Information Problem in Labour Market and Big Data: Colombian Case

Jeisson Cárdenas

COLOMBIA CIENTÍFICA
Conocimiento Global para el Desarrollo

ALIANZAEFI
economía formal e inclusiva

Information Problem in Labour Market and Big Data: Colombian Case[1][2]

Jeisson Cárdenas Rubio
Institute for Employment Research
University of Warwick
Coventry, United Kingdom

Abstract

This working paper discusses the concepts and theoretical framework to analyse the labour market, based on the information found on online job portals. Based on a model with imperfect information (which seems more appropriate to describe Colombian labour market outcomes), the first section explains how skill mismatches can arise and their consequences for informality and unemployment rates. In second section, presents evidence that skill shortages, unemployment and informality are highly occurring phenomena in Colombia; and, it is argued that workers, educational and training providers and the government can do little to address these issues because of a lack of proper information to monitor and identify employers' requirements and possible skill shortages at an occupational level. In section 3 the concept of Big Data is introduced, with its advantages and limitations outlined for labour market analysis, this section explains the limitations and caveats to be considered when online vacancy data are used for economic analysis.

Key words: Skill, Skill shortages, big data, online job portals, informality, unemployment.

JEL classification: J24

---

[1] This working paper is part of the author's PhD thesis at the University of Warwick.
[2] E-mail: j.cardenas-rubio@warwick.ac.uk (J.Cardenas).

1. The Labour Market and Skill Mismatches

1.1 Introduction

The labour market can be defined as a "place" (not necessarily a physical place) where employers (the "demand") and workers (the "supply") interact with each other. The dynamic of this labour market is relevant for an economy as it determines different socio-economic outputs, such as productivity, unemployment, wages, and poverty, among others. Provided the labour market influences various outcomes and different disciplines (e.g. sociology, economy, etc.) address these issues, this working paper narrows and discusses labour market definitions and the theoretical framework adopted to analyse labour demand based on the information found on online job portals.

The next section explains what is understood in the academic literature in economics by labour supply and labour demand, and the possible ways to measure these concepts statistically. Moreover, the informal economy is defined and highlighted as a key issue, especially in Latin American countries such as Colombia. Subsequently, the concept of skills is introduced and its possible implications for unemployment and the informal economy. With these basic definitions outlined, the third section of this working paper describes a labour market and its main outcomes, such as unemployment, wages, etc., under the assumption of perfect competition.

However, the assumptions of perfect competition are substantial and might not be appropriate for different economies such as the Colombian economy. Consequently, it is necessary to consider labour market failures, for example imperfect information, that might appropriately explain the comparatively high rates of informal economy and unemployment levels in Colombia. Thus, the fourth section focuses on explaining how imperfect information might increase skill mismatching, and, consequently, it might create labour market segmentation between formal and informal workers along with a comparatively high unemployment rate. Furthermore, it is highlighted that failures of information might be one of the leading causes of high unemployment and informality rates; especially in developing countries such as Colombia.
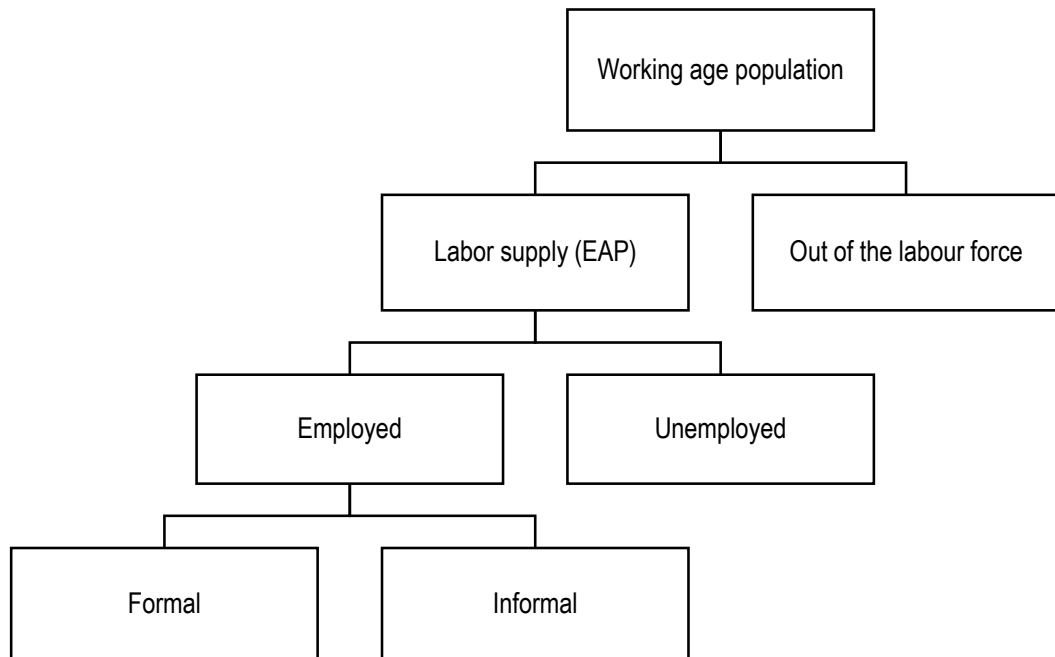
1.2 Basic definitions

Comparable to other markets (e.g. financial markets, physical consumer markets, etc.), the labour market is composed of supply and demand (Cahuc et al. 2014). The merchandise to be exchanged consists of "labour services" that represent human activities (distinguished by numbers of workers or hours of work); these human activities are one of the inputs in the production of goods and services (ILO, 2018). Consequently, the dynamic between supply and demand have various implications for a range of individuals, for instance, people with different characteristics (i.e. skills), employers that create job offers with certain requirements, and the government, among others. Thus, this section explains who composes the labour demand and the labour supply (e.g. unemployed, formal and informal workers) and the relevance of skills in the labour market outcomes.

1.2.1   Labour Supply

In a basic economic model, people or households possess a limited quantity of "labour" that they can offer in the labour market in order to have an income to acquire goods and services (Cahuc et al. 2014). Therefore, the labour supply or labour force is composed of people who offer their "labour". As shown in Figure 1.1, the labour supply (or the Economically Active Population [EAP]) is composed of: 1) people who do not have a job but are looking for one (unemployed); and 2) people who are part of the working age population hired by employers (employed) and the self-employed (ILO, 2017a).

For statistical purposes—according to the International Labour Organization (ILO)—all working-age people that did not participate in the production of goods and services for at least one hour in the reference week because they did not need to, cannot or are not interested in earning a labour income, and are considered out of the labour force (or inactive) (ILO, 2017a). An unemployed individual is a person without work that has sought a job during the last four weeks and is available for work within the next fortnight; or is currently without a job but has accepted a job to start in the next fortnight. An employed individual is employed when he/she has worked for at least one paid or unpaid hour in the reference week (one week before the survey is conducted). These employed and unemployed individuals are considered as the labour force (EAP).

Figure 1.1: Labour market structure

```
                    ┌──────────────────────────┐
                    │ Working age population   │
                    └──────────────────────────┘
                                 │
                  ┌──────────────┴──────────────┐
        ┌───────────────────┐        ┌───────────────────────┐
        │ Labor supply (EAP)│        │ Out of the labour force│
        └───────────────────┘        └───────────────────────┘
                  │
         ┌────────┴────────┐
  ┌──────────────┐   ┌──────────────┐
  │   Employed   │   │  Unemployed  │
  └──────────────┘   └──────────────┘
         │
   ┌─────┴─────┐
┌──────────┐ ┌──────────┐
│  Formal  │ │ Informal │
└──────────┘ └──────────┘
```

1.2.2   Labour demand

In contrast, companies or establishments require "labour services" as an input to produce goods and services in the private and public sector. Consequently, labour demand refers to the demand for workers (or hours of work) in an economy. This demand consists of the level of employment (satisfied labour demand) plus the number of available job vacancies which equates to the labour required but not filled by an employee over a certain period (unsatisfied labour demand or unmet demand) (Farm and Sweden, 2003; Williams, 2004).

In this sense, a job vacancy is defined as a "paid post that is newly created, unoccupied, or about to become vacant:

a) for which the employer is taking active steps and is prepared to take further steps to find a suitable candidate from outside the enterprise concerned; and

b) which the employer intends to fill either immediately or within a specific period" (Eurostat, 2017).

Therefore, the total number of vacancies in an economy is determined by the number of unfilled job openings and, additionally, the number of jobs that are temporary filled by internal substitutes (Farm and Sweden, 2003).

To conclude this subsection, the classic economic model (Cahuc et al. 2014) describes the labour market in the following way: people (or households) offer a certain quantity of their "labour" at a certain level of labour price (wages) in order to generate income and acquire different goods and services available in other markets. Establishments in this model require a certain quantity of "labour" at a certain level of labour price (wages) to produce goods and services, and while some workers have a job and are employed, others are looking for one and are unemployed. Nevertheless, as shown in Figure 1.1, the fact that people are working does not imply that they are working in regulated and good working conditions (e.g. informal economy).

1.2.3 Informal economy

To measure the informal economy, the ILO (2003) recommends making a distinction between the informal sector and informal employment. On the one hand, the informal sector is an enterprise-based definition which considers people working in units that have "informal" characteristics regarding their unregistered and/or unincorporated legal status, small size, the non-registration of their employees, their lack of formal labour relations, without bookkeeping practices, and the under-payment/non-payment of taxes, among others. On the other hand, informal employment is a job-based definition and covers individuals whose main job lacks basic legal and social protections (or employment benefits). For example, a lack of social protection, no income taxation, and so forth[3].

The above sector and informal employment definitions highlight different aspects of an informal economy, and can be used for various public policy targets such as payroll taxes, social protection, among others (ILO, 2003). Consequently, it is possible that people work informally for enterprises that operate in the formal economy or workers might have formal jobs (e.g. with social security) for enterprises in the informal sector (see Figure 1.2).

---

[3] It is necessary to clarify that both these informal economy concepts do not refer directly to underground, illegal production and non-market production.

Figure 1.2: Composition of informal economy
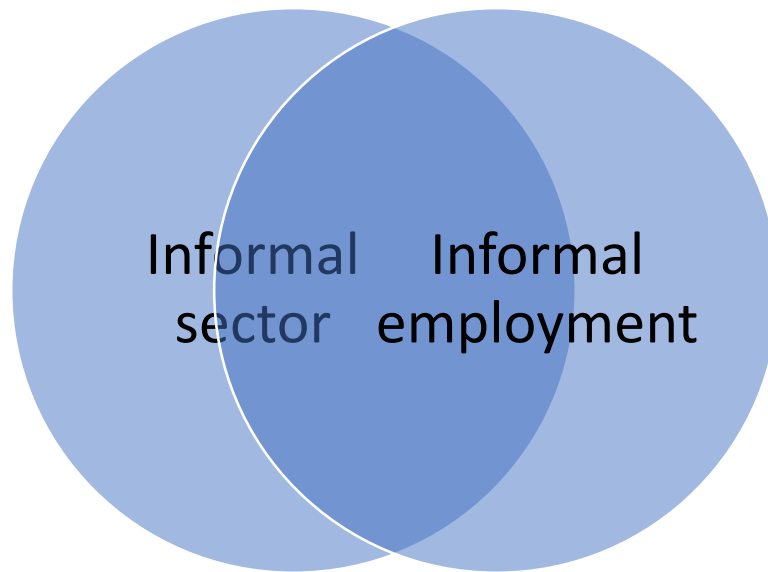
Based upon the ILO's recommendations, countries such as Colombia in national household surveys consider the following individuals to be informal workers: private employees and workers based in establishments, businesses or companies that occupy up to five people (in all their agencies and branches), including the employer and/or partner; unpaid family workers; domestic employees; self-employed workers, except independent professionals; while government employees are excluded from this definition (Hussmanns, 2004). Evidently, Colombia's household surveys classify informal workers according to the concept of the informal sector[4].

This way of measuring informality has some limitations. As mentioned above, informal employees may be formally working in large factories and, in consequence, the way that Colombia measures informality might underestimate the phenomenon (ILO, 2012a). However, using a measure employed by the statistics office of Colombia (DANE) to calculate informality via social security contributions (pensions) and firm size, Bernal (2009) found that—at least for

---

[4] Even though the official informality statistic is based on the concept of the informal sector, it is possible to calculate informality using an informal employment approach (e.g. pension and/or health contributions, among other benefits). Moreover, Colombia excludes agricultural activities from its official informal sector statistics, since including such activities requires developing a more robust definition; especially regarding jobs held by own-account workers, and members of producers' cooperatives in the agricultural industry (ILO, 2003).

the Colombian case—the size of the informal sector is remarkably similar between the social security and firm size informality measurements. The same author found workers who pay for social security contributions (pension and/or health) are less likely to belong to small firms.

In addition, the ILO (2011) studied 47 medium and low-income countries and concluded that almost all workers employed by the informal sector are also in informal employment. This result suggests that criteria based on firm size (in the informal sector) are a suitable approach to calculate the informality issue, at least for the Colombian case.

The magnitude of the informal economy problem depends on different processes. On one side, there is an "exclusion" process. More specifically, workers and companies would prefer formal jobs with state mandatory benefits; however, some barriers restrict agents' access to the formal economy. These restrictions or barriers can take different forms, such as excessive taxation or lack of certain workers' characteristics (e.g. skills), that make it difficult to enter the formal economy. This framework suggests that informal firms and workers are a disadvantaged group.

On the other side, some workers and firms voluntary choose to remain in the informal economy, based on their preferences for working, and the net benefit of being in the informal versus the formal economy. To belong to the formal economy, workers and firms need to incur certain costs, such as tax revenue, health and work insurance, and in return the state must provide benefits, such as health care, access to credit, etc. However, these benefits might not compensate for the cost of formality (such as tax revenue). Thus, the informal economy can be an "escape" for workers and firms to avoid the formal economy and its failures related to the provision of services (Perry, 2007). These facts highlight that the benefits of being in the formal economy are not enough to move some agents into the formal economy.

Informal economy is, usually, a term that describes individuals working in unregulated jobs, and is associated with inadequate working conditions, a lack of social security, lower productivity, limited access to the financial system, etc. As Perry (2007) pointed out, the size of the informal economy is relevant because it affects a country's productivity and growth. Informal firms might experience more barriers to access credit, broaden their sale markets and innovate, which might reduce their potential productivity. For instance, the lack of social protection and other work risks might result in a lower incentive for establishments to invest in human capital (see Section 1.4), and lead to lower worker productivity.

The informal economy, along with unemployment, is considered one of the most important indicators to measure the well-being in the labour market (ILO, 2015; Mondragón et al. 2010). Both phenomena are prevalent in Latin America economies, and reflect the high degree of labour supply underutilization. This result reveals the inability of the Latin America economies to generate "quality" employment for those persons who want to work and can work (ILO, 2017b). For these reasons, it is essential to measure and consider the informal economy in the analysis of any country's labour market, especially in countries such as Colombia where the informality rate is comparatively high, at around 49.4% in 2016 (DANE, 2017a) (See section 2).

To conclude this subsection, the informal economy is a relevant phenomenon which affects different socio-economic outcomes, such as productivity, social protections etc. This high incidence of the informal economy in Latin American countries such as Colombia makes it an important factor to be considered in Colombian labour market analysis. However, this term might cover a variety of activities that can be measured in different but correlated ways. Despite some limitations, the Colombian literature suggests that a valid criterion to classify workers into informality is based on company size— which is the one adopted in the official Colombian labour market statistics and this paper.

Related to unemployment, the informal economy phenomenon might arise due to an extended number of factors: rigid wages, comparatively high non-wage costs, technological shocks, and discrimination (e.g. gender preferences), are examples of such factors, and a vast body of theoretical frameworks have been developed to analyse the role of these elements. One of these theoretical frameworks stresses the importance of skills on the labour market outcomes such as unemployment and informal economy. Individuals possess different labour characteristics that make them more or less productive for specific jobs (Albrecht et al. 2007), so while companies hire labour with different attributes to perform different tasks and produce their products, the misallocation between the skills possessed by workers and the skills demanded by employers might influence unemployment and informality rates.

This framework might be appropriated in a context such as Colombia where there is a comparatively high portion of companies complaining about the skills possessed by the labour supply, and at the same time there is a high proportion of workers desiring formal jobs (section

2 provides a detailed discussion of the Colombian context). Thus, worker skills are important for an economy (the following subsection defines this concept in more detail).

1.2.4 Skills

Skills are a relevant factor that have strong implications for employment outcomes such as productivity, wages, job satisfaction, turnover rates, unemployment, informal economy, etc. (Acemoglu, and Autor, 2011; Kankaraš et al. 2016). However, the skills concept can be understood and interpreted from different perspectives: social constructionist, positivist, and ethnomethodological, among others (Attewell, 1990; Green, 2011; Warhurst et al. 2017). Additionally, there are multiple skill typologies (e.g. workers' skills and skills as attributes of jobs). Thus, this section discusses the skill definition adopted in this paper to analyse labour demand based on online job portal information.

1.2.4.1 Defining skills

Each school of thinking emphasises the importance of different elements that should be considered by the concept of "skill". Within the social constructionist school, for instance, skills are a complex construction of job tasks, labour supply and demand, and certain social conditions (Vallas, 1990). Consequently, skills are defined by the tasks associated with each job, together with the capacity to restrict a number of people into a profession or career. The more time it takes to train for a profession, the dedicated effort involved creates a perception that the job requires higher skills (Attewell, 1990). Therefore, as Gambin et al. (2016) pointed out, from a social constructionist perspective social "norms" and task complexity determine what a valued skill means. This approach is part of an ongoing, subjective and extended debate in which it is difficult to delimit what social processes might affect the construction of skill in a particular society. Consequently, the social constructionist school often finds it challenging to generalise and compare skills between different societies or groups (Green, 2011).

The ethnomethodological and positivist approaches emphasises other aspects. On the one hand, the former points out that any human activity is complex and the required skills to carry out those activities are multifaceted. When a task is easy to carry out it is considered "unskilled". Consequently, what the observer takes for granted (easy to carry out) will determine the skill complexity of a certain task and, thus, generalisations of about skills are difficult to make with an

ethnomethodological method. On the other hand, positivism states that skills are objective attributes of individuals or jobs which are independent of the observer. This view focuses on obtaining uniform skill measures to provide the most precise skills indicator for positivist-based research (Attewell, 1990).

Even though there are different ways to approach how to define "skills", most perspectives agree that the concept of skills is strongly related to the task complexity required to carry out a particular job. In concordance with Green (2011, p.11): "all skills are social qualities, yet are rooted in real, objective, processes not in perceptions". Thus, this paper interprets skills as attributes of people or jobs which are required to perform certain tasks in the labour market. Consequently, in this document, skill refers to any measurable quality that makes a worker more productive in his/her job, which can be improved through training and development (Green, 2011). Simply put, according to Gambin et al. (2016) a skill refers to "the ability to carry out the task that comprises a particular job".

This perspective might be particularly helpful to ease the operationalization of skills into quantitative measurements (to provide easily measured variables), and to enable policymakers and researchers to obtain precise quantitative results to produce straightforward public policy recommendations (Attewell, 1990)—which is also one of the main objectives of this paper. However, this positivist viewpoint has some limitations; for instance, to measure a skill with a variable such as years of education could be considered reductionist. As will be discussed in the next subsection, variables such as education might fail to properly measure skill acquisition and job performance (Attewell, 1990). Consequently, the challenge of this positivist school of thought, and, hence, in this paper, is to find a set of indicators sufficiently reliable and valid that measure individuals' skills and skills mismatches—be it imperfectly.

Despite the limitations which are present in all schools of thinking, a positivist perspective (frequently presented in economic studies) is adopted in this paper in order to provide imperfect but sufficiently reliable and valid indicators for public policy recommendations regarding skills within vacancy data on online job portals. This definition of "skills" still encompasses many elements such as qualifications, competences, education, and aptitudes, among others (Green, 2011), which can be measured by different indicators depending on the typology used and the tools available to measure those qualities (skills). The economic literature has used a variety of

proxies to measure the different dimension of skills in the labour market, some of which are limited because while some typologies overlap others do not make a clear separation between skill categories (as will be explained in more detail in the next subsection).

Given the multiple skill typologies used even within the same economic discipline, it is necessary to discuss which are the most appropriate for this paper. The different typologies can be organised into two groups: those focused on the worker's skills and those which use a task-based approach.

1.2.4.2 Workers' skills

At an early stage, human capital theory stated that the necessary skills for working could be obtained with education (Becker, 1962; Mincer, 1958). In consequence, educational attainment is seen as a way to define skills. The educated worker is considered highly skilled and, thus, more productive if he/she accumulates more years of education and experience. Accordingly, increased human capital through education (the main source of scientific knowledge) is thought to increase employees' productivity in a range of tasks (Attewell, 1990; Becker, 1962).

Consequently, the accumulation of skills (in terms of knowledge) rather the use of skills towards particular jobs has been the focus of analysis for academics and policymakers. For example, Becker (1994) has pointed out that schooling fosters higher workforce productivity and has a positive impact on wages. Likewise, Psacharopoulos (1985; 2006) has shown the nonmarket benefits of schooling, such as lower crime rates, social cohesion among others, in advanced[5], intermediate[6], Latin American[7], Asian[8] and African[9] countries. The important point to take away from this research on schooling is that it is supposed to determine the results in the labour market, and other aspects of life outcomes.

However, the economic literature has found that education attainment only explains a relatively small fraction of the variance of life accomplishments between individuals. Indeed, Kautz et al. (2014, p.9) points out that adolescent (knowledge) test scores, in a best-case scenario, explain

---

[5] Australia, Canada, France, Germany, Japan, Sweden, the United Kingdom and the United States

[6] Cyprus, Greece, Iran and Portugal.

[7] Brazil, Chile, Colombia, Costa Rica, El Salvador, Guatemala, Mexico and Venezuela.

[8] Hong Kong, Malaysia, Pakistan, Singapore, South Korea, South Vietnam, Sri Lanka, Taiwan and Thailand

[9] Ethiopia, Kenya, Morocco and Tanzania.

just 17% of an individual's earning variability in later-life. Additionally, to measure skills by observing educational levels has several limitations.

Firstly, education attainment might be a weak indicator to measure knowledge levels. Education (or qualification) is acquired before people participate in the labour market; however, those qualifications might not be appropriate or might depreciate over time, compared to other skills learnt in the workplace[10].

Secondly, Becker (1994) recognises educational measures ignore some sources of learning, and Cunha and Heckman (2007) suggest that skill formation/acquisition occurs through a variety of processes and situations. For instance, skills can be acquired outside of schools, through on-the-job training (such as apprenticeships, coaching, etc.) and/or off-the-job training (such as lectures, simulations, etc.). Extended literature in labour economics shows the effects of job training on different outcomes. Bassanini et al. (2007, p.128) completed an exhaustive review of data resources (Continuing vocational training survey—CVTS, the International Adult Literacy Survey—IALS, among other data) for on-the-job training in Europe. The authors found evidence that on-the-job training has a positive correlation with private returns for employees and employers[11]. Likewise, Asplund (2004), Barrett et al. (1999) and Blundell et al. (1999), among others, extensively reviewed the different effects of off-the-job training on social and private outcomes. Most of the studies reviewed found a positive impact on social and private returns[12]. Hence, on-the-job and off-the-job training are relevant sources of knowledge for workers and employers which are not measured by education variables.

Thirdly, education variables do not take into account other skills generated via learning-by-doing in the production process. People continue to learn new skills and reinforce them through

---

[10] For instance, with the emergence of modern devices (e.g. computers) have been introduced in the labour market, along with new technologies to perform different jobs (such as programming, social media manager, and so forth), which, in general, were not taught by the educational system years ago. Thus, for some jobs, to be up-to-date and to be able to use these new technologies can be considered more valuable for the labour market compared to previous years spent in education.

[11] For instance, short on-the-job training increased hourly earnings by approximately 2% in Denmark and the UK, while in countries such as Portugal this increase was around 10% (Bassanini et al. 2007, p.128).

[12] Even if there are studies such as Black and Lynch's (1995) that found off-the-job training might have greater impacts on productivity than on-the-job training in US manufacturing industries.

repetition (Dehnbostel, 2002; Rutherford, 1992). From a theoretical perspective, Arrow (1962) demonstrates that technological charge (which is a key aspect for economic growth) can be endogenously determined via learning-by-doing processes. Different empirical studies show that these learning processes increases a firm's productivity. For instance, Bahk and Gort (1993) observe that in 15 industries in the US, leaning-by-doing generates skills (knowledge) and reduces the production costs of incumbent, established organisations. Similarly, Dasgupta and Stiglitz (1988) highlight that learning-by-doing implies economies of scale in the production of products, which increase a firm's productivity.

Finally, employers not only require cognitive and academic skills (qualifications) they also consider personal characteristics as important elements to perform a job. As Green (2011) and Grugulis et al. (2004) note, companies have labelled behavioural characteristics (e.g. reliability, responsibility, leadership, motivation, politeness, and compromise, among others.) as skills needed in the production process. It is not just the knowledge learnt through formal education, job training or learning-by-doing that produces more skilled workers, in addition personal characteristics, such as traits, attitudes and attitudes towards work, are also considered as skills (Grugulis et al. 2004; Kautz et al. 2014). Recent evidence in the economic literature shows that personal characteristics have direct implications for the labour market. For instance, Brunello and Schlotter (2011) and Lindqvist and Vestman (2011) note that wages tend to be higher for workers with higher non-cognitive skills, while people with low non-cognitive skills are significantly more likely to become unemployed. In contrast, when Cunningham and Villaseñor (2016) reviewed 27 studies about the skills-demand profiles of employers in developed and developing economies, they found a greater demand for socio-emotional[13] and higher-order

---

[13] Socio-emotional skills are behaviours, attitudes and traits that are considered necessary complements to cognitive skills in the production process.

cognitive skills[14] than for basic cognitive[15] or technical skills[16]. This evidence is remarkably consistent across the world regarding which skills are demanded by employers.

Due to the importance of workers' behavioural characteristics and to analyse these skills, broader typologies have been recently adapted to measure more of these skill dimensions. For instance, Green (2011) notes that contemporary approaches favour the categorisation of cognitive[17], physical and interactive skills[18] [19].

Other useful typologies for considering workers skills include "basic", "generic" or "specific" skills. The adoption of one or another of these typologies depends on the focus of each research study. Nonetheless, some typologies are not well defined or are inconsistent and should be avoided. This is the case for typologies such as "core" and "non-core" or "hard" and "soft" skills. The concept of "soft" skills might refer to work attitudes and social interaction (e.g. communication). This terminology leaves the wrong impression that "soft" characteristics are easier to learn, or less critical than "hard" skills. However, these terms are misleading because "hard" skills might be no more challenging to generate than "soft" skills, and, indeed, the inverse is also true and "soft" skills might have an important effect on different outcomes, compared to other cognitive or detailed technical skills (Cunningham and Villaseñor, 2016). Thus, these kind of typologies are not sufficiently precise and should not be considered for social analysis (Green, 2011).

However, even well-defined measures of workers' skills have limitations and these dimensions are not always easy to measure given their subjective nature. The categorisation of skill groups depends on a researcher's criteria and is constantly under debate. The important issue that can

---

[14] Higher-order cognitive skills comprise the capacity to deal with complex information processing. These tasks include such as critical thinking, application of knowledge, analysis, problem-solving, evaluation, oral and written communication, and adaptive learning.

[15] Basic cognitive skills comprise fundamental academic knowledge and comprehension, including literacy and mathematics.

[16] Technical skills are defined as the specific knowledge required to carry out an occupation.

[17] Cognitive refers to areas where thinking activities such as reading, numeracy and IT, among others, are required

[18] Physical skills are task-related, referring to dexterity and strength; and interactive skills comprise all forms of communication, including emotional and aesthetic behaviour.

[19] For a more detail description of other typologies used to categorize the behavioural characteristics of workers see Green (2011).

be taken from this debate is to select typologies that do not lead to the misinterpretation of a study's results.

1.2.4.3 Skills as attributes of jobs

Alternatively, to the above workers' skills approach, other typologies focus on the attributes of jobs rather than the attributes of a person to measure job complexity. More complex activities require greater skills (Attewell, 1990; Green, 2011), such task-based typologies have become widely used in the labour market economic literature because these typologies provide a framework to describe processes and changes of job tasks, such as job polarization[20] and the effect of implemented new technologies in the occupational structure (Acemoglu and Autor, 2011; Autor and Dorn, 2012).

Occupation classifications appear to be the most common task-approach used in the economic literature. According to the ILO (2012b, p.59), an occupation can be defined as a "set of jobs whose main task and duties are characterised by a high degree of similarity". Occupational groups or titles are constructed by a group of experts who survey different workplaces and observe workers doing their jobs. For instance, the ILO consult different workers' and employers' organisations, specialised agencies and stakeholders to keep the International Standard Classification of Occupations (ISCO) updated (ILO, 2012b)[21].

Nevertheless, this occupation approach has limitations. Within occupations, skill levels or the kinds of skills being utilised can differ depending on the sector, the company size or by country (Dickerson et al. 2012). Moreover, occupation classifications are not updated as fast as labour market changes. For instance, the ISCO has been updated approximately every ten years; yet, among these processes and periods many changes in terms of skills can occur. So, prevailing occupation classifications can be found to be obsolete when analysing actual labour market skills.

---

[20] Job polarization consists of a decline in the employment share of middle-skilled cognitive and manual jobs characterized by routine tasks.

[21] While different occupational classifications exist, like the SOC (Standard Occupational Classification) in the US, every system of classification agrees with the ILO's basic definition of an occupation. The main differences emerge in the grouping of each occupational category.

Another limitation worth considering, is that most occupation classifications do not take into account personal features, such as attitudes, traits and values. Occupation classifications describe the hierarchical structure of occupations and the tasks required within each one, thereby ignoring the personal characteristics of workers. An exception can be seen in the O*NET system in the US, which contains information on hundreds of standardised and occupation-specific descriptors. It describes occupations in terms of the knowledge, skills, and abilities required by workers, as well as how the work is performed in relation to tasks, work activities, and other descriptors (onetcenter.org, 2016).

Given the above labour market concepts such as supply, demand, unemployment, informal economy, and skills, among others, the literature has provided a theoretical framework with which to understand the labour market dynamics of interest for this paper. The following two sections present the main theoretical model for this study to explain why skill mismatches might arise, their relevance, and the consequences of this phenomenon on labour market outcomes such as unemployment and informality.

1.3 How the labour market works under perfect competition

This section describes a labour market and its main outcomes, such as unemployment, wages, etc., under the assumption of perfect competition. At an early stage, to analyse the matching problem between the demand for skills and the supply of skills, scholars in the field of economics have developed a basic theoretical framework based on the assumptions of perfect competition (Cahuc et al. 2014). A framework which outlines that employers are faced by a certain need for labour services (a derived demand, based upon the demand for their product) and these employers create job offers with certain requirements (skills), and that existing employees and new applicants with those characteristics accept the job when the wage offered is more than their reservation wage[22].

---

[22] Capelli (2015) points out that another theoretical framework exists to understand the relationship between labour supply and employer demand. Employers can select general skills at entry-level positions, and train their employees over a working lifetime to develop specific skill needs for the company. However, the same author notes that this approach has become less plausible in recent years because employers tend to hire applicants who already have the specific skills they require.

1.3.1 Labour demand

The labour market works under perfect competition when employers and workers are perfectly informed about the quality and the price of "labour", and all agents are price-takers (which means that there are no monopolistic/monopsonistic powers). Given these assumptions, what defines a labour market can be expressed as follows.

On one side, picture a representative firm which produces goods and services by using two inputs Labour (L) and Capital (K) at a certain technology level. Consequently, the production function of representative firm is given by:

$$Y = F(L, K)$$

Where Y denotes the physical output of the firm and pY is its value-added, where p is the market price of the product. The cost of labour used in production takes the form of wages and other on-costs, such as National Insurance (the price per hour of hiring a unit of labour services), while the cost of capital is the price of renting a unit of capital[23].

In the short run, when capital is fixed, the marginal product of labour falls as the number of individuals employed rises. The initial condition for employing anyone at all, is that the value of the marginal product of the first worker exceeds their going wage; if so, the firm expands on its number of employees until the marginal return to the last unit of that labour equals the marginal wage (cost of labour):

$$pF'(L) = w$$

On the other side, there are a large number of workers which offer a certain quantity of labour and will receive a wage if they are hired.

---

[23] This is a first approximation because it may be cheaper for the firm to buy capital goods to avoid paying profit to those who rent the goods out. The rental price is the rental firm's estimate of the forgone interest, plus depreciation of the capital plus their profit.

1.3.2 Labour supply

The utility function of a representative worker is composed of two parameters: first, income (R)[24] which is equal to their wage (times the number of hours worked if the worker is hired, and zero if the worker is not hired)[25]; second, the individual's leisure time (the number of hours not spent at work). There is decreasing marginal utility on income (which is spent on goods and services or saved) and leisure time; so, the line in the utility function that joins combinations of income and leisure to yield a given level of utility (i.e. an indifference curve) is convex to the origin (zero income and leisure). Indifference curves further from the origin are associated with higher levels of utility.

1.3.3 Market equilibrium

An equilibrium in the market is achieved where the upward sloping labour supply curve cuts (in other words, it equals) the downward sloping labour demand curve at a certain level of wages (w*) (labour supply equals demand.). Only individuals whose reservation wage, θ (reflecting their disutility of work), is greater than the equilibrium wage, do not participate in the labour market (inactive). In a perfect competition model, as there is perfect information in terms of the supply and the labour demand, all individuals who wish to participate in the labour market (θ ≤ w*) will find a job, and firms will find a worker to fill their vacancies.

The model does not explicitly talk about the role of skills in the labour market; yet, it is relatively easy to incorporate this aspect into the labour market model. As mentioned in Section 1.2 above, Mincer (1958) and Becker (1962) introduced the idea that education is an investment into the economic model. Thus, education makes an individual more productive and might create wage differentials.

To be more specific, people can choose a general level of training (i) which will increase their production function: y(i). Firms will demand workers with a certain level of training (i) until the marginal benefit of using one unit of that labour equals the marginal wage: w(i). Consequently, the wage of a worker, w(i), will be a function of the level of qualification, all other things being equal, and the possibility of a higher wage acts as an incentive to train. Thus, individuals will

---

[24] For simplicity it is assumed that other forms of income do not exist.

[25] It is assumed that the total workers' incomes are consumed by different goods and services.

train until the marginal cost of training equals the marginal return of this investment. Once more, under perfect competition assumptions, an equilibrium is reached when labour demand equals the supply of labour, and all individuals who wish to participate ($\theta i \leq wi^*$) will find a job.

Therefore, one of the most remarkable results from this model is that under perfect competition there is no structural unemployment, instead all workers receive a wage (w^*) at their level of employment (E^*) (Figure 1.3). Nevertheless, there is a possibility that unforeseen impacts on the supply of labour might create disequilibrium in the short run (Bosworth et. al., 1996, p.200). For instance, as shown in Figure 1.3, improvements in technologies such as computers might increase the demand for people who know how to use that technology (from D_0 to D_1), and consequently wages will rise from W_0 to W_1. This situation might create a scarcity (E_1^*-E_0^*) of those people for a period. However, as all agents are (somehow) well informed, workers will start offering labour according to employers' requirements. Workers will find a job according to their characteristics (skills), and employers will find workers according to their requirements. Hence, the unemployment rate remains comparatively low under perfect competition, and there are no barriers that force a worker to work involuntarily in the informal economy.

Figure 1.3: Labour market equilibrium under perfect competition
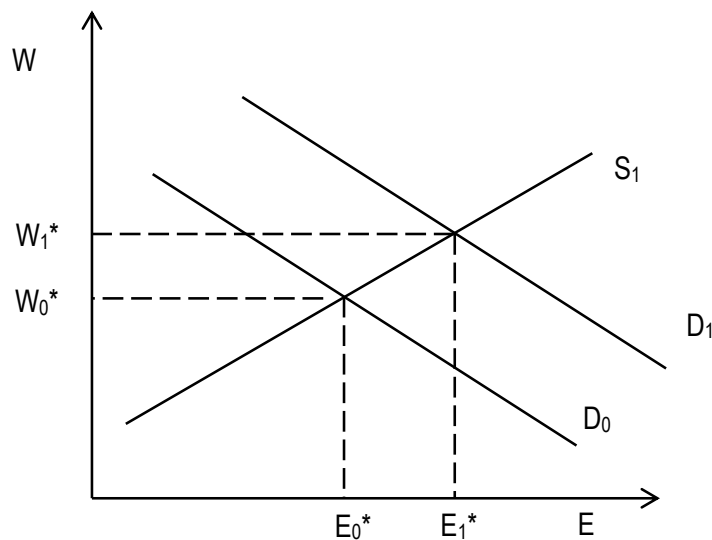
1.4 Market imperfections and segmentation

Developing the above ideas further, Section 1.4 explains how imperfect information (e.g. labour market failures) might increase skill mismatching, and, consequently, it might create labour market segmentation between formal and informal workers along with a comparatively high unemployment rate.
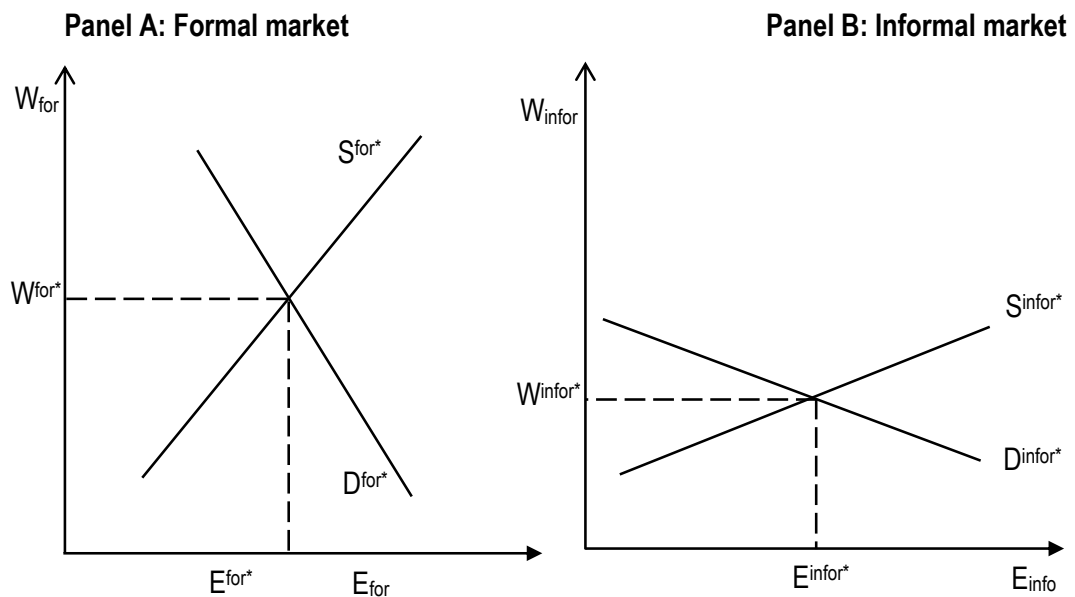
1.4.1 Segmentation

The above assumptions—about perfect information (defined in Section 1.3), where all agents in the model are price-takers—are too simplistic (Garibaldi, 2006). An extended literature has shown that the high incidence of informality in countries such as Colombia can be due to labour market segmentation (Doeringer and Piore, 1971; Reich et al. 1973) (see Section 2). Specifically, barriers might exclude some workers from the comparatively high productive sector (e.g. formal sector) and drive those individuals excluded to a more disadvantaged sector such as the informal market (Gambin et al. 2016; Palmer, 2017).

This duality of the labour market is represented in Figure 1.4. Panel A depicts the more productive formal market sector in which equilibrium wage is $W_{for}*$ at a level of employment $E_{for}*$. While panel B illustrates the more disadvantaged segment in which the equilibrium wage is $W_{inf}*$ at a level of employment $E_{inf}*$.

By comparing panels A and B, two aspects arise. First, labour demand and supply in the formal sector is comparatively more inelastic than the informal sector. This result reflects the fact that in the formal market there are more labour regulations (such as minimum wages, non-wage labour costs, etc.), and more training time, among other entry costs, that make supply and demand less responsive to changes in wages than in the informal sector. Second, wages in the formal sector are higher than in the informal (disadvantaged sector) ($W_{for}* > W_{inf}*$); consequently, this outcome shows that there are incentives to being part of the formal sector. However, there are some barriers that prevent people entering the more advantageous segment of the labour market.

Figure 1.4: Labour market segmentation

Panel A: Formal market　　　　　　　　　　　**Panel B: Informal market**

Source: Bosworth at al. (1996, p.199)

The economic literature reveals several barriers that might explain this labour market segmentation (Reich et al. 1973). One of these barriers is the imperfect information that potential workers possess about the skills required to fulfil employers' requirements. Imperfect information might explain why even when there are incentives (e.g. higher wages) to belong to the formal segment of the labour market some workers remain outside of this more advantageous market, while some vacancies remain unfilled[26]. Thus, as $W_{for}^* > W_{inf}^*$ and the labour conditions $_{for}$ > conditions $_{inf}$, there is an incentive for workers to develop the skills to transfer from the informal to the formal sector, although doing so might take time.

1.4.2　Imperfect market information

Perfect information is one of the necessary conditions to achieve a market under perfect competition. This assumption supposes that all workers know the particularities (e.g. skills required, wages, among others) of all available jobs, and they only need to decide the quantity (number of hours) of labour offered that they are prepared to work, while firms know the characteristics of all potential workers and can choose the one who most suits their job

---

[26] As will be shown in more detail in section 2, the evidence suggests that this situation is prevalent in countries such as Colombia.

requirements. However, labour market failures arise due to imperfect information which occurs when the agents in the economy (in this case employers, employees and training centres) are not fully informed about the price or quality of the product which they are going to buy or sell. As a consequence, agents might not make optimal decisions (Stiglitz et al. 2013), and this limitation might create phenomena such as skill mismatches. In particular, a skill mismatch occurs when imperfect information exists in the job search process or the workplace about the particularities of jobs, mismatches that misalign labour demand and labour supply for skills (UKCES, 2014). These phenomena can acquire different forms, such as skill gaps, skill surpluses and skill shortages, with various consequences on the economy such as unemployment, informality, job dissatisfaction, among others.

Once a job match has been completed, employers can realise that their current employees have need of more skills to be completely proficient in their jobs; this problem is called a skill gap and considered part of the phenomenon of skills mismatch[27]. Nevertheless, the definition of skill gaps per se does not capture the entire skill mismatch phenomenon. For instance, a skill surplus might occur within workplaces. This term refers to a situation where a certain job does not require the highest level of an employee's competences (McGowan and Andrews, 2015)[28].

However, given the multiple configurations that the skill mismatch problem encompasses and the labour market data available to analyse an economy such as Colombia, hereinafter this study will focus on skills shortages. This term refers to the issues that arise in the job searching process when there are no applicants, or applicants do not have the minimum level of skills required to carry out the tasks required by employers. There is a skill shortage when the labour supply lacks

---

[27] Several economic studies have shown the importance of skill gaps in the economy. For instance, in one Irish-based study McGuinness and Ortiz (2016, p.19) suggest that the phenomenon of skills mismatch increases labour costs by approximately 25%, and thus negatively affects the competitiveness of Irish firms.

[28] According to Green and Zhu (2008), graduate over-qualification (which is a way to measure skill surpluses) was about 33% in the UK in 2006. This underutilisation of labour supply creates a misallocation of education and training resources; it increases job dissatisfaction (underemployment) and employee turnover, which might be due to a loss of pay from being overqualified (Green and Zhu, 2008; Okay-Sommerville and Scholarios, 2013).

skills in relation to what employers currently demand to fill their vacancies (Green et al. 1998)[29] [30].

Claims of skills shortages have been made globally. For instance, the European Company Survey for Spring 2013 report that around 39% of firms in Europe experienced difficulties in finding workers according to skills requirements (Cedefop, 2015, p.20). Similarly, the Manpower Group (a well-known international consulting firm) carries out the Talent Shortage Survey, where employers around the world are asked if they have difficulties in filling their jobs (Mazza, 2017). As reported in 2016, due to skills shortages, 40% of the companies interviewed worldwide faced difficulties to fill their vacancies (Manpower, 2016). However, in countries such as Colombia this phenomenon is even larger (as will be shown in more detail in Section 2).

The human capital framework in economics has developed different theories to take into account the possibility of imperfect information, and to explain labour market outcomes in a more realistic way. Search and matching theory, for example, has become one of the most prominent theories to explain skill mismatches and their relation to unemployment (Andrews et al. 2008). This model states that vacancies and workers are heterogeneous in terms of one characteristic such as skills. However, to obtain information about the price and the quality of labour can be costly, and not everyone has access to that information, and this limitation might affect the behaviour of workers and firms.

With imperfect information, the opportunity cost (θ parameter) is not the only relevant parameter to determine if a person is employed or not. In addition, individuals need to devote time to find a job and firms might need to wait or search actively for the candidate that suits their requirements. Thus, included here is the possibility that the labour market does not instantly correct mismatches such as skill shortages (hereinafter skill mismatches refer to skill shortages). The efficiency at which the market matches vacancies and workers depends on the matching function (the formation of new relationships such as job formation), which can be expressed as follows (Mortensen and Pissarides, 1994):

---

[29] This definition excludes other causes of shortages such as firm size and a lack of union recognition, among other causes (Green et al. 1998).

[30] Cárdenas (2020a) discusses the different possible ways to measure skills shortages.

$$m = m(u, v)$$

Where v represents the number of vacancies, u represents unemployed workers and m the rate of job matches (number of people hired, and vacancies filled) over a given time period. Moreover, m is assumed to be homogenous of degree one, which means that if u and v are doubled, the number of matches (m) will increase by the same proportion.

Using equation one can derivate the probability that a vacancy is filled:

$$q = \frac{m(v,u)}{v}$$

As vacancies are filled at the Poisson rate, equation two can be expressed as follows:

$$\frac{m(v,u)}{v} = m\left(\frac{u}{v}, 1\right) \equiv q(\alpha)$$

Where α is v/u, and it is interpreted as labour market tightness—an indicator to identify possible difficulties to fill vacancies, or whether it takes a relatively long time to fill an available job.

Employees also make decisions about educational (skills) investments and where to look for a job according to available information. Subsequently, job opportunities reach jobseekers with a certain probability given by the following:

$$p = \frac{m(v,u)}{u} = \frac{v}{u}m\left(\frac{u}{v}, 1\right)$$

Thus, the probability that a worker finds a job and a vacancy is filled is a function of market tightness, which depends on the quality of labour (skills) offered and demanded—among other characteristics. Individuals whose skills are in demand will find a job of a certain quality, such as health insurance, vacations, etc. (e.g. a formal job), in a relatively short period, and vacancies will be filled.

However, to access and process all relevant employers' requirements could be costly in terms of time and money for a worker. Vacancies are offered in different places, such as newspapers or online job portals, and the detailed information available on them might restrict the numbers of job advertisements that a person screens to make decisions about which roles to apply for. Also, individuals might not have access to, or not use certain sources displaying vacancy information. Consequently, workers' decisions can be based on imperfect information, hence they might or might not properly anticipate an employer's requirements to fill certain vacancies (Mortensen, 1970).

Therefore, according to employers' requirements a lack of skills (which decrease employment probabilities) might affect the labour market matching function, and create labour market segmentation. If the likelihood of finding a formal job is relatively low (which might mean that companies are not demanding the skills some workers have attained) it can take time to find a job. Individuals whose skills are not in demand in the labour market have two options: 1) continue searching for, or create a job for themselves through self-employment, or, 2) take an informal job as a way to earn an income and fulfil personal and family responsibilities. Those individuals who value an informal job more than the expected value of searching and taking a job in the formal sector will be part of the informal economy (Albrecht et al. 2007).

Unemployment and segmented labour might arise as a result of imperfect information about skills demanded by employers (barriers). One segment of the market offers formal jobs, and the other segment offers unregulated jobs with comparatively lower wages (evidence of this can be seen in Section 2), and without the guarantee of social benefit, among other disadvantages. Therefore, the inability (through a lack of information) to train in skills needed by the formal sector locks individuals into the informal sector.

From another aspect of the labour market, firms might not gather perfect information about the skills possessed by potential individuals and where they can be found (Desjardins and Rubenson, 2011; Oyer and Schaefer, 2010). According to this view, employers will hire an

individual when the expected value of matching that individual exceeds the cost of posting a vacancy[31] [32] (Burdett and Smith, 2002).

As a consequence, hiring is an important and costly selection process for heterogeneous productive individuals and firms, and its efficiency depends on the research behaviour of employers, job searchers and the information available to them (Banfi & Villena, 2018). In this sense, companies can face some difficulties in finding people that meet their requirements. Due to that, they spend significant resources on advertising, posting job vacancies and screening to select appropriate workers (Autor, 2001).

Even with those strategies in place, it is possible to reach a situation where unemployed or informal workers with certain characteristics (skills) are willing to work in formal jobs and vacancies available to be filled. This situation can occur because the skills possessed by job seekers are not those required by the companies resulting in skill shortages (or a skill mismatch). Thus, a share of unemployment and informal economy issues can be explained by a mismatch between supply and demand in terms of skills (Gambin et al. 2016; Palmer, 2017).

Provided that companies require different skill combinations and workers have restricted access and limited capacity to respond to those requirements, one straightforward solution to tackle this phenomenon and its consequences is to lower the cost of having (relevant) information about the current labour demand for skills. By doing so, workers have the proper insights about current job roles, which might shape their decisions to acquire skills according to employers' requirements. The matching function will become more efficient if workers have less imperfect information about the employers' needs, and thus unemployment and (involuntary) informality will be reduced.

---

[31] When the cost of posting a vacancy exceeds the profit to be gained from the match, employers do not post vacancies (Burdett and Smith, 2002).

[32] Other models also recognise that employers might not possess perfect information about workers' skills. For instance, Spence (1973) developed a job-market signalling model where employers are not sure about the "productive capabilities" (skills) of a potential employee. To overcome this issue, employers believe that credentials such as higher education are positively correlated with a worker's "productive capabilities". Consequently, potential employees need to send a signal about their skill levels to potential employers by acquiring credentials. In this case, credentials are considered as a proxy to measure skills and help employers and employees in the hiring process.

However, until this point, nothing has been said about the role of education and vocational education (VET) systems on the labour market. Educational and VET systems are one of the main ways to prepare (deliver skills) to people for work (Green 2011; OECD, 2014a), and they also might be affected by a restricted access and a limited capacity to analyse and anticipate employers' requirements. Consequently, it is almost pointless that workers have the right information about current employers' requirements for skills; that is, if there are not educational and training systems in place that provide them. As mentioned by Cedefop (2012a), the capacity of a good "match" between employers and employees depends considerably on how education and VET systems respond and adapt to employers' requirements. In consequence, the better it is understood how to adopt, and develop a capacity for, this understanding into educational and training programs, and into workers decisions the better the match will be between workers' skills and vacancies (See working papers Cárdenas, 2020a and Cárdenas, 2020c).

1.5 Conclusion

This section has outlined the basic labour market framework in order to properly use vacancy data and address unemployment and informal economy phenomena. The labour market is a space where workers (labour supply) offer a quantity of "labour services" with certain qualities to fill vacancies, and employers (labour demand) hire that merchandise at a certain price (wages). In terms of the labour market, people can be divided into three groups: 1) workers whose labour services are bought by employers in the formal economy, 2) workers employed in the informal economy which are characterised by lack of social security, limited access to the financial system, etc., and, 3) workers that offer their labour services but are not hired by employers (unemployed). The size of each group depends on different elements. However, the literature discussed in this section stresses that skills are a relevant factor to determine labour outcomes, such as unemployment and the size of the informal economy.

Due to skills importance and its multiple dimensions (e.g. qualifications, competences, education, aptitudes, etc.) the term skills can be defined in different ways; nevertheless, most of those definitions link the task complexity attached to each job and the characteristics that each worker needs to successfully carry out job tasks. For this reason, this document considers a skill to be any measurable quality that increases workers' productivity, and can be improved by training and/or development. With this definition, it is possible to analyse and subtract information

from vacancy information to construct more reliable indicators of the level of skills required by employers (e.g. qualifications) and address possible skill mismatch issues.

Under perfect competition, the over or undersupply of skills (skill mismatches) only arise over the short term, and have relatively small implications for unemployment and informality rates (exclusion). However, the conditions required for perfect competition rarely exist because agents have imperfect information about offered and demanded skills. This imperfection in the labour market might create a situation where there is a lack of skills in relation to what employers currently require to fill their vacancies: a skills shortage. Skills shortages might create labour market segmentation where workers with the "right" skills have more probabilities to belong to the formal economy, while workers without the "right" skills (according to demand) have more chances of being in the informal economy or unemployed. Consequently, unemployment and the informal economy might increase and/or persist over time.

The above skill mismatch problem involves the coordinated actions of at least three different agents in the economy: employers, workers, and educational and VET systems. The level of coordination between these three groups determines the extent of skill mismatch. This coordination depends on the availability of information about skill requirements and the capability of the workers to process and adopt that information into their decisions, as well as the availability of educational and training systems.

In this sense, one way to tackle the skill mismatch phenomenon is to gather information about labour demand, and extract meaningful information to address workers' decisions, and educational and VET systems' decisions, according to different companies' requirements. New technological developments offer new opportunities in this respect. This particular theoretical framework and straightforward solution might be especially useful for countries such as Colombia where: 1) informality and unemployment rates are high, 2) complaints about skill shortages (skill mismatch) are relatively high, 3) information about companies' requirements is available from resources such as job portals, and, 4) Educational and VET institutions have difficulties to adapt their programs according to labour demand.

For these reasons, the next two sections show that in the context of the Colombian economy, novel sources of information and data analysis regarding the labour demand for skills might have

an important affect on public policy, and reduce unemployment and the informal economy at a lower cost in terms of time and monetary sources.

2. The Colombian Context

2.1 Introduction

Skill mismatches are a widespread phenomena that have strong implications on unemployment and informality rates, among other variables[33] (McGuinness and Pouliakas, 2017). Nevertheless, some countries display a higher incidence of these issues, which might have severe effects on local labour outcomes. This section presents evidence that Colombia is one country where the degree of skill mismatches (skill shortages), unemployment and informality is relatively high. However, public policies that tackle those outcomes are limited, and, consequently, this makes Colombia a relevant case of study to develop novel ways to analyse and reduce skill mismatches.

Based on the concepts discussed in section 1, this section, firstly, provides an overview of the main characteristics of the Colombian labour market and its evolution over time. Secondly, it shows that the issue of skill mismatches and their possible incidences has a relatively high impact on the national economy, which needs to be addressed by public policies. Subsequently, it explains the importance of maintaining systems with accurate labour market information to address these phenomena. Finally, it is argued that the lack of information about skills requirements together with an institutional disarticulation, especially in Colombia (and developing countries), makes it difficult to develop well-orientated public employment policies to deal with the skill shortages phenomenon. For that reason, there is a need to find novel solutions to provide systematically accurate information, and analyse employers' requirements and possible skill mismatches.

---

[33] See section 1 for a more detail discussion.

## 2.2 The characteristics of the Colombian labour market

This section describes the main characteristics of the Colombian workforce and labour demand in order to present the structure of one of the most relevant labour market issues that Colombia has been facing: unemployment and informality.

2.2.1 Labour supply

Figure 2.5 shows the structure of labour supply, in Colombia, at a macroeconomic level. In 2016, the Colombian working age population was composed of around 37,851,500 people, while 64.4% of the working age population participate in the labour market (approximately 24,405,000 people) and represent the current Colombian labour supply.

As mentioned in Section 1, labour supply is composed of: 1) people in the working age population that do not have a job but are looking for one (unemployed), and, 2) people who are in the working age population and hired by employers (employed) and the self-employed. According to Figure 2.5, around 90.7% of the economically active population (EAP) have a job, however, 5,776,750 people work in informal jobs. In addition, around 9.2% of the Colombian workforce is unemployed.

Figure 2.5: Colombian labour structure
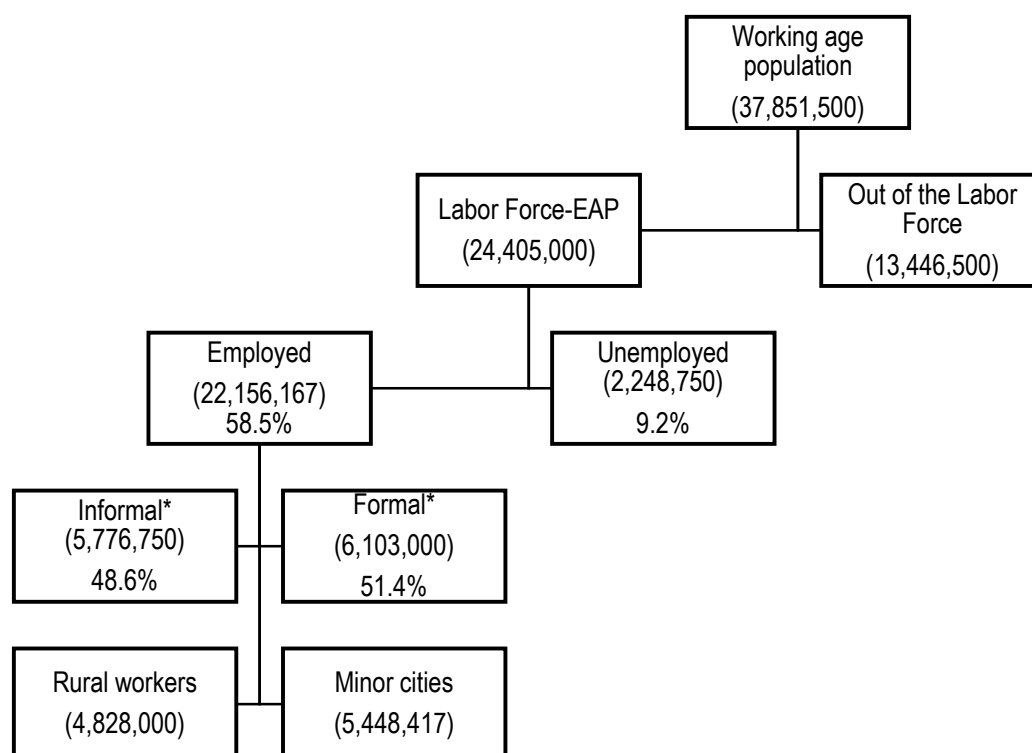
Source: DANE 2017a. Own calculations.

* Informality is only calculated for urban areas. As explained in Section 1, for rural areas, the definition of informality (a company's size) is not accurate. By the time this section was written, there was not an official measure of informality for those rural areas.

These indicators highlight a key point: in Colombia, the labour participation rate is relatively high. Indeed, it is 2.6 percentage points above the Latin-American average (ILO 2016, p.29). However, only 51.4% of the employed population has a formal job (Figure 2.5).

Moreover, high unemployment and informality rates are persistent over time in Colombia. As is shown in Figure 2.6, in 2001[34] the annual national unemployment rate was approximately 15%, and the participation rate was 62.4%. In the same period, the informality rate decreased from 50.4% in 2006[35], to 47.5% in 2016 (DANE, 2017a). This result means that during the last fifteen years more people have participated in the Colombian labour market. Formal labour demand

---

[34] In 2001, there were changes in the household survey methodology, which affect the comparison of labour market indicators before 2001.

[35] Due to methodological changes, informality rates are not comparable before 2007.

has absorbed a considerable proportion of labour supply to the point that unemployment and informality rates have declined, even with more people participating in the labour market.

Figure 2.6: Participation, employment, unemployment and informality rates trends 2001 - 2018
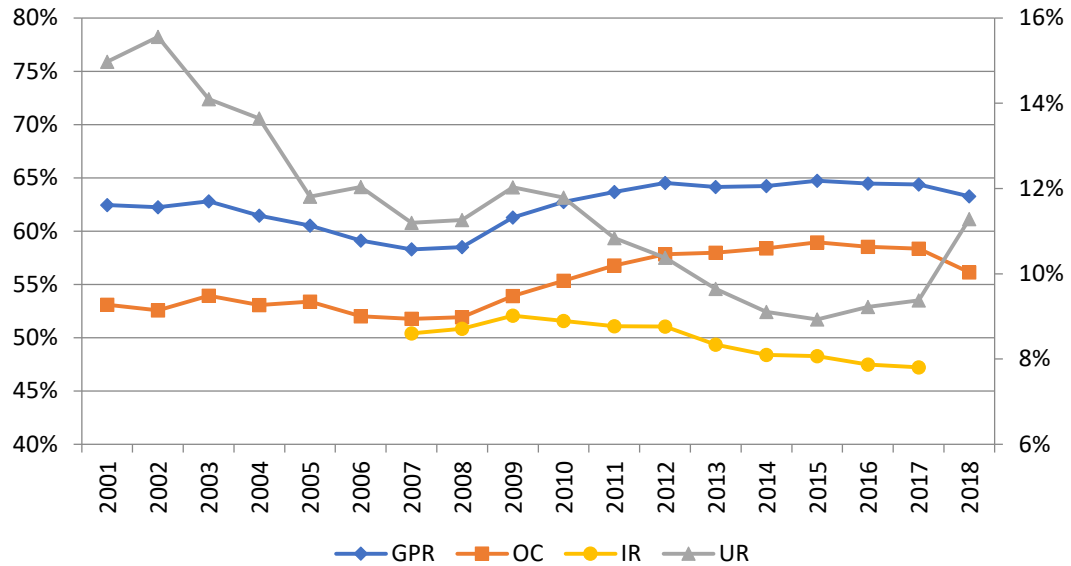
Source: DANE 2017a.

*Unemployment rates are graphed on the right-hand scale

However, Colombia took a relatively long period (15 years) to decrease unemployment and informality rates to 5.8 and 2.9 percentages points, respectively. Additionally, informality and unemployment trends changed in 2017 and 2018, where the unemployment rate increased by 0.2 and 0.3 percentage points respectively, and informality rates stagnated around 47%. Although these informality and unemployment rates have declined in recent decades, Colombia's unemployment and informality rates are above the World average, and even above the Latin American average (World Bank, 2018a). In particular, in 2015 Colombia was the second economy in the Latin American region with the highest unemployment rate (only surpassed by Brazil), and its informality rate was around 1.4 percentage points more than the regional average (ILO, 2016).

Moreover, informality and unemployment do not affect all workers equally. Table 2.1 shows the general characteristics of the Colombian workforce between 2016 and 2018. According to the first column, 56.7% of formal workers are male, while the second column indicates that 53.9%

of informal workers are male. This result is because in the Colombian labour market more men are working than women. However, the presence of women in the informal market is 2.8 percentage points higher than women in the formal market. Moreover, the third column shows that 55.7% of unemployed individuals are women. These results suggest that unemployment and informality issues are comparatively higher for women than for men.

According to the age distribution of all workers (males and females combined)[36], 30.5% of formal workers were less than 29 years old, compared to 23.3% of informal workers. In contrast, only 4.6% of formal workers were over the age of 58 years, compared to 14.4% of informal workers. However, almost half (49.1%) of Colombian unemployed population were less than 29 years old, followed by people between 29 and 58 years old (45.7%), and over 58 years old (5.2%). Consequently, older Colombian workers tend to be more exposed to informality, while young workers are more likely to experience unemployment issues.

The educational distribution shows that higher the level of education (lower and higher vocational education, graduate or postgraduate) the higher the proportion of formal workers is compared to the proportion of informal workers. Moreover, more than half of unemployed individuals in Colombia have just a high school certificate. Indeed, most formal and informal workers and those who are unemployed only have a high school certificate (42.5%, 53.4% and 53.3%, respectively).

The monthly average wage of a formal worker is around 1,511,246 pesos (around £377), while the average salary of an informal worker is about 910,508 pesos (around £227). In accordance with Mondragón-Vélez et al. (2010), a formal worker earns 1.6 times more than an informal worker. In contrast, an informal person works 3.4 hours less per week than a formal worker. More than one-third of workers are underemployed because of the underutilization of their skills (skill surpluses—see section 1). However, this percentage is higher for informal workers.

Around 31.9% of formal workers are in companies related to community, social and personal service activities, followed by the wholesale and retail trade, hotels and restaurants (18.9%) and manufacturing (16.0%). In contrast, most informal workers are in wholesale and retail, or in the hotels and restaurants sector (42.1%), followed by community, social and personal service

---

[36] The age distribution presented in Table 2.1 follows the age bands indicated by DANE, which define a person as a young if she/he is less than 29 years old.

activities (14.5%) and transport, storage and communications (11.5%). Additionally, most unemployed individuals used to work in the wholesale and retail trade, hotels and restaurants sector (30.0%), community, social and personal service activities (24.9%), construction (11.3%) and manufacturing (11.3%). Therefore, the sectors that concentrate most of the informal and unemployed people are the wholesale and retail trade, hotels and restaurants sector, and companies related to community, social and personal service activities. The last row of Table 2.1 shows that the average duration of unemployment was around 4.7 months (20.2 weeks), the Colombian duration of unemployment is above average compared to the average of the OECD countries which was 3.6 months between 2016 and 2017[37].

---

[37] See https://stats2.digitalresources.jisc.ac.uk/Index.aspx?DataSetCode=AVD_DUR

Table 2.1: Characteristics of the Colombian workforce

Variables	Formal workers	Informal workers	Unemployed
% General characteristics			
Males	56.7%	53.9%	44.3%
Less than 29 years old	30.5%	23.3%	49.1%
Between 29 and 58 years old	64.9%	62.3%	45.7%
More than 58 years old	4.6%	14.4%	5.2%
% Educational levels			
None	7.0%	29.1%	14.5%
High school	42.5%	53.4%	53.3%
Lower and higher vocational education	21.3%	11.3%	18.4%
Graduate	19.5%	5.2%	11.4%
Postgraduate	9.8%	1.0%	2.4%
Labour market outcomes			
Mean wage (Colombian pesos)	1,511,246	910,508	-
Mean hours worked per week	47.2	43.8	-
Underemployment	31.7%	35.6%	-
Agriculture. hunting and forestry	2.5%	5.2%	3.4%
Mining and quarrying	1.0%	0.2%	1.0%
Manufacturing	16.0%	11.0%	11.3%
Electricity, gas and water supply	1.3%	0.0%	0.5%
Construction	5.3%	8.4%	11.3%
Wholesale and retail trade, hotels and restaurants	18.9%	42.1%	30.0%
Transport, storage and communications	6.7%	11.5%	6.8%
Financial intermediation	3.3%	0.4%	1.6%
Real estate, renting and business activities	13.1%	6.7%	9.2%
Community, social and personal service activities	31.9%	14.5%	24.9%
Duration of unemployment (weeks)	-	-	20.2

Source: DANE-GEIH. Own calculations.

The results from Figure 2.1, Figure 2.2 and Table 2.1, confirm that informality is a widespread and persistent problem in the Colombian economy. However, these outcomes can be explained by two different phenomena with different implications for public policy and economic research. As pointed out in section 1, informality be might be explained by "exclusion" and "exit" processes. The first term, "exclusion", refers to the situation where there is labour market segmentation and

barriers which prevent informal workers from taking formal jobs (with state-mandated benefits). In consequence, decreasing those barriers might reduce the size of the informal sector. The second term, "exit", occurs when workers and firms decide to stay outside of formality when the cost of being formal overcomes the benefits of belonging to this sector. Consequently, to reduce the size of the informality sector is necessary improvements of the state's service provision and re-enforcement of legislation (Perry, 2007). Economic dynamics, legislations, institutions, and so on, differ among countries; these differences might make exclusion mechanisms more relevant in some regions, while exit mechanisms are more important in others.

Even though in Colombia the two views are important (exclusion and exit), the evidence suggests that exclusion mechanisms are more relevant for the Colombian context. According to Perry (2007), a significant fraction of the informal, self-employed population is involuntary so in Latin America. To be specific, the fraction of informal and independent workers who would rather be formal employees is around 40% in Argentina, 59% in Colombia, and 25% in Bolivia and the Dominican Republic. When informal self-employed workers were asked about their motivations/reasons for being in their current job as an independent worker (such as autonomy, flexible hours, could not find a salaried job, higher wages) the main response to working as informal and self-employed was because they could not find a salaried job: 59% in Argentina and 55% in Colombia gave this response (Perry, 2007, p.66). Additionally, Perry (2007) found similar results for informal salaried workers; thus, difficulties in finding a formal salaried job constitute a much higher fraction of the reported reasons for being in informal salaried jobs than the other possible responses. According to the same study, amongst informal salaried workers, 48.4% in Argentina, 64% in Bolivia, 43% in Colombia and 40% in the Dominican Republic reported that they would rather take a salaried job with benefits over than their current job.

In consequence, evidence in Latin America shows that a significant proportion of informal workers would prefer to work in a formal job but cannot find one. Furthermore, the majority of the Colombian unemployed population (36%) reported in 2016 that the scarcity of available jobs, according to their occupation, is the main reason why they stop looking for formal employment.

This evidence reveals a number of relevant facts: 1) informality and unemployment are relatively high in Colombia, even compared to the country's regional counterparts, 2) labour supply trends reveal that both informality and unemployment rates are explained by structural rather than a

cyclical component; that is, there is a significant and persistent portion of people who are looking for a job, however, they are not hired by the Colombian labour demand, 3) most people affected by informality and unemployment phenomena are the following groups: less than 29 years old, more than 58 years old, women, characterised by a low level of education, 4) a significant share of the workforce employed in informal jobs desires to work as formal workers.

As discussed in section 1, there a set of different reasons that explain why the labour demand is not absorbing the labour force of the groups mentioned in the previous paragraph creating high informality and unemployment rates. To understand the potential causes of the results explained above, it is important to also analyse the Colombian labour demand.

2.2.2   Labour demand

With a GDP per capita of 14,181.406 US dollars in 2016 (World Bank, 2018b) (three times less than the OECD average), Colombia is an economy in which employment is high in the service sector. Indeed, this sector encompassed 57.4% of Colombia's GDP in 2013 and employed around 63% of the labour workforce in 2016 (as mentioned in Subsection 2.2.1). Moreover, most employment is offered by micro, small or medium-size enterprises. According to the ILO (2014), micro-enterprises (defined as units with up to 10 employees) account for 96% of the country's companies, small enterprises (defined as units between 11-50 employees) represent 3%, while medium (between 51 to 200 employees) and large enterprises (>200 employees) represent 0.5% and 0.1%, respectively. Consequently, 80.8% of the Colombian workforce is employed by micro-enterprises and SMEs (small and medium-sized enterprises[38]) and these enterprises contribute to approximately 40% of Colombia's GDP (OECD, 2017a). However, around 60% of those micro-enterprises were in the informal sector in 2010 (ILO, 2014). All these indicators reveal that there is an important informal economy in Colombia that employs a high number of people in the service sector; specifically, in activities related to sales and retail[39]. However, a majority of these people would like a job in the formal sector.

---

[38] According to OECD measures, SMEs refers to companies with fewer than 50 employees, and micro-enterprises which have, at most, 10 employees, or in some cases 5 employees (Stats.oecd.org, 2018).

[39] The national statistical office carries out annually a specific survey to measure the economic activity of companies related to sales and retail because they possess such a high level of importance in the Colombian market.

Many factors might explain why labour demand does not fully utilise the Colombian labour force. For instance, the high cost of hiring is one of the main factors that prevent formal companies from hiring more personnel (Bell, 1997; Kugler and Kugler, 2009; Mondragón-Vélez et al. 2010). Mondragón-Vélez et al. (2010) observe that in the Colombian labour market there are comparatively high non-wage costs (payroll taxes, health and pension contribution, among others), and a high minimum wage relative to the productivity level. These labour market rigidities restrict the formal sector to adapt to the business cycle, thus the size of the informal sector, and unemployment, increases.

Despite the high cost of hiring in Colombia, there is a relatively high vacancy rate. According to the Human Capital Formation Survey carried out by the DANE in 2012[40], around 80.4% of opened vacancies related to sales and retail activities, and 87.6% and 94.4% to the service and industrial sector (excluding sales and retail activities). Moreover, most new vacancies related to sales and retail activities were generated in the area of marketing and sales (68.6%), while in the industrial and services sector (excluding sales and retail activities) most new vacancies were generated in the production area (66.9% and 82.2%, respectively).

Thus, Colombia's labour demand suggests that (even with the relatively high cost of hiring) while there are formal vacancies available there are also a high number of unemployed and informal individuals who are willing to work in formal jobs, but who do not do so. Consequently, there is a mismatch between supply and labour demand.

2.3 Skill mismatches in Colombia

As presented in section 1, skill mismatches occur where the demand for skills and the labour supply for skills are not aligned (UKCES, 2014). In other words, companies require skills that workers do not possess. This misallocation of skills might explain why some countries face high unemployment and informality rates, and, at the same time, a relatively high portion of companies complain about the scarcity of accurate human resources. Consequently, skill mismatches framework might explain a considerable portion of the labour market outcomes in Colombia (as presented in the previous section).

---

[40] See http://www.dane.gov.co/files/investigaciones/boletines/capital_humano/BOL_CH_12.pdf

Globally, Latin America possesses the largest gap between labour demand and supply for skills (OECD, 2017b). In this region, around 44% of companies in 2016 experienced difficulties finding accurately trained candidates (skill shortages) (Manpower, 2016). Consequently, Latin American companies are 3 times more likely than firms in the OECD, and 13 times more likely in the Pacific-Asian region, to experience recruitment difficulties (OECD, 2017b). For Colombia this rate is even worse, as around 50% of companies face problems filling vacancies due to a shortage of skills.

The Colombian Beveridge curve (that depicts the relationship between unemployment and vacancies to determine how well, or not, job vacancies correspond to unemployed workers) illustrates a deep and constant labour market mismatch (Álvarez and Hofstetter, 2014; Blanchard et al. 1989). According to Álvarez and Hofstetter (2014), Colombia has a relatively high level of vacancies and unemployment which suggests that a lack of skills in the workforce (skill shortages) is one of the main reasons for Colombia's labour market mismatches. Moreover, the entrepreneur survey of Human Capital Formation carried out by DANE in Colombia in 2012 shows that around 62.1%, 67.2% and 61.7% of employers in the industrial and service sector, and sales and retail activities, respectively, cited skill shortages[41] as the leading cause of difficulties to find suitable workers. In addition, low productivity/poor performance and lack of specific competences were selected as main reasons to fire workers (around 34.4%, 40.9% and 33.1% in the industrial and service sector, and sales and retail activities, respectively). Thus, a lack of workers' skills is a key problem in Colombia, especially in the service sector. In particular, there is a large shortage of technical specialists, and a surplus of unskilled workers and middle management professionals (OECD, 2015a).

Although the average year of educational attainment has increased to around six years during the last four decades for all age ranges (World Bank, 2018c), Colombia remains a country with relatively low levels of education: in 2012, only 42% of Colombian people between 25–64 years old attained at least their upper secondary school education, around 33 percentage points below the OECD average and just above Mexico in Latin America; whereas only 20% of adults completed a tertiary level of education (12 percentage points below the OECD average) (OECD,

---

[41] Sub-qualified, over-qualified, low performing, gave a bad impression during the interview, lack of candidate experience, lack of reliable information about qualifications and experiences, the candidates did not speak other languages.

2014b). In addition, the Programme for International Student Assessment (PISA) which evaluates education systems worldwide by testing the skills and knowledge of 15-year-old students, reveals a low Colombian student performance in mathematics. Almost 75% of students fail to achieve the baseline level of knowledge in mathematics, which contrasts with the OECD average of 23%. A low proportion of students (around 0.3%) are top performers, 12 percentage points below the OECD (OECD, 2014b). Moreover, based on the Colombian household survey, the "*Gran Encuesta Integrada de Hogares*" (GEIH), only 9% of the working age population during 2014 took a technical or vocational education and training course.

It is not only companies that have observed a large deficiency of skills. Arango and Hamann (2013) consulted an important group of labour market analysts (15 experts) in Colombia about the leading causes of unemployment. The majority (67%) agreed that the skill mismatch between labour demand and supply was the main unemployment cause in the country. Consequently, 60% of the experts recommended strengthening information systems to improve the efficiency of matches between employers and employees.

Thus, there is a generalised consensus between labour market experts and national and international institutions that a lack of skill is one of the main reasons for skill mismatches in Colombia. Consequently, as explained in section 1, one of the main issues that Colombia is facing is that based on the labour market information currently available, people, education and training providers, and the government are making decisions about human capital investments. However, these agents are not accurately anticipating employers' requirements to fill their vacancies. Those workers whose skills are not in demand might choose between being outside of the labour market (being inactive), being unemployed or being employed in the informal sector. Based on the Colombian evidence (discussed above), a high proportion of people select the last two options: the informal sector or unemployment.

At the same time, a relatively high proportion of companies in Colombia complain about the scarcity of workforce according to their needs which leads to the situation where there are vacancies to be filled. However, due to skill mismatches the Colombian labour supply does not have the necessary characteristics to fill these vacancies[42]. This context might explain an

---

[42] See section 1 for a more detailed theoretical discussion about skill mismatches.

important proportion of unemployment and informality rates, and the high relative rate of companies' complains about the scarcity of human resources in Colombia. As a consequence, to reduce unemployment and informality problems the information asymmetries between supply (individuals) and demand (employers) for labour must be addressed. Tackling these problems might have a large positive impact on regions such as Colombia where unemployment and informality rates are relatively high, and there is a large gap between labour demand and supply for skills.

As the OECD (2017b) has pointed out, to tackle informality and improve economic stability Latin American countries such as Colombia should invest in human capital. The same organisation argues that more education in terms of quantity and quality increases a person's likelihood of finding a job and reduces the probabilities of being unemployed or working in the informal sector. Moreover, to guarantee the effectiveness of human capital investments and to avoid any labour market mismatches as described in section 1 (e.g. over-education), governments and other institutions need to promote skills that meet companies' requirements (Gambin et al. 2009; OECD, 2012).

Given the importance of skill mismatches, institutions such as the World Bank (2010), the OECD (2016a), and the ILO (2017b) agree that fostering education and suitable skills (to strength human capital) might have a large positive impact on the main employment problems of Latin America (e.g. Colombia). Thus, it is essential for Colombia to achieve at least the minimum skill levels in its population, and to improve the relevance of education and training systems so as to reduce unemployment and promote well-being (OECD 2015a).

As González-Velosa and Rosas-Shady (2016) mentioned, advance educational and training systems achieve the above by encompassing tools to identify current and future skill requirements for the productive sector. With these tools, curriculum contents can be updated, and the relevance of education and training increased. Consequently, approaches that identify possible skill mismatches when combined with a functional system of active labour market policies can ensure better matches between employers and workers (Escudero et al. 2016).

2.4 An International example of skill mismatch measures

Examples of the above can be found in different regions. As Mavromas et al. (2013) highlight, the most developed approaches to measure skill mismatches (skill shortages) can be found in the UK. For instance, the Migration Advisory Committee (MAC) built 12 indicators[43] of shortage using data for labour demand and supply. In addition, the UK Commission for Employment and Skills (UKCES) and the Department for Education (DfE) carried out a biennial Employer Skill Survey (ESS), which provides insights into the skills problems employers are facing to fill their vacancies and the actions they are taking to solve them. The survey contributes to public policy decisions when addressing the skills challenge and prompting people to adopt relevant skills for the workplace (Vivian, 2016).

Other valuable efforts include the O*NET system launched in 1998 which is updated annually by the US Department of Labour, and Cedefop in Europe which is updated annually under the jurisdiction of the European Union. The O*NET system consults a variety of different resources—such as a national sample of establishments and their workers, occupational experts and analysts, among others—to collect information on hundreds of standardised and occupation-specific descriptors, e.g., knowledge, skills, tasks, work activities, and other descriptors (National Research Council, 2010). The O*NET and its collected data facilitates government officials to understand ongoing changes in the nature of work and their implications on the US workforce. Consequently, it facilitates the government to develop and train the workforce depending on their skill needs. In addition, Cedefop has made important advances towards quantifying skill needs in Europe. For example, the Occupational Skills Profiles (OSP) approach aims to integrate and complement several European sources of skills requirements information in order to provide updated occupational profiles for the region (Cedefop, 2012b).

---

[43] They can be enumerated as follows: percentage change of median real pay; percentage change of median real pay (3 yrs); return to occupation; change in median vacancy duration (1 yr); vacancies/claimant count; percentage change of claimant count (1 yr); percentage change of employment level (1 yr); percentage change of median paid hours worked (3 yr); change in new hires (1 yr); skill-shortage vacancies/total vacancies; skill-shortage vacancies/hard-to-fill vacancies and; skill-shortage vacancies/employment.

2.5 Lack of accurate information to develop well-orientated public policies

In contrast with the advanced systems mentioned in the US and Europe, Colombia does not have these kinds of tools to base their educational and training policies on (González-Velosa and Rosas-Shady, 2016). Some approaches exist to analyse the labour market in terms of skills, but there is not one integrated system of information for skill mismatch analysis (Saavedra and Medina, 2012). Institutions that have tried to measure, directly or indirectly, human capital characteristics have used different statistical approaches and skills concepts.

Since 2006, the Colombian statistics office of Colombia (DANE) has carried out a monthly cross-sectional household survey, the GEIH, to measure the characteristics of the Colombian workforce. The GEIH is nationally representative, and the main source for official labour market information in Colombia. For instance, based on GEIH each month the national government publish the unemployment rate and other relevant labour market indicators for Colombia. In this survey, people are asked about their current level of education and occupation, among other characteristics. As pointed out in section 1, the level of education and the occupation of the labour force are two of the most common indicators to measure skill levels in a country.

However, for the Colombian case, this occupational analysis is limited for two reasons. Firstly, the occupational classification used to classify people's occupations has not been updated since 1970. The DANE uses in its household surveys the Standard Occupational Classification (SOC) which was established in 1970 by the Minister of Labour and Social Protection, and the vocational education and training institution in Colombia (*Servicio Nacional de Aprendizaje*: SENA) (Cabrera et al. 1997). The usage of such outdated classifications might distort any subsequent statistical analysis due to labour market changes and new occupations that emerge or disappear over time. Occupations related to Big Data technologies (machine learning engineers, data scientists and big data engineers) are representative examples, as these kinds of "Big Data" occupations did not exist 50 years ago, yet nowadays these are one of the top emerging jobs on LinkedIn (Economicgraph.linkedin.com). Consequently, the use of outdated classifications such as SOC might lead to misclassification and/or the underestimation/overestimation of certain categories. Secondly, for analysis the occupation variable is aggregated to 2 digits, which means that for statistical purposes DANE aggregates the data into an "occupational area" which groups different occupations together depending on

their qualification level (defined by the complexity of their functions, their level of autonomy and responsibility, their level of education, training and experience) (Sánchez, 2013). However, as mentioned in section 1, the human capital concept has evolved and encompasses different elements—such as socio-emotional, higher-order cognitive, basic cognitive, technical skills, among others—that are relevant for the labour market, and cannot be measured with the usage of an outdated and aggregated classification systems. Consequently, occupational data from GEIH is useful as it provides insights about the general labour market structure, but it does not convey detailed information about skills and important human capital characteristics so as to develop national or local public policies on human resources.

The World Bank carried out the Skills Measurement Program (STEP) to measure skills in low and middle-income countries in 2012, which included Colombia (Pierre et al. 2014). This program consisted of a longitudinal household-based survey and an employer-based survey. Nevertheless, for Colombia only the household survey is available in which people were asked about (self-reported) personality, behaviour, and time and risk preferences, among other personal characteristics, as well as measuring reading proficiency and related competencies according to PIAAC scores to allow international comparison. Questions regarding skills make the STEP a valuable source of human capital information in Colombia. The survey sought to be representative for non-institutionalized people from 15 to 64 years of age, living in private dwellings in the thirteen major urban areas of the country. However, the general sample is composed of only 9,960 people, and after a short questionnaire, a member of the household was randomly selected to answer a more detailed individual questionnaire which contained questions regarding skills. The total number of people who answered the skills modules is about 2,617 (Pierre et al. 2014).

Consequently, one of the main limitations of the STEP is the sample size; indeed, it only represents 0.02% of the target population. Thus, the data sample cannot be disaggregated into different levels (i.e. different occupations) to make national or regional inferences due to the lack of observations. Additionally, the survey has not been updated: the first wave of information gathering was conducted in 2012, and the second wave in June 2014; however, Colombia was not part of the second wave[44]. Thus, the STEP gives valuable, detailed information about the

---

[44] The following countries were included in the second wave: Armenia, Georgia, Macedonia, and Kenya.

human capital characteristics of a reduced sample during a single period in Colombia. Therefore, as noted by the OECD (2017b), the STEP approach can be used as an instrument to understand some of the general structure in the skills performance of people aged between 15 to 64 years old in each country, and allows international comparison, especially with OECD countries. However, as the labour market is dynamic and skills performance change over time, the survey needs to be updated—at least for the Colombian case.

Additionally, both surveys GEIH (DANE) and STEP (World Bank) are based on what people (labour supply) report. Consequently, they do not directly consider one essential part of the labour market: employers' requirements. To analyse labour demand based on what people report in household surveys is limited because it only takes into account the skills or characteristics that people possess for the labour market, but employers' requirements (what is needed to fill their vacancies) remain unknown, which is an important aspect of the labour demand to understand in order to reduce possible mismatches (Autor, 2001; Mavromas et al. 2013).

The DANE carries out sectorial surveys (e.g. industrial, services and sales-retail activities) to measure basic information, such as national account statistics, the composition of production and consumption lines, the amount of labour employed in each sector, among other indicators. Subsequently, these surveys are not designed to obtain detailed information about human capital such as occupational structure, nor the skills required for each position. For example, with regard to human capital characteristics, with these sectorial surveys it is only possible to distinguish the number of people employed by different functional areas (e.g. production, marketing and sales, investigation and development, among others). Additionally, in 2012, DANE carried out other cross-sectorial survey named the Human Capital Formation (EFCH, by its Spanish initials) where companies in the three sectors mentioned above were asked about job training and productivity. Although the EFCH provided valuable insights about job training, selection and hiring practices, and productivity, the data are still aggregated by functional areas and does not capture employers' requirements.

For its part, the institution in charge of delivering vocational education and training in Colombia (SENA) conducts small, voluntary employers' surveys (semi-structured survey questionnaires) in order to identify the occupational requirements of the private sector. However, González-

Velosa and Rosas-Shady (2016) argue that these surveys do not have enough financial resources to guarantee the effectiveness of their results. Indeed, the same authors highlight that employers' survey results are significantly affected by a lack of standard procedures, clarity in their objectives and incentives for companies to participate.

In 2015, SENA surveyed employees and employers to build employability, performance and relevance indices of its vocational programs. SENA tried to evaluate the skills performance of its graduates, such as communication, adaptation to changes, responsibility, teamwork, among others. Around 4,502 people who graduated from that institution (in the second semester of 2013 and in the first semester 2014) were interviewed. In addition, employers that hired those graduated were interviewed (SENA, 2015). The survey attempted to evaluate the content of vocational programs by measuring skill performance in people's jobs. However, even for that purpose, the results from these surveys are limited. Indeed, they are representative of only 13% of the total number of vocational programmes (SENA, 2015), and employers were not asked about their skill requirements to fill vacancies. Moreover, SENA information (microdata) is not available to the public.

Thus, in Colombia the main sources of information used in the analysis of labour demand have come from sectorial (entrepreneur) surveys or household surveys. These data have strengths, such as national standardisation and global representativeness, but the collection of labour demand information through surveys is limited as it can be costly, both in terms of resources and time, to collect. Above all, these sources might not provide enough detailed information about which skills (or occupations) are in demand among different industries or regions (Handel, 2012; OECD, 2016a).

The Colombian law[45] has stated that all job portals and companies need to report their vacancies to the Unidad del Servicio Público de Empleo (UAESPE - Public Employment Service) from 2013. Potentially, UAESPE can provide a vacancy data analysis for Colombia. However, the UAESPE approach has different limitations that affect the robustness of vacancy analysis. First, job portals and companies do not report all the information that describes a vacancy to the UAESPE. There is a predefine format where companies and job portals complete with certain information that partially describes the vacancy. Second, the UAESPE does not know if

---

[45] Decreto 2852 de 2013

companies are reporting the total number of vacancies available. For instance, employers might underreport the number of vacancies because it might be time-consuming to fill and send the information to the UAESPE. Moreover, the UAESPE does not have a methodology to systematically verify that employers have reported the total number of job vacancies advertised. Third, the inclusion or the exclusion of some employer or job portals over time might affect the vacancy time series. An increase in the number of vacancies might be due to the inclusion of a new job portal (with not necessarily different and new vacancies). Fourth, as will be discussed in more detail in Cárdenas, 2020b, the problem of duplication increases by adding more websites. The UAESPE collects information from a different range of job portals and employers. However, a job vacancy can be published on various web sites. Given that employers are not required to report the full vacancy details, it is more difficult for the UAESPE to determine whether a vacancy is duplicated. Finally, the database and the UAESPE methodology to compile, clean, classify, etc. are not available and hence, the vacancy analysis conducted by this institution lacks robustness.

These problems have made employers' requirements or vacancy information scarce (Allen and Velden, 2013). As Álvarez and Hofstetter (2014) mention, vacancy data to study the labour market is scarce in developing countries like Colombia. As a result, the human resource needs of the country have remained unknown until this paper was conducted. As a consequence, Colombia lacks a human capital formation system with accurate tools (among others instructional agreements) to address public policy, educational and job training programs; so far these aspects have remained unaddressed and have not been aligned with employers' needs, and a low standard of quality education has instead proliferated. For instance, only 4% of 1,576 technological training programs, and 3% of 740 professional technical training programs offered by private institutions were accredited (in terms of content and infrastructure, among other characteristics) in terms of quality by the Ministry of Education in 2013 (González-Velosa and Rosas-Shady, 2016). Likewise, Regional Centres of Higher Education (CERES) have been reported to teach their students with outdated technologies and at an insufficient educational quality level (OECD, 2016b). Given the low standards of training and education quality, even the Technical and Vocational Education and Training system (TVET) has not grown enough in the last years due to lost prestige (OECD, 2015b).

Given these facts, it has become necessary to seek new and novel ways to assess what labour supply is needed by companies. One promising approach to address this issue is the provision and analysis of detailed labour demand information with the use of Big Data techniques. As will be discussed in the following section, the building of a web-based model of skill mismatches (skill shortages) for Colombia (and potentially for its regional counterparts) might have a large impact, considering its potential use as a tool for public policy related to the better management of human resources, related to the better management of human resources (i.e. the reduction of informality and unemployment rates), and also to assist in the allocation of skill development and educational budgets.

2.6 Conclusion

Despite the socio-economic improvements of the last decades, the Colombian labour market faces important challenges. The proportion of people participating in the labour market has considerably increased since 2008. Therefore, the labour market needs 1) to engage new job seekers into the formal economy, 2) to retain workers in the formal economy, and, 3) and move informal workers into the formal sector.

While other countries have created systems with statistical tools in order to measure skill mismatches, and thus orientate public policies to decrease this phenomenon, different barriers might prevent the pursuit of that goal in Colombia. According to the evidence discussed in this section, skill mismatches are one of the most important barriers to reduce unemployment and increase employment in the formal sector; consequently, skill mismatches might explain the high incidence of informality and unemployment in Colombia. A revision of the most important sources of information regarding human capital in Colombia shows that 1) available information sources are aggregated at levels that do not enable a detailed knowledge of existing occupations or skills, 2) there are difficulties in updating surveys or classifications (e.g. SOC 1970), 3) there are representative problems in the data gathering process (e.g. limited sample sizes), and, 4) no information sources collect employers' vacancy requirements. Thus, the available data indicates that there is a skill mismatches problem which means that it is not possible to know in enough detail which skills are needed in the Colombian labour market.

The above analyses in combination with institutional efforts, shows the interest Colombia has in measuring and tackling skill mismatches. However, the absence of an accurate tool to measure

the multiple dimensions of human capital together with an institutional disarticulation are one of the most critical factors that complicate the design of public policies, policies that need to be well-orientated to reduce the skill mismatches phenomenon in Colombia. Thus, a web-based model of skill shortages might provide valuable information to policymakers about employers' requirements, and might connect the various efforts that other institutions have made regarding skill mismatch analysis.

3. The information problem: Big data as a solution for labour market analysis

3.1 Introduction

"More and better data" is a common claim that researchers and policymakers make as a prerequisite to design public policies such as tackling skill mismatches issues (Cedefop, 2010; OECD, 2017b; Williams, 2004). To collect information about labour demand through surveys involves statisticians, interviewers, and a sample of companies or individuals available to respond. The cost of this kind of project is relatively high, in terms of resources and time, and can discourage countries (especially with low budgets) from collecting and analysing vacancy data. Additionally, even if a survey is carried out, the information obtained might not be detailed enough to analyse which skills or occupations are in demand among different industries or regions (Handel, 2012; OECD, 2016c).

Currently, with the proliferation of the Internet and electronic devices with higher capacities, large amounts of information about the behaviour of different agents are been stored daily. The storage of all this information has unlocked new borders for research in various areas of knowledge. For instance, Edelman (2012) and Askitas and Zimmermann (2015) detail several research examples using Big Data that have provided different applications for research in micro and macroeconomics, labour and demographic economics, public economics, health, education, and welfare, among others.

Big Data may be a way to overcome the limitations of existing skills analysis. More specifically, online job portals are a promising source of valuable information about labour demand. Thus, the second section of this section defines Big Data. Subsequently, it highlights how Big Data might fill informational gaps in supply and labour demand to address labour market policies and research. The fourth section discusses the potential uses of job portal information to tackle skill

mismatches (skill shortages). Big Data in specific job portals has limitations and for this reason, the fifth section discusses these limitations and indicates some caveats when using this kind of data for analysing the labour market. Finally, the section describes how Big Data sources might facilitate the analysis of the labour market in a context such as the Colombian economy.

3.2 A definition of Big Data

Increased internet speed, the increased use of smartphones, tablets, cameras, computers etc., technology with increasing capacities to store information, have favoured the creation and storage of computerised or digital information on a large scale. Cisco (an important multinational technology conglomerate) estimates that 96 exabytes (1 EB = 10^{18} bytes) was the average monthly amount of data traffic in 2016 and it is expected to increase three times by 2021 (278 EB per month) (see Figure 3.1). This era of massive information has unlocked opportunities for private and public institutions to compile, link and analyse relatively large flows of data produced by different sources to better orientate important decisions and strategies. This set of massive information, including the techniques to process and analyse the information, is commonly labelled as "Big Data".

Figure 3.1: IP Traffic, 2016 by source
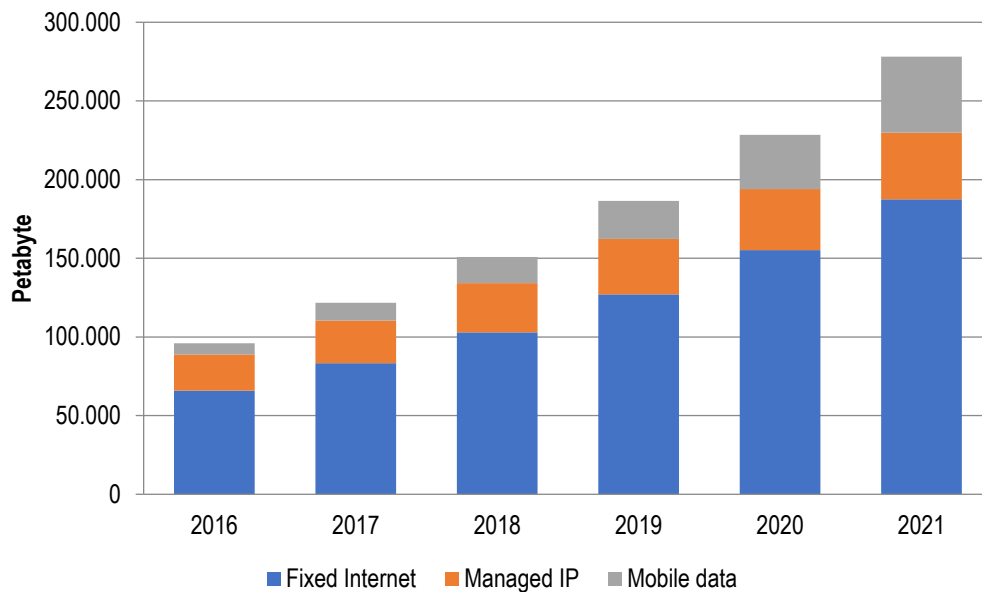
However, there is still an extensive debate about what can or cannot be considered as Big Data. Perhaps one of the most common conventions, defines this term according to three properties: volume, variety and velocity (Laney, 2001). Each of these properties will be discussed in turn. The former refers to the most obvious property that to be considered as Big Data the size (or volume) of data matters. In a simple way, data with a large volume of information might be a candidate to be called Big Data. However, individuals might consider different volumes of data differently because there are different computer capacities available in the market (with more or less data storage capacity, processing, etc.) which allow people to handle a certain number of bits per second. Consequently, it is necessary to determine a standard threshold which classifies data according to its size. One way to do this is by classifying data whose size represents a challenge to be processed and analysed within the average range of computer technologies available as "big".

Note that the threshold to consider if data have a high volume of information might change over time. Average computer capabilities increase over time, as technology improves so does its capacity to process a high volume of information. Hence, what was once considered as Big Data when this paper was started might have altered by the time this paper is finished. Despite the

changing nature of data, this criterion is useful because volume allows a researcher to distinguish between data sources in a technological environment that is constantly changing.

"Variety" refers to data structure. Unlike the information that comes from surveys, information from Big Data might not possess a well-defined structure to organise the different variables in specific spaces (columns) within a database. The information might instead come from a range of unstructured or semi-structured sources and in different formats, such as social media, sensors, websites, mobiles, videos, etc[46]. This characteristic makes data processing a challenge. Algorithms need to be developed to identify patterns (such as tags, keywords, among others) to obtain meaningful information. Thus, it is essential to note that the Big Data concept is not just related to volume, this concept also includes complex data qualities which make it necessary to have access to a higher capacity to store, process and analyse the gathered information.

Finally, "velocity" refers to the speed that the data are generated. Nowadays, information is generated in seconds, people can share an opinion to thousands through platforms such as Twitter or Facebook, and generate different reactions in an instant. Likewise, companies can post their current vacancies in real-time on various websites to quickly attract potential workers. This speed presents a challenge and an advantage for data processing and data analysis[47].

For the purposes of this paper, "Big Data" is considered as relatively high volume of information which is produced in a relatively fast way by different unstructured or semi-structured sources, and might be available in different formats, and where the three characteristics of volume, variety and velocity makes processing and analysing information processes a challenge *per se* with the

---

[46] For instance, every website (as will be seen in more detail in Cárdenas, 2020b) has its own HTML (Hypertext Markup Language), XML (Extensible Markup Language), or Javascript structure, etc., to create its structure, which means that every website can organise the same information (e.g. vacancies) in diverse forms according to their design and purpose (Aguilar, 2016).

[47] There are cases where information is not quickly generated (e.g. on a daily basis), nevertheless they (e.g. medical records) might be considered as Big Data given the size of the database which overpasses the current average computer capabilities.

average technologies available in this given moment (2017)[48]. Thus, it is necessary to implement new techniques to access, store, compile, link and analyse the information.

Despite many challenges, Big Data is expanding or opening a new frontier of knowledge (Askitas and Zimmermann, 2015; Edelman, 2012). Indeed, Big Data might fill the information gaps in different fields and regions where information to carry out or well-oriented public policies was frequently scarce in the past (Azzone, 2018). In the particular case of the labour market in Colombia, this information might give insights about the characteristics of the labour supply; and, more importantly, due to the general scarcity of labour demand data (especially in countries such as Colombia), Big Data offers the possibility of having for the first time a detailed picture of employers' requirements in real-time. The following section discusses in more detail how Big Data has provided new valuable information to analyse the labour market in different areas.

3.3 Big data on the labour market

"Good" data are a requisite to develop well-orientated policy and academic research, where "good" refers to data which involves the representativeness of the population being analysed and, thus, involves a minimum standard of quality during the collection process. As mentioned in section 2, labour supply data usually emanates from household surveys, while labour demand data emanates from employers' surveys. Supply and demand information from surveys has limitations that Big Data might alleviate in order to form a better picture of the labour market. Thus, different efforts from both sides of supply and demand that involve the usage of Big Data have started to be applied in some countries and areas.

---

[48] The debate about what constitutes Big Data is still open. Özköse et al. (2015) or BBVA (2018) add other characteristics such as "veracity" and "value" into the Big Data concept. The former refers to the trustworthiness or credibility of the data, nevertheless this is an implicit characteristic that any data should have; while the latter term establishes that information needs to provide some profit (usually measured in terms of money) to a certain institution. Nonetheless, not every institution or person seeks monetary profit from information. For instance, non-profit institutions might benefit from Big Data information in order to provide goods and/or services for free or at prices that are not economically competitive. Moreover, the value of information depends on the observer: data that for a company might not produce any value for a researcher or other institution might hold some value for that company.

3.3.3   Labour supply

3.3.3.1 Household surveys for the analysis of the labour supply

Traditionally, on the labour supply side, information has emanated from household surveys (e.g. employment rates by age, region, gender, etc.). Generally, these household surveys are characterised by a sampling frame (based on a census) representative of a specific population, a set of questions, and flows which customise the sections participants complete. Such surveys collect the main characteristics of the labour supply over a certain period. In most cases, the surveys are carried out by the Office for National Statistics (ONS) of each country who follow certain quality standards provided by an international institution such as the ILO. These standard procedures make household surveys one of the main sources of information to calculate indicators of the labour market, such as participation and unemployment rates, wages, etc.

Despite the indisputable advantages of household survey information, they have some limitations that might be overcome with Big Data. First, to collect information thorough surveys requires time for design, validation, collection, consolidation, among other processes, that might delay the publication of the resulting database for analysis. When data are available, the researcher needs time to process the information, to analyse possible alternatives, to address a specific issue. However, the disadvantage which such methods is that time will elapse from the moment the survey is designed to the final database, and during this time the data analysis might become outdate and invalidate the research findings due to changes in the socio-economic environment. Indeed, Reimsbach-Kounatze (2015) highlights that many OECD countries only have access to labour supply information after several weeks (at best) after the data was collected. This issue raises an important question about how private and public institutions can respond to ongoing or unexpected changes in the economy with information that might not be updated.

Second, another limitation with household surveys is their fixed structure as a pre-designed questionnaire which collects information on a variety of topics from people for various monitoring, planning and policy purposes. Surveys also have budget and time constraints. For instance, the UK Labour Force Survey (LFS) aims to measure "economic activity and inactivity, all aspects of people's work, job-search for the unemployed, education and training, income from work and benefits" (ONS, 2018a). Similarly, the main purpose of the GEIH is "To provide basic information

about the size and structure of the Colombian labour force" (DANE, 2009). Clearly, variables that are beyond this scope are not measured. Moreover, to add one single question increases the survey's cost and might also affect the structure, flux, response rate and results. This makes it difficult for survey designers to include other relevant labour supply questions.

Given this fixed structure, household survey designers find it challenging to update labour categories, such as occupations or industrial classifications. For instance, as mentioned in section 2, the household survey in Colombia classifies people's occupations based on SOC 1970. Consequently, the usage of this outdated classification might lead to the misclassification and/or the underestimation/overestimation of certain job positions. Thus, household surveys are a rigid tool that attempt to measure social issues, which dimensions might change over time.

Third, due to sample constraints, household surveys have a statistical limit. The more the data are disaggregated (e.g. region, sector, age, education, etc.) the more imprecise the estimates. For instance, the GEIH survey has available labour market results, such as employment or unemployment shares, disaggregated by city and SIC (Standard Industrial Classification, revision 3). This information is useful to analyse unemployment rates by region, major occupational groups, etc; nevertheless, the level of detailed information (granularity) obtained by household surveys might not be sufficient to cover topics which might be particularly useful for institutions and individuals (e.g. sector employment composition, the skills possessed by individuals, and occupations).

Thus, as stated above, household surveys have important limitations: 1) a time lag between designing, collecting, data processing and analysing the results; 2) a fixed structure that makes it difficult to include or modify questions and update categories; and, 3) a household survey is designed to be representative for a certain population at a desegregation level, so when this kind of data are disaggregated beyond the survey's limits its estimates start to be imprecise.

Consequently, several variables of interest for policymakers and researchers are not provided by household surveys. Such is the case for job networking, among the other behaviours of job seekers. Therefore, although household surveys are one of the main sources of labour supply information, relevant uncovered information exists which might be provided by Big Data information.

3.3.3.2 Big Data and labour supply

So far, the contribution of Big Data information to knowledge about labour supply has come from two sources. The first source uses search engines such as Google Insights, and the second source uses social media and networking sites to monitor (over a relatively short period) the behaviour of job seekers. Regarding the former, search engines track millions of searches in real-time concerning different topics such as weather, news, products, and importantly, for this paper, job searches. Consequently, these word searches can be used to identify trends in people's behaviour. For instance, Askitas and Zimmermann (2009, p.6) found the usage of certain keywords — such as "unemployment office or agency", "unemployment rate", among others — by German people on Google to have a strong correlation with, and therefore useful as predictor of, the German unemployment rate. The underpinning idea is that people will use certain words related to job searches in Google if they are (or are likely to be) fired, or when it is difficult to find a job. Thus, access to people's searches on this kind of search engine can provide information before the results from official surveys are available. This might allow public policy to react in a shorter period of time, and might fill the information gaps (e.g. unemployment information) in countries where there is no data being collected, or the periodicity of collection data are infrequent.

Regarding the latter, social media and networking sites might be a source of labour supply information. Specialised social media platforms and websites, such as "LinkedIn" and "BranchOut", have arisen in the last decade. For instance, LinkedIn is one of the most well-known professional networks as it is present in more than 200 countries and has more than 552 million users (with around 250 million users active every month), users who make their curriculum vitae public in order to be contacted or contact potential employers (LinkedIn, 2018). The information available through these social media platforms might provide insights about the skills and other characteristics of the labour supply[49].

---

[49] For instance, Rodriguez et al. (2014) used LinkedIn as a new source of information to analyse professional migration patterns. With this information the authors were available to analyse migration flows by region of destination between 1990 and 2012. They concluded that the United States is the most prominent destination country for professional migrants. However, this flow of people to the US has decreased over the years of study,

Interestingly, information from social media platforms has helped researchers to build or further refine their employment indicators. Such is the case for Antenucci et al. (2014) who created indexes of job loss, job searches, and job postings in real-time by tracking keywords such as "lost job," "laid off," and "unemployment", among others. The authors found that social media brought reliable information in real-time about the effect of government shutdowns, natural disasters, etc., on the labour market. Thus, concerning labour supply, social media and networking sites and search engines have created the opportunity for researchers to deepen understanding in certain topics.

3.3.4   Labour demand

Perhaps the use of Big Data for labour demand analyses has raised higher expectations among researchers, policymakers, etc., than the use of Big Data for labour supply. These expectations might be motivated by the fact that, traditionally, labour demand information has been scarcer than labour supply information (Kureková et al. 2014). As explained in more detail in this section, labour demand information shares many of the same limitations as labour supply information, such as sample constrains and granularity. However, unlike labour supply information, labour demand surveys and the analysis of employers' requirements tend to be less frequent (especially in countries such as Colombia—see section 2). Paradoxically, as Hamermesh (1996) emphasises, one reason that explains why studies about labour demand have been relatively ignored or scarce is due to the "creation of large sets of microeconomic data based on household surveys has spurred and been spurred by development of new theoretical and econometric techniques for studying labor supply" (Hamermesh, 1996, p.6).

Consequently, the main sources of information used for the analysis of labour demand have come from sectoral surveys (such as industry surveys) or even from household surveys. Even though these data sources have strengths, such as national standardisation and representativeness, the collection of labour demand information through surveys is likely to be costly, both in terms of resources and time, and these surveys might not provide enough information to workers, governments and other institutions about human resources needs.

---

while for Australia and Oceania, as well as African, Asian and Latin American countries, this flow of migration has increased.

3.3.4.1 Sectoral surveys

In the UK's "Vacancy Survey", carried out by the Office for National Statistics (ONS), around 6,000 trading businesses[50] are interviewed monthly to provide "an accurate and comprehensive measure of the total number of vacancies across the economy and fills a gap in the information available regarding the demand for labour" (ONS, 2018b). The survey's main results are published in the "Labour Market Statistical Bulletin" within six weeks of the reference date of the survey, and reveal the monthly number of vacancies in the UK. Additionally, there is a time series available regarding the total number of vacancies (seasonally adjusted) by industry which are aggregated (according to SIC's 2007 sections —22 groups[51]) by the size of businesses[52], and a time series comparison between the total number of vacancies and the total number of unemployed people (Beveridge curve)[53]. Moreover, the UK Employer Skills Survey (carried out by the DfE) provides detailed information about job requirements; specifically, skills and occupations demanded by employers (at a 4 SOC digit level if possible), and industries (22 major groups at a one-digit level according to SIC 2007). This survey is a biennial study, and its main results are published over the months following each survey (Vivian, 2016)[54].

Likewise, less developed regions such as Colombia have made different efforts to collect and analyse labour demand. As mentioned in section 2, Colombia has conducted sectoral surveys, such as the annual industrial survey and services survey. To measure the country's human resources needs in detail, in 2012 the Colombian statistic department completed a cross-

---

[50] Excludes agriculture, forestry and fishing.

[51] For more information see: https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/datasets/vacanciesbyindustryvacs02

[52] 1–9 employed; 10–49 employed; 50–249 employed; 250–2,499 employed; 2,500 + employed. See: https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/datasets/vacanciesbysizeofbusinessvacs03

[53] For more information see: https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/datasets/vacanciesandunemploymentvacs01

[54] At the time this paper was written, the last report available was published in 2015, and the results for 2017 survey are going to be available in summer 2018 (Skills survey, 2018).

sectorial survey (industrial, services and sales-retail activities) named the Human Capital Formation (EFCH, by its initials in Spanish). This survey was conceived to fill information gaps about human resource needs in the country, and, consequently, sought "to obtain indicators on the market and other effects on vocational training, competitiveness gaps and personnel appraisal that companies deal with" (DANE, 2018a)[55].

Despite these considerable efforts, these kinds of sectoral or cross-sectorial surveys (e.g. EFCH) present severe limitations for the analysis of labour demand and, hence, skill mismatches. First, as the name suggests, sectoral surveys are applied for a specific sector. The EFCH survey in Colombia is only applied to companies related to industrial services and sales-retail activities. Consequently, some sectors might be excluded, hence their labour demand composition and dynamic will remain unknown. Second, not all types of companies are included in the sampling frame. The industrial EFCH survey interviews establishments with 10 or more employees and those whose annual production is above £125,000. Moreover, the EFCH survey's results are available at "functional areas" such as "Production", "Management", and "R&D". Thus, these sources might not be enough to provide detailed information about which skills (or occupations) are in demand among different industries or regions (Handel, 2012; OECD, 2016c).

Likewise, the Colombian annual industry survey, which is one of the main sources for labour demand information, interviews establishments with the same criteria than the EFCH. Indeed, the EFCH survey is a subsample of the annual industry survey sample. Consequently, many companies (generally small or medium-size companies) in a sector might not be included in the sample, so even within the relevant subsample part of the labour demand is ignored.

Perhaps, more advanced regions such as the UK are less exposed to this aggregation problem. With a greater budget, these regions can design surveys with a higher disaggregation level, such is the case of the UK Skills Survey mentioned above. Nevertheless, even with a larger budget, the results from industry surveys might be produced with a relatively low frequency. For instance, the main findings from the UK Skills Survey are released every two years. Policymakers, educational institutions, and researchers, among others, need to wait at least two years to

---

[55] To carry out this survey took an invest of $397,349 (from the Ministry of Labour and Interamerican Bank of Development — IDB), plus DANE provided survey preparation in terms of sample designs, logistics and advice since 2010 (CONPES, 2010; DANE, 2018b).

access the information that the survey collects about labour demand requirements. In the period when the survey is carried out, the data are processed, cleaned and released, and labour conditions might have changed during the two years it takes to prepare data reports; consequently, some results might be outdated. Regarding this problem, less advanced countries such as Colombia are in a worse situation. For instance, in 2019, at the time this paper was written, the last EFCH survey to be conducted was in 2012 (a period of 7 years).

In some industry surveys, companies or a group of experts are asked about the number of vacancies that opened in the last period (e.g. within last year), the number of vacancies that each company is expected to have in the next period (e.g. within next year), the expect volume and some general characteristics (e.g. experience) of people that they will need in a certain period of time (e.g. the following three months, six months, a year, etc.). By providing information about current and future labour demand dynamics such questions address the problem of the low frequency of data results.

Based on this labour demand information, two different approaches have been developed to anticipate the future labour market's needs: skill forecast and skill foresight. The former term refers to forecast exercises which "use available information or gather new information with the specific aim of anticipating future skills needs, mismatches and/or shortages. Forecast results are meant to provide general indications about future trends in skill supply and/or demand in the labour market" (OECD, 2016c, p.39)[56]. The latter term, skill foresight aims to "provide a framework for stakeholders to jointly think about future scenarios and actively shape policies to reach these scenarios" (OECD, 2016c, p.39)[57]. In both these exercises are valuable because

---

[56] An example of this kind of exercise is the UK Local Economy Forecasting Model (LEFM) developed by Cambridge Econometrics (CE) in collaboration with the Institute for Employment Research at the University of Warwick (Cambridge econometrics, 2013). Based on the ONS's 2011-based Sub-National Population Projections (SNPP) — and assuming that the historical relationship between growth in the local area compared to the region or the UK economy will hold in the future — this model allows researchers to project/anticipate different economic scenarios, and evaluate possible implications for labour demand composition at occupation or qualifications levels among other outcomes (Cambridge econometrics, 2013).

[57] Examples of such skill foresight exercises are the Future of Work: Jobs and Skills in the UK (Störmer et al. 2014), or the Labour prospective in Latin America (ILO, 2013).

they estimate future employers' requirements and address the education and VET system according to possible future needs.

Nevertheless, once again, efforts such as skill forecast and skill foresight are relatively expensive in terms of money and time, and their results are too specific to be of used to the broader job labour market. For instance, in Colombia, labour prospective studies focus on specific sectors, such as coffee production and building construction. Moreover, projections from skill foresights or skill forecasts might be biased or mistaken. For example, companies might experience unexpected period expansions (or contractions) which can unexpectedly increase (or decrease) the creation (or destruction) of future vacancies. Thus, labour demand estimates might under or overestimate the number of vacancies and their characteristics. Likewise, experts might not accurately predict the course of a sector over the long term. Additionally, parameters to make economic projections might be outdated[58]. Consequently, projections based on these data would ignore economic changes that have occurred between 2005 and the date when a new census is conducted, and economics projections are re-estimated.

Therefore, sectoral surveys and exercises derived from them have several limitations, they 1) require large logistical operations and a considerable quantity of money to conduct a labour demand survey. Consequently, 2) time is needed to design, collect, process and release the information, 3) given budget constraints and survey designs, some companies or sectors might be excluded from labour demand analysis. For the same reasons, 4) it is, frequently, unlikely to be able to desegregate survey results at numerous levels: occupational, skills, industry, region, etc. Given the limitations mentioned above, labour demand information is scarce and less frequent (e.g. monthly) than household surveys. Finally, 5) Skill forecast or skill foresight methods might not properly foresee economic changes and their implications for skills (labour demand). Consequently, labour demand estimates based on sectoral surveys might be biased. Therefore, due to the limitations mentioned above, it is relatively common to find labour demand studies in the economic literature whose main sources of information are household surveys.

---

[58] For instance, in Colombia economic projections such as labour demand are based on the 2005 population census. A new population census commenced in 2018. Consequently, economic projections are supposed to be updated in the following years.

3.3.4.2 Household surveys for labour demand analysis

Traditionally, household surveys have functioned as inputs to analyse labour demand issues. These sources provide information about the intersection between labour supply and labour demand (filled labour demand) over a certain period of time. Household surveys provide information about labour demand in the following way: employed people can occupy one or more job vacancies, consequently, the total number of employed weighted by the number of jobs held by each one of them is equal to the total number of vacancies filled (satisfied demand — see section 1).

This information about the filled labour demand has been used in different studies as an approach to analyse the labour demand dynamic. Moreover, the availability of a relatively long series of household data have allowed analysing relevant trends and changes of the (filled) labour demand. For instance, this source of information is used to evidence processes such job polarisation which consist a simultaneous growth of labour demand for high-skilled (high wage) and low-skilled (low wage) workers experienced in the last decades (Acemoglu and Autor, 2011; Autor and Dorn, 2012; Autor et al. 2006; Salvatori, 2015).

Even though household surveys tend to be more frequent than vacancy surveys, being able to analyse the labour demand based on what people report on household surveys is limited. First, as explained above, survey constraints (e.g. money and time) might not allow disaggregating the results at a skill or occupational level (e.g. 4-digit level ISCO). Second, household surveys only take into account the current/past skills or characteristics of the workforce; what it is unknown are employers' requirements to fill their vacancies, which is an important aspect of labour demand to reduce possible mismatches (Autor, 2001; Mavromas et al. 2013)[59]; nevertheless, the acquisition of information is based on what people (labour supply) report, and does not consider one essential part of the labour market: the employers' requirements. As explained in section 1, skill mismatches occur when labour demand and labour supply for skills are misaligned. With household surveys, at most, it is possible to know which skills or occupations are not demanded, however, the employers' requirements remain ignored.

---

[59] For instance (as mentioned in section 2), the World Bank has conducted a Skills Measurement Program to assess skills in low- and middle-income countries (Pierre et al. 2014)

This issue is an important limitation when considering the employment share as a proxy of the labour demand. Total employment is at the intersection between labour supply and demand. Nevertheless, the level of employment might significantly differ from the true level of demand because of unfilled labour demand (vacancies). For instance, employers might demand high-skilled jobs, but there is no labour supply to fill them; consequently, by only using the employment total the fact there is an important demand for high-skilled workers would be ignored.

Therefore, household surveys are a valuable input to analyse filled labour demand and its long-term changes. Nevertheless, this information is limited in the following aspects: 1) there are constraints (e.g. time and money) that affect the level of aggregation and the frequency of data collection; 2) these surveys do not capture information about employers' requirements which is essential to address issues such as skill shortages. Consequently, all the issues mentioned above for sectorial and household surveys restrict the capacity of researchers and policymakers to tackle skill mismatches.

3.3.4.3 Big data and labour demand

As previously mentioned, the collection of labour demand information is relatively less systematic than labour supply information. Moreover, even when labour demand information is available, different limitations make skill mismatch analysis a challenge. However, it seems that the proliferation of a high volume of information (such as the Internet) and techniques to analyse it have brought the opportunity to evaluate possible skills mismatch (skill shortage) through the analysis of employers' requirements.

Nowadays, one important source of information is the Internet. This source is widely used for different purposes, and it stores relevant information regarding the behaviour of agents such as employers. As Autor (2001) highlights, the Internet provides an opportunity to collect more and possibly better labour market data. Indeed, online information contains a large number of detailed observations about labour demand, and it can be accessed, mostly, in real-time at a relatively inexpensive cost (Barnichon, 2010; Edelman 2012).

Moreover, employers' use of the Internet for advertising and finding suitable applicants, and for individuals to find a job, has dramatically increased. As mentioned by Maurer and Liu (2007) and Kuhn (2014), both employers and job seekers have increasingly used the Internet to find a

vacancy or to advertise. In fact, by 2007, more than 110 million vacancies and 20 million unique resumes were stored in online US sources (Maurer and Liu, 2007, p.1). Likewise, the number of job seekers looking for a job using online sources has increased. For instance, in the US, the share of unemployed people who used the Internet to find a job increased from 24% in 2000 to 74% in 2009 (Kuhn, 2014, p.2).

The use of the online job portals as a source of information has grown amongst researchers and has also attracted the attention of policymakers because they seem to provide quick and relatively inexpensive access to analyse information about employers' requirements. Job portals are websites where companies make public their current (or future) vacancies. Companies describe, to some extent, the job position and the attributes that a potential worker should have to be considered as a candidate. Additionally, job seekers can screen and select vacancies, and contact potential employers. In other words, job portals help to connect employers with job seekers and vice versa.

Job portal information, however, is not produced for the purpose of economic analysis (indeed in most cases it is posted online by private businesses). Yet job advertisements can potentially function as an essential input to analyse employers' needs. The systematic collection of information from job portals might help to diagnose the performance of an economy in real-time (e.g. at the level of available vacancies), and to understand employers' requirements and how these requirements change over time. Consequently, along with the increasing usage of the Internet and job portals, studies have used online job vacancy data to provide insights about the labour demand in different countries, such as in the US, Slovak, and Colombia (Cárdenas et al. 2014; Carnevale et al. 2014; Marinescu and Wolthoff, 2016; Štefánik, 2012; Tjdens et al. 2015).

In this sense, Kureková et al. (2014) have emphasised that job portals can be useful to generate a better understanding of companies' needs, which might enrich labour market policies. In contrast with household surveys (filled demand), job portal information (unfilled labour demand) might be useful to reveal which occupations or types of skills are currently in demand. Moreover, this kind of data might be of more relevance in contexts where employers experience difficulties to fill job vacancies, and job portal information might be the only the data available to analyse the labour demand for skills to address the labour supply according to employers' requirements. As mentioned in section 2, according to the Talent Shortage Survey conducted in 2016 by

Manpower it was revealed that 40% of companies worldwide experienced difficulties in filling their positions given skills (talent) shortages (Manpower, 2016). Besides, this phenomenon might be worsening due to the fact that skill requirements are changing rapidly because of structural shifts (technology and work organisation), given that the labour supply is not updating at the same pace (OECD, 2016c).

Consequently, in less advanced regions such as Latin America (e.g. Colombia) where the biggest skill mismatch exists, there is a lack of labour demand information (see section 2) and the usage of job portal information to measure employers' requirements might have a high impact on different labour demand outcomes.

3.4 The potential uses of job portal information to tackle skill shortages

Targeted vacancy information gathered from online job portals might improve informational and public policy deficiencies regarding skill shortages problems in the following ways: 1) to maintain an accurate estimation of vacancy levels, 2) to identify skills and other jobs requirements, 3) to recognise new occupations or skills, and, 4) to update occupational classifications.

3.4.1 Estimation of vacancy levels

The number of job offers together with other labour markets indicators (such as unemployment levels) help to determine the business cycle and possible mismatches in an economy. High vacancy rates might mean that the economy is in a stage of economic expansion and/or there are mismatches between supply and labour demand[60].

In this sense, online job vacancy advertisements might provide real-time access to job offers in an economy and public policymakers might react or re-design public policies in a shorter period aligned to the current economic changes. Given the advantages of collecting online information, different countries have started to create job vacancies databases based on information from the Internet. For instance, in the US there is the Help Wanted Online data series created by the Conference Board (The Conference Board, 2018) and in Australia, there is the Internet Vacancy Index developed by the Australian Department of Education, Employment and Workplace

---

[60] An example of the above is the Beveridge curve, which relates unemployment and vacancy to determine how well, or not, vacancies match with unemployed workers (Blanchard et al. 1989) (see Cárdenas, 2020a).

Relations (DEEWR) (Australian Government, 2018). Both provide measures of labour demand (advertised vacancies) at various levels, including at a national, state, regional, and occupational level (Reimsbach-Kounatze, 2015).

Moreover, online job vacancy information is not limited to counting the number of job offers in the economy. Indeed, one of the most important advantages of online job vacancy advertisements is that they provide detailed information about employers' requirements. This aspect allows researchers, policymakers, among others, to delve into topics (which before were relatively difficult or costly to obtain updated information on) and identify the demand for skills and other job requirements.

3.4.2   Identify skills and other jobs requirements

Perhaps, one of the most promising uses of online vacancy information is the identification of job requirements in a relatively short time duration to enable public policy design. As will be seen in more detail in Cárdenas (2020b), companies post job vacancies on job portals along with detailed candidate requirements to fill each position (skills, education, experience, etc.). This detailed information creates an opportunity to monitor job requirements at a disaggregated level (e.g. 4-digits occupation level) and, for instance, advise VET institutions in which skills they need to train people to increase their employability.

In this sense, one of the most important ongoing projects, at the time this paper was written, is the "Big data analysis from online vacancies" project carried out by the European Centre for the Development of Vocational Training (Cedefop). Cedefop combines its efforts with Eurostat and DG Employment, Social Affairs and Inclusion to collect data on skills demand using online job portals. With this information, Cedefop attempts to monitor skills and other job requirements at an occupation level, and identify emerging skills and jobs in Europe to advise training providers to revise or design new curricula according to current labour demand requirements in Europe (Cedefop, 2018).

Moreover, private companies such as Burning Glass Technologies provide and analyse labour demand information using job portals for countries such the US and the UK. For instance, this company has reported that 80% of middle-skill job advertisements demanded digital skills in 2016, which represents an increase of 4% compared with 2015 (Burning Glass, 2017, p.3). Thus,

job portals have become a relevant source of information to identify which skills are most in demand.

3.4.3 Recognising new occupations or skills

As was mentioned in section 2, the labour market changes rapidly and new occupations or skills might emerge or disappear over time. The identification of these new patterns in labour demand is relevant because it allows curricula to be adapted by training providers, and, as a consequence, prepares people for technological change. Patterns in labour demand can be identified by recording labour demand information from job portals. For instance, Emsi (2018) a labour market analytics company has started to build a skill taxonomy which has identified the growing demand for relatively new skills, such as "Cloud Engineer Architects" and "Cloud computing". Emsi (2018) mentions this information might be useful to understand how to adapt the labour supply according to changes in labour demand—especially for the most innovative sectors such as IT and tech.

3.4.4 Updating the occupation classification

With a demand for identification of occupations and skills, and new emerging patterns for job requirements, job portal data might facilitate the updating of occupational classifications with real-time information. As mentioned in section 1, usually occupational classifications are not updated as fast as labour market changes. A significant amount of time and financial resources are required to analyse information collected from companies and other stakeholders to update an occupational classification. However, with the relatively quick and inexpensive collection of online job advertisements it is now possible to identify job requirements (skills, educational level, tasks, etc.) of each occupation, hence this information might become an essential contribution to update occupational classifications according to changes in labour demand.

Moreover, the analysis of job advertisements might help to further adapt occupational classifications according to the context of a region or a country. For instance, as recognised by the ILO (p.2, 2008) "some countries may not have the capacity to develop national classifications in the short to medium term. In these circumstances it is advisable for countries initially to focus limited resources on the development of tools to support implementation of ISCO in the national context, for example a national index of occupational titles". In this circumstance, online job

advertisements might provide relevant information to adapt ISCO classifications according to a regional context.

Consequently, job portal information can be used for a range of different topics[61]. One of the most promising uses of this information is the identification of skill mismatches. The study of labour demand for skills is a key input to overcome informational barriers between labour demand and supply (Kureková et al. 2016); especially in a context (such as in Colombia) where employers are experiencing difficulties filling job vacancies. Yet, as the next section will address, despite the potential of vacancy information it is essential to take into account its possible limitations, so as to avoid potential bias when analysing job portal information.

3.5 Big Data limitations and caveats

It is important to note that despite the advantages of Big Data such as the greater volume of information it allows researchers to analyse, limitations exist that might affect the analysis of labour demand via job portal information. Consequently, any study that uses online job advertisements should consider the following issues: 1) data quality; 2) that job postings do not necessarily represent real jobs; 3) data representativeness; 4) Internet penetration rates, and, 5) data privacy.

3.5.1 Data quality

Data quality is one of the most important factors that determines the reliability of any database for statistical purposes. According to the quality framework and guidelines provided by the OECD, data quality is a multi-faced concept within which the relative importance of each dimension depends on user needs. These dimensions are: relevance, accuracy, credibility, timeliness, accessibility, interpretability, and coherence (OECD, 2011, pp.7–10):

---

[61] Clearly, the possible uses of job portal information are not limited to the ones mentioned above. Authors such as Turrell et al. (2018) use job vacancy information to understand the effects of labour market mismatch on UK productivity. Moreover, Rothwell (2014) employ advertisement duration as a proxy of vacancy duration in order to determine skill shortages in the US. Additionally, Marinescu and Wolthoff (2016) and Deming and Kahn (2018) use online job advertisements to determine the portion of wage variance explained by employers' skill requirements (e.g. cognitive, social, writing, and so on) in the US.

Table 3.1: OECD quality framework and guidelines

Criteria	Description
Relevance	"Degree to which the data serves to address the purposes for which they are sought by users. It depends upon both the coverage of the required topics and the use of appropriate concept".
Accuracy	"Degree to which the data correctly estimate or describe the quantities or characteristics they are designed to measure".
Credibility	"Refers to the confidence that users place in those products based simply on their image of the data producer … This implies that the data are perceived to be produced professionally in accordance with appropriate statistical standards, and that policies and practices are transparent. For example, data are not manipulated, nor their release timed in response to political pressure".
Timeliness	"Reflects the length of time between their availability and the event or phenomenon they describe, but considered in the context of the time period that permits the information to be of value and still acted upon".
Accessibility	"Reflects how readily the data can be located and accessed".
Interpretability	"Reflects the ease with which the user may understand and properly use and analyse the data. The adequacy of the definitions of concepts, target populations, variables and terminology, underlying the data, and information describing the limitations of the data, if any, largely determines the degree of interpretability".
Coherence	"Degree to which they (data) are logically connected and mutually consistent".

Source: OECD, 2011

With regards to these conditions (Table 3.1), given the nature of Big Data in specific job portals as sources of information, this source has a clear advantage in terms of "timeliness" compared with other sources of information such as sectoral surveys. However, as mentioned in Subsection 3.3.4, job portals and, in general, Big Data sources (such as LinkedIn) were not initially created for policy or academic purposes. This makes the data available through websites seem relatively disorganised; for example, without standardisation, with duplication issues and/or with a relatively high portion of missing values. Hence, data quality (relevance,

interpretability, coherence, accuracy and credibility) and the analysis of labour demand with Big Data sources might be affected or limited by these issues of organisation.

As will be seen in Cárdenas (2020b) in more detail, employers do not follow a specific format when they advertise vacancies[62]. This unstructured way of announcing vacancies can make statistical inference difficult. For instance, to generate a simple tabulate of a particular variable (e.g. wages), it is necessary to first identify where all (or most) of the information is located on the website and put only this information together to form the corresponding tabulate. Moreover, companies' use their own "language" when providing information, such as job descriptions, titles and the required skills; thus, employers might use different words to define a similar job position.

Additionally, companies are not required to provide a standard set of detailed information about the vacancy. As will be discussed Cárdenas (2020b) in more detail, the high presence of missing values might create bias in the analysis of a certain database. For instance, employers might reveal the wages offered for low-skilled jobs while they might not reveal the wages offered for high-skilled jobs. In consequence, when the mean of wages offered are estimated from any subsequent database the results would underestimate the average of real wages due to missing information of a specific (high-skilled) occupation group. Hence, these kinds of biases might affect the data quality of job portals as a source of information for statistical inferences about labour demand.

Likewise, duplication issues might affect the data quality of job portals. There are two possible types of duplication: in and between job portals. The first type ("in") refers to the situation where companies might advertise the same job position in the same job portal more than once. The second type ("between") occurs when employers advertise the same vacancy on more than one website. Consequently, when collecting information about labour demand using different job portals, the number of job vacancies might be overestimated, hence any statistical inference might be biased. Therefore, both duplication problems (in and between job portals) might affect the credibility and the interpretability of job vacancy data.

---

[62] For instance, where the online content signifies the presence of a "job title", there may also be information regarding company's name, location, etc.

Finally, employers might make mistakes when typing in information, and, in some cases, the information provided might be contradictory. If, for example, an employer writes in the job description that work experience is not required, but in the job title it states that some work experience is required.

All the problems cited above, show that when working with job portal information important issues need to be addressed to guarantee a certain level of data quality (some of these issues are also true of survey information). Clearly, the problems mentioned above can be reduced with the use of data mining techniques such as data cleaning, classification and imputation, among others, but they might not be completely eliminated. This result depends on the effectiveness of the algorithms used and the information provided by the employer (Cárdenas, 2020d, discusses whether or not the vacancy database for Colombia fulfil the quality requisites established by the OECD).

Thus, the level of these data quality problems and the techniques implemented to tackle them will determine the extent to which job portal information can be used to analyse labour demand. However, data quality is not the only concern when job portal information is used for analysis. There are other issues: job postings might not necessarily be real jobs, data representativeness, Internet penetration, among others issues, might limit the usage of Big Data for the analysis of labour demand.

3.5.2 Job postings are not necessarily real jobs

Given the nature of job portals, any company or individual can post a vacancy[63]. However, job portals do not have the means to verify if the advertisement corresponds to a real vacancy—or might not be interested in doing so. It is logical to suppose that most job portal users are companies or individuals that are potentially interested in hiring a worker; nevertheless, this potential interest does not necessarily imply that a job vacancy is, or will, open. As Emsi (2013) remarks, when using job portal information there are difficulties in making a one-to-one comparison between job advertisements and a real job vacancy. For instance, companies might

---

[63] Depending on the job portal, advertising a vacancy might be free or associated with a cost which generally depends on the time the advertisement is active on the website.

post more job advertisements than available positions in order to receive more applications, and then hire the candidates who best fit their requirements[64].

If job portals can post jobs which are not real, or companies can open vacancies without posting them, it is then difficult to precisely determine the number of job vacancies for an occupation, sector, etc., using job portals. These issues do not mean that job portal information cannot be used as a source to analyse labour demand. With this information in mind to utilise the proper statistical techniques, it is possible to comprehend the structure and trends of labour demand (see Cárdenas, 2020c); although it may be challenging to determine the exact number of real vacancies available in a period through job portal information.

Moreover, as Emsi (2013) discusses, even with the above problems, job vacancy advertisements are useful to understand current skills demands, such as who is (or interested in) hiring and where the most employee rotation (turnover) is occurring. For instance, an employer might advertise ten job positions for accountants in a single job advertisement when he/she will eventually only hire five of them. Despite the possibility that job advertisements might overestimate the number of available vacancies, this information might reveal occupations and the demand for skills associated with those occupations. Even where companies advertise vacancies for a certain job position to collect CVs and store them in their databases, this a relevant labour demand signal that predicts what kind of occupations companies are aware of that might soon be demanded by employers. Therefore, job portal information is a valuable resource to support the analysis of labour demand; even if, not all advertisements correspond to a real job position.

3.5.3   Data representativeness

Even though job portal information contains a considerable amount of data, this does not guarantee that this information is representative of the whole economy. On one side, some companies with a specific characteristic (sector, localisation, etc.) might not commonly use job portals to advertise vacancies. On the other side, even in the unlikely situation that every

---

[64] Another alternative is that companies (such as recruitment agencies) might advertise vacancies to collect CVs and store them in their databases. With this technique companies have already collected the data of potential workers and have the ability to quickly start the screening process in the eventually of a job opening.

company used job portals, some specific job positions might exist that are not advertised on websites. For instance, companies might recognise that people with low skills do not tend to use the Internet to find a job, and the most effective way to recruit such candidates is through informal channels, such as one-to-one or personal references (e.g. friends). In consequence, depending on the available information on job portals, in some cases, it is not possible to make any statistical inference for a labour market segment or, in other cases, there might be some restrictions when the data are analysed.

Thus, when using job portal information it is relevant to understand which segments of the market are properly represented by these sources of information. This discussion of data representativeness is one of the main concerns regarding the use of job portal information for policy recommendation. The representativeness issue determines whether or not it is possible to analyse and make public policy recommendations for labour markets with job portal information. However (as will discussed in more detail in Cárdenas, 2020d), to test data representativeness is a complex task. To illustrate this point, it is important to consider how household surveys or sectoral surveys guarantee data representativeness. As mentioned in Subsection 3.3.3, household surveys are based on a population census. This census enables researchers to obtain information about the total number of individuals ("universe"), and their main characteristics over a certain period. Once the population and its characteristics are known, it is possible to draw a sample for household. In this way, the information from household surveys guarantees that their sample results are as close as possible to the required population parameters (age, gender, etc.).

However, usually, in the case of vacancy analysis, the "universe" is unknown: for instance, the total number of vacancies available in a period by population groups (sector, occupation, localisation, etc.). Therefore, in this case, it is more difficult to know which population is represented by job portal sources. Paradoxically, the relative absence of vacancy information motivates researchers' use of job portal information; nevertheless, this absence of representativeness might limit or put in doubt the usefulness of data from job portals.

Some authors have addressed this issue. For instance, Štefánik (2012) used job vacancies and CV data of a private job portal in Slovakia to examine whether or not Internet job search data can be used a possible source of information to analyse labour demand for skills. Basically, the

representativeness test consisted of comparing the occupational (one-digit level-major groups) and sectorial (one-digit level) structure from vacancy and CV data with the employment structure from the EU Labour Force Survey (LFS). Occupational groups that presented a different structure between job portal information and LFS were considered to be represented inadequately by job vacancies and CV data[65].

However, as pointed out by Kureková et al. (2014) most of the studies that have used job advertisements (printed or online) do not discuss or test data representativeness, and their findings are generalised for occupational or sectorial groups. The absence of discussion aimed at identifying data representativeness might affect the reliance of many studies[66].

Therefore, to discuss and test the data representativeness of job portal data for academic and public policy purposes is a key issue when considering the use of these sources of information. The validity and the generalisation of results from the analysis of online job advertisements depends on the population being represented by job portal sources. For this reason, Cárdenas (2020d) discusses and tests data representativeness for the Colombian case.

3.5.4 Limited Internet penetration rates

Related to the above point, the usefulness of job portal information and, hence, their representativeness depends on Internet penetration rate (the percentage of the total population that uses the Internet). Although Internet usage has increased (see Section 4.2), this growth might not cover some sectors, regions, etc. For instance, in Colombia, there is a remarkable

---

[65] Additionally, Kureková et al. (2012) used online job advertisements to identify the skills and other characteristics more frequently demanded by employers in Slovak for low- and medium-skilled occupations. Based on the comparison of online data with the LFS, and given the high reputation of job portals (many employers and employees use job portals) in Slovakia, the authors argued that their results are generalisable for the Slovak population.

[66] For instance, given the nature of the Internet, occupations related to Internet Technologies (IT) tend to be overrepresented in online job advertisements. Consequently, a study that does not account for this source bias might conclude that IT skills are one of the most relevant skills required to find a job, while considering the total number of real vacancies (those advertised and not advertised on the Internet), the actual share of IT occupations might be minimal. Hence, the study might mislead any public policy recommendation, or any other conclusions, derived from it.

disparity between rural and urban zones in terms of Internet access[67]. Given this limited access, employers might tend towards the use of other job advertising channels such as asking friends or colleagues to recruit potential workers. Consequently, job portal information in Colombia might not cover a considerable portion of the labour demand in rural zones.

In regions where the growth of Internet access has not occurred or has occurred at a slower pace, the inferences that can be drawn from job portal information might be more restricted than in areas where Internet access is more widespread. Places where there is less Internet access tend to be poorer, and information about labour demand tends to be scarcer due to the prohibited cost of doing a vacancy survey. In consequence, even where the Internet is not widely used, paradoxically, it might be the only reliable source of information to analyse labour demand. Hence, the statistical inference from job portals depends on the Internet penetration rate; however, even when Internet access is relatively low online job advertisements might be a rich source for analysing important segments of the labour market.

Additionally, as Kureková et al. (2014) mentions, it is highly likely that the Internet continues to spread across different regions and socio-economics groups, so that the reliance on Internet-based recruitment methods will increase over time. In consequence, Internet penetration rates limits the statistical inferences that can be drawn from job portal information; however, those limits are becoming less relevant due to technological advances.

3.5.5 Data privacy

Online job vacancy advertisements belong to job portals or to other platforms where employers have decided to make their vacancies public. Provided that job vacancy information is shared and is administrated by a third party, this issue might affect the statistical inferences that can be drawn from those sources. First, the availably of information might change due to changes in the platforms. As private administrators, job portals might unexpectedly change the number of vacancies or the number and/or kind of variables available on their websites, which in turn affects

---

[67] According to the Economic Commission for Latin America and the Caribbean (ECLAC, 2016, p.12) around 10% of households in rural areas had access to the Internet in 2014, while around 50% of households in urban zones had access to the Internet in 2015.

what information is available for researchers, especially when attempting to analyse the changes in the economic environment (e.g. number of vacancies, wages, etc.)[68].

Second, job portals can restrict the usage of vacancy information. In most cases, job portals prohibit the storage and the usage of job advertisements for commercial purposes; however, for statistical purposes there does not seem to be any legal restriction. For instance, the Cedefop project "Big data analysis from online vacancies" has started to collect information from different job portals in Europe. Cedefop has informed these portals that information is going to be collected for statistical purposes, and most of the job portals have not denied access to the data. Nevertheless, as mentioned above, the project has required new statistical legislation to delineate the use of job portal information and other non-traditional information sources.

Table 3.2 summarises the main advantages and disadvantages of the different data sources for the analysis of labour demand. Both traditional (sectoral and household surveys) and non-traditional sources of information (online job portals) have advantages and disadvantages regarding the study of labour demand. Consequently (at this point), non-traditional surveys cannot replace traditional sources of information. Although non-traditional sources such as Big Data might complement and support sectoral or household surveys and vice-versa (see Cárdenas, 2020a).

---

[68] Some websites might adjust the number of variables displayed, such as wages, because potential workers might not apply for the job given the previous characteristics of the vacancy.

Table 3.2: Advantages and disadvantages of data sources for the analysis of labour demand

Source	Advantages	Disadvantages
Sectoral surveys	- Guarantee a certain level of data representativeness - Provide (usually macro) indicators of labour demand	- Aggregated data - Time consuming - Relatively expensive - Fixed structure - Less frequent than household surveys
Household surveys	- Guarantee a certain level of data representativeness -Provide (aggregated occupational or skills) indicators about the labour force - Generally available as long-term time series	- Aggregated data - Time consuming - Relatively expensive - Fixed structure - Information from the labour supply
Job portals	- High volume of data - Information in real time - Inexpensive - Disaggregation level - Detailed information - Useful for different purposes (e.g. the estimation of vacancy levels, to identify skills and other job requirements, etc.)	- Data quality issues - Job postings are not necessarily real jobs - There is not a priori guarantee of a certain level of data representativeness -Depends on Internet penetration rates

3.6 Big Data in the Colombian context

As was mentioned, in some contexts, Big Data sources might be the only ones available to analyse different labour market topics (Kureková et al. 2014). Specifically, in Latin American countries such as Colombia the use of information from online job portals can provide valuable first insights about "skill-shortage vacancies".

Therefore, given the potential for vacancy analysis in Colombia and the high expectations that this topic generates, to understand the potential scope of these data sources it is first necessary to answer the following questions: 1) how and to what extent a web-based system for monitoring skills and skill mismatches based on job portals and household surveys could be developed for Colombia? Specifically, how may job portals information be used to inform policy

recommendations, primarily to address two of the major labour market problems in Colombia which are its high unemployment and informality rates?; To what extent can these sources be used together (job portal information [unsatisfied demand] and national household surveys [labour supply]) to provide insights into skills mismatch (skill shortage) in a developing economy?

By answering the above questions, my research will contribute to current knowledge of the advantages and the limitations held by novel sources of information which attempt to address public policy issues and/or academic research problems. It will provide a methodological and analytical model for countries with scarce information regarding occupations and skills in the labour market by taking into account possible limitations and biases surrounding vacancy data. It will provide an analysis of the labour market in occupational and skills terms. Importantly, this research will be useful to institutions to match disadvantaged workers (especially unemployed and informal workers) to jobs for which they have the potential capabilities to fill, or could be used to help employees develop certain skills which might not be easily transferable through the formal educational system, or programs such as VET (Kureková, 2014).

As previously mentioned, the most important ongoing project similar to this paper is the "Big Data analysis from online vacancies" project conducted by Cedefop. So far, this project is focused on analysing skills and job requirements in Europe from job portals. A remarkable task given the necessity to capture and analyse online sources from more than 24 official EU languages, since April 2018 (Cedefop, 2019).

However, as summarised in Table 3.3, this project is distinct from the Cedefop initiative in eight respects: 1) this research investigates advantages, limitations and uses for job portal information for a country such as Colombia that does not belong to the EU, and is affected by serious skill mismatch issues (section 2). Moreover, 2) this research provides a theoretical framework regarding labour market mismatches and the protentional usefulness of job portals to tackle these phenomena. 3) the Cedefop project only collects and processes job titles, skills and sector information, while this paper considers and proposes various methods to collect and process a wider number of variables. These variables aim to determine the consistency of the vacancy database and to accurately describe Colombian labour demand (Cárdenas, 2020c and Cárdenas, 2020d). 4) new methods are proposed to classify job titles into occupations and identify skills for a country that does not have national skill dictionaries (Cárdenas, 2020b). 5)

this paper avoids a sole focus on processing and analysing occupational and skill variables; thus, variables such as educational requirements, wages, and sector, among others, become integrated into Colombian labour market analysis. 6) the time period of the Colombian analysis is longer than the Cedefop data. This longer period enables an analysis of trends and seasons within the Colombian labour market. Importantly, 7) this paper will provide a framework to test the validity and consistency of job portal information. 8) Finally, this study will combine job portal and household survey data to determine what skill shortages exist in Colombia.

Table 3.3: The main differences between the Cedefop and Colombian vacancy projects

Source	Cedefop	Colombian vacancies
Region	European Union	Colombia
Theoretical framework regarding labour market mismatches and the protentional usefulness of job portals to tackle skill mismatches	The project is in a stage where the vacancy data have begun to download and be processed (exploration stage). It has not been exhaustively discussed and tested to determine the usefulness of job portals to tackle skill mismatches.	It provides a theocratical framework and concepts which highlight the benefits of analysing job portal information for tacking skill mismatches.
Extraction of information	Job title, skill and sector variables are collected and processed.	This study considers and proposes various methods to collect and process a wider number of variables, such as job title, labour experience, educational requirements, (imputed and non-imputed) wage, skills, among others.
Methods to classify job titles into occupations and identify skills	Machine learning algorithms and the use of a European skills dictionary.	It proposes new mixed methods to proper classify job titles into occupations and identify skills for a country that does not possess national skill dictionaries.
Analysis of variables such as educational requirements, wages and sector, among others	The project (so far) is focused on describing the skills and occupations most demanded.	This study uses variables, such as occupations, skills, wages, educational requirements, etc., to exhaustively validate and analyse the vacancy data.

Period of analysis	April 2018 – Ongoing.	January 2016 – Ongoing.
Framework to test the validity and consistency of job portal information	The consistency of the results has not been tested yet.	The vacancy database is exhaustively tested. In fact, a framework is suggested to evaluate the representativeness of the vacancy database for each occupation at a different level of disaggregation.
Combination of job portal and household survey data to determine skill shortages	The vacancy data have been used to provide a preliminary overview of demanded occupations and skills.	Provides a descriptive and detailed analysis of occupations, skills, educational requirements, wages, among others. The vacancy data are combined with household data to monitor skill shortages.

3.7 Conclusion

Technological changes have eased the generation and storage of large amounts of information at a low cost in terms of time and money. Together with the increase of a large volume of information, a set of different techniques has been developed to process and analyse the massive information generated and available for research and analysis. This large amount of information and the techniques to manage this kind data have been named "Big Data". As the name suggests, this term refers to a relatively high volume of data; nevertheless, this is not the only characteristic of "Big Data", indeed, the most common three properties assigned to this term, as described in Section 4.2, are volume, variety and velocity (Laney, 2001).

The Big Data phenomenon has attracted the attention of private and public companies, and researchers (among others) because Big Data might provide relevant information for the analysis of individual behaviour, especially in contexts where there was previously a lack of data. The labour market is one of these scenarios where traditionally limited information was available or the required information was relatively absent, especially for the analysis of labour demand requirements. To collect information regarding labour demand by traditional methods (e.g. surveys) is relatively costly in time and monetary terms. Moreover, even in cases where there is information about labour demand, this information might not be disaggregated (or well-designed)

enough to analyse employers' requirements. This absence of information and, hence, labour demand analysis is one of the main obstacles to tackle possible skill mismatches. Individuals and training providers unaware of employers' requirements might offer skills that are not required by the labour demand.

Consequently, Big Data, specifically job portals, might provide in real-time and at low cost, valuable information for the analysis of labour demand, and thus the identification of skill shortages. Compared with traditional sources of information, such as sectoral or household surveys, job portals 1) provide labour market information in a short period of time (real-time); 2) enable the relatively inexpensive collection of information from job portals; 3) provide a high volume of detailed information and, hence, 4) their data might be disaggregated to skills and occupational levels. Given these advantages and the potential use of job portal information, there has been an increasing interest from researchers and policymakers to utilise online job advertisements.

However, little attention has been paid to the possible limitations and biases of job portal information, and how these issues might affect labour demand analysis. As a source of information, job portal data have limitations such as 1) the data quality; 2) job postings are not necessarily real jobs; 3) data representativeness; 4) Internet penetration rates, and, 5) data privacy. This section has discussed the need for labour demand information that job portals might fill. However, before making any statistical inferences for these sources of information, first it is necessary to know as much as possible about the biases and limitations of the data. Consequently, Big Data has considerable limitations and, as with household or sectoral surveys, is necessary to evaluate the scope of these sources of information.

Therefore, at this point, Big Data is a complement rather than a substitute for traditional data collection, such as household and employer surveys, among others. Yet, in a context where information is scarce, Big Data might be the only "reliable" source available for labour demand analysis. This is the case for Colombia (and Latin America), where there are high complaint rates about the quality of the workers by companies, and there is not enough labour demand information to address workers' skills according to employers' requirements. Consequently, the next working papers present a methodology to collect and analyse labour demand information taking into account possible information biases.

4. References

Acemoglu, D., & Autor, D. (2011). *Skills, tasks and technologies: Implications for employment and earnings*. Handbook of labor economics, 4, 1043-1171.

Aguilar, L. J. (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones.* Alfaomega Grupo Editor.

Albrecht, J., Navarro, L. and Vroman, S. (2007*). The Effects of Labor Market Policies in an Economy with an Informal Sector*. IZA DP No.2141

Allen, J., Levels, M., & van der Velden, R. (2013). *Skill mismatch and skill use in developed countries: Evidence from the PIAAC study*. Maastricht: Research Centre for Education and the Labour Market.

Álvarez, A., & Hofstetter, M. (2014). *Job vacancies in Colombia: 1976–2012*. IZA Journal of Labor & Development, 3(1), 1-11.

Andrews, M. J., Bradley, S., Stott, D., & Upward, R. (2008). *Successful employer search? An empirical analysis of vacancy duration using micro data*. Economica, 75(299), 455-480.

Antenucci., D. Cafarella., M. Levenstein., M. Ré., C. and Shapiro., M. (2014*). Using Social Media To Measure Labor Market Flows*. NBER Working Paper No. 20010.

Arrow, K. J. (1962). *The economic implications of learning by doing.* The review of economic studies, 29(3), 155-173.

Askitas, N., & Zimmermann, K. F. (2015). *The internet as a data source for advancement in social sciences*. International Journal of Manpower, 36(1), 2-12.

Asplund, R. (2004). *The provision and effects of company training: A brief review of the literature* (No. 907). ETLA Discussion Papers, The Research Institute of the Finnish Economy (ETLA).

Australian Government (2018). *Internet Vacancy Index*. [online] Available at: https://data.gov.au/dataset/internet-vacancy-index [Accessed 20 August. 2018].

Attewell, P. (1990). *What is skill? Work and occupations*, 17(4), 422-448.

Autor, D. H. (2001). *Wiring the labor market*. The Journal of Economic Perspectives, 15(1), 25-40.

Autor, D., Katz, L. F., & Kearney, M. S. (2006). *The polarization of the US labor market* (No. w11986). National Bureau of Economic Research.

Autor, D. & Dorn D (2012). *The growth of low skill service jobs and the polarization of the US labor market.* American Economic Review 2013, 103(5): 1553–1597

Azzone, G. (2018). *Big data and public policies: Opportunities and challenges.* Statistics & Probability Letters, 136, 116-120.

Bahk, B. H., & Gort, M. (1993). *Decomposing learning by doing in new plants.* Journal of political economy, 101(4), 561-583.

Banfi, S., & Villena-Roldan, B. (2018). *Do high-wage jobs attract more applicants? Directed search evidence from the online labor market*. Journal of Labor Economics, Forthcoming.

Barnichon, R. (2010). *Building a composite help-wanted index*. Economics Letters, 109(3), 175-178.

Barrett, Alan; O'Connell, Philip J. (1999). *Does Training Generally Work? The Returns to In-Company Training*. IZA Discussion paper series, No. 51, Institute for the Study of Labor (IZA), Bonn

Bassanini, A., Booth, A. L., Brunello, G., De Paola, M., & Leuven, E. (2007). *Workplace training in Europe.* In G. Brunello, P. Garibaldi, & E. Wasmer (Eds.), Education and Training in Europe. Oxford: Oxford University Press.

BBVA (2018). *The five V's of big data*. [online] Available at: https://www.bbva.com/en/five-vs-big-data/ [Accessed 05 May. 2018].

Becker, G. S. (1962). *Investment in human capital: A theoretical analysis.* Journal of political economy, 70(5, Part 2), 9-49.

Becker, G. S. (1994). *Human capital revisited. In Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education* (3rd Edition) (pp. 15-28). The university of Chicago press.

Bell, L. A. (1997). *The impact of minimum wages in Mexico and Colombia*. Journal of labor Economics, 15(S3), S102-S135.

Bernal S, R. (2009). *The informal labor market in Colombia: identification and characterization*. Revista Desarrollo y Sociedad, (63), 145-208.

Black, S. E., & Lynch, L. M. (1995*). Beyond the incidence of training: evidence from a national employers survey* (No. w5231). National Bureau of Economic Research.

Blanchard, O. J., Diamond, P., Hall, R. E., & Yellen, J. (1989). *The Beveridge curve.* Brookings papers on economic activity, 1989(1), 1-76.

Blundell, R., Dearden, L., Meghir, C., & Sianesi, B. (1999). *Human capital investment: the returns from education and training to the individual, the firm and the economy.* Fiscal studies, 20(1), 1-23.

Bosworth D., Dawkins P. and Stromback T. (1996). *The Economics of the Labour Market*, Addison Wesley, Harlow.

Brunello, G., & Schlotter, M. (2011). *Non-cognitive skills and personality traits: Labour market relevance and their development in education & training systems.* IZA Discussion Papers, No. 5743.

Burdett, K., & Smith, E. (2002). *The low skill trap*. European Economic Review, 46(8), 1439-1451.

Burning Glass (2017). *The Digital Edge: Middle-Skill Workers and Careers.* September 2017, [online] Available at: https://www.burning-glass.com/wp-content/uploads/Digital_Edge_report_2017_final.pdf [Accessed 10 August. 2018].

Cabrera, A., Rodríguez Mora, D., Vargas Zúñiga, F., Barragán, Á., Rubiano, E., & Cifuentes Arango, C. (1997). *Clasificación nacional de ocupaciones.*

Cahuc, P., Carcillo, S., & Zylberberg, A. (2014). *Labor economics.* MIT press.

Cambridge econometrics (2013). *Assumptions for the Baseline and 'Smart Efficiency and Growth' Scenarios for Worcestershire Districts.* [online] Available at: http://www.swdevelopmentplan.org/wp-content/uploads/2013/02/Jobs_Forecast_Briefing_Note_-_CE_ScenarioAssumptions_2011131.pdf [Accessed 12 July. 2018].

Cappelli, P. H. (2015). *Skill gaps, skill shortages, and skill mismatches: Evidence and arguments for the United States*. ILR Review, 68(2), 251-290.

Cárdenas R., Jeisson & Guataquí R., Juan Carlos & Montaña, Jaime. (2014). *La problemática del análisis laboral de demanda en Colombia.* Perfil De Coyuntura Económica, Universidad De Antioquia - CIE, issue 24, pages 71-107, October.

Cárdenas R., Jeisson. (2020a). Possible uses of labour demand and supply information to reduce skill mismatches. Universidad del Rosario. Working Paper No. WP2-2020-002.

Cárdenas R., Jeisson. (2020b). Extracting value from job vacancy information. Universidad del Rosario. Working Paper No. WP2-2020-003.

Cárdenas R., Jeisson. (2020c). Descriptive analysis of the vacancy database. Universidad del Rosario. Working Paper No. WP2-2020-004.

Cárdenas R., Jeisson. (2020d) Internal and external validity of the vacancy database. Universidad del Rosario. Working Paper No. WP2-2020-005.

Carnevale, A. P., Jayasundera, T., & Repnikov, D. (2014). *Understanding online job ads data: a technical report.* Georgetown University, McCourt School on Public Policy, Center on Education and the Workforce, April.

Cedefop. (2010). *The skill matching challenge: analysing skill mismatch and policy implications.* Office for Official Publications of the European Communities, Luxembourg.

Cedefop (2012a). *Skill mismatch.* Luxembourg: Publications Office of the European Union.

Cedefop (2012b). *Quantifying skill needs in Europe*. Luxembourg: Publications Office of the European Union.

Cedefop. (2015). *Quantifying skill needs in Europe*. Luxembourg: Publication Office of the European Union.

Cedefop (2018). *Big data analysis: online vacancies.* Available at: http://www.cedefop.europa.eu/en/events-and-projects/projects/big-data-analysis-online-vacancies. [Accessed 23 Jun. 2018].

Cedefop (2019). *Online job vacancies and skills analysis: a Cedefop pan-European approach.* Luxembourg: Publications Office of the European Union, 2019.

Cisco (2017). *Cisco visual networking index: global mobile data traffic forecast update, 2016–2021*. Technical Report. Available at: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html. [Accessed 20 Apr. 2018].

Conpes (2010). *Lineamientos De Politica Para El Fortalecimiento Del Sistema De Formación De Capital Humano SFCH.* [online] Available at: http://www.oas.org/juridico/pdfs/mesicic4_col_con3674.pdf. . [Accessed 09 Jul. 2018].

Cunha, F., & Heckman, J. (2007). *The technology of skill formation*. American Economic Review, 97(2), 31-47.

Cunningham, W. V., & Villaseñor, P. (2016). *Employer voices, employer demands, and implications for public skills development policy connecting the labor and education sectors.* The World Bank Research Observer, 31(1), 102-134.

DANE. (2009). *Metodología Gran Encuesta Integrada de Hogares.* Available at: https://www.dane.gov.co/files/investigaciones/fichas/Gran_encuesta_integrada_hogares.pdf. [Accessed 13 Jun. 2018]

DANE. (2017a). *Empleo Informal y Seguridad Social.* [online] Available at: http://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-informal-y-seguridad-social [Accessed 05 Dic. 2017].

DANE (2018a). *Human Capital Formation Survey.* [online] Available at: https://www.dane.gov.co/index.php/en/statistics-by-topic-1/education/human-capital-formation-survey [Accessed 09 Jul. 2018].

DANE (2018b). *Informe Gestion Dane-Fondane.* [online] Available at: http://www.dane.gov.co/files/control_participacion/rendicion_cuentas/InformeGestionDane_Fondane_13.pdf. [Accessed 09 Jul. 2018].

Dasgupta, P., & Stiglitz, J. (1988). *Learning-by-doing, market structure and industrial and trade policies.* Oxford Economic Papers, 40(2), 246-268.

Dehnbostel, P. (2002). *Bringing work-related learning back to authentic work contexts. Transformation of learning in education and training: Key qualifications revisited*, 190-202. Transformation of learning in education and training. Cedefop.

Deming, D., & Kahn, L. B. (2018). *Skill requirements across firms and labor markets: Evidence from job postings for professionals.* Journal of Labor Economics, 36(S1), S337-S369.

Desjardins, R. & Rubenson, K. (2011). *An Analysis of Skill Mismatch Using Direct Measures of Skills*. OECD Education Working Papers, No. 63, OECD Publishing. http://dx.doi.org/10.1787/5kg3nh9h52g5-en

Dickerson, A., Wilson, R., Kik, Genna., and Debra, D. (2012). *Developing Occupational Skills Profiles for the UK: A Feasiblity Study. UK Commission for Employment and Skills*. UKCES Evidence report, 44

Doeringer, P & Piore, M. (1971). *Internal labor markets and manpower analysis.* Lexington, Mass., D.C., Heath.

ECLAC (2016). *Estado de la banda ancha en América Latina y el Caribe 2016.* CEPAL

Edelman, B. (2012). *Using internet data for economic research.* Journal of Economic Perspectives, 26(2), 189-206.

Emsi (2013). How Should We Look At Jobs? *A Discussion of Labor Market Data and Job Postings.* [online] Available at: https://www.economicmodeling.com/2013/04/09/how-should-we-look-at-jobs-a-discussion-of-labor-market-data-and-job-postings/. [Accessed 05 Dec. 2016].

Emsi (2018). *New skills taxonomy update.* Available at:https://www.economicmodeling.com/2018/06/14/new-skills-taxonomy-update/. [Accessed 20 August. 2016].

Escudero, Veronica & López Mourelo, Elva & Pignatti, Clemente. (2016). *What works: Active labour market policies in Latin America and the Caribbean.* Studies on Growth with Equity.

Eurostat (2017). *Job Vacancies.* [online] Available at: http://ec.europa.eu/eurostat/web/labour-market/job-vacancies [Accessed 10 Jan. 2017].

Farm, A., & Sweden, S. (2003). *Defining and Measuring Unmet Labour Demand.* WP, 1, 2003.

Gambin, L., Green, A. and Hogarth, T. (2009). *Exploring the links between skills and productivity. Final report.* Warwick Institute for Employment Research.

Gambin, L., Hogarth, T., Murphy, L., Spreadbury, K., Warhurst, C., & Winterbotham, M. (2016). *Research to understand the extent, nature and impact of skills mismatches in the economy:* May 2016.

Garibaldi, P. (2006). *Personnel Economics in Imperfect Labour Markets.* Oxford University Press.

González-Velosa, C., & Rosas-Shady, D. (2016). *Avances y retos en la formación para el trabajo en Colombia.* BID

Green, F., Machin, S., & Wilkinson, D. (1998). *The meaning and determinants of skills shortages*. Oxford Bulletin of Economics and Statistics, 60(2), 165-187.

Green, F. & Zhu, Y. (2008). *Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education.* Department of Economics Discussion Paper, University of Kent, No. 08,03

Green, F. (2011). *What is skill? An inter-disciplinary synthesis*. Centre for Learning and Life Changes in Knowledge Economies and Societies.

Grugulis, I., Warhurst, C., & Keep, E. (2004). *What's happening to 'skill'. The skills that matter,* 1-18. London: Palgrave, 288-314.

Hamermesh, D. S. (1996). *Labor demand.* princeton University press.

Handel, M. (2012). *Trends in Job Skill Demands in OECD Countries.* OECD Social, Employment and Migration Working Papers, No. 143, OECD Publishing, Paris. http://dx.doi.org/10.1787/5k8zk8pcq6td-en

Hussmanns, R. (2004). *Statistical definition of informal employment: Guidelines endorsed by the Seventeenth International Conference of Labour Statisticians (2003).* In 7th Meeting of the Expert Group on Informal Sector Statistics (Delhi Group) (pp. 2-4).

ILO (2003). *Guidelines concerning a statistical definition of informal employment, endorsed by the Seventeenth International Conference of Labour Statisticians* (November-December 2003); in: Seventeenth International Conference of Labour Statisticians (Geneva, 24 November - 3 December 2003), Report of the Conference; Doc. ICLS/17/2003/R; International Labour Office, Geneva, 2003

ILO. (2008). *ISCO-08: Introductory and methodological notes.* International Labour Office.

ILO (2011). *Statistical update on employment in the informal economy.* Geneva: ILO Department of Statistics.

ILO (2012a). *Measurement of the Informal Economy.* International Labour Office, Geneva, 2003

ILO (2012b*). International Standard Classification of Occupations Structure, group definitions and correspondence tables*. Geneva.

ILO (2014). *Policies for the formalization of micro and small enterprises in Colombia.* ILO

ILO (2015). *Panorama Laboral 2015. America Latina y el Caribe.* Oficina Regional para América Latina y el Caribe.

ILO (2016). *Labour Overview 2016.* Latin America and the Caribbean ILO Regional Office for Latin America and the Caribbean.

ILO (2017a). *Laborsta*. [online] Available at: http://laborsta.ilo.org/applv8/data/c2e.html [Accessed 10 Jan. 2017].

ILO (2017b). *The Future of work, employment and skills in Latin America and the Caribbean*. ILO for Latin America and the Caribbean.

ILO (2018). *ILO Thesaurus*. [online] Available at: http://ilo.multites.net/defaulten.asp [Accessed 15 Mar. 2017].

Kankaraš, M., Montt, G., Paccagnella, M., Quintini, G., & Thorn, W. (2016*). Skills Matter: Further Results from the Survey of Adult Skills.* OECD Skills Studies. OECD Publishing.

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success (No. w20749).* National Bureau of Economic Research.

Kugler, A., & Kugler, M. (2009). *Labor market effects of payroll taxes in developing countries: evidence from Colombia.* Economic development and cultural change, 57(2), 335-358.

Kuhn, P. J. (2014). *The internet as a labor market matchmaker.* IZA World of Labor.

Kureková, L. M., Beblavy, M., & Thum, A. E. (2014). *Using internet data to analyse the labour market: a methodological enquiry* (No. 8555). IZA Discussion Papers.

Kureková, L. M., Beblavy, M., Haita, C., & Thum, A. E. (2016). *Employers' skill preferences across Europe: between cognitive and non-cognitive skills.* Journal of Education and Work, 29(6), 662-687.

Laney, D. (2001). *3-D Data Management: Controlling Data Volume, Velocity and Variety.* Gartner Blog Network. Meta Group inc. Stamford, CT. Web. 10 Jul. 2014.

Lindqvist, E., & Vestman, R. (2011). *The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment.* American Economic Journal: Applied Economics, 3(1), 101-28.

Manpower (2016). *Global press release. Talent shortage.* [online] Available at: https://www.manpowergroup.com/wps/wcm/connect/8ccb11cb-1ad4-4634-84ea-1656ee74b3ed/GlobalTalentShortageSurvey-PressRelease.pdf?MOD=AJPERES&ContentCache=NONE& [Accessed 10 Dec. 2016].

Marinescu, I., & Wolthoff, R. (2016). *Opening the black box of the matching function: The power of words* (No. w22508). National Bureau of Economic Research.

Maurer, S. D., & Liu, Y. (2007). *Developing effective e-recruiting websites: Insights for managers from marketers.* Business horizons, 50(4), 305-314.

Mavromaras, K., Healy, J., Richardson, S., Slonane, P., Wei, Z., and Zhu, R. (2013). *A System for Monitoring Shortages and Surpluses in the Market for Skills.* Final Report to the Australia Workforce and Productivity Agency (AWPA) by the National Institute of Labour Studies (NILS). National Institute of Labour Studies Flinders University, Australia.

Mazza, J. (2017). *Jobs and Job Search in Developing Countries: Nice Work if You Can Get it! In Labor Intermediation Services in Developing Economies* (pp. 1-18). Palgrave Macmillan US.

Mincer, J. (1958). *Investment in Human Capital and Personal Income Distribution*. Journal of Political Economy, 66(4), 281-302. Retrieved from http://0-www.jstor.org.pugwash.lib.warwick.ac.uk/stable/1827422

McGowan, M. A., & Andrews, D. (2015). *Skill Mismatch and Public Policy in OECD Countries.* Economics department working papers no. 1210

McGuinness, S., & Ortiz, L. (2016). *Skill gaps in the workplace: measurement, determinants and impacts. Industrial relations journal,* 47(3), 253-278.

McGuinness, S., & Pouliakas, K. (2017). *Deconstructing Theories of Overeducation in Europe: A Wage Decomposition Approach. In Skill Mismatch in Labor Markets* (pp. 81-127). Emerald Publishing Limited.

Mondragón-Vélez, C., Peña, X., & Wills, D. (2010). *Labor market rigidities and informality in Colombia*. Economía, 11(1), 65-95

Mortensen, D. (1970). *Job Search, the Duration of Unemployment, and the Phillips Curve.* The American Economic Review, 60(5), 847-862. Retrieved from http://0-www.jstor.org.pugwash.lib.warwick.ac.uk/stable/1818285

Mortensen, D. T. and Pissarides, C. A. (1994) *Job Creation and Job Destruction in the Theory of Unemployment.* Review of Economic Studies, July 1994, 61 (3), 397–415.

National Research Council. (2010). *A database for a changing economy: Review of the Occupational Information Network (O* NET).* National Academies Press.

OECD. (2011). *Quality Framework and Guidelines for OECD Statistical Activities.* [online] Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=std/qfs(2011)1&doclanguage=en. [Accessed 19 Jan. 2018].

OECD. (2012). *Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies*. OECD Publishing. http://dx.doi.org/10.1787/9789264177338-en

OECD (2014a). *Preventing unemployment and underemployment from becoming structural.* [online] Available at: http://www.oecd.org/g20/topics/employment-and-social-policy/OECD-Preventing-unemployment-and-underemployment-from-becoming-structural-G20.pdf. [Accessed 23 Sept. 2017].

OECD (2014b). *Education at a Glance 2014. Colombia.* OECD Publishing. http://www.oecd.org/education/Colombia_EAG2014_CountryNote_ENG.pdf

OECD. (2015a). Colombia Policy Priorities For Inclusive Development, OECD Publishing. Available at: https://www.oecd.org/about/publishing/colombia-policy-priorities-for-inclusive-development.pdf. [Accessed 29 Nov. 2017].

OECD (2015b). *Latin American Economic Outlook 2015: Youth, Skills and Entrepreneurship.* ECLAC.

OECD (2016a). *OECD Reviews of Labour Market and Social Policies: Colombia 2016*. OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264244825-en

OECD (2016b). *Education in Colombia*. OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264250604-en

OECD (2016c). *Getting Skills Right: Assessing and Anticipating Changing Skill Needs, Getting Skills Right.* OECD Publishing, Paris, https://doi.org/10.1787/9789264252073-en.

OECD (2017a). *Financing SMEs and Entrepreneurs 2017: An OECD Scoreboard*. OECD Publishing, Paris. http://dx.doi.org/10.1787/fin_sme_ent-2017-en

OECD (2017b). *Latin American Economic Outlook 2017: Youth, Skills and Entrepreneurship.* OECD Publishing, Paris, https://doi.org/10.1787/leo-2017-en.

Okay-Somerville, B. and Scholarios, D. (2013). *Shades of grey: Understanding job quality in emerging graduate occupations.* Human Relations, 66(4), pp. 555-585. (doi:10.1177/0018726712465094).

Onetcenter (2016*). About O*NET.* [online] Available at: https://www.onetcenter.org/overview.html [Accessed 06 Dec. 2016].

ONS (2018a). *Labour Force Survey (LFS).* [online] Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveylfsqmi. Accessed 01 Jul. 2018].

ONS (2018b). *Vacancy survey.* [online] Available at: https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/vacancysurvey. Accessed 01 Jul. 2018].

Oyer, P., & Schaefer, S. (2010*). Personnel economics: hiring and incentives* (No. w15977). National Bureau of Economic Research.

Özköse, H., Ari, E., and Gencer C. (2015). *Yesterday, Today and Tomorrow of Big Data.* Procedia-Social and Behavioral Sciences, 195, 1042-1050.

Palmer, R. (2017). *Jobs and skills mismatch in the informal economy.* International Labour Organization 2018

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

Pierre, G., Puerta, M. L. S., Valerio, A., & Rajadel, T. (2014). *STEP Skills Measurement Surveys.* The World Bank.

Psacharopoulos, G. (1985). *Returns to education: a further international update and implications.* Journal of Human resources, 583-604.

Psacharopoulos, G. (2006). *The value of investment in education: Theory, evidence, and policy.* Journal of Education Finance, 113-136.

Reich, M., Gordon, D. M., & Edwards, R. C. (1973). *A theory of labor market segmentation.* The American Economic Review, 63(2), 359-365.

Reimsbach-Kounatze, C. (2015), *The Proliferation of "Big Data" and Implications for Official Statistics and Statistical Agencies: A Preliminary Analysis.* OECD Digital Economy Papers, No. 245, OECD Publishing. http://dx.doi.org/10.1787/5js7t9wqzvg8-en

Rodriguez, M., Helbing, D., & Zagheni, E. (2014). *Migration of Professionals to the US. In International Conference on Social Informatics* (pp. 531-543). Springer, Cham.

Rothwell, J. (2014). *Still searching: Job vacancies and STEM skills*. *Report.* Washington: Brookings Institution.

Saavedra, J. E., & Medina, C. (2012). *Formación para el Trabajo en Colombia*. Universidad de los Andes, Facultad de Economía, CEDE.

Salvatori, A. (2015). *The anatomy of job polarisation in the UK*. Technical report, IZA Discussion Papers.

Sánchez Molina, E. (2013). *Clasificación nacional de ocupaciones*. Versión 2013.

SENA (2015). *Informe: Operación Estadística Para Seguimiento A Las Condiciones De Empleabilidad Y Desempeño De Los Egresados Del Sena.* Bogotá. SENA.

Skills survey (2018). *Vacancy Survey.* [online] Available at: https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/vacancysurvey. [Accessed 01 Jul. 2018].

Spence, M. (1978). *Job market signalling. In Uncertainty in Economics.* Quarterly Journal of Economics, 90(4), 591-597

Stats.oecd.org (2018). *Glossary of statistical terms.* [online] Available at: http://stats.oecd.org/glossary/detail.asp?ID=3123 [Accessed 10 Mar. 2018].

Štefánik, M. (2012). *Internet Job Search Data as a Possible Source of Information on Skills Demand (with Results for Slovak University Graduates).* In Building on Skills Forecasts — Comparing Methods and Applications, edited by CEDEFOP. Luxembourg: Publications Office of the European Union. http://www.cedefop.europa.eu/EN/Files/5518_en.pdf

Stiglitz, J. E., Walsh, C. E., Gow, J., Guest, R., Richmond, W., & Tani, M. (2013). *Principles of economics.* John Wiley & Sons.

Störmer, E., Patscha, C., Prendergast, J., Daheim, C., Rhisiart, M., Glover, P., & Beck, H. (2014). *The future of work: jobs and skills in 2030.* UKCES February 2014

The Conference Board (2018). *The Conference Board Help Wanted OnLine® (HWOL)* . [online] Available at: https://www.conference-board.org/data/helpwantedonline.cfm [Accessed 20 August. 2018].

Tjdens K., Bebblavý M., & Thum-Thysen A (2015). *Do educational requirements in vacancies match the educational attainments of job-holders? An analysis of web-based data for 279 occupations in the Czech Republic.* InGRID working paper.

Turrell, A., Speigner, B., Djumalieva, J., Copple, D., & Thurgood, J. (2018). *Using job vacancies to understand the effects of labour market mismatch on UK output and productivity.* Bank of England Working Paper No. 737

UK Commission for Employment and Skills (UKCES). (2014). *The labour market story: the state of UK skills.* UKCES July 2014.

Vallas, S. P. (1990). *The concept of skill: A critical review.* Work and Occupations, 17(4), 379-398.

Vivian, D. (2016). *Employer Skills Survey 2015: UK results.* UKCES

Warhurst, C., Mayhew, K., Finegold, D., & Buchanan, J. (Eds.). (2017). *The Oxford Handbook of Skills and Training.* Oxford University Press.

Williams, R. D. (2004). *The demand for labour in the UK. An introduction to the topic illustrated with data from two regions.* Labour Market Trends, 112(8), 321-330.

World Bank (2010). *Informality in Colombia: implications for worker welfare and firm productivity* (English). Washington, DC: World Bank. http://documents.worldbank.org/curated/en/385321468243561833/Informality-in-Colombia-implications-for-worker-welfare-and-firm-productivity

World Bank (2018a). *International Comparison Program database.* [online] Available at: https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?locations=ZJ-CO-1W [Accessed 15 Ene. 2018].

World Bank (2018b). *International Comparison Program database.* [online] Available at: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD?locations=CO-DE-OE [Accessed 15 Ene. 2018].

World Bank (2018c). *Education Statistics: Education Attainment.* [online] Available at: http://databank.worldbank.org/data/reports.aspx?source=Education-Statistics:-Education-Attainment [Accessed 05 Apr. 2018].

Agradecimientos

Acknowledgments