the Internet between treated and control groups, which Dehejia and Wahba (2002) call a sub-set of units. Figure A-2 shows the propensity scores for Internet users vs. non-users, where 90% of the sample is located at the intersection between the probabilities of both groups. In this common support, the matching method looks for treated and control individuals who are as similar as possible in all the covariates.

TABLE 4
PROPENSITY SCORE: PROBIT MARGINAL EFFECTS MODEL

<i>Dep. Var:</i>	Internet use (1)	Internet at work (2)	Internet at home (3)	Productive users (4)
Schooling	0.058*** (0.000)	0.086*** (0.000)	0.086*** (0.000)	0.055*** (0.000)
Age	-0.026*** (0.000)	-0.015*** (0.000)	-0.025*** (0.000)	-0.029*** (0.000)
Age squared	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Women	0.028*** (0.000)	0.027*** (0.000)	0.064*** (0.000)	0.017*** (0.000)
Telephone at home	0.065*** (0.000)	0.091*** (0.000)	0.171*** (0.000)	0.058*** (0.000)
Satellite TV at home	0.059*** (0.000)	0.082*** (0.000)	0.129*** (0.000)	0.028*** (0.000)
PC at home	0.191*** (0.000)	0.237*** (0.000)	0.624*** (0.000)	0.158*** (0.000)
Property owner	-0.039*** (0.000)	-0.047*** (0.000)	-0.036*** (0.000)	-0.018*** (0.000)
Agriculture	-0.002*** (0.000)	-0.019*** (0.001)	0.015*** (0.001)	-0.037*** (0.000)
Mining	-0.002** (0.001)	0.015*** (0.002)	0.159*** (0.002)	0.014*** (0.001)
Manufacturing	-0.007*** (0.000)	-0.068*** (0.000)	0.017*** (0.000)	-0.019*** (0.000)
Electricity	-0.058*** (0.000)	-0.077*** (0.000)	-0.029*** (0.001)	-0.034*** (0.000)
Construction	-0.021*** (0.000)	-0.131*** (0.000)	0.011*** (0.000)	-0.055*** (0.000)
Commerce	-0.021*** (0.000)	-0.084*** (0.000)	0.001*** (0.000)	-0.031*** (0.000)
Transport	0.024*** (0.000)	0.074*** (0.000)	0.043*** (0.000)	-0.007*** (0.000)
Financial service	0.043*** (0.000)	0.062*** (0.000)	0.084*** (0.000)	0.084*** (0.000)
Housing sector	0.040*** (0.000)	0.078*** (0.000)	0.091*** (0.000)	0.030*** (0.000)
2-5 employees	-0.087*** (0.000)	-0.197*** (0.000)	-0.133*** (0.000)	-0.062*** (0.000)
6-10 employees	-0.051*** (0.000)	-0.089*** (0.000)	-0.091*** (0.000)	-0.047*** (0.000)
11-50 employees	-0.031*** (0.000)	-0.031*** (0.000)	-0.049*** (0.000)	-0.033*** (0.000)
51-100 employees	-0.032*** (0.000)	-0.032*** (0.000)	-0.040*** (0.000)	-0.012*** (0.000)
Professional Specialty	0.143*** (0.000)	0.316*** (0.000)	0.238*** (0.001)	0.179*** (0.001)
Executive and Managerial	0.192*** (0.000)	0.426*** (0.000)	0.290*** (0.001)	0.153*** (0.001)

<i>Dep. Var:</i>	Internet use (1)	Internet at work (2)	Internet at home (3)	Productive users (4)
Administrative	0.175*** (0.000)	0.368*** (0.000)	0.301*** (0.001)	0.153*** (0.001)
Support Service	0.072*** (0.000)	0.202*** (0.001)	0.133*** (0.001)	0.008*** (0.001)
Sales	0.101*** (0.000)	0.256*** (0.000)	0.174*** (0.001)	0.039*** (0.001)
Machine operator	0.042*** (0.001)	0.103*** (0.001)	0.074*** (0.001)	-0.030*** (0.001)
Equip. cleaners and laborers	0.036*** (0.000)	0.106*** (0.001)	0.058*** (0.001)	0.012*** (0.001)
Precision production	0.026*** (0.000)	0.049*** (0.001)	0.033*** (0.001)	-0.027*** (0.001)
Transportation	0.015*** (0.000)	0.019*** (0.001)	0.054*** (0.001)	-0.045*** (0.001)
Municipality and time FE	✓	✓	✓	✓

Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. FE denotes municipality and time fixed effects.

Second, the matching method has to close most of the gaps in observable and unobservable characteristics between Internet users and non-users. Table 5 presents the differences in averages between treated and controlled individuals before and after the matching process, using two different methods: the nearest-neighbor distance with five neighbors and the Mahalanobis distance. In the unmatched sample, all the differences are statistically different from zero. After using the nearest-neighbor method, there are still many differences between Internet users and non-users; even using 1 neighbor or 10 neighbors the gaps do not close, as shown by Table A-2. Using the Mahalanobis distance, however, closes the gaps for all the variables, showing that the matrix of variances and covariances, in this case, is much more informative for matching individuals. Closing the gaps reduces the possibility of reverse causality, because the small cells compare a clone in socio-demographic characteristics, in the same firm, performing the same occupation, and it is unlikely that workers were already rich before accessing the Internet.

Finally, estimated coefficients have to be robust to changes including controls and the matching method used. Table 6 presents a statistically significant positive correlation between the Internet and labor income in two matching methods and different specifications. Column 5 shows that Internet users earn 6.5% of one standard deviation more than non-users, using the Mahalanobis distance method and all the covariates. Is this coefficient large or small in magnitude? Considering that 54% of Colombia's population lives with one minimum wage income, which is defined as the minimum amount of money to have a basic standard of living

TABLE 5
DIFFERENCE IN MEANS BETWEEN TREATED AND CONTROLS BEFORE
AND AFTER MATCHING SELECTED VARIABLES

Variable	Unmatched (U) Matched (M)	Mean		Difference
		Treated	Control	
Schooling more than 12	U	0.603	0.190	0.412***
	M-NN	0.503	0.504	-0.001
	M-MD	0.503	0.503	0.000
Age	U	32.989	35.543	-2.554***
	M-NN	33.856	33.565	0.291***
	M-MD	33.847	33.871	-0.024
Age squared	U	1194.20	1374.30	-180.1***
	M-NN	1259.70	1242.90	16.80***
	M-MD	1259.10	1258.90	0.200
Woman	U	0.493	0.525	-6.800***
	M-NN	0.492	0.509	-0.017**
	M-MD	0.492	0.492	0.000
Telephone at home	U	0.603	0.399	0.205***
	M-NN	0.567	0.563	0.003***
	M-MD	0.566	0.566	0.000
Satellite TV at home	U	0.812	0.651	0.161***
	M-NN	0.792	0.787	0.006**
	M-MD	0.792	0.792	0.000
PC at home	U	0.608	0.239	0.369***
	M-NN	0.528	0.517	0.011***
	M-MD	0.529	0.528	0.001
Property owner	U	0.406	0.400	0.006**
	M-NN	0.409	0.409	-0.000
	M-MD	0.409	0.409	0.000
White collar	U	0.430	0.139	0.291***
	M-NN	0.337	0.347	-0.010***
	M-MD	0.338	0.337	0.000

Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. The row M-NN shows the result using the Nearest-Neighborhood matching method with 5 neighbors. The row M-MD presents the results using the Mahalanobis distance method.

(MinTrabajo, 2014). An increase of 6.5% of one standard deviation is equivalent to 11% above one minimum salary. This means that using ICTs can definitely help workers pass this threshold. How does this coefficient change over time? Using the available information, Table A-3 estimates the matching model for each year separately. Whereas Internet users earned 7.8% more than non-users in 2009, the coefficient is 22% less in 2011.

The literature identifies a wage premium depending on where people use the Internet. The matching method presented in Table 7 shows that differences in wages amount to 28.4% and 17.4% of one standard deviation from using the Internet at work and at home, respectively. This means that the labor market

TABLE 6
NON-PARAMETRIC MODEL: PROPENSITY SCORE MATCHING FOR INTERNET USE

	<i>Dep. Var.: Standardized labor income</i>				
	(1)	(2)	(3)	(4)	(5)
Nearest Neighbor	0.166*** (0.005)	0.123*** (0.005)	0.115*** (0.004)	0.094*** (0.004)	0.068*** (0.005)
Mahalanobis Distance	0.165*** (0.005)	0.122*** (0.005)	0.114*** (0.004)	0.092*** (0.004)	0.065*** (0.004)
<i>Controls</i>					
Socio-demographic	✓	✓	✓	✓	✓
Firm size		✓	✓	✓	✓
Economic sector			✓	✓	✓
Occupation				✓	✓
Municipality and time FE					✓
R-squared	0.369	0.383	0.394	0.419	0.429
No. of observations	421,349	421,349	418,798	415,813	415,813

Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. FE denotes municipality and time fixed effects. The nearest-neighbor method uses 5 neighbors.

TABLE 7
PLACES AND ACTIVITIES

	<i>Dep. Var.: Standardized labor income</i>			
	Use (1)	Treatment: Internet		Productive users (4)
		At work (2)	At home (3)	
OLS	0.069*** (0.005)	0.291*** (0.006)	0.183*** (0.006)	0.076*** (0.005)
Matching Mahalanobis Distance	0.065*** (0.005)	0.284*** (0.005)	0.174*** (0.005)	0.071*** (0.004)
<i>All controls</i>	✓	✓	✓	✓
R-squared	0.429	0.439	0.432	0.430
No. of observations	451,813	451,813	451,813	451,813

Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. All the controls are socio-demographic, firm size, economic sector, occupation, municipality and time fixed effects. The control group for columns 1, 2 and 3 is Internet non-users, and for column 4 it is Internet users who use the Internet for non-productive activities. Internet users are workers accessing from any of the following places: at home, at work, in educational institutions, in free public access centers, in paid access centers, in the house of another person (relative, friend, neighbor). The number of observations shows the individuals in the common support for the PSM using Mahalanobis Distance. The OLS number of observations is 462,014.

(imperfectly) identifies what people do on the web, and higher incomes relate to using the Internet for productive activities. What about actions using this technology? The last column in Table 7 shows that workers who use the Internet for education and financial services earn 7.1% of one standard deviation more than those who use the Internet for less productive tasks, such as searching for music and videos. An additional dimension analyzed in this paper is the intensity of using this technology. Table A-4 estimates the four treatments only for people who claim to use the Internet every day. Internet users and productive users who access it daily earn 18% and 13%, respectively, more than those who do not necessarily use it daily⁶.

Is there a heterogeneous effect depending on the workers' skill levels? Table 8 estimates the results of using the Internet divided into occupations. Those who benefited the most from using the Internet are the lowest-skilled workers in farming or mining, who earn 20% of one standard deviation more than non-users in the same occupation, followed by the highest-skilled workers in professional specialties, who earn 0.2% less than farmers or miners who use the Internet. Seven sectors out of ten have a positive and statistically significant coefficient between 2.7% and 10% of one standard deviation. Figure A-3 summarizes the results of this table, showing that the effect of using the Internet in the Colombian labor market has a u-shaped distribution depending on occupation.

5. CONCLUSION

Although a large international literature demonstrates that ICTs have a positive impact on salaries, the evidence in the developing world is scarcer. This paper uses the Colombian case to analyze the heterogeneous effects of Internet use in the wage distribution curve, using matching methods. It finds a positive and significant effect of using the Internet. Whereas the lowest wage premium is for Internet users in the middle of the skill distribution, the highest increase in income is for Internet users who work in the lowest-skilled occupations, farming or mining, followed by highly skilled professionals. In relation to other international experiences, the effect of using ICTs has decreased over time due to an increase in the supply of workers who know how to use these tools.

Colombia presents the same pattern as developed countries: even though the United States has more Internet users than Colombia, both present a decline in the effect of Internet use over time. In terms of heterogeneous effects across abilities, the evidence for the developed world partially holds in a developing country such as Colombia. On the one hand, there is a complementarity between

⁶ Guataqui, Martin, and Porto (2016) find that self-employed workers and salaried workers are very different in several characteristics, including who pays for the Internet. Whereas the former probably pay directly for this service, the latter do not. Thus, Table A-5 shows that self-employed workers earn around twice as much as salaried workers who use the Internet.

TABLE 8
INTERNET USE BY SKILL LEVEL

	<i>Dep. Var: Standardized labor income</i>									
	Treatment: Internet use									
	Farming and mining (1)	Transportation (2)	Precision production (3)	Cleaners and laborers (4)	Machine operator (5)	Sales (6)	Service (7)	Administrative support (8)	Executive administrative (9)	Professional specialty (10)
OLS	0.207*** (0.058)	0.049*** (0.012)	0.060*** (0.011)	0.027*** (0.007)	0.074*** (0.037)	0.087*** (0.011)	0.101*** (0.009)	0.096*** (0.010)	0.156*** (0.031)	0.184*** (0.020)
Matching Mahalanobis Distance	0.196*** (0.061)	0.048*** (0.013)	0.058*** (0.014)	0.023*** (0.006)	0.071*** (0.039)	0.084*** (0.012)	0.098*** (0.009)	0.091*** (0.011)	0.152*** (0.033)	0.178*** (0.021)
All controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.472	0.288	0.285	0.247	0.348	0.241	0.397	0.351	0.336	0.361
No. of observations	3,888	33,010	27,618	63,671	3,029	72,478	45,160	33,937	64,225	68,796

Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. All the controls are socio-demographic, firm size, economic sector, occupation, municipality and time fixed effects. Internet users are workers accessing from any of the following places: at home, at work, in educational institutions, in free public access centers, in paid access centers, in the house of another person (relative, friend, neighbor). The number of observations shows the individuals in the common support for the PSM, using Mahalanobis Distance. The OLS number of observations is around 10% more of the mentioned. Occupation is built using the Standard Occupational Classification System (SOCPC, 2010).

higher-skilled workers and the Internet. On the other hand, this relationship is not unique and lower-skilled workers also gain from using this technology. As shown above, there is a low supply of Internet users in this occupation, compared with other sectors, which makes Internet users very valuable in the labor market.

One of the key findings of the paper is that the Internet seems to have the smallest impact on the middle of the skill distribution. One possible reason for this is the lack of complementarity of ICTs with these occupations. Transportation workers, cleaners and laborers are at the bottom of the wage distribution presumably because the Internet is not frequently used in these occupations. There are some companies looking for ingenious solutions to reduce costs in the communication between demand and supply in these sectors, which may suggest that perfecting these applications could increase the magnitude of the effect. Another possible reason for this finding is that these occupations are at a high risk of disappearing in the future. Workplace automation could decrease the demand for workers in these tasks, at the expense of an increase in the demand for these new technologies. It is therefore necessary to delve more deeply into the possible impacts of these applications in the developing world, which despite going a step backwards in terms of absorbing ICTs can lead to undesired effects.

These results show that the Internet could reduce the income inequality gap between occupations, which is already large in Latin America. Colombia, for example, has one of the highest Gini coefficients in the world, around 51. Some previous studies show that ICTs can reinforce pre-existing socio-economic inequalities, although the u-shape curve in the Colombian labor market wage distribution shows that the Internet serves as an equalizing factor in the wage distribution, at least between the highest- and lowest-skilled workers. The developing world needs public policies on ICTS to help overcome poverty.

Since the worldwide trend is to increase access to ICTs, future studies should examine the heterogeneous effects of such access depending on the place of access and particularly the effects of activities on the web. It is very important to understand which access-use patterns are more directly associated with an increase in productivity, and which sectors benefit the most from these policies.

REFERENCES

- Abadie, A., & Imbens, G. W. (2006). "Large sample properties of matching estimators for average treatment effects". *Econometrica*, 74 (1), 235-267.
- Abadie, A., & Imbens, G. W. (2016). "Matching on the estimated propensity score". *Econometrica*, 84 (2), 781-807.
- Acemoglu, D. (1999). "Changes in unemployment and wage inequality: An alternative theory and some evidence". *The American Economic Review*, 89 (5), 1259-1278. Retrieved from <http://www.jstor.org/stable/117057>
- Acemoglu, D., & Autor, D. (2011). "Skills, tasks and technologies: Implications for employment and earnings". *Handbook of Labor Economics*, 4, 1043-1171.

- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: an empiricist's companion* (Vol. 1). Princeton, NJ: Princeton University Press.
- Arabsheibani, G. R., Emami, J., & Marín, A. (2004). "The impact of computer use on earnings in the UK". *Scottish Journal of Political Economy*, 51 (1), 82-94.
- Arceneaux, K., Gerber, A. S., & Green, D. P. (2006). "Comparing experimental and matching methods using a large-scale voter mobilization experiment". *Political Analysis*, 14 (1), 37-62.
- Autor, D. H. (2015). "Why are there still so many jobs? The history and future of workplace automation". *The Journal of Economic Perspectives*, 29 (3), 3-30. Retrieved from <http://www.jstor.org/stable/43550118>
- Autor, D. H., & Dorn, D. (2013). "The growth of low-skill service jobs and the polarization of the US labor market". *The American Economic Review*, 103 (5), 1553-1597. Retrieved from <http://www.jstor.org/stable/42920623>
- Autor, D. H., Levy, F., & Murnane, R. J. (2003). "The skill content of recent technological change: An empirical exploration". *The Quarterly Journal of Economics*, 118 (4), 1279-1333. Retrieved from <http://www.jstor.org/stable/25053940>
- Barrera-Osorio, F., & Linden, L. L. (2009). The use and misuse of computers in education: evidence from a randomized experiment in Colombia.
- Bartel, A., Ichniowski, C., & Shaw, K. (2007). "How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills". *The Quarterly Journal of Economics*, 122 (4), 1721-1758. Re-trrieved from <http://www.jstor.org/stable/25098887>
- Benavente, J. M., Bravo, D., & Montero, R. (2011). "Wages and workplace computer use in Chile". *The Developing Economies*, 49 (4), 382-403.
- Blanco, M., & Vargas, J. F. (2014). "Can SMS technology improve low take-up of social benefits?" *Peace Economics, Peace Science and Public Policy*, 20 (1), 61-81.
- Borghans, L., & Ter Weel, B. (2004). "Are computer skills the new basic skills? The returns to computer, writing and math skills in Britain". *Labour Economics*, 11 (1), 85-98.
- Card, D., & DiNardo, J. (2002). "Skill-biased technological change and rising wage in- equality: Some problems and puzzles". *Journal of Labor Economics*, 20 (4), 733-783. Retrieved from <http://www.jstor.org/stable/10.1086/342055>
- Chigona, W., Beukes, D., Vally, J., & Tanner, M. (2009). "Can mobile internet help alleviate social exclusion in developing countries?" *The Electronic Journal of Information Systems in Developing Countries*, 36.
- Chiswick, B. R., & Miller, P. W. (2007). "Computer usage, destination language proficiency and the earnings of natives and immigrants". *Review of Economics of the Household*, 5 (2), 129-157.
- Colecchia, A., & Schreyer, P. (2002). "ICT investment and economic growth in the 1990s: is the United States a unique case? A comparative study of nine OECD countries". *Review of Economic Dynamics*, 5 (2), 408-442.

- Cortés, G. M., *et al.* (2016). "Where have the middle-wage workers gone? A study of polarization using panel data". *Journal of Labor Economics*, 34 (1), 63-105.
- Dehejia, R. H., & Wahba, S. (2002). "Propensity score-matching methods for nonexperimental causal studies". *The Review of Economics and Statistics*, 84 (1), 151-161.
- DiMaggio, P., & Bonikowski, B. (2008). "Make money surfing the web? The impact of internet use on the earnings of us workers". *American Sociological Review*, 73 (2), 227-250.
- DiNardo, J. E., & Pischke, J.-S. (1997). "The returns to computer use revisited: Have pencils changed the wage structure too?" *The Quarterly Journal of Economics*, 112 (1), 291-303. Retrieved from <http://www.jstor.org/stable/2951283>
- Dobransky, K., & Hargittai, E. (2016). "Unrealized potential: Exploring the digital disability divide". *Poetics*, 58, 18-28.
- Entorf, H., Gollac, M., & Kramarz, F. (1999). "New technologies, wages, and worker selection". *Journal of Labor Economics*, 17 (3), 464-491.
- Foley, P. (2004). "Does the internet help to overcome social exclusion?". *Electronic Journal of e-government*, 2 (2), 139-146.
- García-Murillo, M. (2013). "Does a government web presence reduce perceptions of corruption?" *Information Technology for Development*, 19 (2), 151-175.
- 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. (2009). "Job polarization in Europe". *The American Economic Review*, 99 (2), 58-63.
- Goss, E. P., & Phillips, J. M. (2002). "How information technology affects wages: Evidence using internet usage as a proxy for it skills". *Journal of Labor Research*, 23 (3), 463-474.
- Guataqui, C., Martin, D., & Porto, I. (2016). El raro caso de los trabajadores cuenta propia en Colombia. Unpublished.
- Gust, C., & Márquez, J. (2004). "International comparisons of productivity growth: the role of information technology and regulatory practices". *Labour Economics*, 11 (1), 33-58.
- Heckman, J. J., Ichimura, H., & Todd, P. (1998). "Matching as an econometric evaluation estimator". *The Review of Economic Studies*, 65 (2), 261-294.
- Heckman, J. J., LaLonde, R. J., & Smith, J. A. (1999). "The economics and econometrics of active labor market programs". *Handbook of Labor Economics*, 3, 1865-2097.
- ISIC (2008). International standard industrial classification of all economic activities. Revision 4. *United Nations Publications*. Retrieved from <http://unstats.un.org/unsd/publication/seriesM/seriesm4rev4e.pdf>
- Katz, L. F., & Murphy, K. M. (1992). "Changes in relative wages, 1963-1987: supply and demand factors". *The Quarterly Journal of Economics*, 107 (1), 35-78.

- Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). "Internet paradox revisited". *Journal of Social Issues*, 58 (1), 49-74.
- Krueger, A. B. (1993). "How computers have changed the wage structure: Evidence from microdata". *The Quarterly Journal of Economics*, 108 (1), 33-60.
- LaLonde, R. J. (1986). "Evaluating the econometric evaluations of training programs with experimental data". *The American Economic Review*, 604-620.
- Lee, S.-H., & Kim, J. (2004). "Has the internet changed the wage structure too?". *Labour Economics*, 11 (1), 119-127.
- Lee, S.-Y. T., Gholami, R., & Tong, T. Y. (2005). "Time series analysis in the assessment of ICT impact at the aggregate level-lessons and implications for the new economy". *Information & Management*, 42 (7), 1009-1022.
- McDonald, S., & Crew Jr, R. E. (2006). "Welfare to web to work: Internet job searching among former welfare clients in Florida". *J. Soc. & Soc. Welfare*, 33, 239.
- Michaels, G., Natraj, A., & Van Reenen, J. (2014). "Has ict polarized skill demand? Evidence from eleven countries over twenty-five years". *Review of Economics and Statistics*, 96 (1), 60-77.
- MinHacienda (2014). Presupuesto de 2015. Ministerio de hacienda y crédito público. Ministerio de Hacienda. Colombia.
- MinTrabajo (2014). Definición de salario mínimo. Ministerio de Trabajo. Colombia.
- Navarro, L. (2010). *The impact of internet use on individual earnings in Latin America* (Tech. Rep.). Institute for Advanced Development Studies.
- Oosterbeek, H., & Ponce, J. (2011). "The impact of computer use on earnings in a developing country: Evidence from Ecuador". *Labour Economics*, 18 (4), 434-440.
- Paunov, C., & Rollo, V. (2016). "Has the internet fostered inclusive innovation in the developing world?". *World Development*, 78, 587-609.
- Porto, I. (2016). Impacto del programa de alimentación escolar en el trabajo infantil: Una aproximación desde la toma de decisiones familiares.
- Rosenbaum, P. R., & Rubin, D. B. (1983). "The central role of the propensity score in observational studies for causal effects". *Biometrika*, 70 (1), 41-55.
- Rubin, D. B. (1978). "Bias reduction using Mahalanobis metric matching". *ETS Research Report Series*, 1978(2).
- Smith, J. A., & Todd, P. E. (2005). "Does matching overcome Lalonde's critique of nonexperimental estimators?". *Journal of Econometrics*, 125 (1), 305-353.
- SOCPC (2010). Standard occupational classification. *Bureau of Labor Statistics On behalf of the Standard Occupational Classification Policy Committee*. Retrieved from https://www.bls.gov/soc/soc_structure_2010.pdf
- Steinfeld, C., Ellison, N. B., & Lampe, C. (2008). "Social capital, self-esteem, and use of online social network sites: A longitudinal analysis". *Journal of Applied Developmental Psychology*, 29 (6), 434-445. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0193397308000701> (Social Networking on the Internet) doi: <https://doi.org/10.1016/j.appdev.2008.07.002>
- World-Bank. (2015). World Bank indicators. *The World Bank Group*.

APPENDIX

TABLE A-1
DIFFERENCES IN MEAN BY TYPE OF WORKER

Variables	Salaried worker	Self-employed worker	Difference
Logarithm of labor income	13.526 (0.0013)	13.111 (0.0024)	0.416*** (0.005)
Age	33.385 (0.0188)	37.010 (0.0243)	-3.625*** (0.054)
Age squared	1,217.343 (1.392)	1,488.661 (1.900)	-271.318*** (4.112)
Schooling	13.035 (0.0047)	12.594 (0.0053)	0.442*** (0.013)
12 or more years of education	0.512 (0.0009)	0.412 (0.0010)	0.100*** (0.003)
Telephone at home	0.622 (0.0009)	0.556 (0.0011)	0.066*** (0.003)
PC at home	0.770 (0.0007)	0.717 (0.0010)	0.053*** (0.002)
Satellite TV at home	0.519 (0.0009)	0.473 (0.0011)	0.047*** (0.003)
House owner	0.399 (0.0009)	0.427 (0.0011)	-0.028*** (0.003)
2-5 employees	0.192 (0.0007)	0.823 (0.0008)	-0.631*** (0.002)
6-10 employees	0.074 (0.0004)	0.033 (0.0003)	0.040*** (0.001)
11-50 employees	0.159 (0.0006)	0.040 (0.0004)	0.119*** (0.001)
51-100 employees	0.056 (0.0004)	0.009 (0.0002)	0.047*** (0.001)
101 and over employees	0.520 (0.0009)	0.095 (0.0006)	0.425*** (0.002)
Agriculture	0.016 (0.0002)	0.021 (0.0003)	-0.006*** (0.001)
Mining	0.001 (0.0000)	0.001 (0.0000)	0.000 (0.000)
Manufacturing	0.171 (0.0007)	0.120 (0.0007)	0.051*** (0.002)
Electricity	0.013 (0.0002)	0.002 (0.0001)	0.010*** (0.000)
Construction	0.033 (0.0003)	0.048 (0.0004)	-0.015*** (0.001)
Commerce	0.235 (0.0007)	0.331 (0.0010)	-0.095*** (0.002)
Transport	0.077 (0.0004)	0.130 (0.0007)	-0.053*** (0.002)
Financial service	0.038 (0.0003)	0.007 (0.0001)	0.032*** (0.001)
Housing sector	0.084 (0.0005)	0.111 (0.0007)	-0.028*** (0.002)

Variables	Salaried worker	Self-employed worker	Difference
Public administration	0.078 (0.0005)	0.031 (0.0003)	0.047*** (0.001)
Education	0.111 (0.0005)	0.038 (0.0004)	0.073*** (0.001)
Health	0.073 (0.0004)	0.069 (0.0005)	0.005*** (0.001)
Domestic service	0.069 (0.0004)	0.090 (0.0006)	-0.021*** (0.001)
Professional specialty	0.169 (0.0007)	0.144 (0.0007)	0.026*** (0.002)
Executive and managerial	0.219 (0.0007)	0.070 (0.0005)	0.149*** (0.002)
Administrative support	0.093 (0.0005)	0.080 (0.0006)	0.013*** (0.001)
Service	0.113 (0.0005)	0.089 (0.0006)	0.024*** (0.002)
Sales	0.112 (0.0005)	0.249 (0.0009)	-0.137*** (0.002)
Machine operator	0.005 (0.0001)	0.014 (0.0002)	-0.009*** (0.001)
Equip. cleaners and laborers	0.151 (0.0006)	0.142 (0.0007)	0.008*** (0.002)
Precision production	0.060 (0.0004)	0.092 (0.0006)	-0.032*** (0.001)
Transportation	0.067 (0.0004)	0.099 (0.0006)	-0.032*** (0.001)
Farming and mining	0.011 (0.0001)	0.022 (0.0003)	-0.011*** (0.001)
White collar	0.398 (0.0009)	0.266 (0.0009)	0.132*** (0.002)

Notes: Standard errors in brackets. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level.

TABLE A-2
DIFFERENCE IN MEANS BETWEEN TREATED AND CONTROLS BEFORE
AND AFTER MATCHING SELECTED VARIABLES

Variable	Unmatched (U) Matched (M)	Mean		Difference
		Treated	Control	
Schooling more than 12	U	0.603	0.190	0.412***
	M-1	0.512	0.513	-0.001*
	M-10	0.503	0.504	-0.001
Age	U	32.989	35.543	-2.554***
	M-1	33.859	33.561	0.298***
	M-10	33.856	33.565	0.291***
Age squared	U	1194.20	1374.30	-180.1***
	M-1	1259.70	1242.90	16.80***
	M-10	1259.70	1254.50	5.70***
Woman	U	0.493	0.525	-6.800***
	M-1	0.493	0.512	-0.019**
	M-10	0.492	0.500	-0.008*
Telephone at home	U	0.603	0.399	0.205***
	M-1	0.567	0.552	0.015***
	M-10	0.566	0.565	-0.002*
Satellite TV at home	U	0.812	0.651	0.161***
	M-1	0.794	0.777	0.017**
	M-10	0.792	0.787	0.005**
PC at home	U	0.608	0.239	0.369***
	M-1	0.530	0.515	0.015***
	M-10	0.529	0.522	0.007***
Property owner	U	0.406	0.400	0.006**
	M-1	0.409	0.409	-0.000
	M-10	0.409	0.409	0.000
White collar	U	0.430	0.139	0.291***
	M-1	0.368	0.317	-0.051***
	M-10	0.338	0.341	0.003*

Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. The row M - 1 shows the result using the Nearest Neighborhood matching method with 1 neighbor and M - 10 the same matching method with 10 neighbors.

TABLE A-3
INTERNET USE OVER TIME

	<i>Dep. Var: Standardized labor income</i>			
	2009-2011 (1)	Treatment: Internet use		
		2009 (2)	2010 (3)	2011 (4)
OLS	0.069*** (0.005)	0.080*** (0.008)	0.072*** (0.008)	0.065*** (0.008)
Matching Mahalanobis Distance	0.065*** (0.004)	0.078*** (0.004)	0.069*** (0.004)	0.061*** (0.004)
<i>All controls</i>	✓	✓	✓	✓
R-squared	0.429	0.425	0.443	0.431
No. of observations	451,813	130,618	138,221	146,966

Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. All the controls are socio-demographic, firm size, economic sector, occupation, municipality and time fixed effects. The number of observations shows the individuals in the common support for the PSM, using Mahalanobis Distance. The OLS number of observations is 462,014 in column 1, 145,212 in column 2, 153,506 in column 3, 163,296 in column 4.

TABLE A-4
PLACES, ACTIVITIES AND INTENSITY

	<i>Dep. Var: Standardized labor income</i>			
	Use (1)	Treatment: Internet daily		
		At work (2)	At home (3)	Productive users (4)
OLS	0.081*** (0.009)	0.295*** (0.006)	0.186*** (0.006)	0.092*** (0.005)
Matching Mahalanobis Distance	0.077*** (0.007)	0.289*** (0.006)	0.178*** (0.005)	0.080*** (0.004)
<i>All controls</i>	✓	✓	✓	✓
R-squared	0.429	0.439	0.432	0.430
No. of observations	451,813	451,813	451,813	451,813

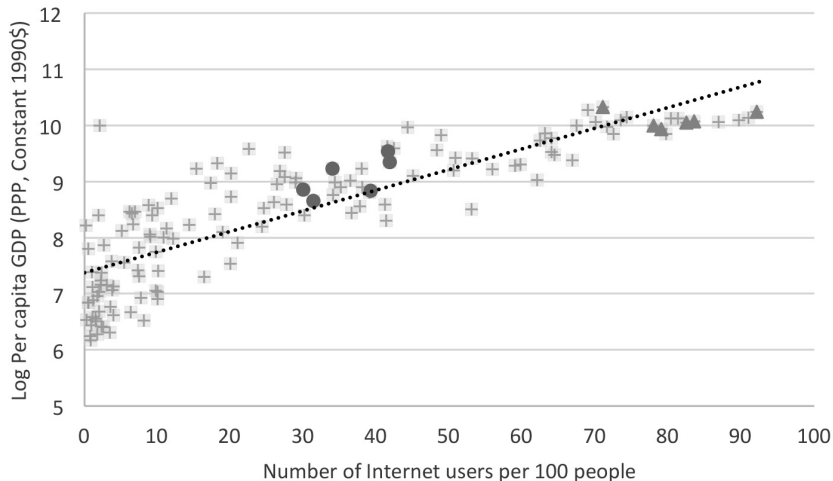
Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. All the controls are socio-demographic, firm size, economic sector, occupation, municipality and time fixed effects. The control group for columns 1, 2 and 3 is Internet non-users, and for column 4 it is Internet users in non-productive activities. Internet users are workers accessing from any of the following places: at home, at work, in educational institutions, in free public access centers, in paid access centers, in the house of another person (relative, friend, neighbor). The number of observations shows the individuals in the common support for the PSM, using Mahalanobis Distance. The OLS number of observations is 462,014.

TABLE A-5
PLACES AND ACTIVITIES BY WORKER TYPE

	<i>Dep. Var: Standardized labor income</i>			
	Use (1)	Treatment: Internet		Productive users (4)
		at work (2)	at home (3)	
<i>Panel A: Salaried Worker</i>				
OLS	0.054*** (0.005)	0.206*** (0.006)	0.159*** (0.007)	0.058*** (0.006)
Matching Mahalanobis distance	0.054*** (0.004)	0.192*** (0.005)	0.145*** (0.003)	0.054*** (0.004)
<i>All controls</i>	✓	✓	✓	✓
R-squared	0.495	0.501	0.498	0.495
No. of observations	246,359	246,359	246,359	246,359
<i>Panel B: Self-employed worker</i>				
OLS	0.098*** (0.009)	0.529*** (0.017)	0.243*** (0.013)	0.112*** (0.011)
Matching Mahalanobis distance	0.092*** (0.008)	0.511*** (0.010)	0.229*** (0.009)	0.097*** (0.009)
<i>All controls</i>	✓	✓	✓	✓
R-squared	0.368	0.387	0.372	0.368
No. of observations	169,462	169,462	169,462	169,462

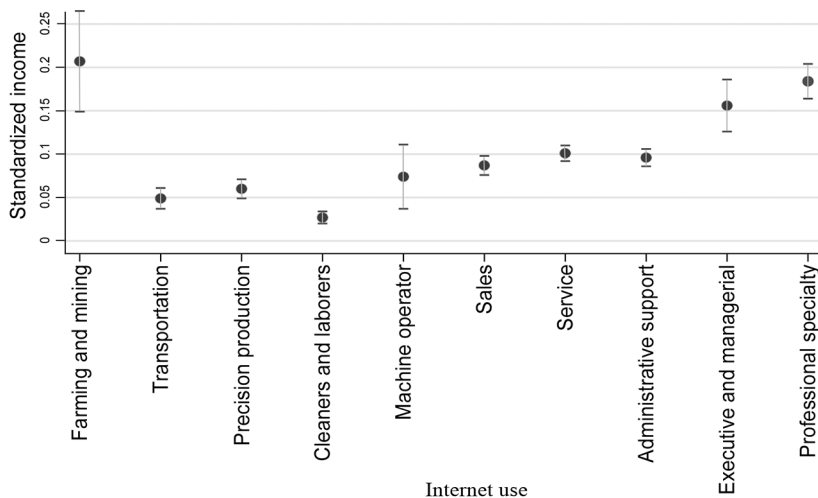
Notes: Robust standard errors are shown in parentheses. *** is significant at the 1% level, ** is significant at the 5% level, * is significant at the 10% level. All the controls are socio-demographic, firm size, economic sector, occupation, municipality and time fixed effects. The control group for columns 1, 2 and 3 is Internet non-users and for column 4 it is Internet users in non-productive activities. Internet users are workers accessing from any of the following places: at home, at work, in educational institutions, in free public access centers, in paid access centers, in the house of another person (relative, friend, neighbor). The number of observations shows the individuals in the common support for the PSM, using Mahalanobis Distance. The OLS number of observations is 273,733 for salaried workers and 188,281 for self-employed workers.

FIGURE A-1
INTERNET AND INCOME IN THE WORLD



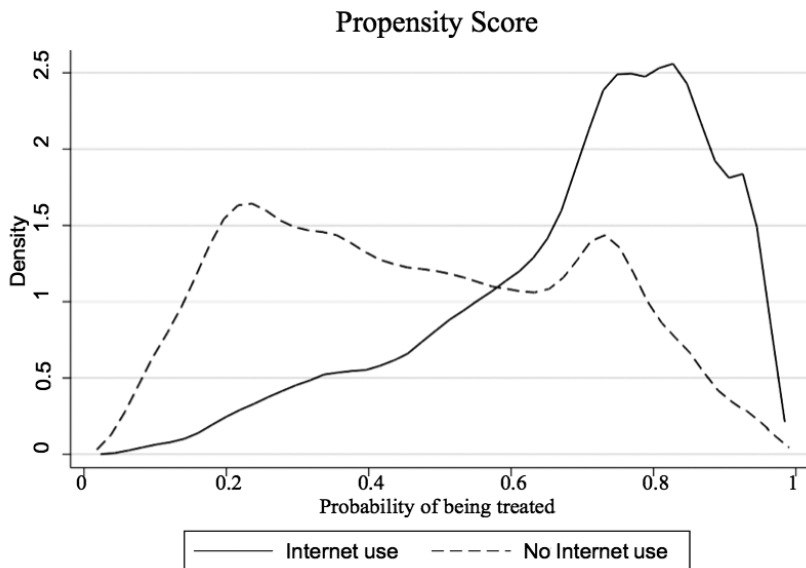
Note: Own calculation the following indicators: the logarithm of per capita GDP (PPP, constant 1990\$) in 2010 and the number of Internet users per 100 people from the World Bank (2015).

FIGURE A-2
CORRELATION BETWEEN INTERNET USE AND STANDARDIZED INCOME



Note: Propensity score built using equation 2.

FIGURE A-3
COMMON SUPPORT BETWEEN TREATED AND CONTROL GROUPS



Note: The figure shows the estimators and interval coefficients using Mahalanobis distance and including the entire set of covariates from Tables 1 and 2, for 10 sub-samples depending on occupation. Each category is built using SOCP (2010).