The **online job market trace** in Latin America and the Caribbean

Martin Hilbert
Kangbo Lu

United Nations

ECLAC

german cooperation
DEUTSCHE ZUSAMMENARBEIT

Thank you for your interest in this ECLAC publication

Please register if you would like to receive information on our editorial products and activities. When you register, you may specify your particular areas of interest and you will gain access to our products in other formats.

Register

UNITED NATIONS

ECLAC

www.cepal.org/en/publications

facebook.com/publicacionesdelacepal

www.cepal.org/apps

Project Documents

The online job market trace in Latin America and the Caribbean

Martin Hilbert
Kangbo Lu

UNITED NATIONS

ECLAC

german cooperation
DEUTSCHE ZUSAMMENARBEIT

This document has been prepared by Martin Hilbert and Kangbo Lu, Consultants of the Economic Development Division of the Economic Commission for Latin America and the Caribbean (ECLAC), within the activities of the project, "Sustainable development paths for middle-income countries within the framework of the 2030 Agenda in Latin America and the Caribbean (2018-2020)", "digital transformation" cluster, implemented by ECLAC together with Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and financed by the Federal Ministry for Economic Cooperation and Development (BMZ) of Germany.

We thank Timothy Zhang for his collabouration and input, and Sonia Gontero and Jürgen Weller for guidance in the design of the study, as well as for feedback on diverse drafts.

United Nations publication
LC/TS.2020/83
Distribution: L

Printed at United Nations, Santiago
S.20-00406

This publication should be cited as: M.Hilbert and K. Lu,"The online job market trace in Latin America and the Caribbean", *Project Documents*, (LC/TS.2020/83), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC), 2020.

Contents

Tables

Figures

Diagrams

Maps

Introduction

Online platforms for labour markets are the most noticeable current innovation in today's job marketplace, and therefore one of the most tangible solutions to tackle several of the objectives outlined in Sustainable Development Goal 8 ("Decent Work and Economic Growth"). While these platforms play an increasingly important role in the day-to-day of workers and employers, very little is known about their constitution and role, and what they can reveal about labour market dynamics in general.

This reports contributes to the following aspects to the ongoing discussion:

- **Tracking new developments:** Labour market online platforms are used as both intermediary systems for offline job delivery and as a contract platform for online service delivery, primarily through the so-called 'online gig economy', 'platform economy', 'digital labour', or 'on-demand economy'. The former digitalizes traditional 'domestic job-call' dynamics, while the latter is an essential innovation of recent years, since, traditionally, service delivery had been limited to geographical boundaries. We are studying both aspects of the labour market in parallel. This differentiates our analysis from others that exclusively focus on online service provision (Graham et al., 2017; Kässi & Lehdonvirta, 2018).

- **Geographical reach:** The digital footprint allows obtaining data from countries that often fall through the cracks in traditional (survey-based) statistics. Many of our exercises include 33 countries from LAC, including many small countries from Central America and the Caribbean.[1] From a methodological standpoint, it is arguable if it justifies to include cases that only have a handful of data points (e.g., some Caribbean countries with only a few dozen available job calls per day). Hand-wavy rule-of-thumb heuristics from introductory

[1] We cover the following 33 countries and, if applicable, group them as follows into different sub-regions: Andean Region: Bolivia, Colombia, Ecuador, Peru; Caribbean: Antigua and Barbuda, Bahamas, Barbados, Cuba, Dominica, Dominican Republic, Grenada, Guyana, Haiti, Jamaica, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago; Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama; North America: Mexico; South America: Brazil, Paraguay, Suriname, Venezuela; Southern Cone: Argentina, Chile, Uruguay.

statistics classes would suggest that 30 is "the magical number" (Kar & Ramalingam, 2013) to determine the minimum sample size requirement. However, we do not do actual statistical tests here, but simply report what is out there, so we decided to keep these numbers, even if they are small.

- **Supply and demand:** Traditionally, labour market analysis in LAC was informed by surveys about participation, which mainly reflects on the supply of labour, neglecting the demand side. The later was scattered across newspaper ads and bulleting boards, practically impossible to gather. Several labour market platforms provide data about both supply (talent profiles of professionals offering themselves), and demand (employers specifying job opportunities). Some have a national focus, while others have international reach. The new information allows us to study the mismatch between supply and demand, which was previously missing. This distinguishes our analysis from others, which exclusively focuses on online job postings, vacancies, and tasks (González Velosa & Peña Tenjo, 2019; Kässi & Lehdonvirta, 2018).
- **Temporal dynamics:** The digital footprint is updated in real-time with every digital step taken by any online user. Continuous observations allow researchers to study ongoing dynamics over time, in a scalable and sustainable manner. The change of information over time allows understanding cycles and periodicities, as well as the identification of trends and tendencies. This is another aspect that distinguishes the current analysis from previous exercises (UN ECLAC, 2019). The real benefits from this opportunity will be seen only over longer and longer periods.
- **Detailed insights:** The digital footprint allows analysts to obtain more fine-grained insights than traditional surveys, which facilitates us to ask new questions. There are no limits to the creativity and the sophistication with which new questions can be pursued. As some exploratory examples, we take advantage of the crowd-sourced nature of skill specifications on some platforms, which leads to distinctions among more than a thousand different skills. We use this unprecedented granularity to study the continuously changing skill set of today's labour markets in real-time. This sophistication of information design follows the lead of several innovative academic studies aimed at understanding the underlying forces of change in the labour market (Alabdulkareem et al., 2018; Frank et al., 2019; Guerrero & Axtell, 2013).
- **Interactive analysis:** In this exercise, we pay special attention to present alternative forms of data visualization. Modern data visualization includes traditional graphs and maps (as shown in this print report), but also interactive online tools in the form of dashboards, which allow a researcher to undertake their own, tailor-made analysis of the gathered data. We complement the static screenshots presented in this print report with interactive dashboards available online.

The methodological contributions of the report stem from the lessons learned from this exercise of working with online trace data. In this sense, the report serves a rough guide for practitioners interested in using modern data science for development policies. Chapter I discusses many of the methodological issues, and several sections come back to the tricky nature of doing measurements within this new paradigm. Making sense of these new data sources in a meaningful way includes computational challenges, but goes much further, and touches on the definition of data science as the convergence between computer science, statistics, and its substantive application area, in this case, labour market economics.

Chapter II discusses the freelancing market. The gig-economy has become an important aspect of modern labour market dynamics. They are on demand side-jobs. Chapter III discusses domestic labour market portals.

For both markets, we analyze labour demand (job calls) and supply (job profiles) and analyze the match between supply and demand. This shows the mismatch between demand and supply per industries and sector in different countries. We find that LAC as a whole has a notable oversupply of professionals focusing on 'Administration, Financing' and 'Management, Consulting', while there is a clear undersupply in 'Commerce and Sales' and 'Media, Marketing and Communication'. We also analyze salaries and hourly rates. Among other things, we find that freelancer jobs are very well paid and therefore likely quite sophisticated, as the freelancing hourly rates pay about 10 times the national minimum wage. Analyzing the work experience of profiles in domestic markets, we find that 54 % of the profiles have less than 5 years of experience, wich seems to suggest that these new opportunities are taken up by a new generation of job-seeking workers. A complementary analysis of educational attainment shows that more than half of the workers on the national online market have a university degree, which suggests that, for now, these new forms of labour intermediation are taken advantage of by the more educated workforce.

Chapter IV takes a look at the temporal dynamics. In contrast to traditional survey-based sources, digital trace data lends itself to time series analysis. We identify a clear weekly cycle, as well as a seasonal cycle.

Chapter V explores two additional topics that take advantage of the fine-grained data available in online trace data. We explore the skills-landscape in the form of a skill-network. We identify clear groups of skills and show that this new datasource makes it possible to track the ongoing skill-evolution in action. Finally, we also discuss the implication of the digital wealth of data for data visualization. We argue that the wealth of data calls for the use of an interactive dashboard, which allows the user to identify the variables of interest. Complementary to the selection of chosen graphs printed in this traditional report, we have developed such interactive dashboards, which add thousands of possible variations to these printed graphs (available online).

In the last seccion, we report some of the lessons learned throughout this exercise. We discuss the benefits of working with these new sources, such as their impressive thematic and geographical coverage, level of detail and timeliness, as well as challenges, such as questions of representativeness and generalizability.

I. Methodological considerations

The involved analysis always follows a similar workflow, independent from the particularities of the source, its content, and the pursued research question. It starts with (1) the collection of online data, followed by (2) methodological decisions and data cleaning, and (3) ending with visualization and analysis (see diagram 1). In reality, this process is more like a circular loop, since the identification of adequate sources always depends on the intended outcome, which, at the same time, is limited by the available data and the technical and skill-based data processing capacities, etc.

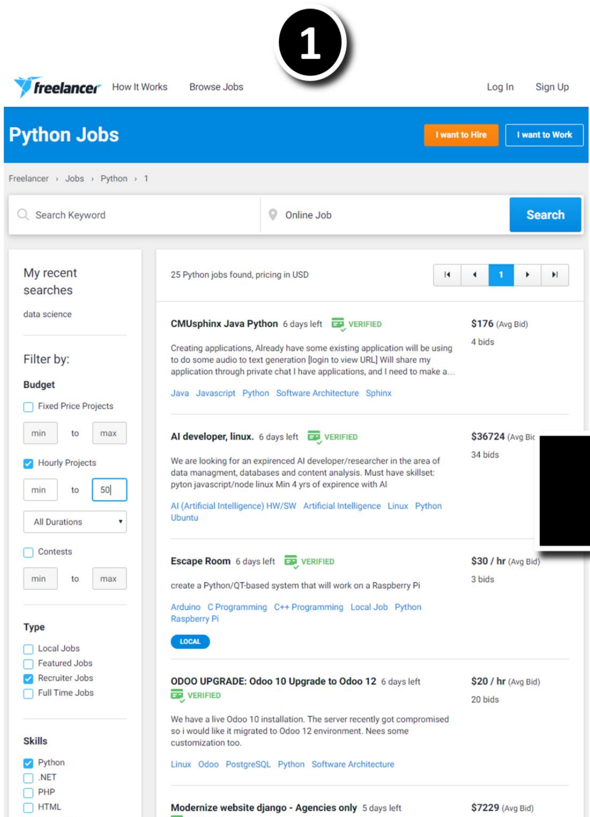
In practical terms, each step can be done in a more or less automated manner and with the help of more or less sophisticated tools. Roughly speaking, the first step of data collection requires the most computational skills (or, at least, the most non-traditional computational skills); the second requires predominantly statistical skills; and the final third step requires analytical skills with specific domain knowledge. The result is the definition of data science as the convergence between computer science, statistics, and its substantial application area.

A. Collection options

There are several possible ways to collect online data. In principle, everyone able to navigate a mouse-curser could execute a simple version of the first step by copy-pasting content from the open internet. This way of collecting data is often a reasonable choice if one is faced with a one-time task, but it does not scale. In our case, gathering data manually for more than thirty countries can be a tremendous amount of labours of copy-paste over time. Among the more automated options are web-crawling, web-scraping, and API (Application Programming Interface) access. The first two are essentially machine-enabled automation of the copy-paste mechanism, while the latter consists of direct access to a database provided by the platform. Naturally, only publicly accessible contents can be collected legally. If the collector does not have access to specific content (e.g., a private forum, a hidden page, or a missing data point in a database), the content cannot be crawled, scraped, or accessed.

Diagram 1

Schematic presentation of the data science workflow with digital trace data from data collection, over data processing, to data visualization



2

	Main category	Skills count
Accounting	2,085	11,904
Engineering	25,386	117,197
HR & Legal	20,733	94,339
Media	7,502	36,429
Translation	3,045	16,904
Business	17,734	79,359
Data Entry	1,278	9,725
Design	77,763	365,857
IT	2,522	24,791
Marketing	24,791	19,859
Science	19,859	7,746
Software	7,746	3,551
Writing	3,551	17,054
Other	17,054	2,240

Source: Author's elaboration

Web-crawling, sometimes called a spider or spiderbot, is the right choice if one does not have a predefined list of sources. For example, web search engines like Google and Bing use distributed crawlers to update their content by collecting and indexing website content with ranking systems. The benefit is that crawlers can collect content from webpages with the most diverse web-designs, while the downside is that the obtained content is usually not harmonized, contains many ambiguities and standardization challenges.

If the focus is set on a platform that hosts data in a harmonized design framework, web-scrapers are the right choice. In the past era of Web 1.0, websites were statically built with XML using conventional technology such as ASP, JSP, and served with Java and PHP. In the present era of Web 2.0, web pages are built using text-based mark-up languages (HTML4 and HTML5) with the addition of CSS3 for styles, and JavaScript ES6 standard for the interactivity and the dynamic loading user experience. Users take advantage of this harmonized design to quickly assess the content without having to orient themselves anew after each click. In the same way, web-crawling programs can take advantage of the harmonized design to collect data. The program then basically accesses page after page, copies the specified content (not necessarily all of it, but the parts specified by the collector), and registers information into a pre-designed database (such as a simple spreadsheet, like an Excel or CSV file). There are several out-of-the-box web scraping tools readily available that can be learned within minutes (a quick online search for 'web scraping tools' will display dozens of options). Some of these open source tools are free, while others offer web-scraping as a user-friendly service. The most common and flexible solution consists of programming scrapers in-house, which is usually done with Python, a modern multipurpose programming language. Popular Python packages for writing web scraping programs include Selenium (www.seleniumhq.org) for the interactive web scraping and Beautiful Soup (www.crummy.com/software/BeautifulSoup) for processing scraper-returned HTML contents.

It is important to note that some platforms specify that crawling or scraping of their sites would be against their **terms of service**. This term is their legal right in most countries, being private companies. It is important to note that such terms of service do not necessarily say anything about the ruling law. In principle, everything on the open internet is accessible by the public, but the portal and its content are still the private property of the provider. Therefore, under most legislations, the owner has the right to ban users who do not comply with their particular rules of service. Still, this term is only a written agreement, and it does not necessarily elabourate on the legal status of that behavior. Think of a private restaurant: it has the right to deny access to an individual whose behavior is not in line with their code of conduct, such as that no shorts are allowed within the premises. While this does not mean that the wearing of shorts would be illegal in that country, the owner still has the right to ban an individual from using the facilities.

Any question regarding **intellectual property rights** depends on the ruling legislation, as well as the nature of the collection and its reuse. Think of job posts presented on public bulletin boards or in paper-based newspaper ads. Counting those ads and displaying summary statistics about them does not automatically infringe on any property rights. Copying them literally and selling the same content in a different venue might pose a legal problem. The same accounts for online job postings. In this sense, it is essential to inform oneself about the particular terms of service of a portal, and the ruling legislation of the domestic context, just in the same way that it is important to inform oneself about the terms of service of a brick-and-mortar store.

The technical specifications regarding crawling and scraping can be verified by consulting the robots exclusion standard, also known as the robots exclusion protocol or simply robots.txt, which is a standard protocol used by websites to communicate with web crawlers, scrapers, and other web robots. The standard specifies how to inform the web robot about which areas of the website should not be processed or scanned. The protocol is stored in the root directory of the website's server. It is consulted by adding the appendix .../robots.txt to the domain of a website without sub-directories or the root directory of a specific website, e.g. see www.workana.com/robots.txt. Some websites prefer not to be indexed or scraped, given that such activity burdens their servers, which compete for processing with requests from human clients. However, many sites prefer to opt for a rather lax set of restrictions when it comes to crawling and scraping since this assures that the platform is present in automated search engines. This automated processing creates traffic on the site, which then again increases the presence of the site in the web-search hierarchy. From 2012 to 2015, an Internet security company estimated that bots produced a quite stable 50% share of online traffic (Zeifman, 2015).

Last but not least, obtaining data directly from the back-office through offered API services as a registered developer is usually considered a more official, controlled, and sustainable entry through the front door of the data-edifice. Online platforms cannot control who scrapes or crawls their content, but they can control who enters their API (through trackable tokens that provide limited access). From the perspective of the data provider, the benefits of providing API access include easing the processing load of the company's servers and providing a certain degree of control over data usage. It also takes advantage of positive network effects that happen when other services connect to the selected services provided by the company (think about the Google Maps and Google Translate APIs, or flight databases from airlines, which provide APIs that allow other services to build onto their data). The downside to APIs is that they require maintenance on behalf of the platform provider. Several important service providers in the region do not provide APIs.

1. Source selection

In our analysis, we focus on both labour market demand and supply. Demand consists of the online posting of job calls, vacancies, and tasks. Labour market supply consists of online postings of worker profiles, curricula, and the proactive offering of certain skills and services by workers.

The period of the data collected for this report was for the nine weeks from Monday, Oct. 14, 2019, to Sunday, Dec. 15, 2019. As a first step in the process of source selection, we reviewed the most important platforms in the region. We evaluated the volume, reach, and scope of the platform's content, as well as its Alexa rank.[2] Our considerations for inclusion were guided both by desires to create an ideal sample in terms of representativeness and diversity, and by practical and pragmatic considerations related to legality, scalability, and sustainability of any data collection effort.

First, we decided against the inclusion of national portals, such as Zonajobs and Postularse from Argentina, BNE.cl from Chile, and OCCMundial from Mexico). The primary reason for this decision is that this study aims at creating a comparative perspective among LAC countries. The data harmonization effort that would be required for the comparative processing of these national portals would be beyond the resource-scope of this current project.

We then identified the main international online labour market platforms active in the region. We found twelve, namely Bumeran, CaribbeanJobs, CaribbeanJobsOnline, CompuTrabajo, Freelancer, Indeed, JobisJob, Jooble, Opcionempleo, Profdir, Upwork, and Workana. Unfortunately, we had to exclude Workana, because it does not allow for web-scraping and its API was not functional at the time of our data collection. We eventually also decided against the inclusion of Indeed, Jooble, and Opcionempleo: Indeed turned out not to be very prevalent in LAC countries; Opcionempleo and Jooble limit public access to data (e.g., only giving access to the first 50 profiles or jobs, without specification

[2] Alexa Internet is an Amazon.com company and provides a ranking designed to be an estimate of a website's popularity. The rank is calculated from a combination of daily visitors and page views on a website over a 3-month period. It is not necessarily representative, since it requires the installation of a plugin from this amazon.com company, but is often used as a proxy for a websites popularity.

of how these were selected); and Jooble seems to duplicate information from other platforms. This left us with eight platforms (diagram 2).

As shown in diagram 2, these sites cater to two different markets. On the one hand, Profdir (acciontrabajo), CaribbeanJobs, CaribbeanJobsOnline, CompuTrabajo, JobisJob, and Bumeran are sites that focus on rather traditional employment in domestic markets. Each of their national portals has a different URL and is adjusted to national markets, including domestic laws, customs, and terminology. These platforms usually specialize in job postings and vacancy announcements and therefore focus on the demand for labour. In short, they act like the digital version of a newspaper with job calls. Among them, Profdir (acciontrabajo) is an exception, as it also acts as a labour supply platform, allowing workers to post individual profiles, and enabling researchers to compare supply and demand.

On the other hand, Freelancer and Upwork are two platforms focused on the online gig-economy, which is characterized by online service delivery. It is an important and growing aspect of the international and national labour market, also in developing countries (Heeks, 2017; Kässi & Lehdonvirta, 2018). However, this is quite a different market, as its focus is set on sporadic work and online delivery. Given the mostly digital nature of service delivery on these platforms, the demand for labour constitutes a global market. Employer postings of jobs and tasks have a global reach. The national location of the worker is one variable that describes the labour supply, which then allows us to compare global demand with national supply.

2. Collection method

As can be seen from diagram 2, the chosen platforms offer their services in different LAC countries. For each of these countries, the sub-domains of each platform mentioned above are built on the consistent site structures. This standardized web-design feature enables us to design and collect the data with one single scraper per sub-domain of each platform. Across different countries, our web-scrapers can obtain comparable data without the need to work with designing web-crawling programs for each site. Some of the platforms also offer APIs, especially Freelancer.com, Upwork, and JobisJob. The other platforms did not provide API services, and they did not impose any limitations to scraping at the time of our exercise, according to the /robots.txt protocol.

It is important to note that web scraping implies that one opens and closes the accessed page. Therefore, one uses the resources of the server of the company that provides the information, which effectively competes with server resources for the use of their actual clients (a certain server can only handle a given number of visits at a time). Mimicking regular web visitors' behavior, the program opens each page (such as a worker profile), then loads its content, and copy-pastes the content into a database. The cost of hosting website depends on the number of requests a website can receive and handle with a period of time. More requests imply a higher cost of hosting and more popularity, but the web-scrapers can hurt the company's business model by not creating real human activities. Therefore, many website owners ban web scrapers, saving hosting money and avoiding miscalculations of their marketing A/B testing strategies caused by web scrapers bombarding a site.

In the view of creating a sustainable collection system that allows us to obtain temporal data systematically, we would like to minimize this negative impact on the data providers, especially on the spending on web hosting resources. Besides, even if the provider would not be worried about this use of its scarce resources, from the collector's perspective, each additional step increases collection time and augments the risk for collection failure, since the automated program can get stuck at specific steps. This evaluation of the collecting strategy further pushes us to reflect on what data we need to obtain the aspired insights in a sustainable fashion, which is distinct from reflecting on what data is available.

Diagram 2
Our sources of transnational online platforms in Latin America and the Caribbean and their national presences

Source: Screenshots from websites of Bumeran, CaribbeanJobs, CaribbeanJobsOnline, CompuTrabajo, Freelancer, JobisJob, Profdir, and Upwork.

With these caveats in mind, we opted for the less invasive collection of pre-aggregated metadata in this exercise and did not open and copy-paste each profile or job call. Metadata is a set of data that describes and gives information about other data; in this case, data that describes the content of each labour market platform. In this way, we take advantage of the platform's categorization system, and merely collect the aggregate numbers of offered jobs and profiles, without the need to open each sub-page that contains the specifications of the call or profile. For example, diagram 3A. shows the Upwork filter for Colombia. The in-house search engine of the platform provides metadata on the different categories of worker profiles. For example, it shows that 270 workers have made more than US$10k, that 5,957 workers offer their services for an hourly rate of US$10 —US$30, and that 420 workers claim to have skills in Data Science & Analytics. Diagram 3B shows job offers on the Argentinean portal for CompuTrabajo. The search engine's metadata indicates that 11,791 job calls are for fulltime commitments ('jornada completa'), 2,316 vacancies are available for sales jobs ('ventas'), and 874 jobs are offered in the province of Santa Fe. Naturally, the search engine allows users to select combinatorial cross-sections of the provided categories. The resulting combinatorial possibilities increase exponentially, so it is worth reflecting on the kind of variables a researcher is interested in.

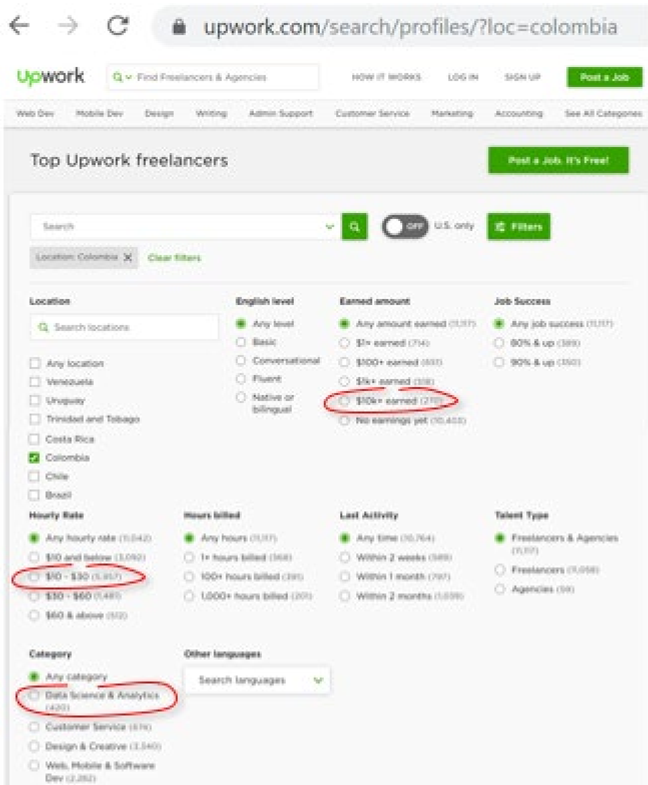
Using this kind of metadata as the primary source reduces computational and time resource involvement significantly. For example, platforms like Freelancer.com and Upwork post some 30,000 job calls per day each, and host some 150,000 profiles each. Collecting metadata for each of the 33 countries of LAC, allows researchers to obtain the data with 33 clicks, not with 150,000 clicks. The question comes down to the decision of what data is needed and for what purpose. As an intermediate step, one can also collect the search result preview data, which is often returned as preview lists by specific searches.

Diagram 4 visualizes the logic of increasing collection depth. On the right-hand side, the search engine specifies several skills, and Argentina as the location of origin for job profiles. The search returns 790 profiles. Each displayed site shows 10 profile previews in a list. Therefore, one would need to load 79 consecutive pages to obtain the summary of each profile. This summary gives useful additional information, such as the aspired hourly rate of the worker, the average review, and some additional skills that are complementary to the ones specified in the search.

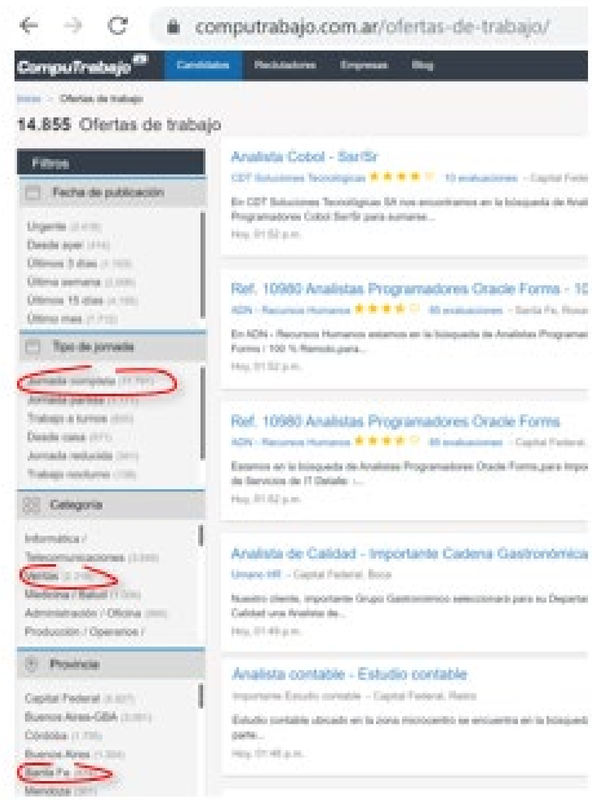
Going one step further, one could also retrieve the data from each worker (left-hand side in diagram 4. Here one obtains even more detailed information about the worker: a verbal statement (which could be processed with a natural language understanding program), details of job performance, the exact location and length of membership, examples from the worker's portfolio, recent reviews, expressed in rich natural language, a series of language certifications, with scores, top skills, previous experience, and official educational titles. This more in-depth collection provides unprecedented levels of insight into thousands of workers. However, it requires more resources (in terms of server capacities, programming tasks, and collection time), as the program will need to obtain the data from each of the 790 workers. If such detailed micro-data is desired, one would naturally prefer to work with the provider's API or ask for the direct deposit of data through the establishment of a partnership with the platform.

Diagram 3
Non-invasive collection of metadata using the platform's categorization system

A. Example of Upwork, labour supply, filter for Location: Colombia



B. Example of CompuTrabajo, labour demand, Argentinean portal



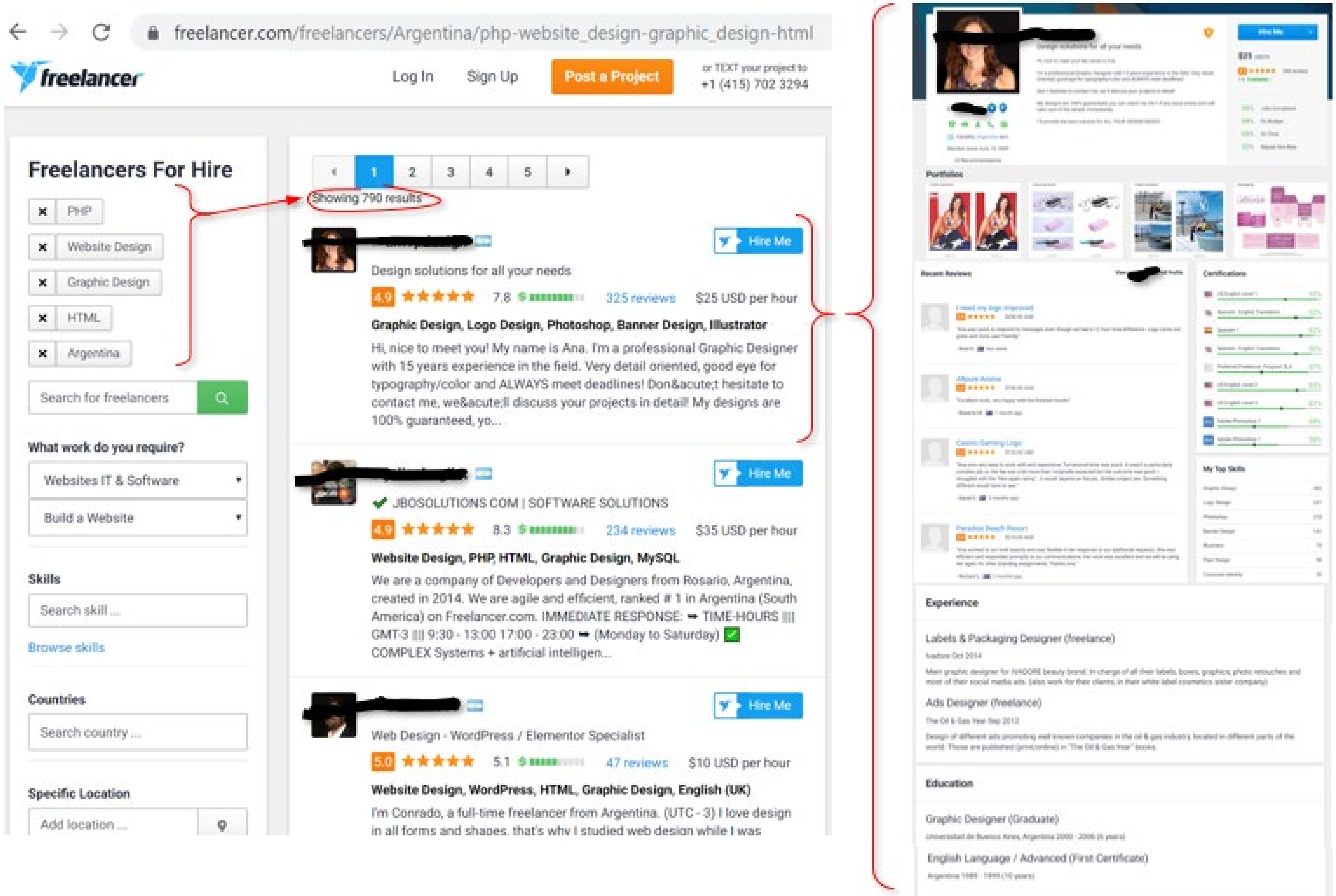
Source: Author's elaboration based on screenshots from Upwork and CompuTrabajo.

For our purposes of this first exploratory study, we do not collect the most detailed micro-data. We stay on higher levels of aggregation, including summary statistics from search results, and eventual summary statistics akin to the information provided by the intermediate aggregation level of the list in diagram 4 (e.g., for selected case studies of salary distributions). This aims at striking a balance between the terms of service of the data provider, and a principle of proportionality between input and required output. As for the former, higher-level collections are usually permissible according to the robots.txt protocol. The exception is Upwork, where we solicited access to the API, and for some cases, also write down search result numbers manually (which is more time-effective). As for the latter reason, collecting more detailed micro-data per job post and worker profile would not only require thousand-fold resource usage (first and foremost on behalf of the data provider). It would also not raise questions about possible negative side effects in terms of user privacy. However, it would also require us to have a very good justification of why we would like to collect this data: what exactly are the questions we would like to answer, and why? At this exploratory stage, we do not consider that our understanding of these dynamics is deep enough to justify any more extensive data collection.

The numbers presented throughout this report represent daily averages of all posts found over the period from Oct. 14, 2019, to Sunday, Dec. 15, 2019. The collection was done daily for all sources, except for Upwork and Profdir supply, for which it was done weekly (given more intense data collection efforts for these sites).

Diagram 4

Deeper collections: from search engine metadata, over search result preview lists, to microdata on individual profiles, on the example of the Freelancer.com interface



Source: Author's elaboration, based on screenshots from Freelancer.

Summing up, our chosen collection method of metadata summaries and preview lists favors sustainable collection aimed at minimizing computational and temporal resource requirements and risks of potential complications. It also considers the fact that concrete needs and well-justified research questions should guide responsible data science, and not merely by the availability of data (which might not be needed but could be abused once collected). This does not imply that more in-depth collection would not be possible, interesting, or valuable for well-justified future studies. Such further studies could, for example, include in-depth analysis of possible geographic discrimination in the gig economy (Galperin & Greppi, 2019), among others.

3. Harmonization of labour categories

One reason that could make it worthwhile to collect more detailed microdata refers to classification into harmonized categories that make data from different sources comparable. Categories provided by different platforms are not automatically harmonized. It is important to remember that the resulting work is nothing new for researchers working with any kind of data classification problem, including ethnographers who convert observations into rows and columns in their notebooks, or international organizations like UN ECLAC, which has worked for many decades on the intricate statistical challenge of harmonizing the results from household surveys from different countries. The ensuing work of data cleaning and preparation includes decisions about how to harmonize discrepancies and if and how to extrapolate over missing values.

Keeping things simple, transparent, and replicable, we opted for a quite traditional solution and harmonized different predefined job and skill categories manually. We looked at specific cases of the microdata, and found that the classification systems of the platforms provide sufficiently consistent results in their classification. In case of doubt, we aggregated categories into larger classes. As a result, we work with the aggregated categories provided in the left columns of table 1 and 2. The classification systems of the sites are exhaustive, so all jobs are covered in some categories.

Table 1
Harmonization scheme of collected job categories for freelancer labour markets

Harmonization	Upwork	Freelancer
Business, Accounting, HR & Legal	Accounting & Consulting	Business, Accounting, Human Resources & Legal
	Legal	
Data Entry & Administration	Admin Support	Data Entry & Administration
Design, Media, Engineering & Science	Engineering & Architecture	Design, Media & Architecture
	Data Science & Analytics	Engineering & Science
	Design & Creative	Product Sourcing & Manufacturing
IT, Networking & Software	Web, Mobile & Software Dev	Websites, IT & Software
	IT & Networking	Mobile Phones & Computing
Sales & Marketing	Sales & Marketing	Sales & Marketing
	Customer Service	
Writing, Translation & Languages	Translation	Translation & Languages
	Writing	Writing & Content
Other		Freight, Shipping & Transportation
		Local Jobs & Services
		Other

Source: Author's elaboration.

Table 2
Harmonization scheme of collected job categories for domestic labour markets

Harmonization	Bumeran	Computrabajo	Job is Job	Profdir	Caribbean jobs	Caribbean Jobs Online
Administration, Financing	Administration, Accounting, Finance	Administration/Office	Administration - Office	Administration	Banking, Financial services & Insurance	Accounting & Finance
	Secretaries and Reception	Accounting/Finance	Finance	Banking and Finance	Actuary, Accountancy & Finance, Secretarial & Administration	Secretarial Administrative & Clerical
Commerce and sales	Customs and Foreign Trade	Shopping/Foreign Trade	Sales	Import-Export	Retailing, Wholesaling & Purchasing	Retail
	Commercial, Sales and Business	Sales	Retail	Commerce	Sales	
	Trades and others			Art, Textile		
Construction and Real Estate	Civil and Construction Engineering	Construction and work	Construction - Real Estate	Building	Construction, Architecture & Property	Real Estate
				Real Estate		Construction & Building Service
				Architecture		
Education and Teaching	Education, Teaching and Research	Teaching	Education	Education	Education, Childcare & Training	Education & Training
Healthcare and Pharmacy	Health, Medicine and Pharmacy	Medicine/Health	Medicine - Health	Health	Medical Professionals & Healthcare	Healthcare & Medical
	Nursing			Pharmacy		
IT, Telecommunication	Technology, Systems and Telecommunication	Computers/Telecommunications	IT - Telecomm	Electronics	Information Technology	Information & Communication Technology
				Internet, IT, Telecomms	Telecoms	
Legal	Legal	Legal/Advice	Legal	Law	Legal	Legal
Management, Consulting, HR	Management and General Direction	Address/Management	Management	Consulting	General Management, Project Management	Consulting & Project Management
	Human Resources and Training	Human Resources	Consultancy	Human Resc	Human Resources	Human Resources & Recruitment
			Human Resources		Procurement & Supply Chain Management	Land Use & Environmental Management

Harmonization	Bumeran	Computrabajo	Job is Job	Profdir	Caribbean jobs	Caribbean Jobs Online
Maintenance and Engineering	Engineering	Engineering	Engineering	Engineering	Engineering & Architecture, Scientific	Engineering & Architecture, Scientific
		Maintenance and Technical Repairs	Science - Research	Energy	Installation, Maintenance & Repair, Research, Monitoring & Evaluation	Installation, Maintenance & Repair, Research, Monitoring & Evaluation
		Research and Quality			Automotive & Vehicle Repair, Quality Assurance	Automotive & Vehicle Repair, Quality Assurance
Media, Marketing and Communication	Customer Service, Call Center and Telemarketing	Call Center/Telemarketing	Marketing - Media	Marketing	Marketing	Advertising & Marketing
	Marketing and Publicity	Marketing/Advertising/Communication	Customer Service	Advertising	Customer Service, Call Centres & Languages	Customer Service & Call Centre
	Communication, Institutional and Public Relations	Customer Service		Public Relations	Publishing, Media & Creative Arts	Media & Corporate Communications
	Design	Design/Graphic Arts		Media-Press		
				Design		
Mining, Manufacturing and Production	Production and Manufactures	Production/Operators/Manufacturing	Production - Manufacturing	Manufacturing	Production, Manufacturing & Materials	Manufacturing
	Mining, Oil and Gas			Oil, Gas, Chemical		Mining, Resources & Energy
Social Service, Catering, Entertainment	Gastronomy and Tourism	Hospitality/Tourism	Leisure, Tourism, Beauty	Tourism	Hospitality & Tourism	Government & Public Sector
	Sociology/Social Work	General Services, Cleanliness, Security	Hospitality - Restoration	Hospitality	Security, Trades & General Services	Hospitality, Tourism & Food Service
			Social Services, security	Entertainment	Science, Pharmaceutical & Food	Security & Armed Forces
				Charity	Social, Voluntary & Vocational	Sports & Leisure
				Food Processing		
Transportation, Logistics	Supply and Logistics	Warehouse/Logistics/Transportation	Logistics - Distribution	Transport	Transport, Warehousing & Automotive	Procurement, Logistics & Supply Chain
Others	Insurance	Others		Defense, Government, Agriculture, Insurance	Environmental, Health & Safety, Public Sector	Agriculture, Forestry & Fishing, Arts, Fashion & Design, Environment, Health & Safety

Source: Author's elaboration.

II. Freelancing labour market

This section analyzed the data of the two largest freelancing platforms, Freeelancer.com and Upwork.com (see diagram 2).

A. From skills to categories

Before getting into the description of freelancing labour markets, one methodological caveat is necessary for the data stemming from Freelancer.com. It is rather technical, and the reader interested in substantial results might want to skip this section at first.

The site provides a large number of specific skills per job and worker (see diagram 4). It does not offer a clear aggregation into a broader industry level category. The 7 categories from Freelancer listed in table 1 represent some 1,665 different skills. We write code to assign skills to the categories offered by Freelancer on their website (https://www.freelancer.com/job). However, it is essential to emphasize that there are different ways of doing that, since a certain job or worker can span several different categories (see related discussion in (Kassi, 2016)). None of the different ways of assigning skills to categories is more or less correct. Each of them simply weighs things differently.

Main category: For example, one could assume that each professional or job has one main focus, some kind of skill set that mostly corresponds to one of our categories. For the sake of demonstration, let [A, A, A, B] represent a project/professional with two different skills from category A (for example 'C+', 'Java', and 'SQL' for the category 'IT') and one skill from category B (for example 'Spanish' for the category 'Language'). This method would categorize the job/professional as 'IT' and neglect the 'minority' category. We did this for one week of data from Freelancer.com, and collect 17,286 jobs and 77,805 profiles, with a distribution of main categories within. In case of a tie in skills, we choose one category at random (which happens less than 10% of our cases).

Covered categories: On the other extreme, we could also be interested in all categories covered, even if they are not the central aspect of the skillset. From this perspective, the job/professional with the skillset [A, A, A, B] would increase the count of both categories, A and B by one. We learned that the average job call asks for skills from 1.6 different categories. The average professional profile offers skills from almost two different categories (1.8 to be precise). This method of counting gives much weight to skills from categories that might not have many different skills but could act as a potential bottleneck (such as a foreign language). Some categories (such as IT) make much more fine-grained distinctions among skills (e.g., differentiating between 'C++' and 'C+') than others (e.g. 'Spanish', not 'Argentinian Spanish' and 'Chilean Spanish').

There are also two in-between approaches between these extremes. **Skills count:** One would simply count the listed numbers of skills ([A, A, A, B] -> 3*A and 1*B), counting 72,501 different skills demanded by the 17,286 jobs, and 365,857 different skills offered by the 77,805 professionals. **Skills fractions:** A similar approach would see the job/professional as a weighted unit, and count the fraction per job/professional ([A, A, A, B] -> 0.75*A and 0.25*B). Note that both last approaches would result in the same total distribution if all jobs/professionals would have the same number of skills (since the latter approach normalizes on skills per job/professional). In practice, they turn out to be similar.

One can think of four ways to analyze the skills requested in each job vacancy:

i) Main category: [A, A, A, B] -> A

ii) Covered category: [A, A, A, B] -> A and B

iii) Skills count: [A, A, A, B] -> 3*A and 1*B

iv) Skills fractions: [A, A, A, B] -> 0.75*A and 0.25*B

The following shows the resulting distributions. Comparing the changing shares of these different approaches for job postings (demand) and professional profiles (supply) underlines that there is no 'right/wrong' way of counting since the different approaches over- /under-represent different categories for demand and supply. For example, "Design, Media, Engineering & Science" is 'under-represented' when neglecting minority categories for job postings (demand, figure 1A.) and overrepresented for professional profiles (supply figure 1B). When accounting for the number of 'covered categories', instead of merely the 'main category', this category gains in share for job postings and diminishes for professional profiles. This suggests that this skill set demands many different other side-skills, while the bulk of the skill supply is rather homogeneous.

Since we adopt the platform's categorization, we do not have access to how exactly each platform assigns skills to categories. We suspect that Upwork using the approach we here called 'covered categories'. In other words, if a job/professional demands/offers at least one skill from the skillset of a certain category, this category is considered as covered. As previously stated, this increases the uniformity among categories, is less sensitive to different fine- and coarse-graining of skillsets, and given special weight to categories with a smaller (but not necessarily less important) skillset. For presentations in the rest of the report, we opt for this 'covered categories' version of how to assign skills to categories.

In practice, in Freelancer.com, we find a daily average of N = 77,757 professionals offering their services during mid-October and mid-December 2019. However, the total supply of freelancers is N = 130,726, which implies that each worker covers on average 1.7 categories. Since freelancers usually do not work full time, this is a justifiable assumption. The same worker can work on satisfying demands from different industries, so the labour supply in freelancing markets is more defined by the available skills, than by the number of workers.

Figure 1
Complementary ways of assigning skills to categories in Freelancer.com, October 2019[a]

Source: Author's elaboration, based on Freelancer.
[a] See text for descriptions of labels.

There would have been other ways of classifying jobs and worker profiles into categories. One approach that has received attention is to use automated classifiers (Amato et al., 2015; Kässi & Lehdonvirta, 2018). This implies that one would collect the micro-data for job descriptions and worker profiles, and then follows explicit-rules, machine learning (often supervised), or other feature extraction algorithms that make use of natural language processing. Amato et al. (2015) obtain an overall precision of about 50 % for their natural language processing approach, and the Online Labour Index from the Oxford Internet Institute (Graham et al., 2017; Kassi, 2016; Kässi & Lehdonvirta, 2018) receive an accuracy of 72%. These are very low level of reliability, which —looking at the data— we suspect mainly stem from the sparse and inconsistent data for both job descriptions and worker profiles. At this point, we, therefore, decided against this approach and stuck to the more transparent and replicable approach of merging predefined categories (as outlined in tables 1 and 2). In a later chapter we further discuss in more detail other network approaches in order to study the sophisticated relationship between the ever-changing skill set and labour market categories.

B. Freelancing demand

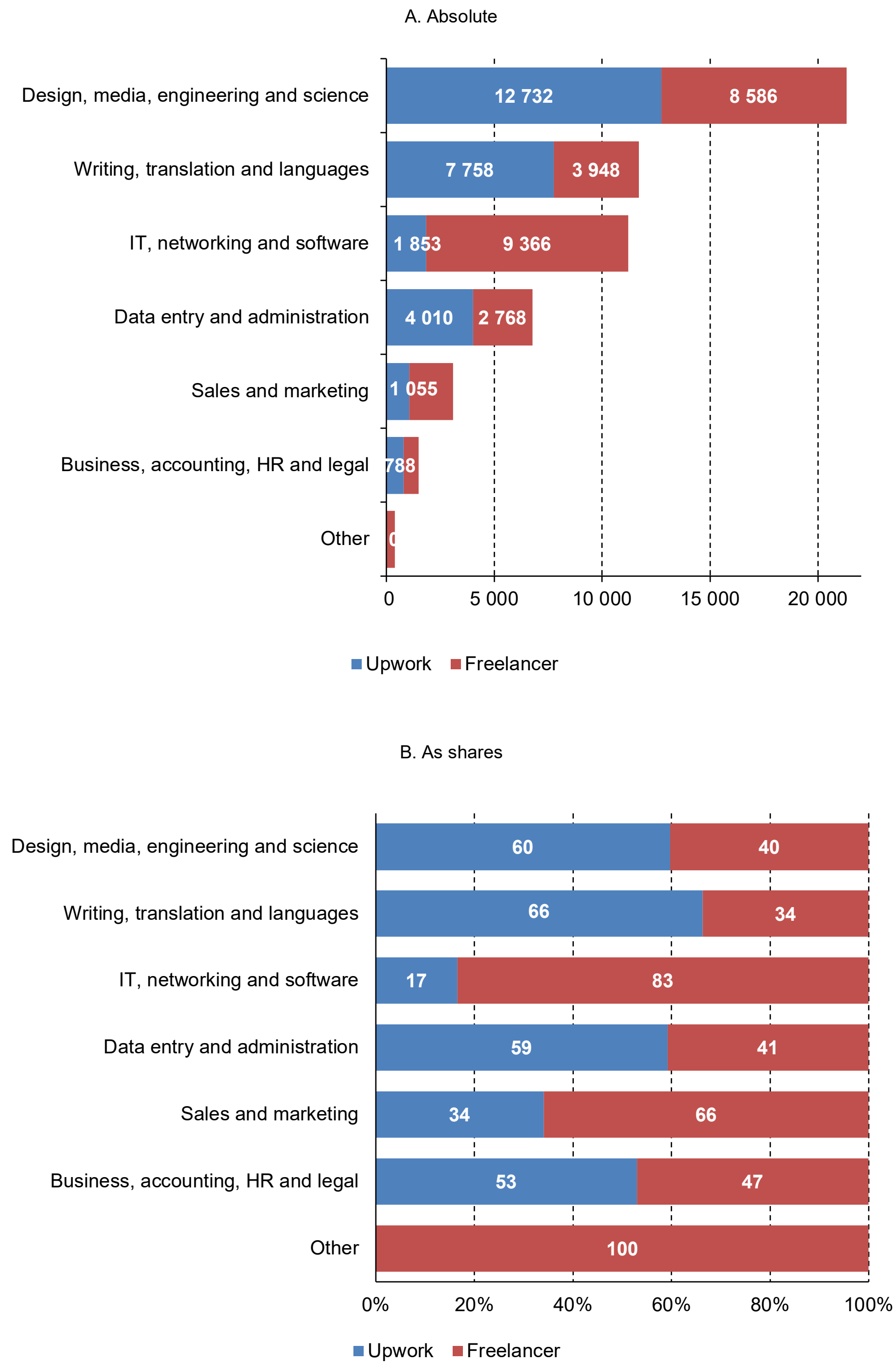
In the freelancing market, demand for labour is global. Jobs and tasks are posted online, and workers compete globally for the projects. Figure 2 shows that the most demanded freelancing sector consists of jobs in the field of 'Design, Media, Engineering and Science' (including product sourcing, see table 1). Together with 'Writing, Translation & Languages' and 'IT, Networking & Software', these sectors represent 79% of the global freelancing demand.

Figure 2 also shows that both of our freelancing sources have a surprisingly equitable market share worldwide. With a daily average 28,196 job posts during the nine weeks from Oct. 14 to Dec. 15, 2019, Upwork represents some 50.3% of the total projects offered on both platforms. It is important to note that we do not have any trustworthy way to eliminate potential duplicate posting among these two platforms. It might be that someone posts the same task on both platforms. The general argument in the 'big data' world is that such possible impreciseness is made up for the sheer amount of data, and while the absolute numbers could be somewhat affected by this fact, the average shares should not be much changed (see figure 1B).

It is also interesting to note some differences between both platforms (figure 1B). Freelancer, an Australian crowdsourcing marketplace website, founded in 2009 (headquartered in Sydney, with offices in Vancouver, London, Buenos Aires, Manila, and Jakarta), is the choice of employers when it comes to job calls for 'IT, Networking & Software'. Upwork, an American company, headquartered in Silicon Valley, is the result of a 2015 merger between ELance (founded in 1999) and oDesk (founded in 2003), and has a clear lead in the segment of 'Writing, Translation & Languages'.

Figure 2
Freelancing market demand of Upwork (N=28 196) and Freelancer (N=27 806)

(Per labour market category and platform)



Source: Author's elaboration, based on Upwork and Freelancer.

C. Freelancing supply

Freelancers do have a specific national origin (where they usually owe taxes). We only track freelancers from the 33 member countries of UN ECLAC (see diagram 2).

The general supply of job profiles from LAC on both platforms is with over 205,000 almost four times as large as the global demand from both (some 56,000) (figure 2A.). Again, also here, we do not have any way to control for duplicates, but since freelancers very rarely work at their available capacity, we consider even double postings of profiles to count as labour supply.

Some of the specializations from the labour demand side carry over to the supply side (figure 2B.). For example, Freelancer.com hosts many more profiles specialized in 'IT, Networking & Software'. Others do not, as Freelancer also has more workers with skills in 'Writing, Translation & Languages'. This is mainly affected by the sheer size of Freelancer's pool with over 130,000 profiles versus some 75,000 for Upwork.

Figure 3C shows important differences among countries.[1] Upwork (U.S. company) seems to the platform of choice for freelancers in the Caribbean, while Freelancer.com (an Australian company) is more popular with Latin American workers. Figure 3D visualizes this clear tendency. The average of the platforms shares over all 33 countries of LAC is very balanced, with 51% Freelancer and 49% Upwork, while there is an important difference in platform preference among different subregions.

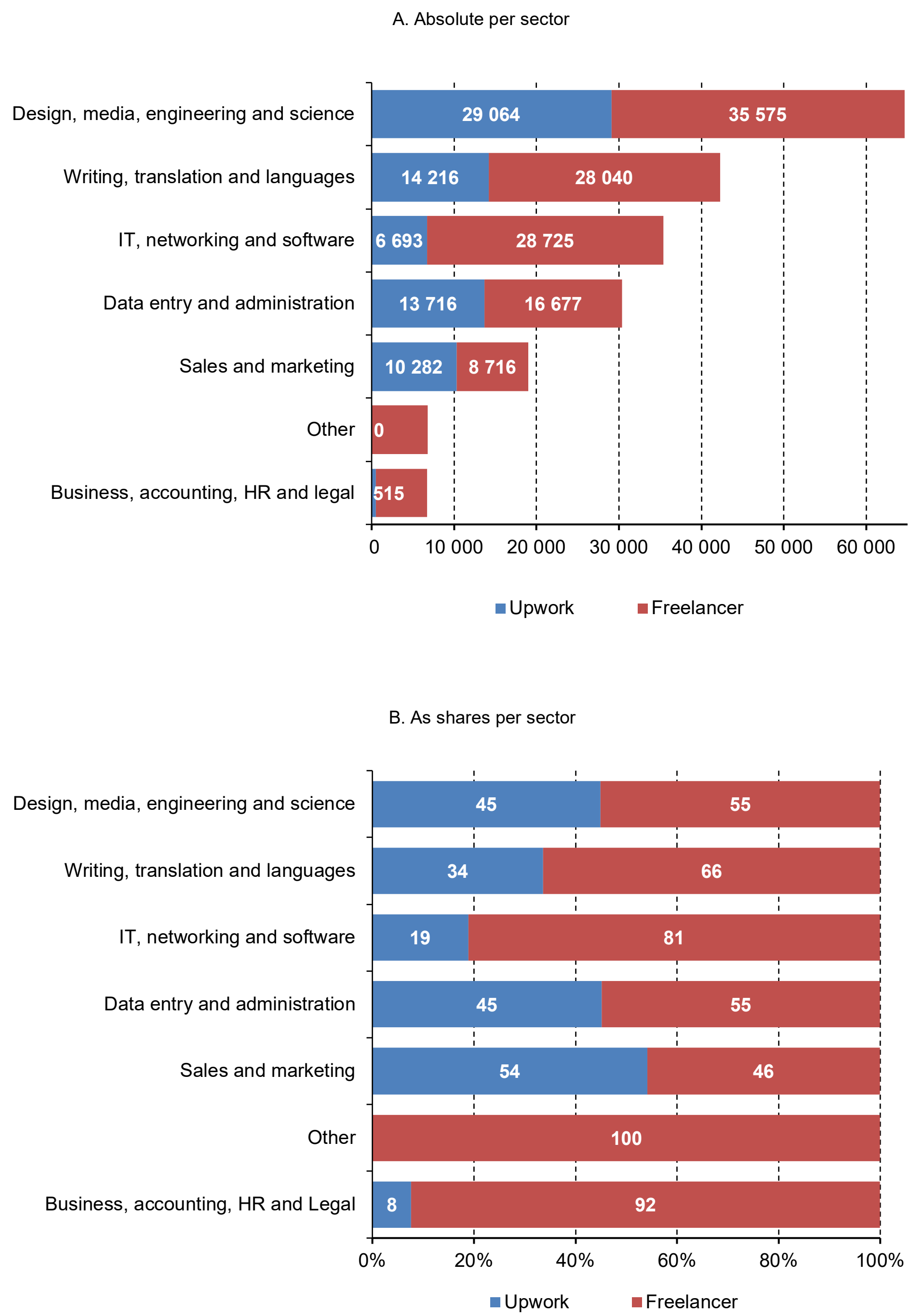
Online facilitated freelancing is a rather new form of work, and the entire concept of the gig-economy is an active area of research (Heeks, 2017; Kässi & Lehdonvirta, 2018). One crucial question refers to the intensity and frequency with which freelancers engage in paid activities. We tackle this question by analyzing the share of freelancer profiles in Freelancer.com that have received a review. We assume that most paid activities receive a review at the end, at least a star rating, which allows us to use this indicator as a proxy for paid freelancing services.

Figure 4 reveals that only 10% of workers with active profiles ever received a review. Since employers are automatically redirected to the feedback form after completion of a project, this can be seen as a lower-bound proxy for effective contracts which is probably not too far from the real number. Across the region, Freelancer.com has helped at least 7,653 workers to get freelancing gigs. The respective distribution among countries (figure 4A) and sub-regions[2] (figure 4B) is quite stable. Freelancers from some countries seem to be in above-average demand, including workers from Venezuela, Barbados, Suriname, Uruguay and Trinidad and Tobago (Dominica and Cuba have small sample sizes, which makes the presented average less stable).

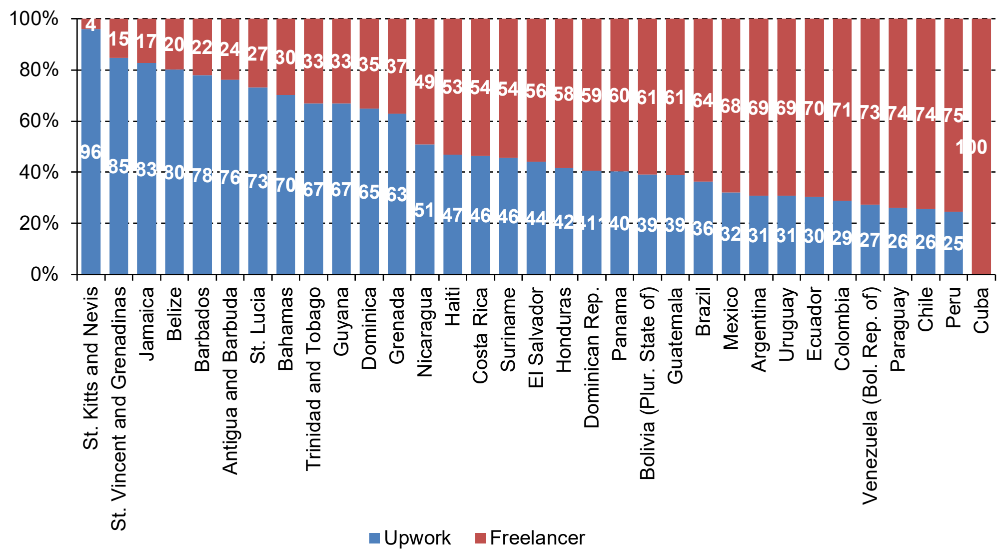
Should the definition of the freelancing labour market be restricted to only those workers who have already earned money on the platform? Data science alone cannot answer this question, as it refers to an economic and eventually political definition. Even workers who have not (yet) been lucky enough to find a job are conceptually part of the labour supply, as long as they offer their services, which is the reason why we included them in the preceding graphs.

Figure 4C adds the average reviews of freelancers per country. The average evaluation score per worker is quite stable at around 4.7 (on a scale from 1 to 5 stars). We checked a series of correlations but did not find any further meaningful correlation of star ratings with other variables. More in-depth research might be useful here (also related to cultural differences).

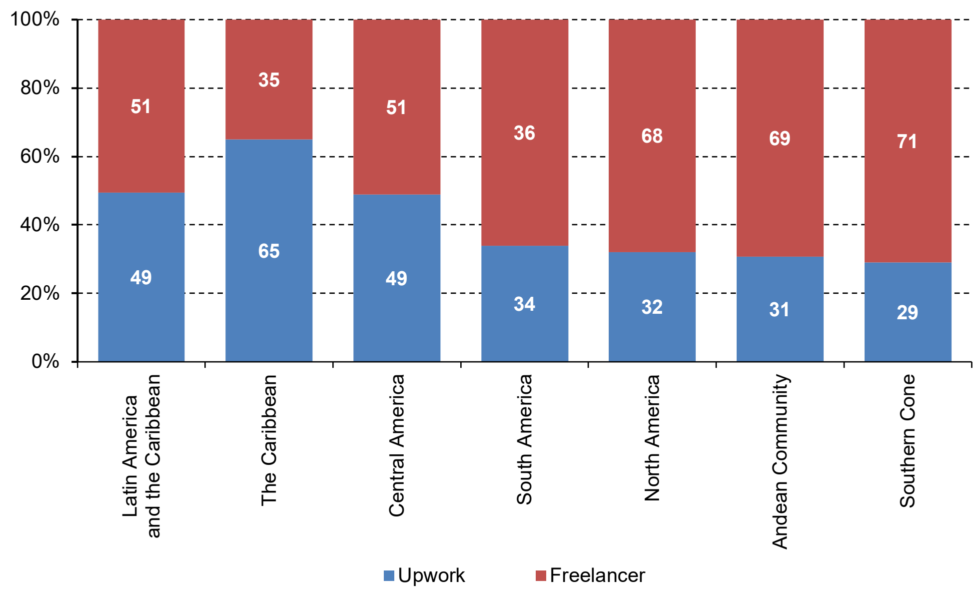
Figure 3
Freelancing market supply of Freelancer (N=130 726) and Upwork (N=74 486)
*(Per labour market category and platform)**



C. As shares per country



D. Simple per country average grouped by subregion1



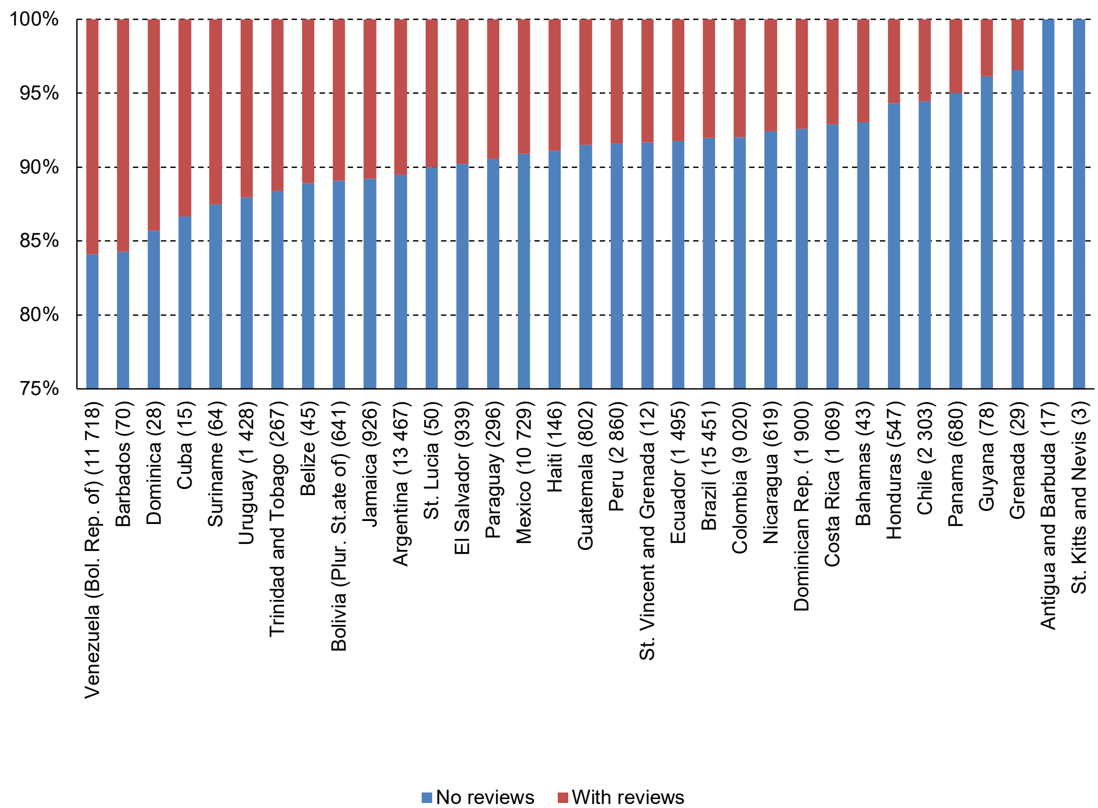
Source: Author's elaboration, based on Upwork and Freelancer.

* Sub-regions are grouped as follows: Andean Region: Bolivia, Colombia, Ecuador, Peru; Caribbean: Antigua and Barbuda, Bahamas, Barbados, Cuba, Dominica, Dominican Republic, Grenada, Guyana, Haiti, Jamaica, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago;Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama;North America: Mexico;South America: Brazil, Paraguay, Suriname, Venezuela;Southern Cone: Argentina, Chile, Uruguay.

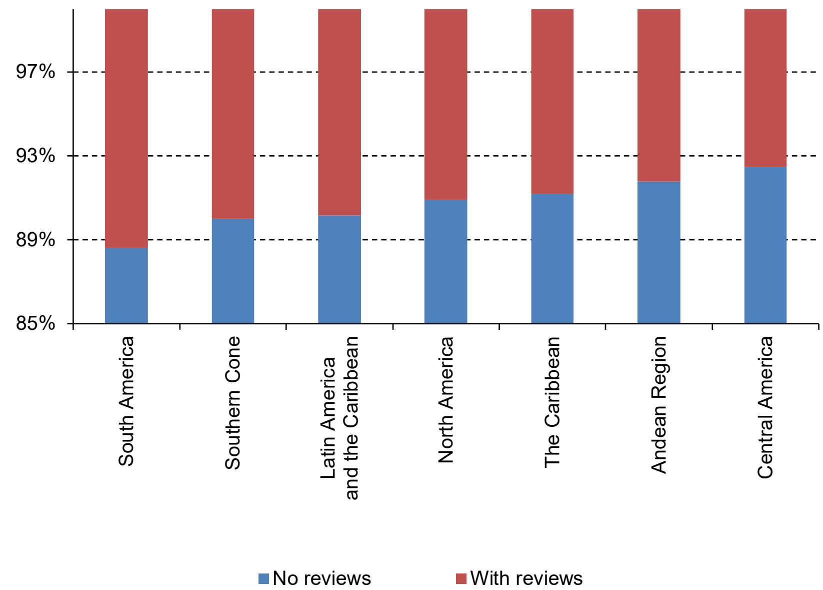
Figure 4D shows how the number of ratings is distributed among freelancers. Among workers, ratings are distributed according to a powerlaw, with exponentially few freelancers receiving exponentially many ratings, and exponentially many freelancers receiving exponentially few reviews (the equation $y = 191.85x^{-0.624}$ fits the data with $R^2 = 0.905$). This is a typical finding for social networks, which follow a preferential attachment logic resulting in scale-free networks (Barabasi & Albert, 1999). It is widespread to find that social media sites organize according to the exponentially skewed distribution. The powerlaw fit is not perfect, which suggests that there are other mechanisms at work besides preferential attachment's 'rich get richer' logic (probably related to ratings and skills). The suggestion that the freelancer market roughly follows a similar logic as social media has very interesting implications for the arising labour market dynamic. We know that income distribution in general also follows a powerlaw distribution (in its cumulative form called a Pareto distribution), where exponentially few have an exponentially high income, and exponentially many have exponentially low income. The fact that online labour markets might add a reputation powerlaw to this existing dynamic is worthy of more in-depth research.

Figure 4
Share of job profiles in Freelancer.com with and without reviews, (N=77 757)[a b]

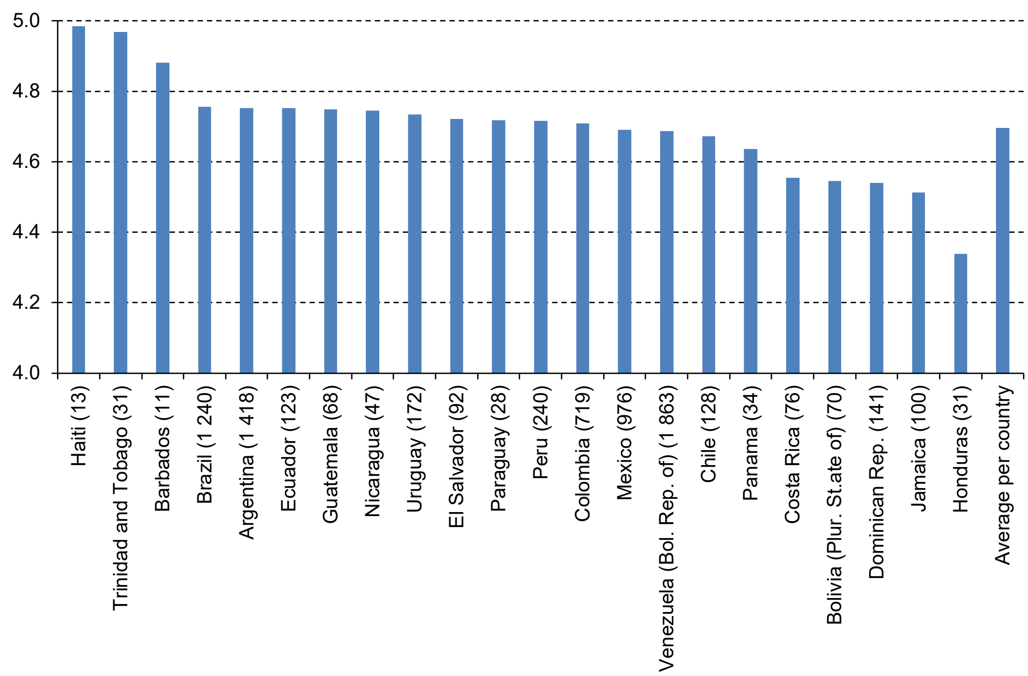
A. Countries
(Parentheses show number of freelancers with reviews)

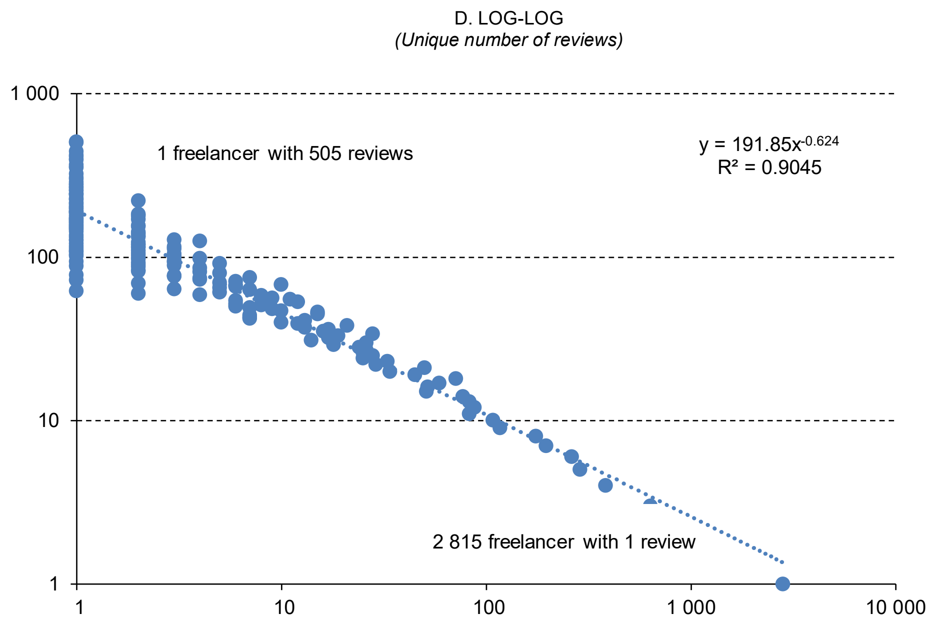


B. Subregions
(As average of respective country shares)



C. Average
(Star-rating per country)





Source: Author's elaboration, based on Freelancer.

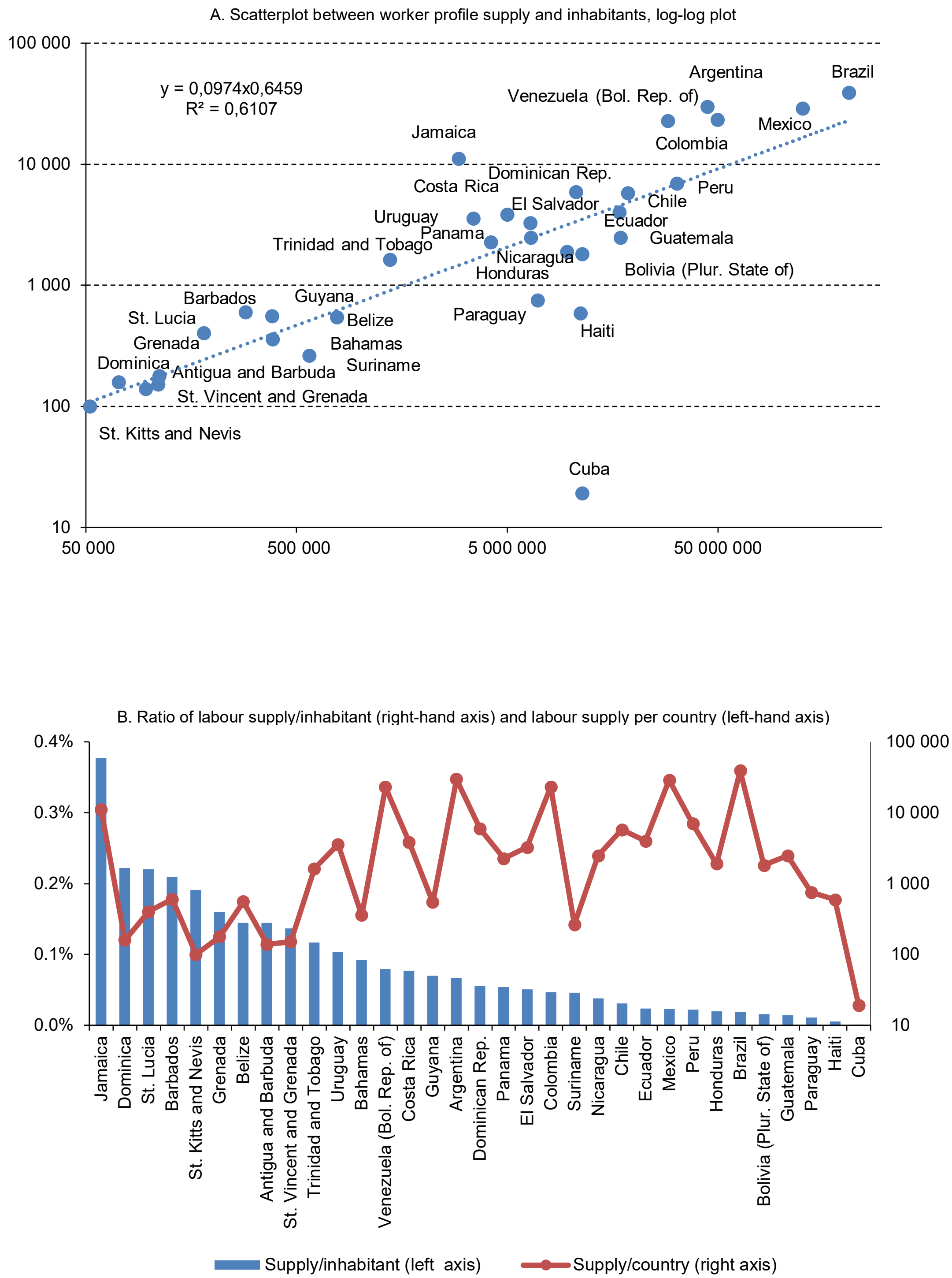
[a] For those who received review; powerlaw fit between number of freelancers with specific numbers of reviews.

[b] Sub-regions are grouped as follows: Andean Region: Bolivia, Colombia, Ecuador, Peru; Caribbean: Antigua and Barbuda, Bahamas, Barbados, Cuba, Dominica, Dominican Republic, Grenada, Guyana, Haiti, Jamaica, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago;Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama;North America: Mexico;South America: Brazil, Paraguay, Suriname, Venezuela;Southern Cone: Argentina, Chile, Uruguay

Figure 5A contrasts the number of freelancer profiles (vertical y-axis) with the number of inhabitants. It shows that Cuba is underrepresented, with a daily average of merely 19 profiles in a country of over 11 million inhabitants. The highest penetration of online freelancing profiles is found in Jamaica, with 11,063 profiles and 2,934,855 inhabitants (or 0.38 % of the population) (see also figure 5B). The large countries of Brazil, Mexico, Colombia, Argentina, and Venezuela are all found above the regional trend line. One hypothesis could be that freelancer jobs require skills that are only available in some larger countries. Another hypothesis could be that there might be some kind of heard-effect for a freelancing culture that kicks in once a critical mass of freelancers is reached. We cannot answer this question for now. Despite these specific cases of interest, the overall log-log powerlaw correlation between inhabitants and the number of freelancer profiles is quite strong, suggesting that most countries of the region take advantage of the new opportunities to a similar degree.

It is important to emphasize that the digital footprint is never a randomly chosen statistical sample, and therefore cannot be directly compared with traditional sample sizes. Hand-wavy rule-of-thumb heuristics from introductory statistics classes would suggest that 30 is "the magical number"(Kar & Ramalingam, 2013) to determine the minimum sample size requirement. However, we do not do actual statistical tests here, but simply report what we find, so we decided to keep these numbers, even if they are small. The graphs show the complementary nature of our sources to get a more complete picture of the entire LAC region, including precisely such Caribbean countries, which usually always fall through the cracks of international reporting.

Figure 5
Freelancing market supply of freelancers (N=205 212) per country

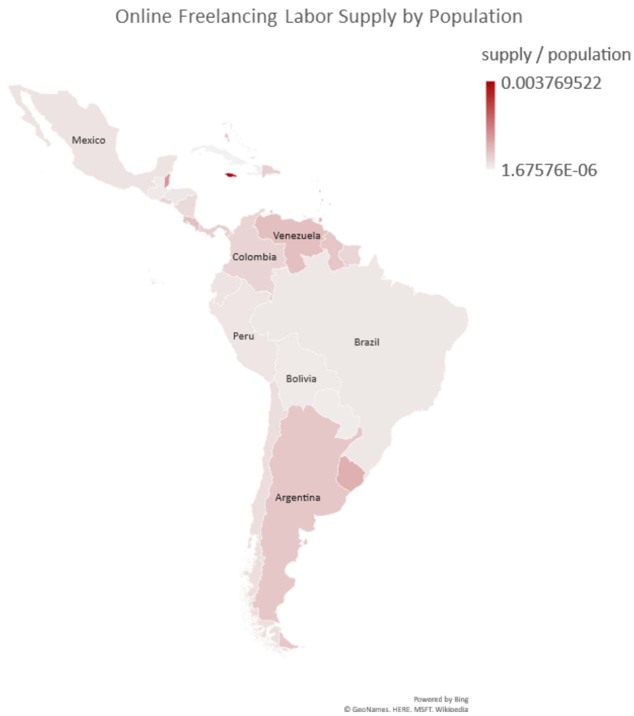


Source: Author's elaboration, based on Upwork, Freelancer and World Bank.

Map 1 presents the same numbers as maps below. In our web-interface, they are interactive maps.

Map 1
Freelancing market supply maps per population (N=205 212)

A. Latin America



B. The Caribbean

C. Central America

D. South America

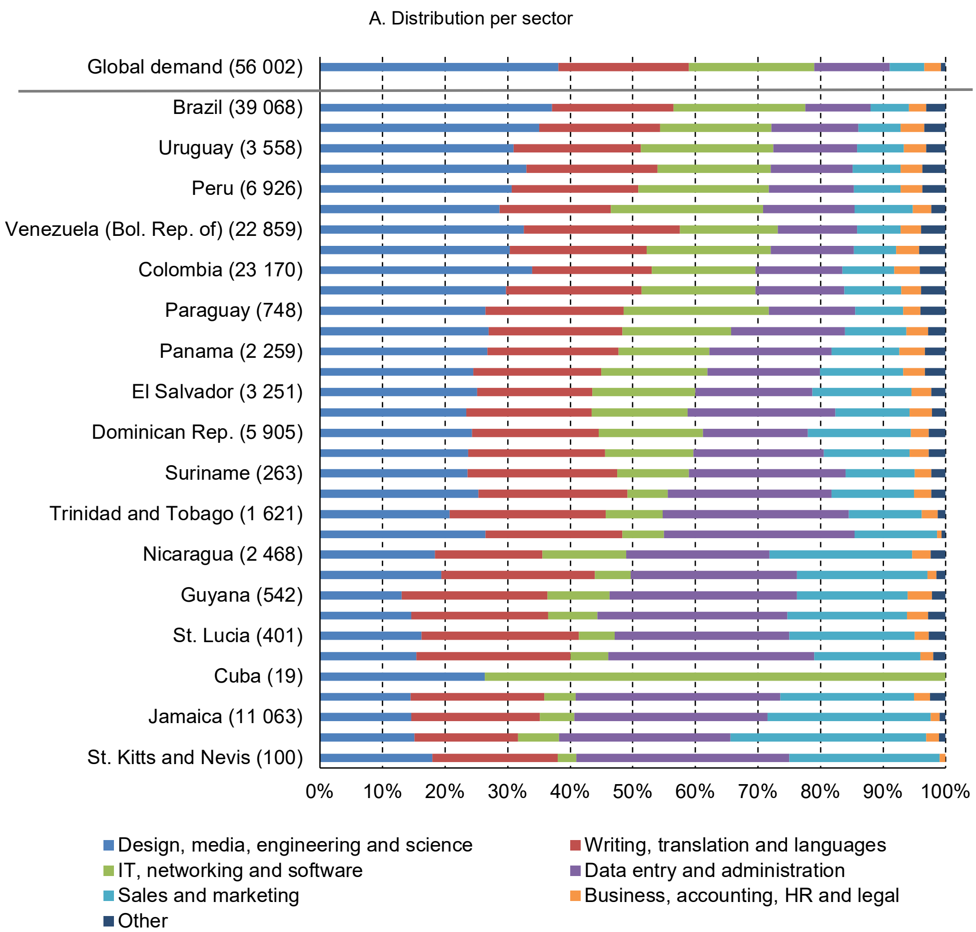
Source: Author's elaboration, based on Upwork and Freelancer.

D. Freelancing supply and demand per industry

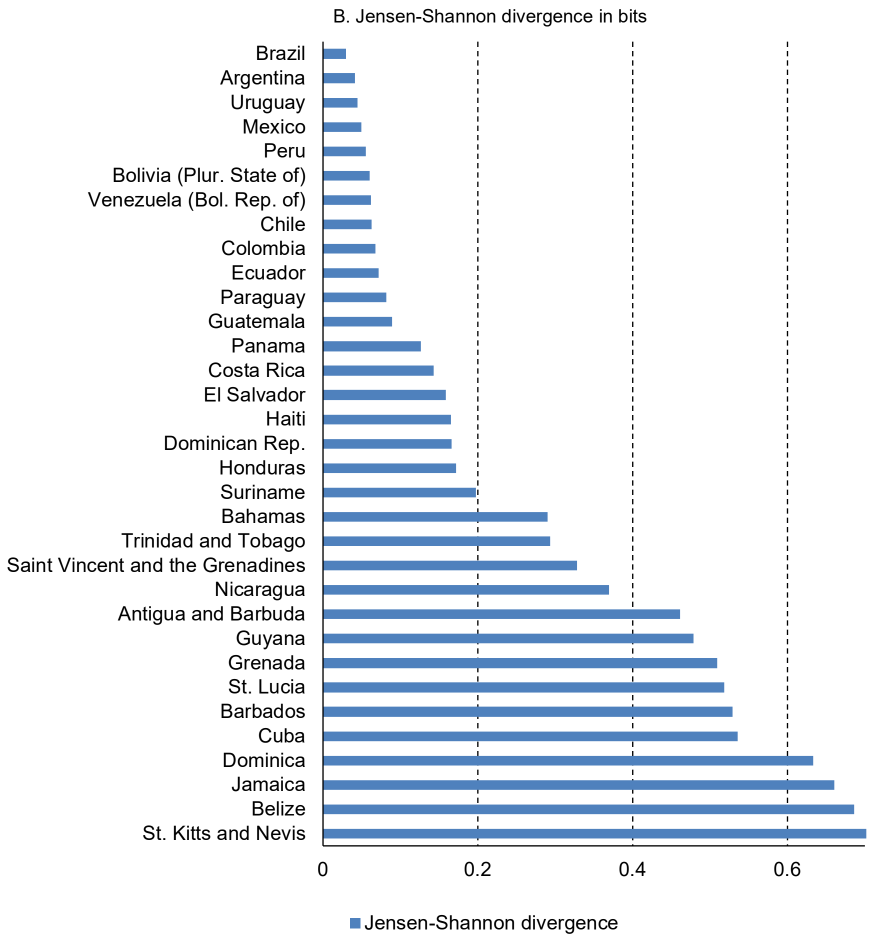
Figure 6 contrasts the global demand per sector by the national supply. We summarize the mismatch between the different sectors in one single number. In probability theory and statistics, the Jensen–Shannon divergence[3] is a method of measuring the similarity between two probability distributions. It is based on the popular Kullback–Leibler divergence (a central measure of accuracy in machine learning) and is also known as the total divergence to the average. The larger the number (measured in bits), the larger the divergence among the different shares. The size of the country does not matter, only the difference between the share-distribution of the global demand, and each of the national supplies (in figure 6, it compares the first row, with each of the country rows, respectively).

It shows that Brazilian freelancers are best prepared to match global demands per sector, and St. Kitts and Nevis has the largest mismatch.

Figure 6
Shares of freelancing global demand versus national supply shares



[3] https://en.wikipedia.org/wiki/Jensen%E2%80%93Shannon_divergence.

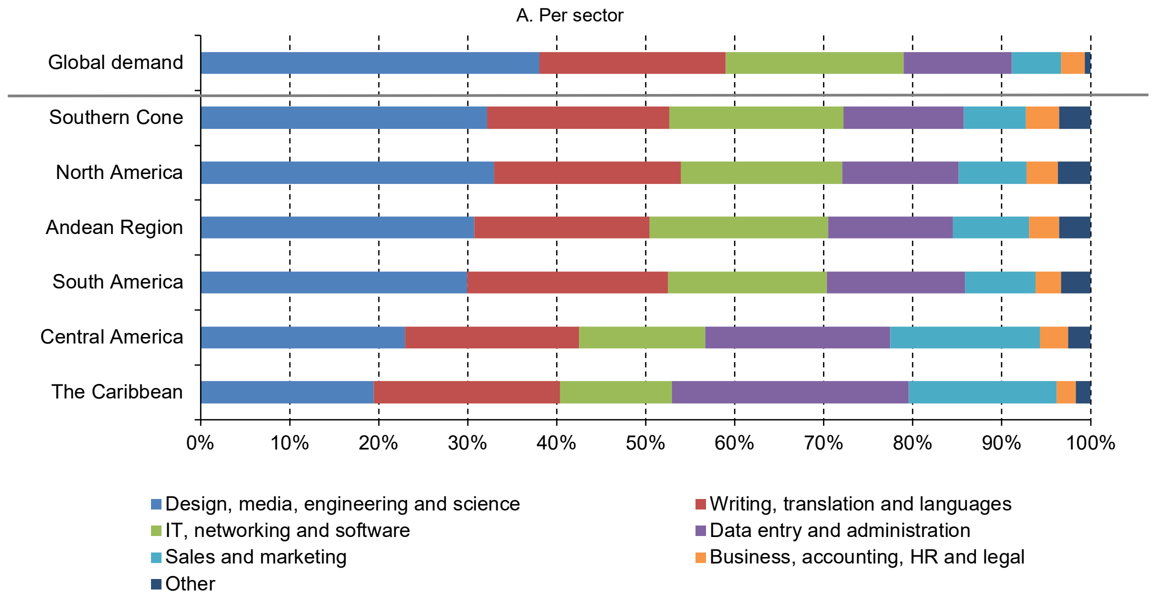


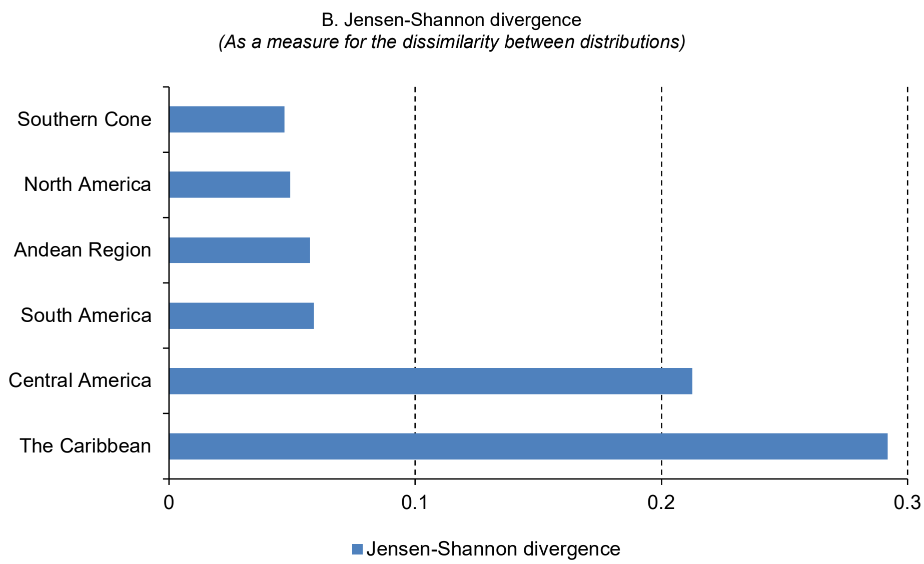
Source: Author's elaboration, based on Upwork and Freelancer.

The Jensen-Shannon divergence consists of a sum,[3] which makes it straightforward to calculate how much the mismatch of each sector contributes to the total divergence between global demand and national supply. In general, the biggest contributor for the mismatch is the oversupply of professionals in 'Sales & Marketing' (contributing 25% to the average Jensen-Shannon divergence). The second and third place is shared by an undersupply of professionals in 'IT, Networking and Software' and 'Design, Media, Engineering & Science' (each adding 23% to the divergence). The best matches of national supply and global demand can be found for 'Business, Accounting, HR and Legal' (a very small share), and also 'Writing, Translation and Languages' (both contributing less than 1% to the average divergence).

Figure 7 reveals that the largest average mismatch between global demand and regional supply stems from the smaller countries of the region in the Caribbean and Central American markets.[1] On average, South American markets are quite well-adjusted to the global demand when comparing the distribution among the seven chosen sectors.

Figure 7
Freelancing global demand versus regional supply[a]
(Simple average per country) [b]





Source: Author's elaboration, based on Upwork and Freelancer.
[a] Sub-regions are grouped as follows: Andean Region: Bolivia, Colombia, Ecuador, Peru; Caribbean: Antigua and Barbuda, Bahamas, Barbados, Cuba, Dominica, Dominican Republic, Grenada, Guyana, Haiti, Jamaica, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago;Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama;North America: Mexico;South America: Brazil, Paraguay, Suriname, Venezuela;Southern Cone: Argentina, Chile, Uruguay.
[b] Central America with Dominican Republic; Caribbean with Belize, Suriname, Guyana.

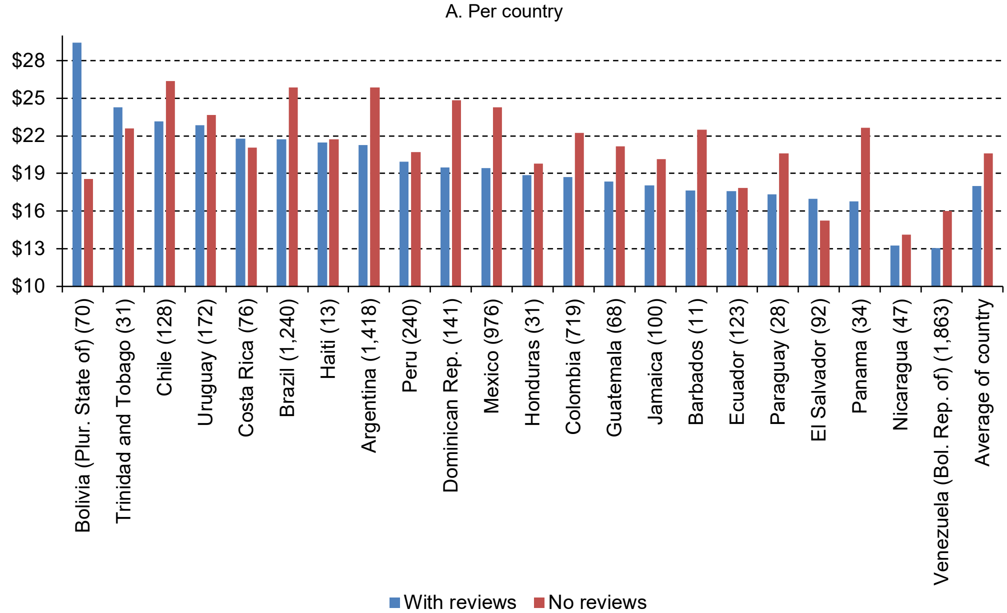
E. Freelancing hourly rates and payments

In online freelancing markets, freelancers post an hourly asking rate, which we collected for the 77,757 freelancer of our sample who maintains a profile in Freelancer.com. Being aware that 90% of freelancers never seem to have obtained any jobs through the platform (see figure 4A), we compare the hourly asking rate of freelancers with reviews and without reviews (figure 8A). In general, those workers with evidenced experience on the platform ask for 12.5% less salary. The most surprising fact relates to the relatively high hourly rate for services offered on Freelancer.com. With an average hourly asking rate of US$ 18.02 (average among the country averages), the hourly rate is notably above the minimum wage of most developed countries (at the time of the study, the minimum wage in the US and Japan was US$7.25, in Canada US$ 7.7, in Germany US$10.8, in California US$12). Figure 8B shows this finding in context with the minimum wage per country. On average, among the 23 countries in our sample, the hourly minimum wage in 2019 was US$ 1.7 (average of countries in figure 8A). This is barely 8% of the average hourly wage collected by the region's freelancers. Turning the ratio around this implies that the average freelancer hourly rate is 15.4 times the minimum wage of the region's countries. Figure 8B shows that this multiplier factor ranges from 5 in Barbados, to 50 in the Dominican Republic. For most countries in the region, freelancing hourly rates pay about 10 times the national minimum wage. This suggests that these jobs demand a relatively sophisticated skill set.

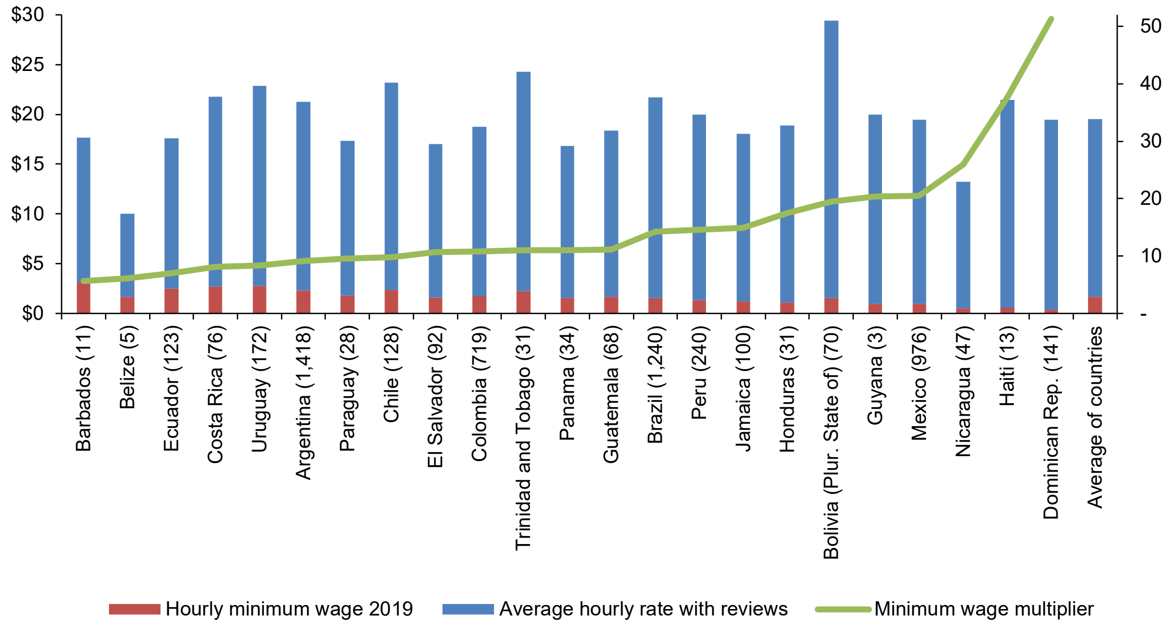
Figure 8C shows that the average hourly asking rate of publicly rated freelancers has no correlation with adjusted net national income per capita. This suggests that the freelancer market is independent of national earning levels. Through freelancer platforms, workers virtually become part of the global economy, which includes a different regiment for hourly rates.

Figure 8
Freelancing average hourly asking rate on Freelancer.com with reviews and without reviews
(N (no reviews) =70 104; N (with reviews) =7 653)
(Number of freelancers with reviews in parentheses)

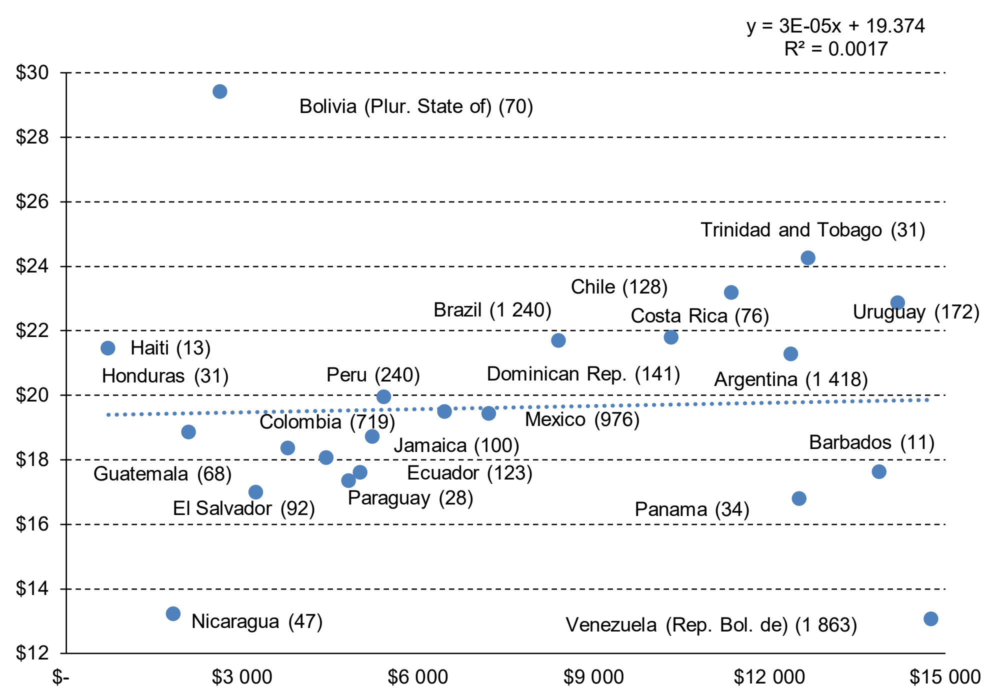
A. Per country



B. In comparison to national minimum wage



C. With GINI p.c.



Source: Author's elaboration, based on Freelancer.

III. Domestic labour market

The domestic labour market is traditionally dominated by a demand logic (aka by 'job calls'), and this is also reflected in the way online labour market platforms work. Most domestic platforms focus exclusively on demand (only one of our platforms also focuses on supply).

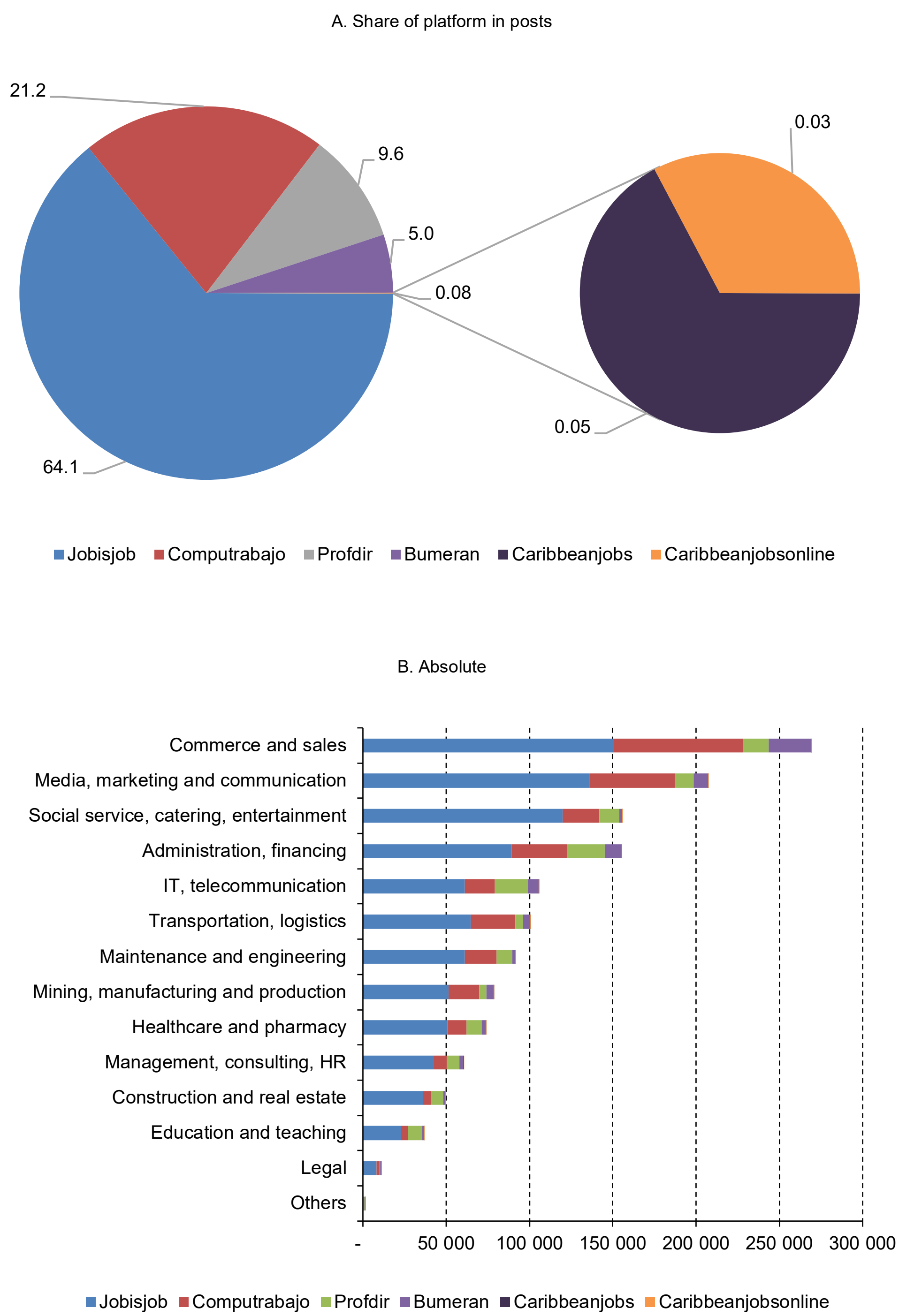
A. Domestic market demand

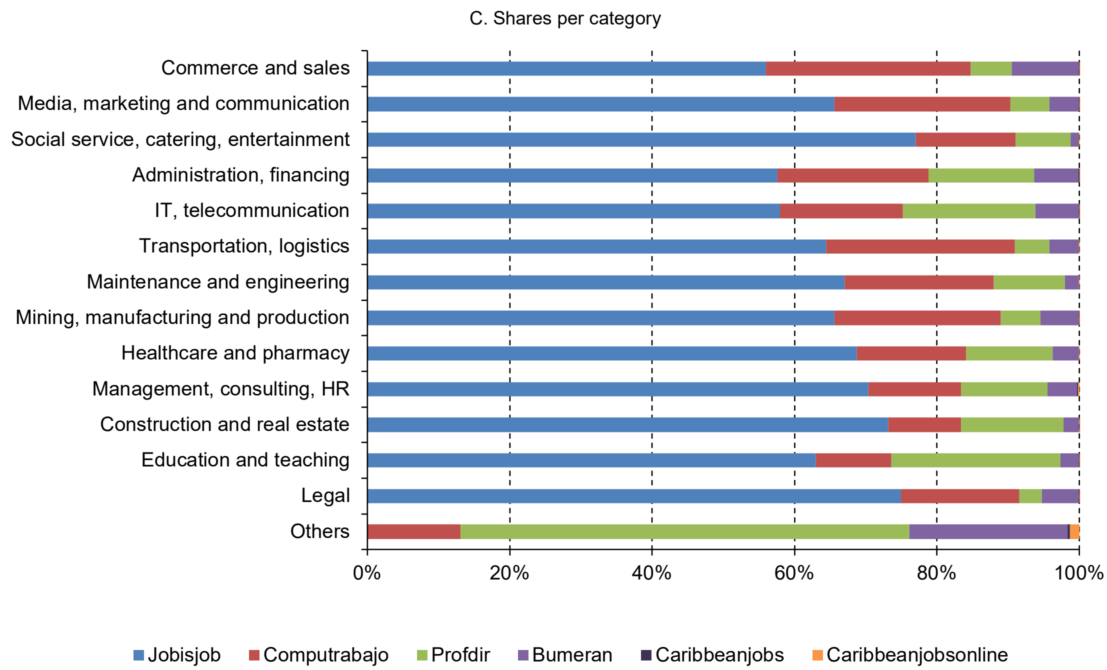
In total, we obtain an average of 1,395,333 available daily job calls across our six platforms. The market shares of our chosen domestic platforms are more skewed, with almost 2/3 of job calls coming from Jobisjob (figure 1A). In terms of the number of available daily job calls, it dominates the online job market geared to domestic jobs in LAC.

Figure 1B shows that 'Commerce and Sales' jobs are in the highest demand in the LAC online labour market, followed by 'Media, Marketing and Communication'.

Figure 9C shows that some platforms seem to specialize on some specific sectors, such as Profdir on 'IT, Telecommunication' and 'Education and Teaching', while CompuTrabajo has a larger proportional share in the leading sectors of 'Commerce and sales' and 'Media, Marketing and Communication'.

Figure 9
Daily averages of posts for local freelancing market demand (total N=1 395 333), Oct. 14 – Dec 15, 2019; with Jobisjob (N=894 901), CompuTrabajo (N=295 751), Profdir (N=133 327), Bumeran (N=70 213), CaribbeanJobs (N=767), and CaribbeanJobsOnline (N=374)
(Per labour market category and platform)





Source: Author's elaboration, based on Bumeran, CaribbeanJobs, CaribbeanJobsOnline, CompuTrabajo, JobisJob, and Profdir.

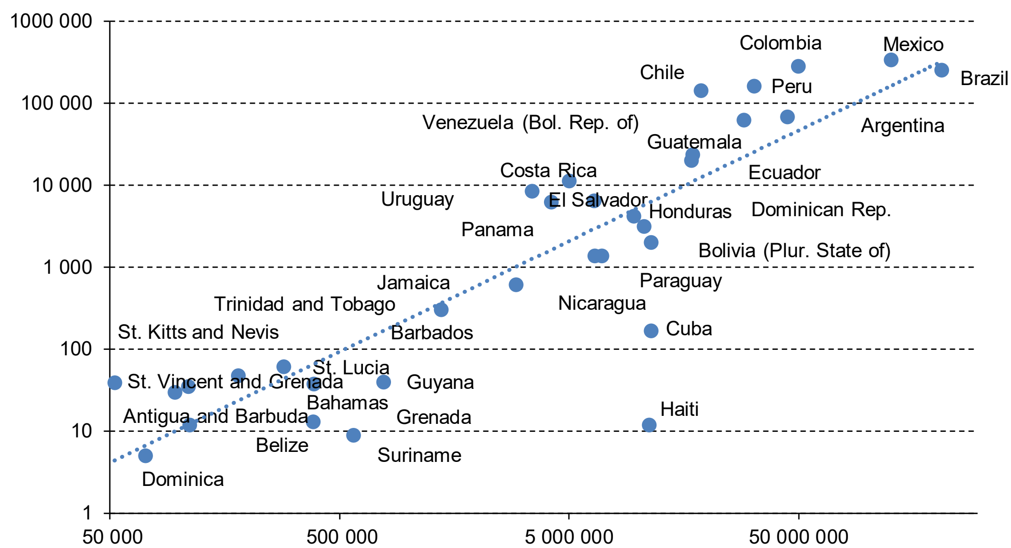
The following figure 10A shows that the number of job posts (demand) correlates strongly with the size of the country (in terms of the number of inhabitants). It also shows that there is some important variation: countries found above the trendline (which has the form of a power law in this log-log plot) make more effective use of the new online possibilities than the average in the region, considering their size; and vice-versa for the underrepresented countries below the trendline. Figure 10B highlights that several countries make particularly effective use of the new online possibilities, especially Chile, Colombia, and Peru. This is reconfirmed by the spatial map in figure 10C.

For reasons of transparency, figure 10D makes explicit the skewed contributions of our different sources. The graphs show the complementary nature of our sources to get a more complete picture of the entire LAC region, including precisely such Caribbean countries, which usually always fall through the cracks of international reporting. Figure 10A confirms that several of them are actually doing better than the average when considering their small size, including St. Kitts & Nevis, St. Vincent and the Grenadines, and Antigua & Barbuda, which —when normalized on the country's inhabitants— have more online job calls than the Dominican Republic, Paraguay, or Bolivia.

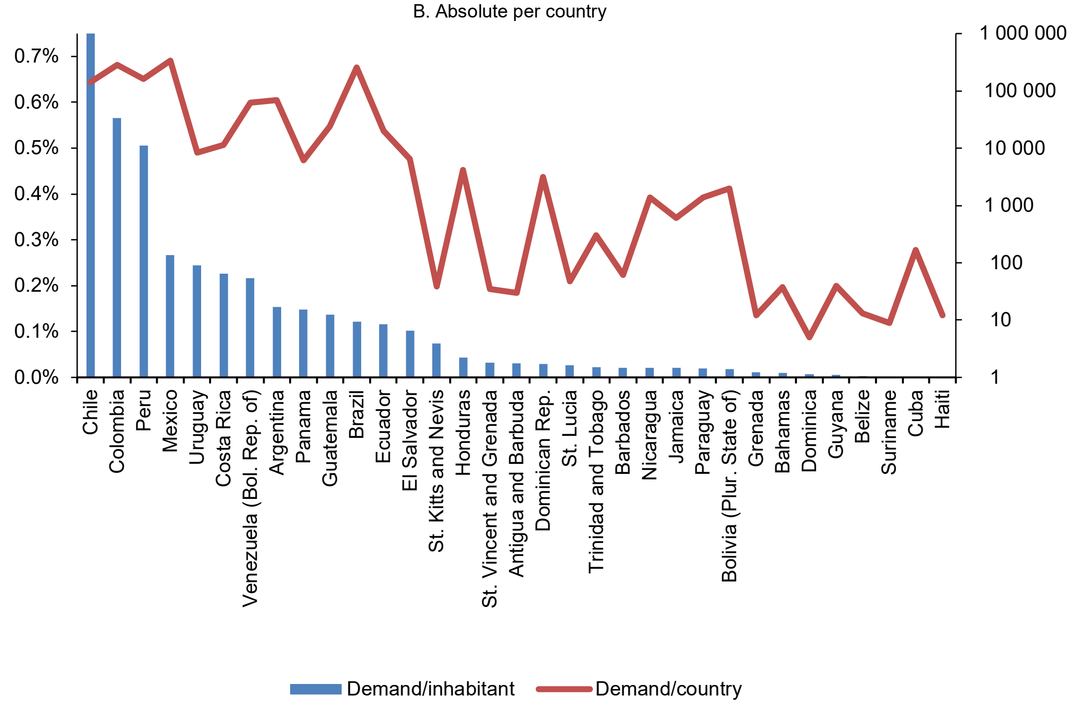
The sectoral distribution reveals some particularities among regions and countries. Figure 11A shows the regional averages of the shares per country.[2] North America (aka Mexico) has a larger than average share of 'Commerce and Sales', while Caribbean online job calls specialize in 'Social Services, Catering, Entertainment' and 'Management, Consulting, HR'. The South American online job market has a larger than average focus on 'IT, Telecommunication'. Some of the same difference can be seen in the respective countries (figure 11B). It is important to note the different sample sizes per country. Observational digital footprints are never randomly drawn and therefore generalizations always need to keep sample sizes in mind.

Figure 10
Domestic market demand for all 6 platforms per country

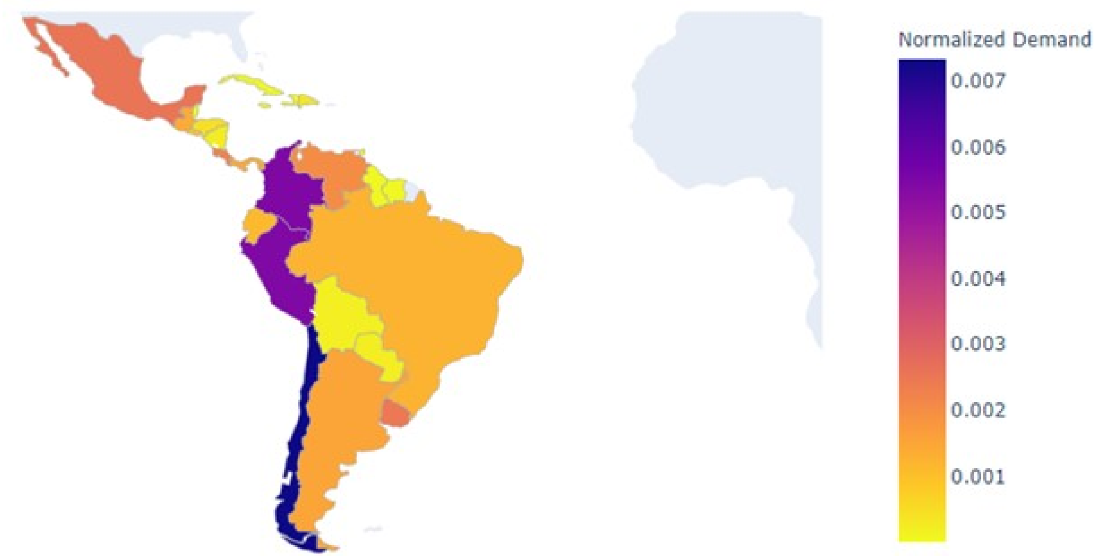
A. Scatterplot between job offers and inhabitants, log-log plot



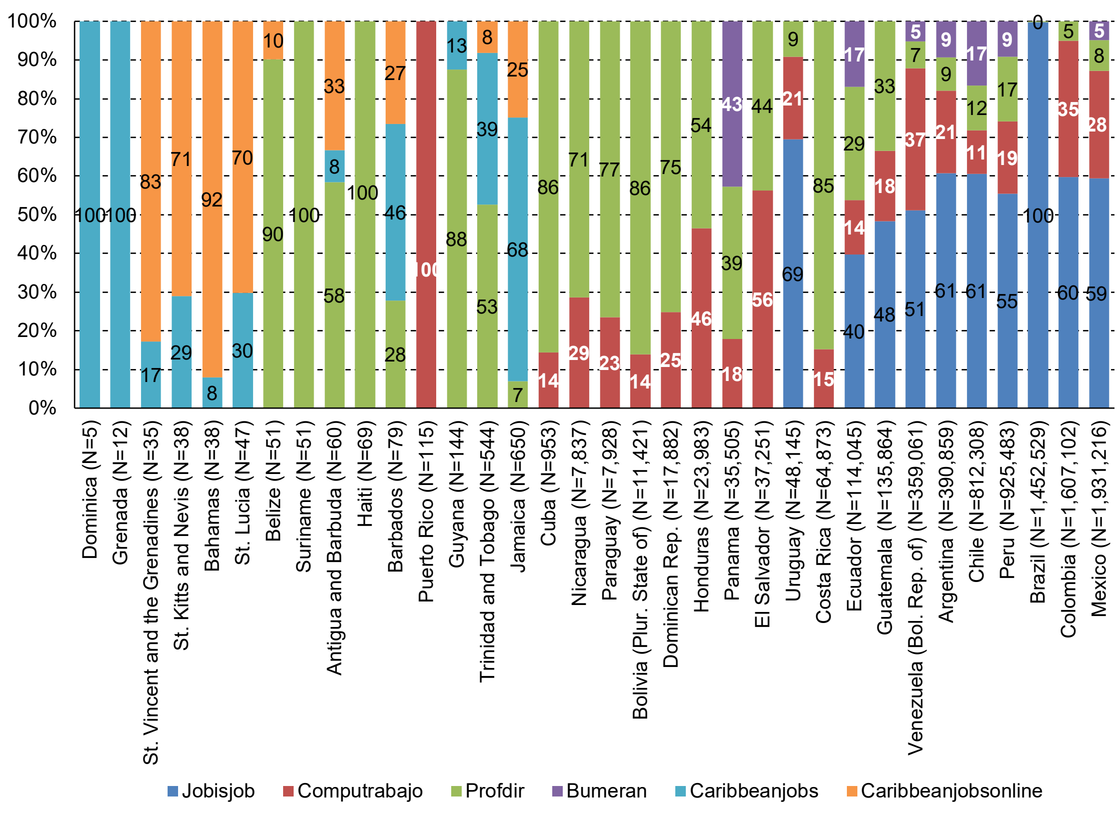
B. Absolute per country



C. Demand normalized by population as spatial map, only including Profdir, Jobisjob, CompuTrabajo, Bumeran

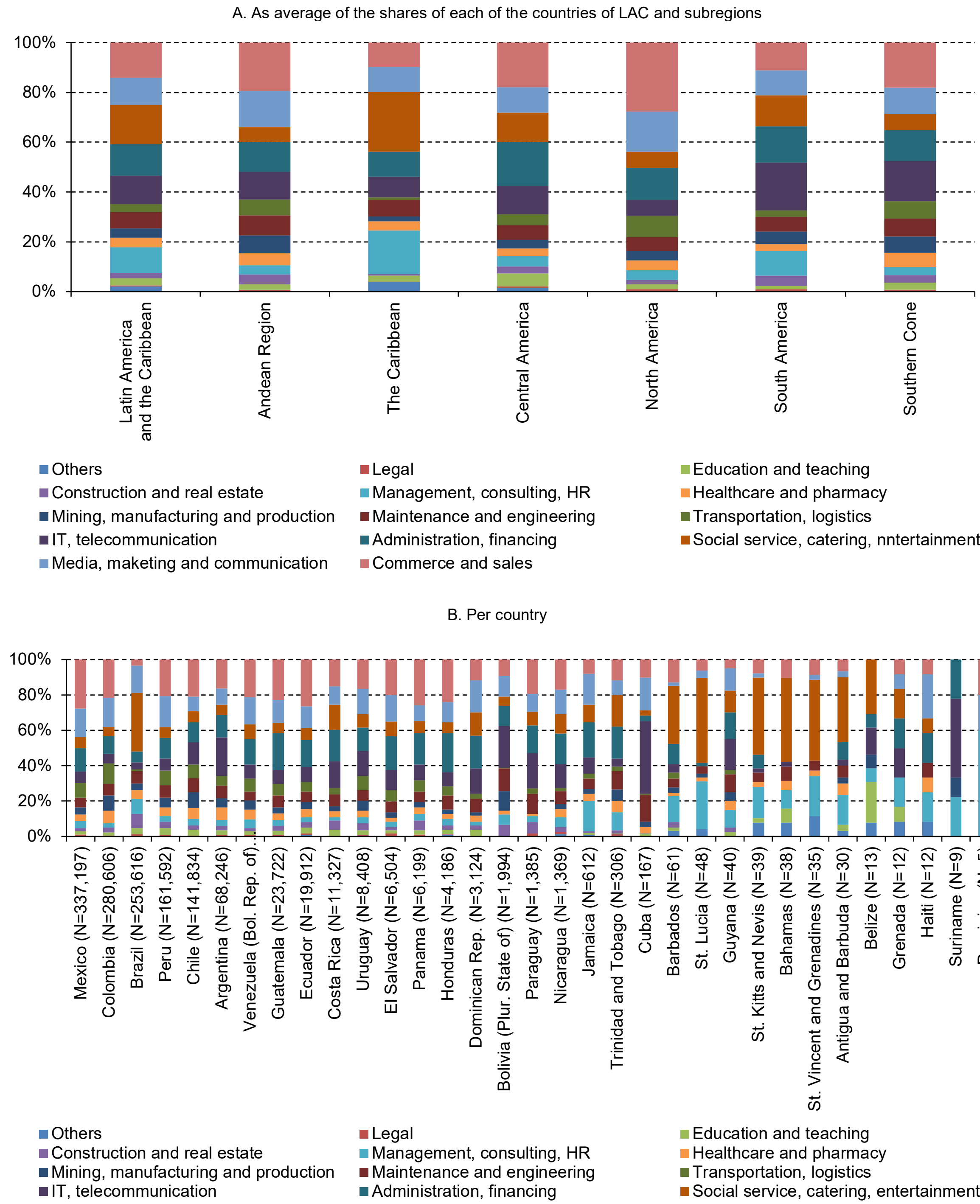


D. As shares per platform per country



Source: Author's elaboration, based on Bumeran, CaribbeanJobs, CaribbeanJobsOnline, CompuTrabajo, JobisJob, and Profdir.

Figure 11
Domestic market demand for all 6 platforms and 33 countries (N=1,395,339)[a]
(Per industry)



Source: Author's elaboration, based on Bumeran, CaribbeanJobs, CaribbeanJobsOnline, CompuTrabajo, JobisJob, and Profdir.
[a] Sub-regions are grouped as follows: Andean Region: Bolivia, Colombia, Ecuador, Peru; Caribbean: Antigua and Barbuda, Bahamas, Barbados, Cuba, Dominica, Dominican Republic, Grenada, Guyana, Haiti, Jamaica, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago;Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama;North America: Mexico;South America: Brazil, Paraguay, Suriname, Venezuela;Southern Cone: Argentina, Chile, Uruguay.

B. Domestic market supply

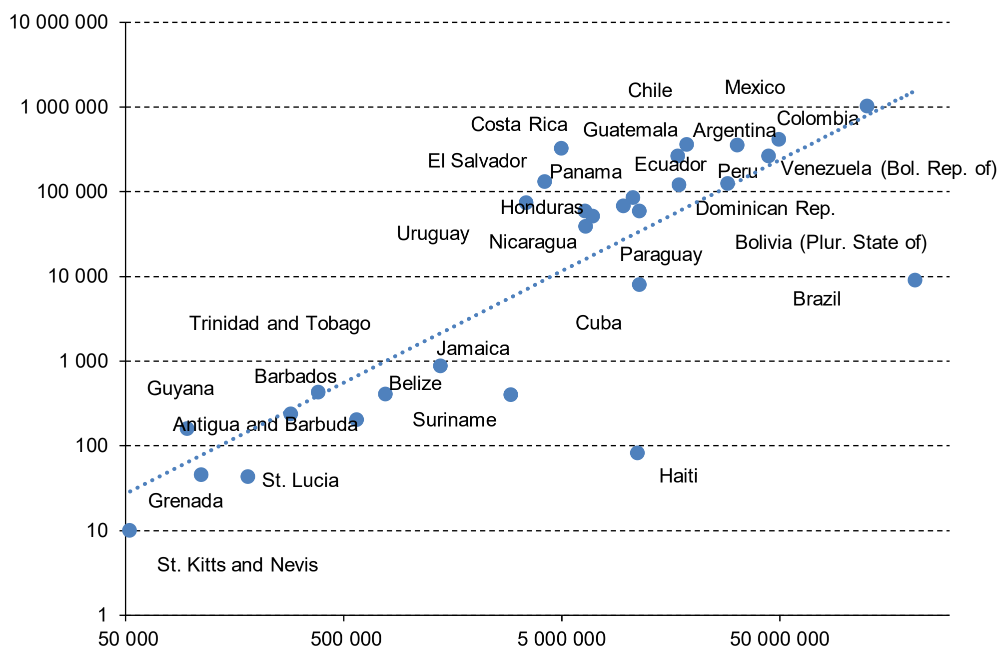
While labour demand refers to job calls, labour supply refers to workers offering their services online. From all of our six our domestic market platforms (see diagram 2), only Profdir (aka 'acciontrabajo') offers both supply and demand data. It is our third biggest source in terms of demand job postings, hosting about 10% of the online job calls in the region (figure 9). On the daily average, we find 3,843,330 workers offering their services between Oct. 14 and Dec. 15, 2019.

The following shows the analysis we did in figure 10 for the supply of domestic labour profiles from Profdir. Those countries above the trendline in figure 12A have an above-average number of worker profiles posted on the online job market. Note that the low numbers from Brazil and Haiti are likely because workers in these countries do not speak Spanish as their first language, so there might be other portals that cater better to their needs. Observational data always has to consider the potential bias in its source. Figures 12B and 12C illustrate that some countries from Central America stand out with above-average use of online possibilities. On the one hand, this could indicate that workers in smaller countries make more efficient use of online options to promote their profiles and skills. Uruguay and Ecuador are also among the top countries.

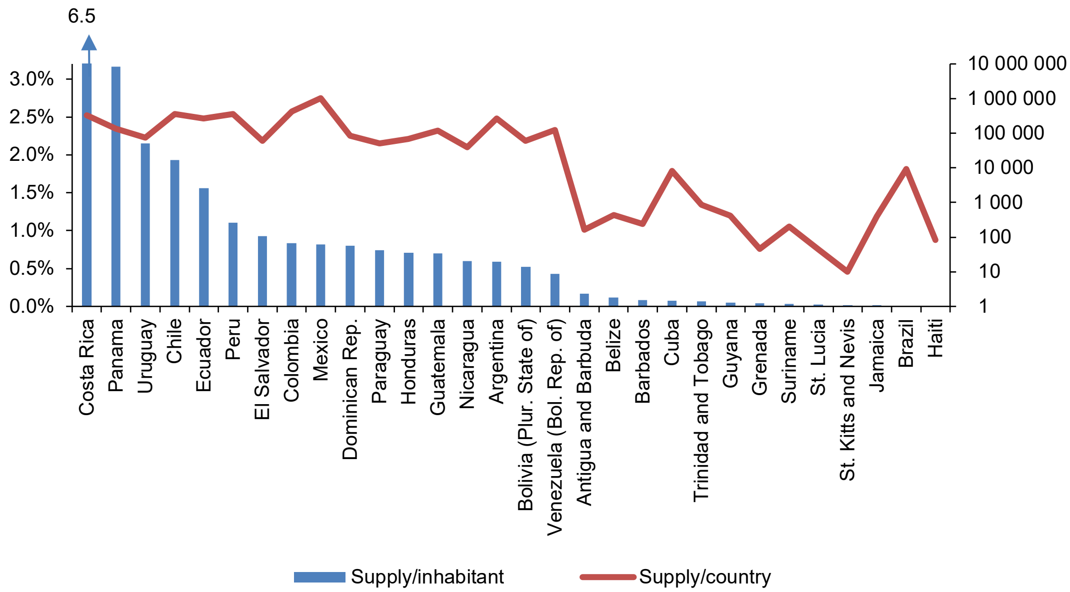
On the other hand, this might not necessarily be something structural, but could also simply stem from the fact that Profdir is more prevalent in some of these Central American countries. Some of the data in figure 10C would support this claim, while some other does not (as evidenced by the case of Uruguay). More in-depth follow-up studies will need to look closer into this issue to understand better the relationship between the nature of the source and generalizable tendencies in this case.

Figure 12
Domestic market Supply for the Profdir platform (N = 3,843,330)
(Per country)

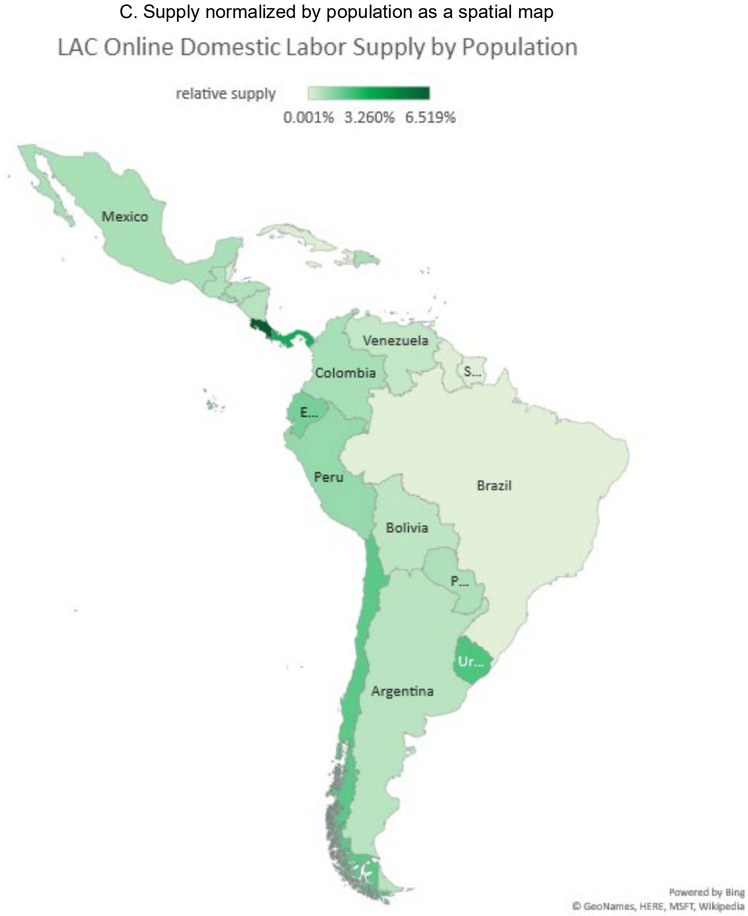
A. Scatterplot between worker profile supply and inhabitants, log-log plot



B. Absolute per country



C. Supply normalized by population as a spatial map



Source: Author's elaboration, based on Profdir.

1. Domestic supply by education attainment

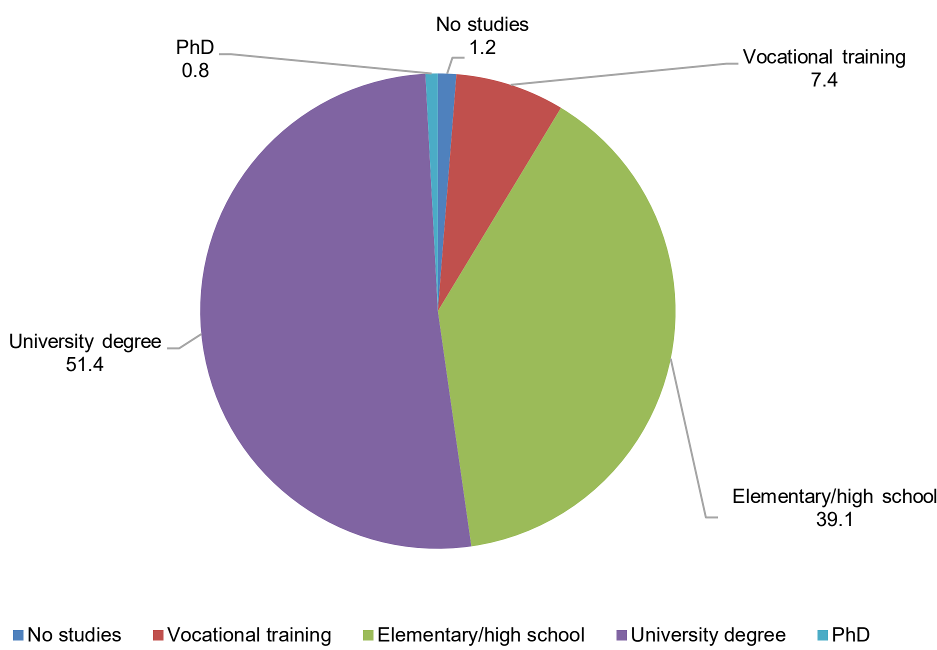
The Profdir source also provides data about educational attainment for almost all of the worker profiles (N = 3,356,240 daily average). More than half of the workers on the platform have a university degree, and about one percent even a PhD. This is certainly not representative for the general population, and shows that for now, rather educated people take advantage of these new possibilities to obtain labour. Still, 39% of the users only have elementary or high school training, and more than one percent has no training, which also goes to show that the technological barriers of entry to access these markets are low, and that even untrained workers can take advantage of them. Figure 13B shows that especially in some Caribbean countries, workers with less education make above-average use of the online possibilities. Besides, the case of Uruguay stand out, as 76.2% of the workers do not have a university degree.

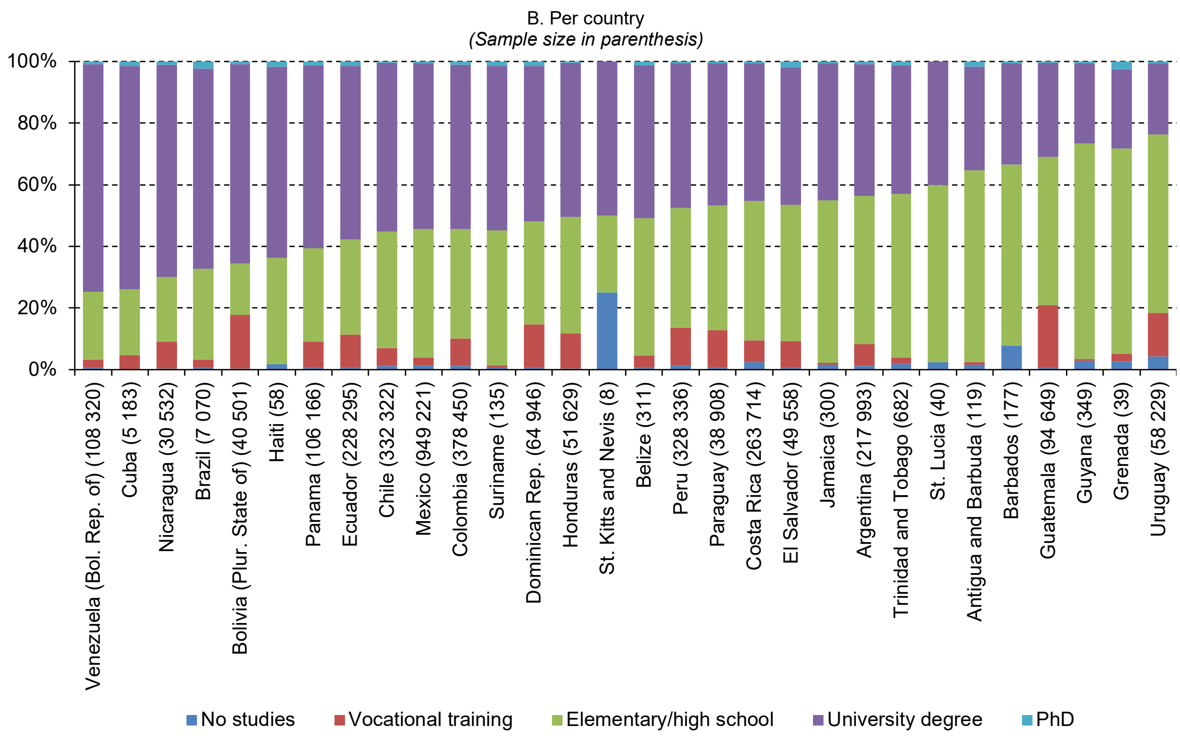
Figure 13C correlates the share of workers with university degree with the Education Index from UNDP's Human Development Index. The index is a mix of 'expected years of schooling' and 'mean years of schooling'. This reconfirms that there is no correlation between the general educational level of a population and the share of educated workers taking advantage of the new opportunities. In other words, there does not seem to be a need for a country to count with a highly educated workforce in order to become part of the online labour market.

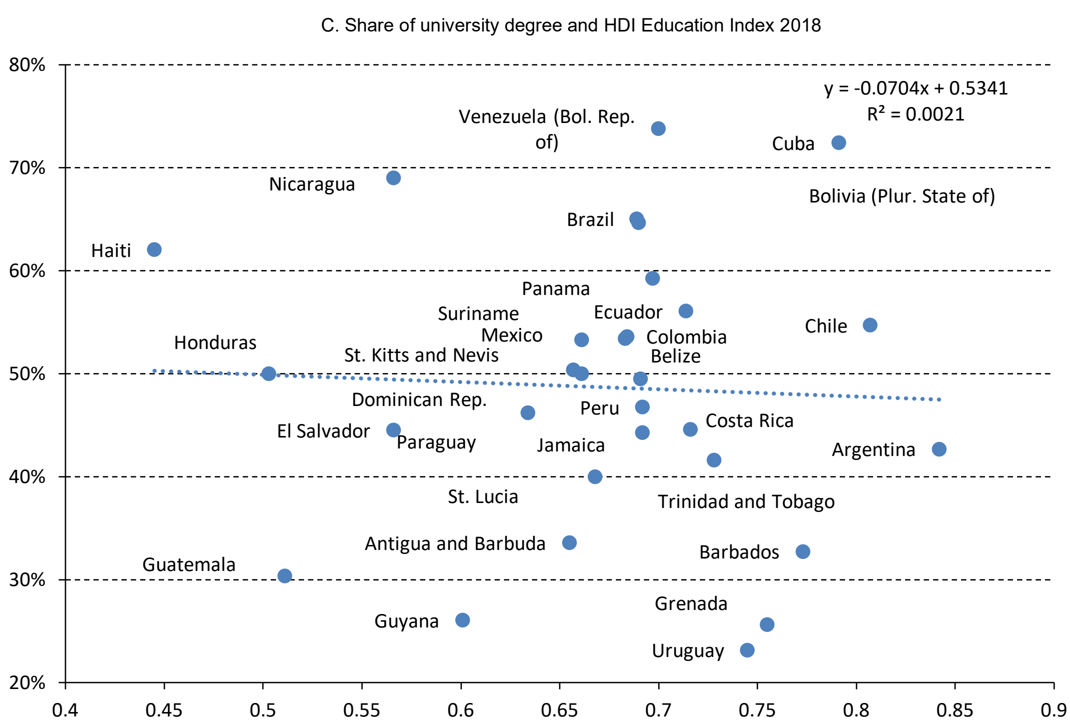
Figures 13D-F show the different shares in form of a map visualization.

Figure 13
Domestic Market supply by educational attainment from Profdir (N=3 356 240)

A. General distribution




B. Per country
(Sample size in parenthesis)
100%
80%
60%
40%
20%
0%
Venezuela (Bol. Rep. of) (108 320)
Cuba (5 183)
Nicaragua (30 532)
Brazil (7 070)
Bolivia (Plur. State of) (40 501)
Haiti (58)
Panama (106 166)
Ecuador (228 295)
Chile (332 322)
Mexico (949 221)
Colombia (378 450)
Suriname (135)
Dominican Rep. (64 946)
Honduras (51 629)
St. Kitts and Nevis (8)
Belize (311)
Peru (328 336)
Paraguay (38 908)
Costa Rica (263 714)
El Salvador (49 558)
Jamaica (300)
Argentina (217 993)
Trinidad and Tobago (682)
St. Lucia (40)
Antigua and Barbuda (119)
Barbados (177)
Guatemala (94 649)
Guyana (349)
Grenada (39)
Uruguay (58 229)
No studies
Vocational training
Elementary/high school
University degree
PhD


C. Share of university degree and HDI Education Index 2018
y = -0.0704x + 0.5341
R² = 0.0021
Venezuela (Bol. Rep. of)
Cuba
Nicaragua
Bolivia (Plur. State of)
Brazil
Haiti
Panama
Suriname
Ecuador
Mexico
Colombia
Chile
Honduras
St. Kitts and Nevis
Belize
Dominican Rep.
Peru
Costa Rica
El Salvador
Paraguay
Jamaica
Argentina
St. Lucia
Trinidad and Tobago
Antigua and Barbuda
Barbados
Guatemala
Grenada
Guyana
Uruguay
80%
70%
60%
50%
40%
30%
20%
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9

D. Map of shares with elementary/high school education

E. Map of shares with university degree

F. Map of shares with PhD

Source: Author's elaboration, based on Profdir.

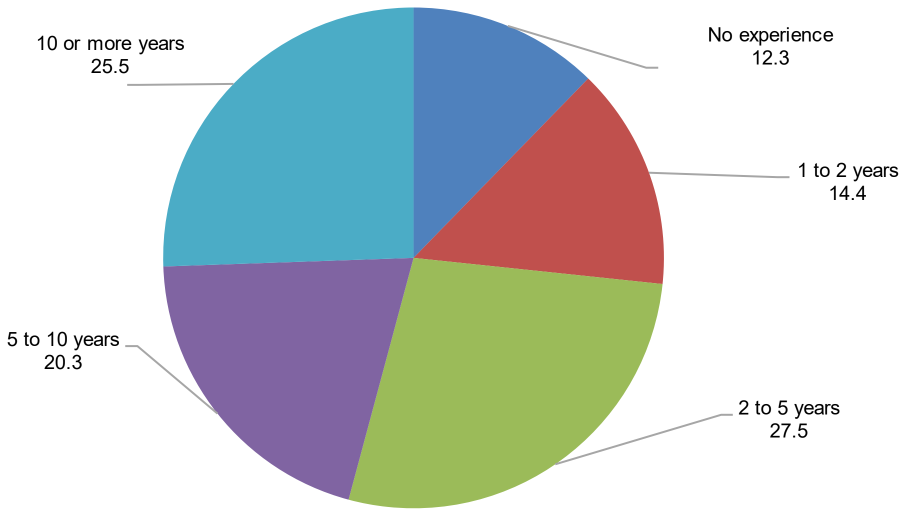
2. Domestic supply by work experience

The Profdir platform also provides data about the extension of previous work experience for most of the worker profiles listed (N = 3,358,271). Some 54% of the profiles have less than 5 years of experience (figure 14A). Figure 14B reveals some interesting differences among countries (neglecting variations introduced by some countries with quite small and unstable sample sizes). For example, in Brazil and Cuba, more than 70% of the worker profiles show more than 5 years of experience, while it is some 40% in Bolivia, Honduras, Peru and Paraguay.

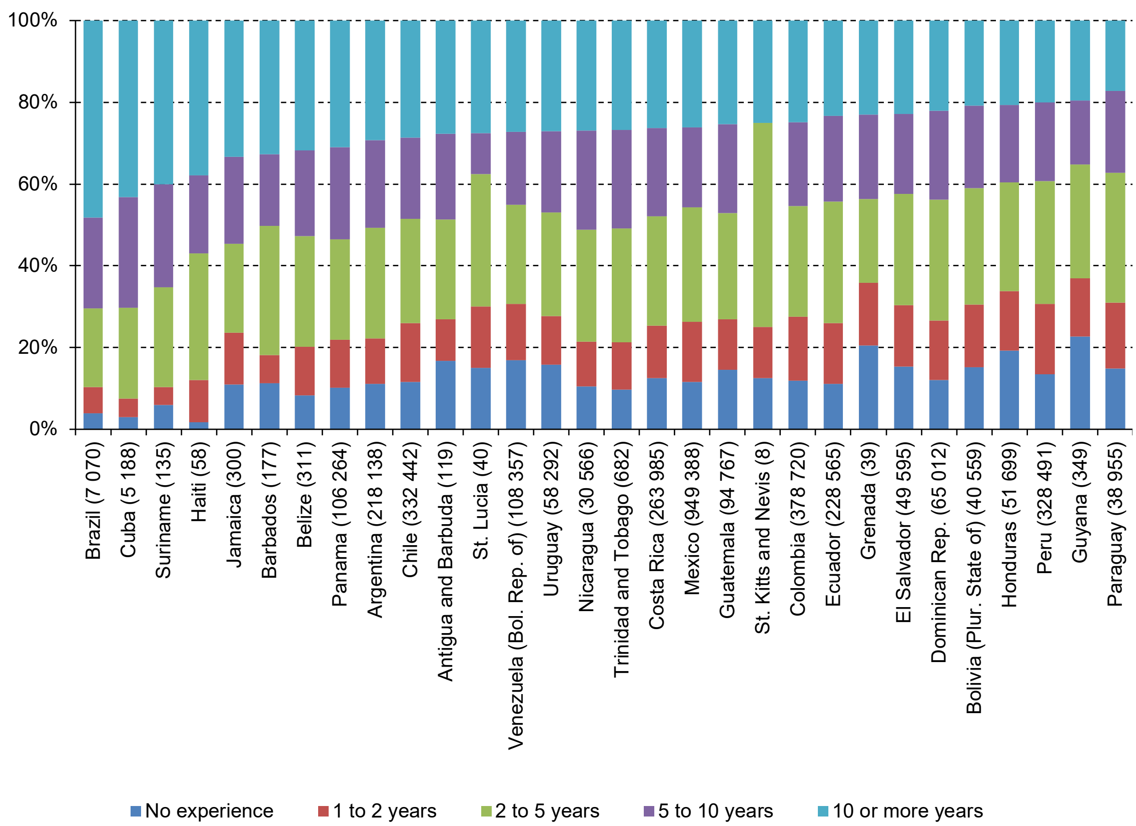
Figure 14C shows that the experience of workers in this online labour market roughly correlates with the percentage of labour force participation of young workers in a country, i.e., ages 15-24 (World Bank, 2019). Some notable exceptions are again Brazil and Cuba, where more experienced workers dominate the online market. This suggests that, contrary to the previous finding on education, the online distribution might simply reflect the off-line distribution of workers in a rather proportional manner.

Figure 14
Domestic Market supply by work experience from Profdir, (N=3 358 271)

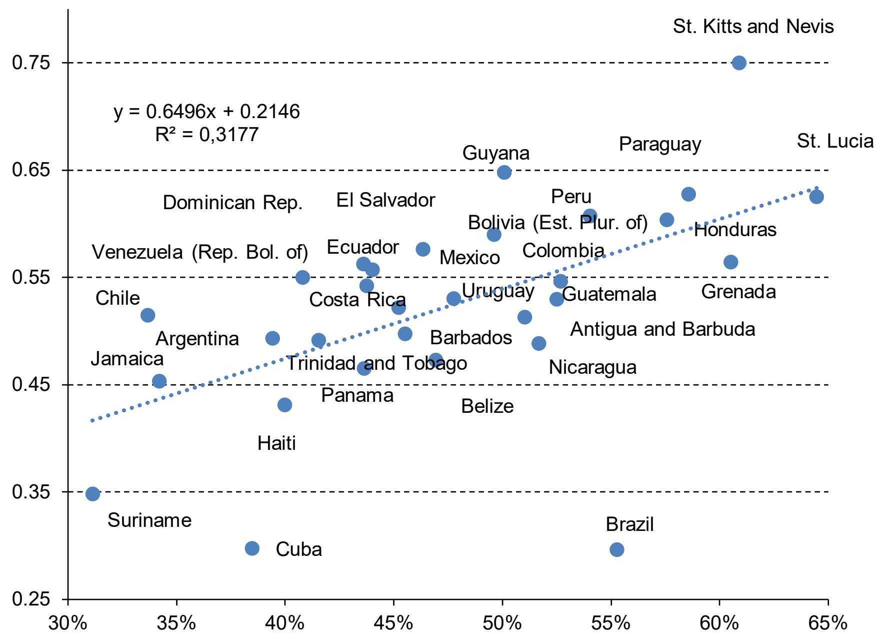
A. General distribution



B. Per country



C. Share of workers with less than 5 years of experience and labour force participation rate for ages 15-24, total percent



Source: Author's elaboration, based on Profdir.

C. Domestic market supply and demand

One interesting opportunity of the obtained online data is to analyze the mismatch between supply and demand of the different sectors. It gives insights into structural challenges in the labour market. Unfortunately, we only have supply data for Profdir, but the results are quite coherent when comparing it to the demand from all six platforms. Figure 15A shows that Latin America and the Caribbean as a whole has a notable oversupply of professionals focusing on 'Administration, Financing' and 'Management, Consulting', while there is a clear undersupply in 'Commerce and Sales' and 'Media, Marketing and Communication'. Disaggregating this finding among sub-regions,[1] it shows that most of the sectoral mismatches are the same across the region, with some notable exceptions. For example, South America has a noteworthy supply of 'Maintenance and Engineering' professionals, and the Southern Cone of 'Social Services, Catering, Entertainment'. In general, it is interesting to note that domestic supply and demand in the sector of 'IT, Telecommunication' seems to be quite well adjusted across the region.

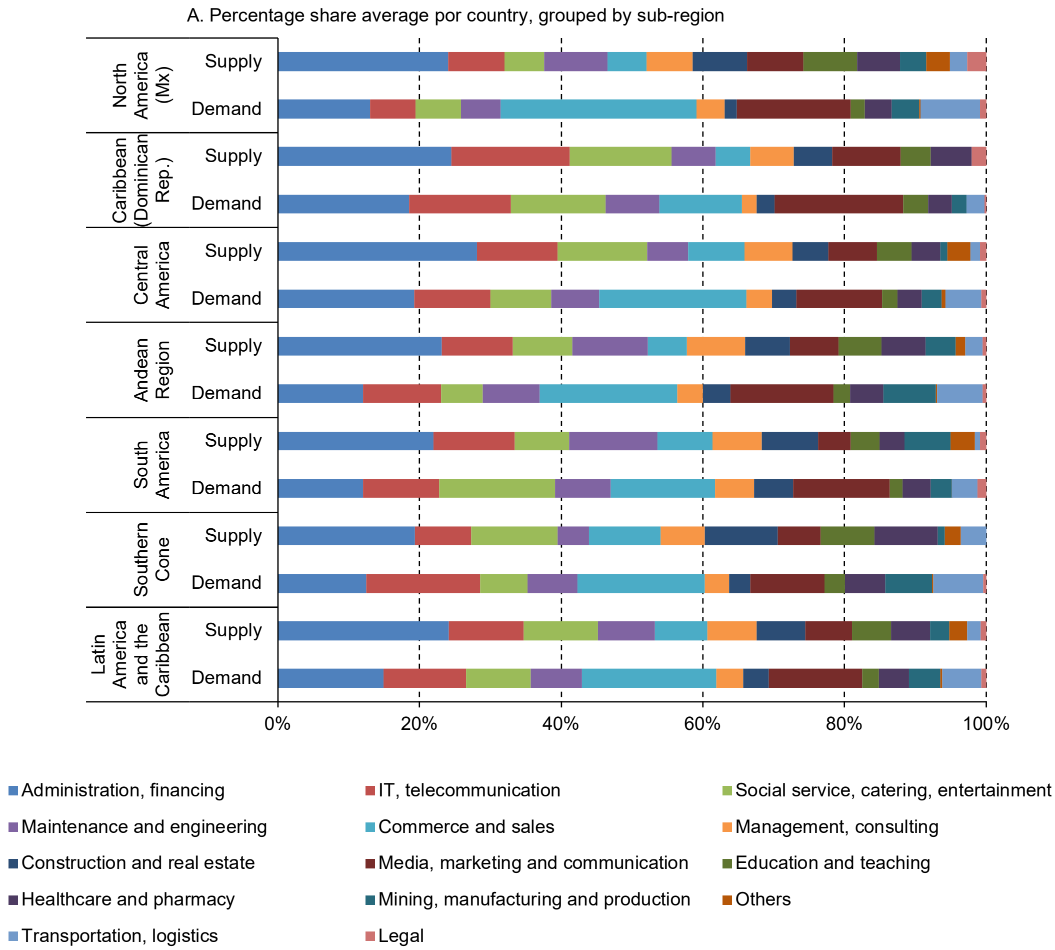
Figure 15B summarizes the mismatch between the different sectors in one single number. We again calculate the Jensen–Shannon divergence (measured in bits).[3] The larger the number, the larger the divergence. This summary statistic reveals that sub-regions with smaller countries achieve a better match between supply and demand, compared with the larger countries of North- and South America and the Southern Cone.[1]

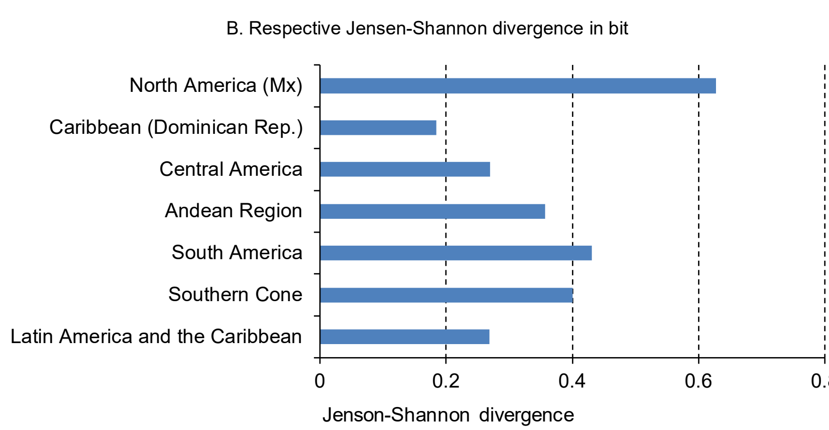
This tendency is emphasized by figures 15C and 15D, which breaks down the same analysis to the country level. Costa Rica, Dominican Republic, Panama, and El Salvador show the best overall match between supply and demand of labour per sector. Venezuela, Mexico, and Colombia, the largest mismatch. It is interesting to note all of these three cases have a large mismatch in 'Commerce and Sales' and in 'Media, Marketing and Communication'. Venezuela also shows an apparent oversupply of workers in 'Maintenance and Engineering', in 'Construction and Real Estate', and 'Mining, Manufacturing and Production'. It is interesting to note that something similar can be noted for the case of Brazil, although to a smaller extent. Among the many other interesting details, it is notable that Bolivia, Argentina, and Chile have an overdemand for 'IT, Telecommunication', and that several countries have a notable overdemand for 'Transportation, Logistics'.

Figure 15

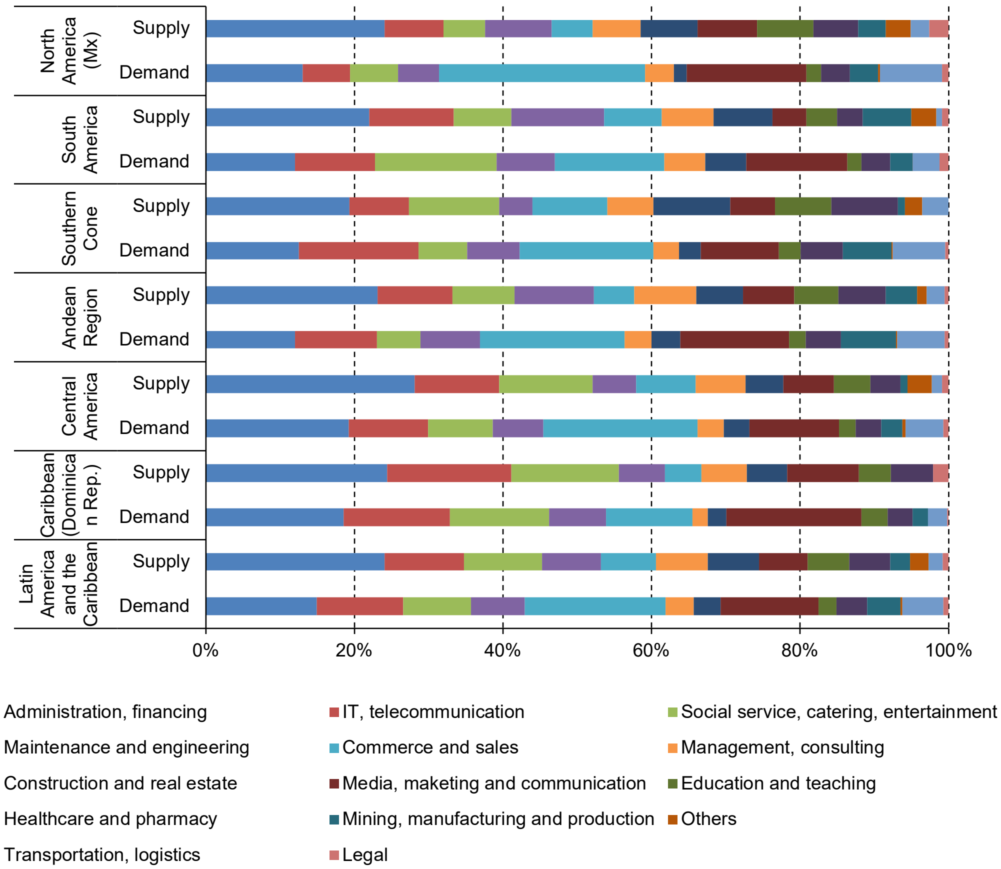
Domestic market supply (worker profiles, Profdir) versus demand[a]

(All six platforms)

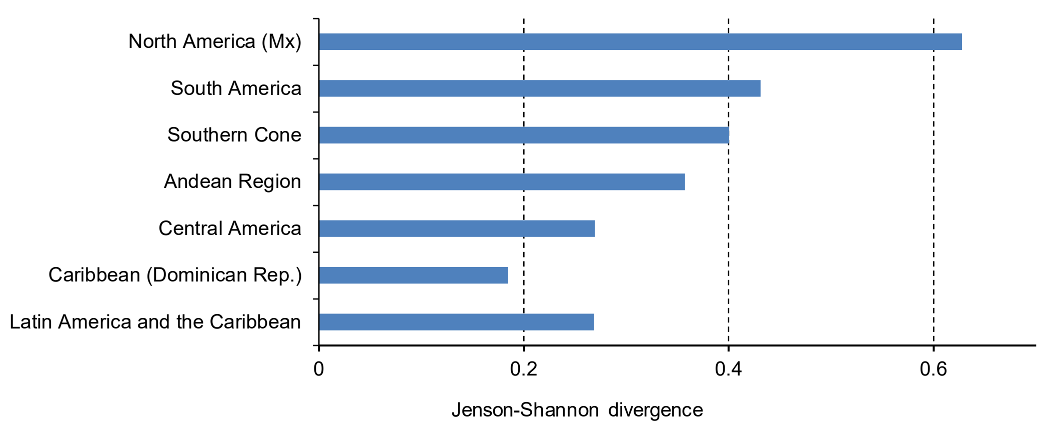




C. Percentage shares for countries



D. Jensen-Shannon divergence in bits for countries



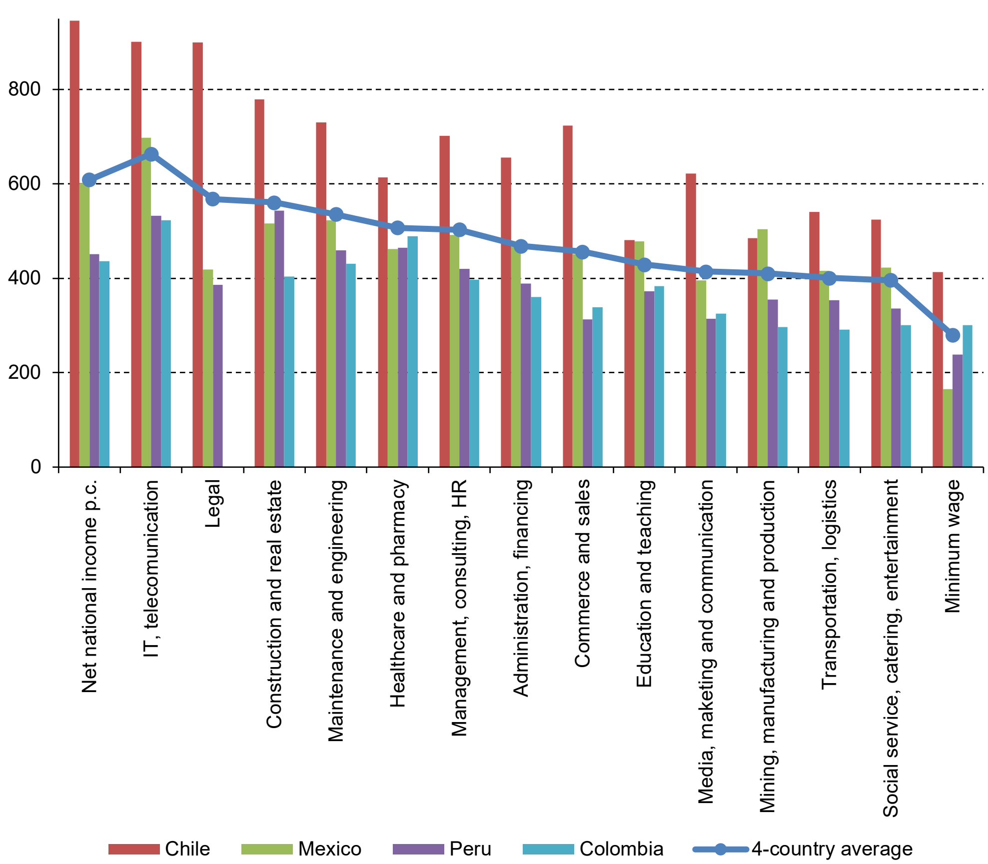
Source: Author's elaboration, based on Profdir.
[a] Sub-regions are grouped as follows: Andean Region: Bolivia, Colombia, Ecuador, Peru; Caribbean: Antigua and Barbuda, Bahamas, Barbados, Cuba, Dominica, Dominican Republic, Grenada, Guyana, Haiti, Jamaica, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago;Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama;North America: Mexico;South America: Brazil, Paraguay, Suriname, Venezuela;Southern Cone: Argentina, Chile, Uruguay.

D. Domestic market salaries

We obtained the salary average estimations published by CompuTrabajo, which is our second biggest source. The salary data of one occupation is calculated from all the monthly salary data of the same occupation from the last 12 months in CompuTrabajo by sector. We converted it to USD. It is important to note that in this observational data, each occupation has a different sample size, and each industry has a different number of occupations.

Figure 16 shows that the average salaries offered on CompuTrabajo are framed by the adjusted net national income per capita (current US$) on the higher end, and the minimum wage on the lower end (see right-most bars in figure 16). Most salaries are below the estimated adjusted income per capita, which has to consider that the total income of an individual might come from more sources than a single job (including investments, etc.). 'IT, Telecommunication' offers the highest salaries. This shows opportunities for those countries and regions that have an undersupply in that sector (see figure 15). 'Social Services, Catering, Entertainment' offers the lowest average services in the region. It is interesting to note that in Colombia, several sectors offer jobs with an average salary close to the national minimum wage (compare Colombia bars with right-most minimum wage bar at $300 in figure 16).

Figure 16
Weighted average monthly salary for domestic labour market job calls in Chile, Mexico, Peru and Colombia, per sector



Source: Author's elaboration, based on CompuTrabajo.
Note: minimum wage assuming 8 hours work per weekday day.

IV. Temporal dynamics

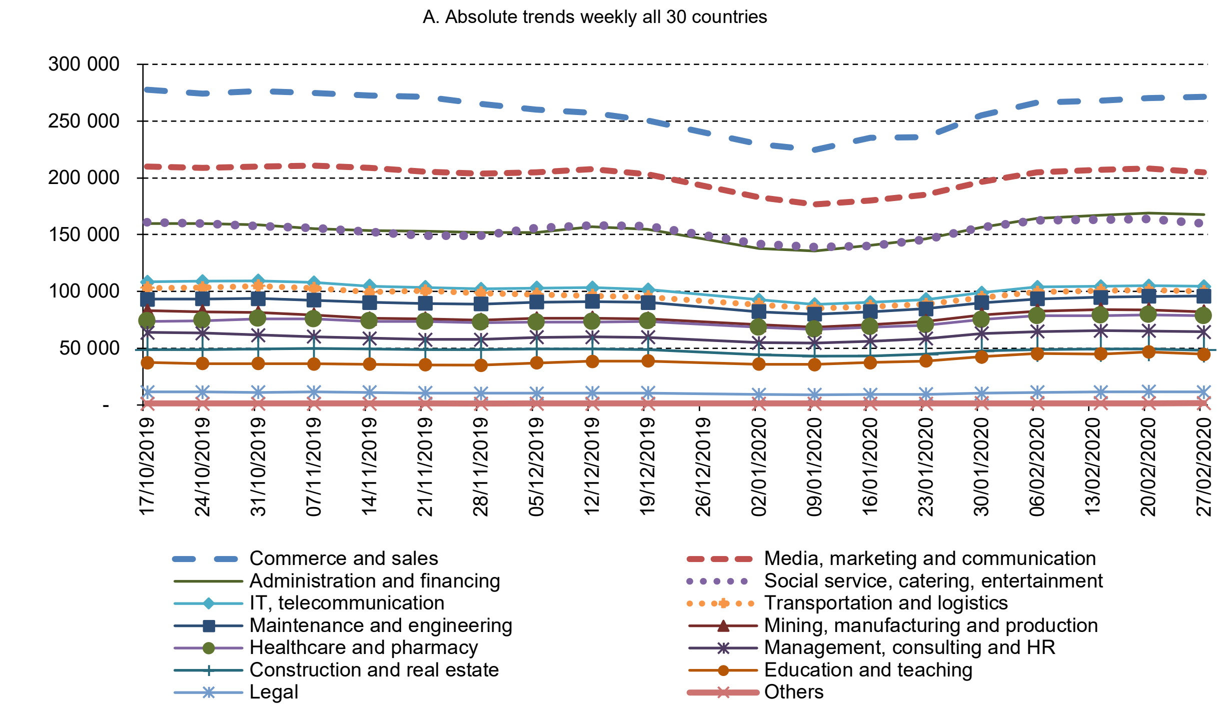
One of the new opportunities arising from work with digital trace data is the creation of time series data. We collected data daily for Profdir, Jobisjob, CompuTrabajo, and Bumeran. We collected the data for the 19 weeks between mid-October 2019, to end of February 2020. Figure 17A shows that the sectorial shares are quite stable over time, with a notable dip over New Years. Figure 17B shows the percentage change of the involved tendencies in labour demand, which makes clear that all sectors experienced the decrease of job postings at the beginning of the year. Some sectors, including ‘Education and Teaching’ and ‘Social Service, Catering, Entertainment’ experienced a notable increase at the beginning of December. Then again, at the end of January, all sectors saw a large increase. A complete yearly analysis will allow analysts to get an idea about such annual cycles.

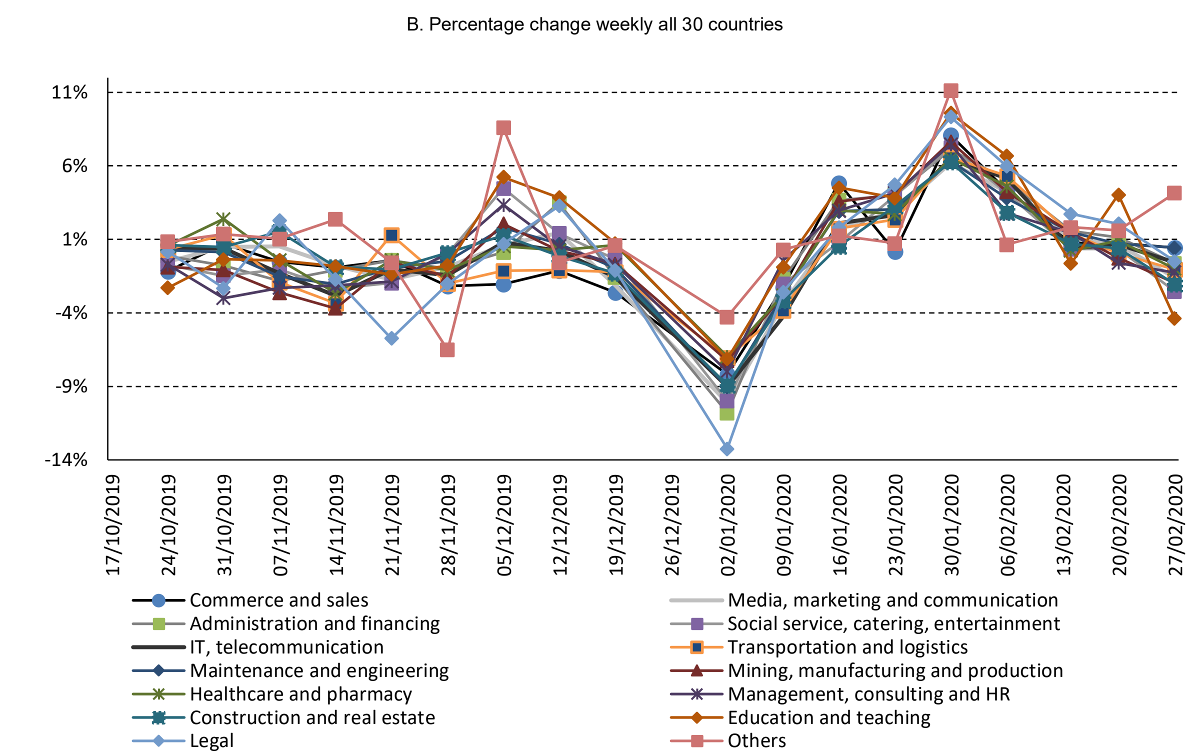
Our interactive dashboard also allows us to focus on individual countries, chosen from the 30 countries included in our collection. Figure 17C shows the case of Chile, which has a slightly different pattern than the regional average, including a notable increase mid-November. This is only one example of a myriad of analytical possibilities to choose from.

Figure 17D creates daily averages per weekday. It shows a clear weekly cycle for all sectors, with the peak of job postings on Friday, and the low on Mondays. This cycle is likely dictated by those posting job calls, which is expected to be Human Resource Departments in most cases.

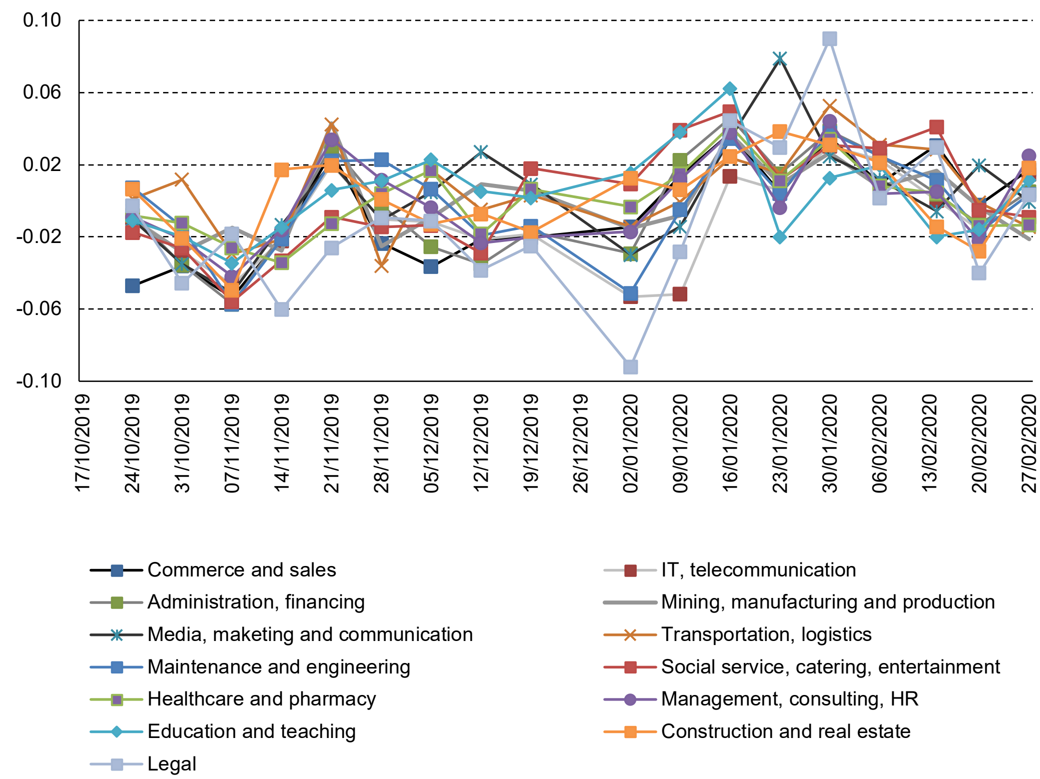
While we carried out daily collections between Oct. 14 to Dec. 15, we reduced our collection effort to weekly collections for the period until Feb. 27, and only collected every Thursday. When comparing the daily and weekly collection for the first period, we find that the average difference is very small (0.23% on average, or some 3k job calls of a total of 1.4M). With weekly collections, one loses the weekly cycle variation, and our choice of Thursday slightly overestimates the daily average (compare with figure 17D), but the overall average seems to provide a reasonable estimate, with 1/7 of the collection effort. Figure 17D suggests that any of Tuesday, Wednesday, or Thursday seems to be adequate for weekly collections.

Figure 17
Temporal dynamics of demand per sector for mid-October 2019 to end-February 2020

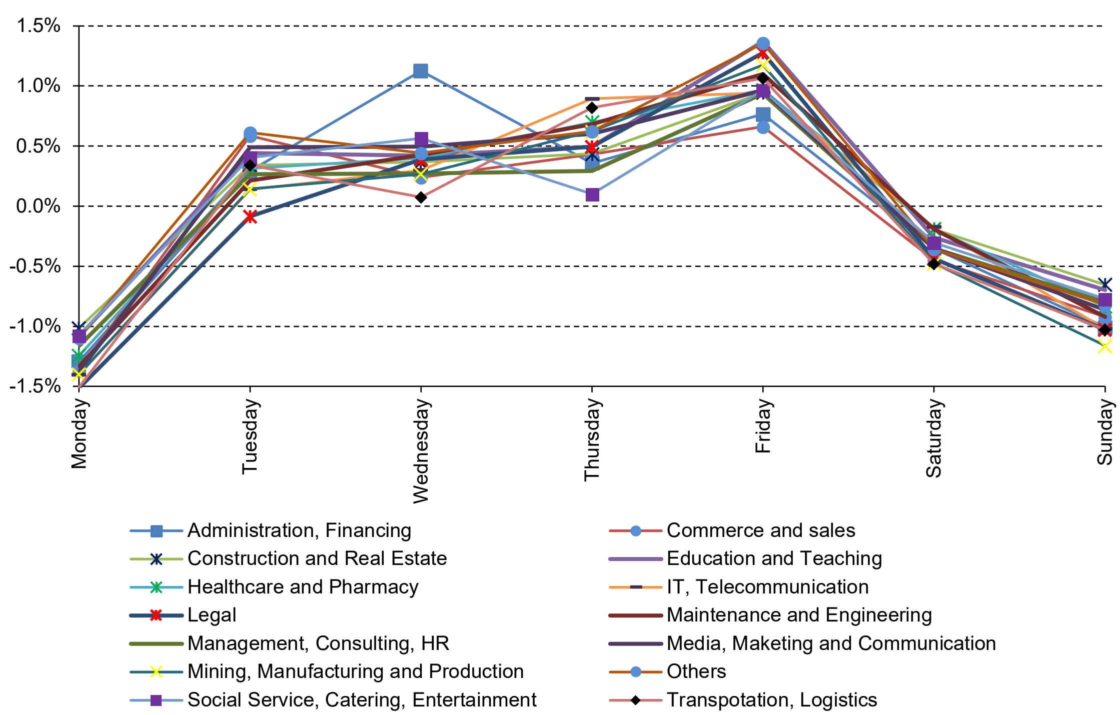




C. Percentage change weekly only Chile



D. Percentage change per weekday all 30 countries (for first nine weeks).



Source: Author's elaboration, based on Bumeran, CompuTrabajo, Jobisjob and Profdir.

Table 3
Weekly versus daily collections for October 14 to December 15

	Weekly collection *(Thursdays)*	Daily collection	Difference in percent
Administration and Financing	155 703	155 053	0.42
Commerce and Sales	269 886	269 047	0.31
Construction and Real Estate	49 150	49 111	0.08
Education and Teaching	36 489	36 493	-0.01
Healthcare and Pharmacy	73 956	73 796	0.22
IT, Telecommunication	105 736	105 439	0.28
Legal	10 922	10 919	0.03
Maintenance and Engineering	91 480	91 333	0.16
Management, Consulting and HR	60 292	60 256	0.06
Media, Marketing and Communication	207 783	207 211	0.28
Mining, Manufacturing and Production	78 446	78 310	0.17
Others	1 356	1 354	0.13
Social Service, Catering, Entertainment	155 583	155 494	0.06
Transpotation and Logistics	100 644	100 394	0.25
SUM	**1 397 425**	**1 394 213**	**0.23**

Source: Author's elaboration, based on Bumeran, CompuTrabajo, Jobisjob and Profdir.

V. Further explorations

A. The skills network

The detail of the obtained data allows researchers, analysts, and policymakers to undertake a large variety of other studies. For example, one pressing issue refers to the question of which skills are in demand, how do skills relate to each other, and how is this dynamic evolving? This is especially pressing in a labour market that is swiftly reshaped by important technological innovations like artificial intelligence (Frank et al., 2019). One fruitful approach to answering this question proposes to create skill networks (Alabdulkareem et al., 2018; Guerrero & Axtell, 2013).

Nodes can represent skills and links some kind of connection between them. As an exemplary exercise, we describe the online freelancing labour market of Freelancer.com in a skill co-occurrence network. We represent skills of profiles on Freelancer.com as nodes, and links are the co-occurrence of two skills in the same profile. In other words, if a worker has two skills, these are connected in the network. The 'strength of the tie' reflects how many workers have this specific skill pair. Every existing skill pair of a worker adds to the connection between this skill pair in our skill network. If only one single worker has this specific skill pair, the 'tie strength' is 1.

The original network contains 1,388 different skills and 1,416,320 co-occurrences of skill pairs, of which, 165,686 are unique pairs of existing co-occurrences of different skills (figure 18A). In other words, the network has 1,388 different nodes, and 165,686 links, each with an average strength of about 8.5 mentions (which in reality are not all equally strong). The size of the nodes in the graph corresponds to the weighted degree of the nodes (the number of co-occurrences of that skill with other skills). The color corresponds to the result of running a community detection algorithm.[4] We detect seven different groups of skills.

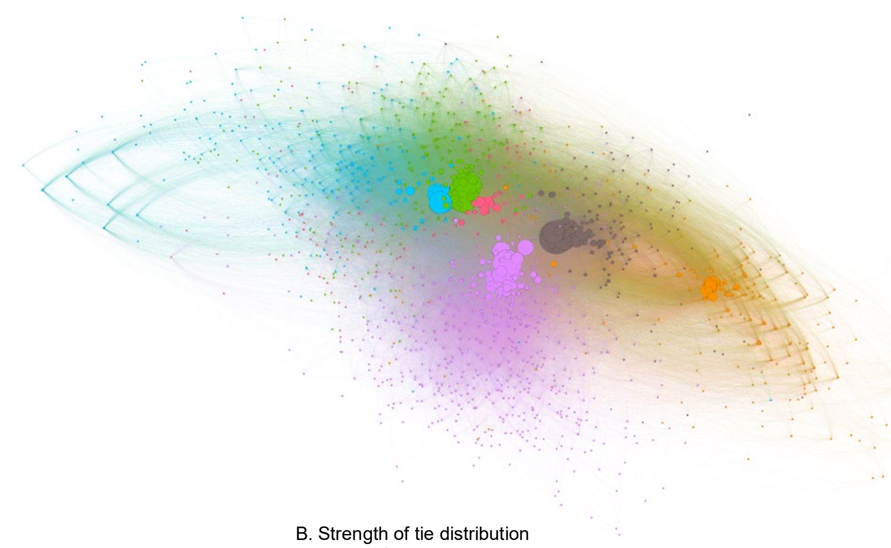
[4] There are many different ways on how one can partition a network in different groups and how one can detect communities in a graph (exponentially many). Which parts of the network create sub-groups? We use a so-called modularity algorithm for community detection. It optimizes the differences of within- and between connections in possible groupings (Blondel et al., 2008).

Figure 18B shows the distribution of 'tie strength'. It shows that most links have less than 5 mentions, which means that those co-occurrences of skills appear less than five times in our dataset. We delete these rather rarely co-occurring skill pairs to simplify the analysis and are left with 827 nodes and 39,628 links, each with an average tie strength of 30. Figure 18C shows the result. The largest resulting group, with 32% of the nodes, refers to web and design skills. Websites, Android, and HTML are part of this skill community. The second-largest refers to rather traditional skills, including Spanish and English, translation and article editing. The third-largest group refers to basic data tasks, call center work, and customer service.

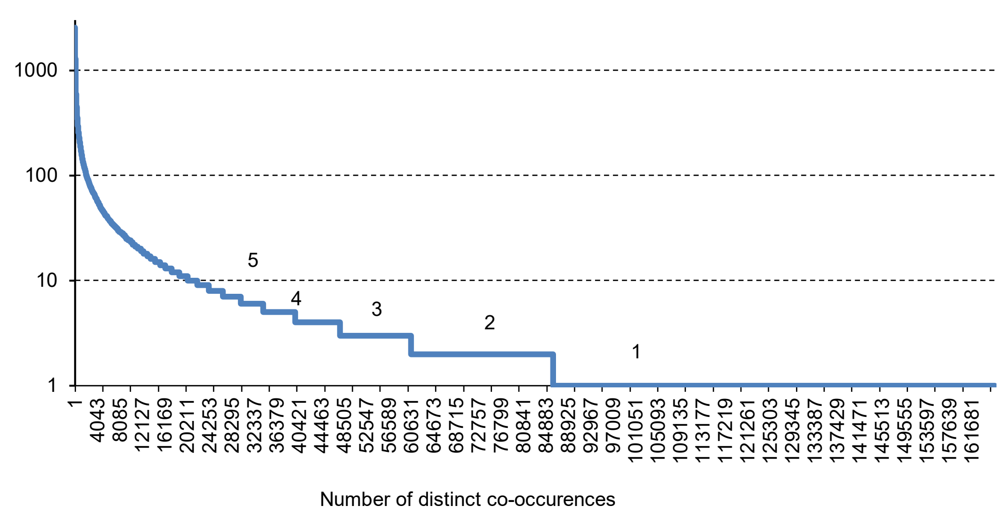
This kind of information becomes very useful when tracking quickly evolving skill landscapes. It can also be used for statistical purposes. In this study, we used rather out-of-the-box, manually harmonized ways to create labour market sectors (see table 1 and 2). However, one could also use a more data-driven approach, like the one presented here, to identify labour market sectors that represent the quickly evolving landscape of sectors and skills more adequately.

Figure 18
Skill co-occurrence network from Freelancer.com

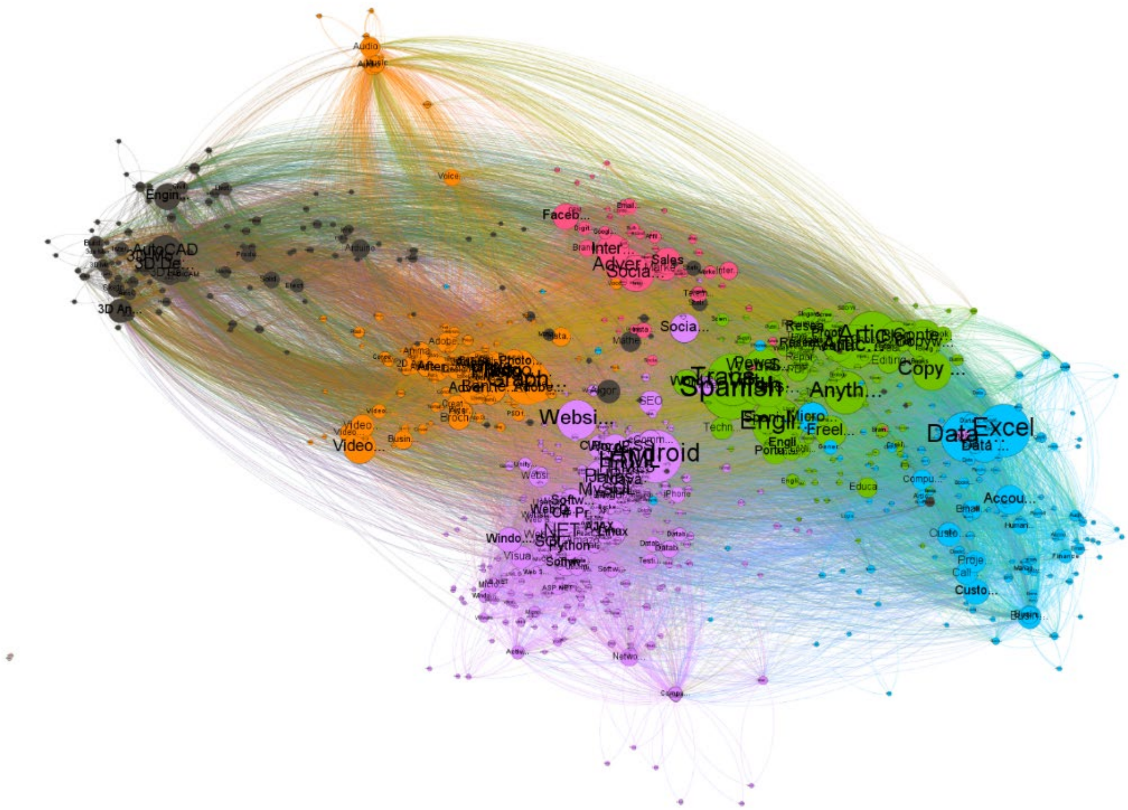
A. Total network



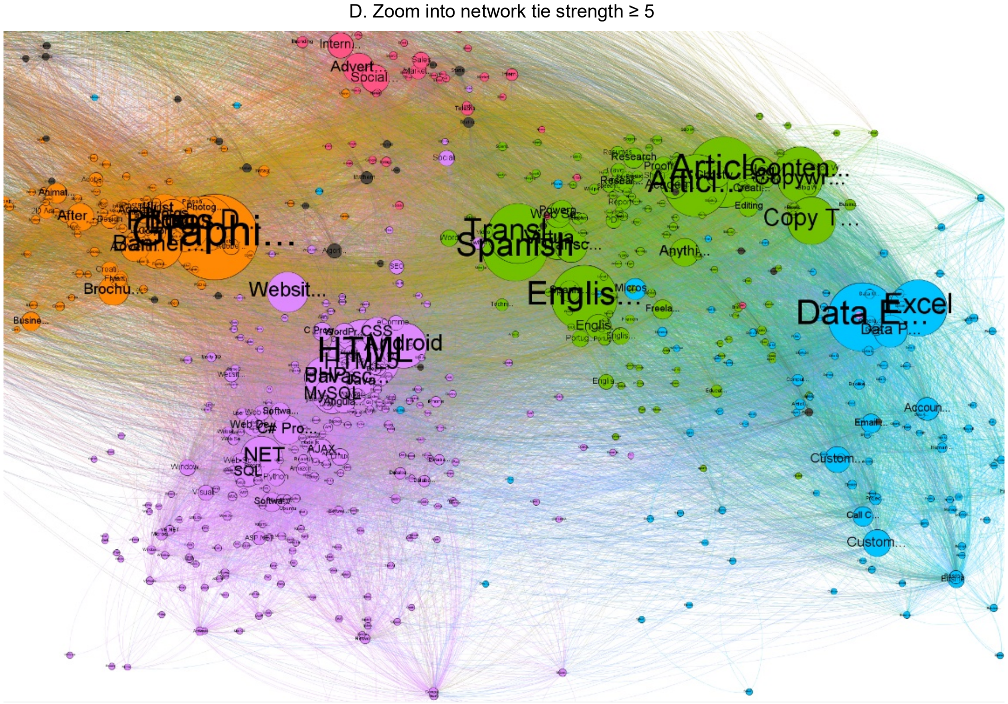
B. Strength of tie distribution
(Repetition of co-occurrence link)



C. Network with tie strength ≥ 5



D. Zoom into network tie strength ≥ 5



Source: Author's elaboration, based on Freelancer.

B. Custom-made data visualization

Returning to where we started with diagram 1, the third and final step from our data science workflow consists of converting raw numbers into some format that lends itself to derive actionable insights. We chose to keep the analysis rather descriptive and straightforward and focus on visual examination. In our case, we opt for the deployment of a series of visualizations that allow analysts, policymakers, scholars, and the general public to gain informed insights promptly. This is still a very superficial level of analysis. Most of the dozens of graphs and figures presented in this report could benefit from a much more in-depth analysis in the form of a case study (e.g., with qualitative insights from specific domains) or more sophisticated analytical methods (e.g., multivariate analysis, like structural equation models, or other econometric examinations).

In the most rudimentary form, the obtained data can be visualized with simple graphs with the help of popular programs like MS Excel or Google sheets, as done in this report. The fundamental limitation with this approach is that it restricts the richness of presentation. The presented graph is only one of all possible charts that could have been displayed based on the same data. Static graphs are useful to answer a specific research question or hypothesis or to formulate a judgment concerning existing arguments, but it goes against the central premises of the 'big data' paradigm: taking advantage of all of the available data.

For example, a typical dataset collected for this exercise might contain 28 countries, tracking activity of 14 sectors, over 9 weeks (see, for example, figure 17 and diagram 5). This implies a total of (28+14+9) = 51 variables. In the most basic form, we could choose a graph that shows all of them together, including all weeks and all sectors (which is what we do in diagram 5), and we could additionally also disaggregate among all countries. This would be a very busy graph, most probably not intelligible. In contrast, we could only present one single number, e.g., one sector for one week, in one country. We have exactly 51 of these insightful numbers. It would hardly be a graph at all and miss important comparative aspects. In between, and in practice, someone with domain knowledge makes a deliberate choice about the selection of variables and their level of aggregation, choosing the most 'interesting', 'insightful', or 'meaningful' dates/sectors/aspects, and/or their level of aggregation. This process of selection is inevitably more an art than a science, because the number of choices is almost unfathomable, which makes the selection of any specific graph so restrictive. The number of possible graphs is the sum of all possible values of the well-known 'n choose k' formula, where n denotes the number of available variables, and k the number of the chosen variables presented in a given graph:

$$Number\ of\ possible\ graphs = \sum_{k=1}^{n} \binom{n}{k} = \sum_{k=1}^{n} \frac{n!}{k!\,(n-k)!}$$

Choosing to present the combination of any two of the total 51 variables allows creating 1,275 different graphs. Creating a graph that focuses on three of them leads to 20,825 different figures. Working with four variables, allows the researcher to choose from 250 thousand different graphs (249,900 to be precise), and selecting any five of the 51 variables makes for almost 2.5 million figures (2,349,060). In total, the sum of all possible choices allows making more than two-quadrillion different graphs (exactly 2,251,799,813,685,250), which is more than 3.5 million different graphs for each and every one of the 637 thousand inhabitants of Latin America and the Caribbean.

A central part of data science is making choices among these unfathomable number of options. The first requirement to make this in any meaningful way is domain knowledge of the area under investigation. Without the guidance of informed research questions, it is unlikely that useful graphs are being selected. Data science goes way beyond computer science and also beyond mere statistics. The second step consists in trying to open the straight jacket provided by data representation tools. Modern data science, therefore, often does not work with static graphs but with so-called dashboards. These are interactive tools that allow users to select and deselect variables and analyze different aspects of the chart. Research from different countries and with different interests have different questions. Interactive dashboards allow answering yet unasked questions based on the same data.

The general public is getting used to such online dashboard assessments, as interactive line-graphs allow them to zoom into a certain time period (e.g., when looking at currency exchange rate histories), or when evaluating the changing number of the total price of their purchase by (de)selecting products in their e-commerce shopping cart.

In our case, we opted to work with a Python based visualization tool, which we host here: https://lac-local-market.herokuapp.com/. The two most common open-source options are Dash (available supports for Javascript, Python, and R) and R-shiny (which works on the basis of the statistical package R, and has commercial extension: https://shiny.rstudio.com). Popular commercial options include Tableau and Qlik, which require a commercial license, that usually costs several US$ per user.

Diagram 5
Example screenshot of one of our Plotly/Dash dashboards that displays domestic labour market demand over 9 weeks

Source: Author's elaboration.

VI. Conclusions

This report presents the main findings of data collected from six major international labour market platforms and two global freelancing sites over the last months of 2019 and the beginning of 2020. These platforms are Acciontrabajo (Profdir), CompuTrabajo, Jobisjob, CaribbeanJobsOnline, CaribbeanJobs, and the two global freelancing sites, Freelancer and Upwork which covered 33 countries in Latin America and the Caribbean.

We conclude that the analysis of on line job vacancies and of labour supply may become a very powerful labour market information tool. The wealth of these kind of data calls for the use of an interactive daily updated dashboard with several variables of interest.

Using digital trace data to monitor labour market dynamics has several benefits such as:

- **Thematic coverage**: including areas that were previously difficult or impossible to measure, like the educational level of workers or supply-demand mismatch per country;
- **Geographical coverage**: our international sources provided sizable and comparable data for up to 33 countries, including many small countries that usually are not included in international inventories;
- **Level of detail:** like the skill network, including the connections among skil sets, or the review of workers by their employers;
- **Timeliness and timeseries**: we produced graphs the same week we collected the data, and are able to quickly create extensive timeseries to start looking into labour market dynamics, which are often not captured in labour-intensive surveys.

However, working with digital trace data instead of traditional survey data does not eliminate the traditional challenges involved when working in the field of international quantitative analysis. Priorities change a bit, but the basic discussions remain the same. Among the main challenges it is important to considered:

- **Representativeness.** While traditional development statistics is mainly concerned with the representativeness of random survey samples, digital trace data is never a random sample. It is observational data, and ambition is usually to obtain the entire population completely. Technical skills shift from mastering how to collect a representative sample and how to perform a significance test, collecting digital traces and how to retrieve, clean and store large amount of data effectively. Practical issues can make a difference. This became obvious when Brazil and Haiti became outliers, which might simply be because these non-Spanish speaking countries are not properly represented within our scope of observation.
- **Generalizability.** While observational data always represents this source very well, it only represents what it represents, and nothing more. While it is tempting to generalize from specific observations of one platform to broader settings, this is often very deceptive. For example, we have seen that our two freelancing platforms focus on slightly different market segments. Freelancer is stronger in 'IT' while Upwork is stronger in 'Writing'. Could we have generalized from either one of them to the entire labour market? What other platform are we missing?
- **Harmonization.** International organizations such as UN ECLAC are used to issues related to the international harmonization of indicators. Digital trace data does not make this work disappear. It adds the challenge of so-called 'data-fusion', the harmonization of different sources. We took advantage of several pan-regional and international platforms, which provide a common outlook. This facilitates international outlooks. But none of them covers all countries and sectors equally well. The Caribbean countries use other portals. We then had to harmonize different sources, which sometimes can be tricky (as shown by our rather lengthy and technical discussion on how to convert skills into categories in the case of Freelancer.com).
- **Data overload.** The wealth of data provided by digital trace data is a clear benefit over the traditional data scarcity in developing countries. We still present it as a challenge because analysts and institutions are not used to work with interactive dashboards. We still lack a standard workflow that would allow researchers, users and policymakers to efficiently and effectively. We proposed to work with interactive dashboards. While this sounds great in theory, the devil is in the detail of its institutional presentation: how to make this wealth of data effective for policymakers to obtain useful insights? How to balance the trade-off between overwhelming users with too many options and limiting the analysis of special cases? Much more experimental work needs to be done in this regard.

.

Bibliography

Alabdulkareem, A., Frank, M. R., Sun, L., AlShebli, B., Hidalgo, C., & Rahwan, I. (2018). Unpacking the polarization of workplace skills. Science Advances, 4(7), eaao6030. https://doi.org/10.1126/sciadv.aao6030.

Amato, F., Boselli, R., Cesarini, M., Mercorio, F., Mezzanzanica, M., Moscato, V., Persia, F., & Picariello, A. (2015). Challenge: Processing web texts for classifying job offers. Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), 460–463. https://doi.org/10.1109/ICOSC.2015.7050852.

Barabasi, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509–512.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. ArXiv:0803.0476. https://doi.org/10.1088/1742-5468/2008/10/P10008.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labour. Proceedings of the National Academy of Sciences, 116(14), 6531–6539. https://doi.org/10.1073/pnas.1900949116.

Galperin, H., & Greppi, C. (2019). Geographic Discrimination in the Gig Economy. In M. Graham (Ed.), Digital Economies at Global Margins (p. 295). MIT Press. https://arnicusc.org/publications/geographic-discrimination-in-the-gig-economy/.

González Velosa, C., & Peña Tenjo, N. (2019). Demanda de trabajo en América Latina: ¿Qué podemos aprender de los portales de vacantes online? Inter-American Development Bank. https://doi.org/10.18235/0001965.

Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. Transfer: European Review of Labour and Research, 23(2), 135–162. https://doi.org/10.1177/1024258916687250.

Guerrero, O. A., & Axtell, R. L. (2013). Employment Growth through Labour Flow Networks. PLoS ONE, 8(5), e60808. https://doi.org/10.1371/journal.pone.0060808.

Heeks, R. (2017). Decent Work and the Digital Gig Economy: A Developing Country Perspective on Employment Impacts and Standards in Online Outsourcing, Crowdwork, etc. Global Development Institute. University of Manchester–2017–Режим Доступу До Ресурсу: Https://Www. Gdi. Manchester. Ac. Uk/Research/Publications/Di/Di-Wp71.

Kar, S. S., & Ramalingam, A. (2013). Is 30 the magic number? Issues in sample size estimation. National Journal of Community Medicine, 4(1), 175–179.

Kassi, O. (2016, Sept. 21). How is online work classified in the OLI? Oxford Internet Institute. https://www.oii.ox.ac.uk/blog/how-is-online-work-classified-in-the-oli/.

Kässi, O., & Lehdonvirta, V. (2018). Online labour index: Measuring the online gig economy for policy and research. Technological Forecasting and Social Change, 137, 241–248. https://doi.org/10.1016/j.techfore.2018.07.056.

UN ECLAC, (United Nations Economic Commission for Latin America and the Caribbean). (2020). Tracking the digital footprint in Latin America and the Caribbean: Lessons learned from using big data to assess the digital economy (Productive Development, Gender Affairs LC/TS.2020/12; Documentos de Proyecto). United Nations ECLAC. https://repositorio.cepal.org/handle/11362/45484.

World Bank. (2019). World Bank Data. https://data.worldbank.org/indicator/SL.TLF.ACTI.1524.NE.ZS.

Zeifman, I. (2015). 2015 Bot Traffic Report: Humans Take Back the Web, Bad Bots Not... Imperva Incapsula. https://www.incapsula.com/blog/bot-traffic-report-2015.html.

Jobs intermediated by online platforms have become a central pillar of labour markets in Latin America and the Caribbean. Public online platforms inevitably leave digital trace data that can be used as a source of information regarding online labour supply and demand. This report explores the opportunities and challenges of the systematic use of these publicly available data. The objective is to give an overview of the volume and nature of these data, and to share the lessons learned in order to develop a research agenda that enables alternative labour market information tools to be created, based on these new sources.

To provide an initial analysis of the type of information that can be drawn from this data source, this report presents the main findings from data collected from six major international labour market platforms and two global freelancing sites in late 2019 and early 2020. These platforms are Acciontrabajo (Profdir), CompuTrabajo, Jobisjob, CaribbeanJobsOnline, CaribbeanJobs, and the two global freelancing sites are Freelancer and Upwork, covering 33 countries in Latin America and the Caribbean. One finding, from comparing the characteristics of vacancies (labour demand) and of people offering services (labour supply), is that Latin America and the Caribbean as a whole has a notable oversupply of professionals focusing on ‘Administration, Financing’ and ‘Management, Consulting’, while there is clear undersupply in ‘Commerce and Sales’ and ‘Media, Marketing and Communication’. However, since this type of data is available in abundance, ideally, an interactive dashboard of several variables of interest is required, with daily updates.

ECLAC

Economic Commission for Latin America and the Caribbean (ECLAC)
Comisión Económica para América Latina y el Caribe (CEPAL)
www.eclac.org

LC/TS.2020/83