InGRID

Supporting expertise in inclusive growth

www.inclusivegrowth.eu

Deliverable 12.3

STATE-OF-THE-ART

Data on the platform economy

Zachary Kilhoffer

April 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 730998

Abstract

Platform work is paid labour intermediated by online labour platforms (OLPs) like Uber, Deliveroo, Upwork, and Amazon Mechanical Turk. This report attempts to map the most relevant measurements and methodologies used to estimate the size and relevance of platform work, and suggest potential avenues for future research. I first discuss the terminology of platform work, arguing that a more inclusive understanding is important. While platforms like YouTube and Etsy are not OLPs per se, they clearly allow the intermediation of paid work. Platform work will continue evolving, and researchers should not preclude the possibility that new types of platform work deserve our attention. Second, I discuss the primary ways we understand the size and relevance of platform work: (1) survey data, and (2) administrative and big data. Survey data is found to be especially relevant for measuring worker sentiment. However, generalising findings to the broader population (e.g. to estimate total platform workers) is very difficult, and researchers must pay special attention to survey mode to correct for sample bias. Administrative and big data are very promising for large-scale analyses and seem particularly valuable for informing evidence-based policymaking. Nevertheless, general reliability issues with big data apply, and accessing useful data usually requires cooperative agreements (e.g. with platforms, financial institutions, or government bodies).Third, I discuss two prominent works from Huws *et al.* (2017; 2019) and the COLLEEM surveys (Pesole *et al.*, 2018; Brancati *et al.*, 2019). Both are highly influential in the literature for estimating the number of platform workers in Europe and their activity. However, their conclusions are starkly different. I highlight the reasons why the COLLEEM survey's more modest estimates are more sound. At EU-level and for individual Member States, I conclude that policymakers should consider requiring platforms to provide administrative data to ensure conformity with regulations and better socio-economic insights.

This report constitutes Deliverable 12.3, for Work Package 12 of the InGRID-2 project.

April 2021

General contact: inclusive.growth@kuleuven.be
p.a. InGRID
HIVA - Research Institute for Work and Society
Parkstraat 47 box 5300, 3000 LEUVEN, Belgium

For more information zachary.kilhoffer@ceps.eu

Please refer to this publication as follows:
Kilhoffer, Z. (2021). State-of-the art. Data on the platform economy, Deliverable n°12.3, Leuven, InGRID-2 project 730998 – H2020

This publication is part of the InGRID-2 project, this project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 730998.

InGRID
Supporting expertise in inclusive growth

Contents

1. Introduction

Platform work was already growing, but since the COVID-19 epidemic, some sectors are positively booming. With restaurants shutdown or at limited capacity, more people are relying on couriers to deliver them meals. A huge fraction of the labour force is working from home or out of work, bringing alternative working arrangements into the mainstream. As such, it is an important time to consider what we do and do not know about platform work.

One narrative welcomes platform work as a flexible work form encouraging innovation, entrepreneurship, and social mobility, while disrupting antiquated business models like the taxi industry and over-reliance on the standard employment relationship.[1] Others see platform work as exploitative, displacing stable employment with low-paid work without social protections, growing the 'cybertariat' class, and entrenching the power of tech giants like Amazon and Uber (Huws, 2014). Some authors see platformisation or 'Uberisation' as the wave of the future, while others dismiss it as a marginal labour market phenomenon.

Unfortunately, we still lack reliable data and statistics to evidence these narratives. While debate about platform work has continued for over twelve years,[2] we still struggle to answer even basic questions about platform workers, such as their share of the labour force. This has hampered policymakers' efforts to apply existing regulatory regimes to platform work, or craft new ones (Kilhoffer *et al.*, 2020). In the meantime, platform work continues to evolve at a rapid pace; it remains an elusive topic and exceptionally difficult to pin down.

We lack quality data and statistics on platform work due to a number of theoretical and methodological factors. The most important **theoretical difficulty** is that platform work is poorly defined and means many different things to different people. The primary **methodological difficulty** is that traditional methods of gathering labour market data rarely isolate platform work from other forms, like self-employed, casual work, or occasional work. This is hardly surprising given divergent understandings of what constitutes platform work.

This report revisits and highlights some of the main issues with platform work data and statistics. While not a comprehensive literature review, it attempts to illustrate the theoretical difficulty of platform work and propose a working terminology (2), examine methodological challenges of different data sources (3), and discuss how these challenges impact two influential estimations (4). Thereafter, we offer conclusions and recommendations for researchers and policymakers (5). Appendix 1 contains an overview of data sources for reference.

1 Standard employment relationship refers to a full-time employment contract between an employee and a single employer of indefinite duration (Kilhoffer *et al.*, 2020).

2 For example, Amazon Mechanical Turk launched in 2005 and critical scholarship began by 2008 (Silberman, 2017).

2. The terminology of platform work

The primary theoretical difficulty with data is that platform work is poorly defined. This is reflected by the wide range of terms used for identical or similar concepts, such as 'platform economy', 'sharing economy', and 'collaborative economy' for the larger phenomenon, and 'gig work', 'crowd work', and 'cloud work' for the labour-intensive part thereof.

These terminological differences have important implications on platform worker statistics. For example, some argue that people selling *goods* on Etsy or eBay, as well as *services* via Uber or Freelancer, qualify as platform workers (Huws *et al.*, 2019). After all, such individuals earn money through an online platform. In recent years, more people have begun to make a living off streaming and video platforms like YouTube and Twitch, bringing new types of platforms under additional scrutiny.

Clarifying terminology has only increased in importance given new legislation. On 20 June 2019, the Platform to Business (P2B) Regulation on promoting fairness and transparency for business users of online intermediated services was adopted.[3] While not aimed at platform work, the P2B Regulation introduces rules for 'online intermediation services' and 'business users of online intermediation services', who provide either goods or services. In some cases, the P2B Regulation could have great relevance for platform workers (Kilhoffer *et al.*, 2020).

It is therefore important to clarify several terms used in this report:
- platform economy;
- platform;
- platform work;
- online labour platform (OLP);
- platform worker.

Starting from the broadest, the *platform economy* is understood as the totality of digital commercial platforms. A *platform* is any digital commercial platform forming part of the platform economy. *Platform work* means the matching of supply and demand for paid work through a platform.[4] Figure 2 illustrates how platform work (which Schmidt calls digital labour) are a subset of the larger platform economy. The particular platforms of interest to this report are *Online Labour Platforms* (OLPs), which are those platforms through which platform work takes place.[5] Finally, a platform worker is any natural person providing paid work through OLPs.[6] This conceptualisation of platform work versus standard work[7] is presented in Figure 1.

3 Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (OJ L186/57 11.07.2019).

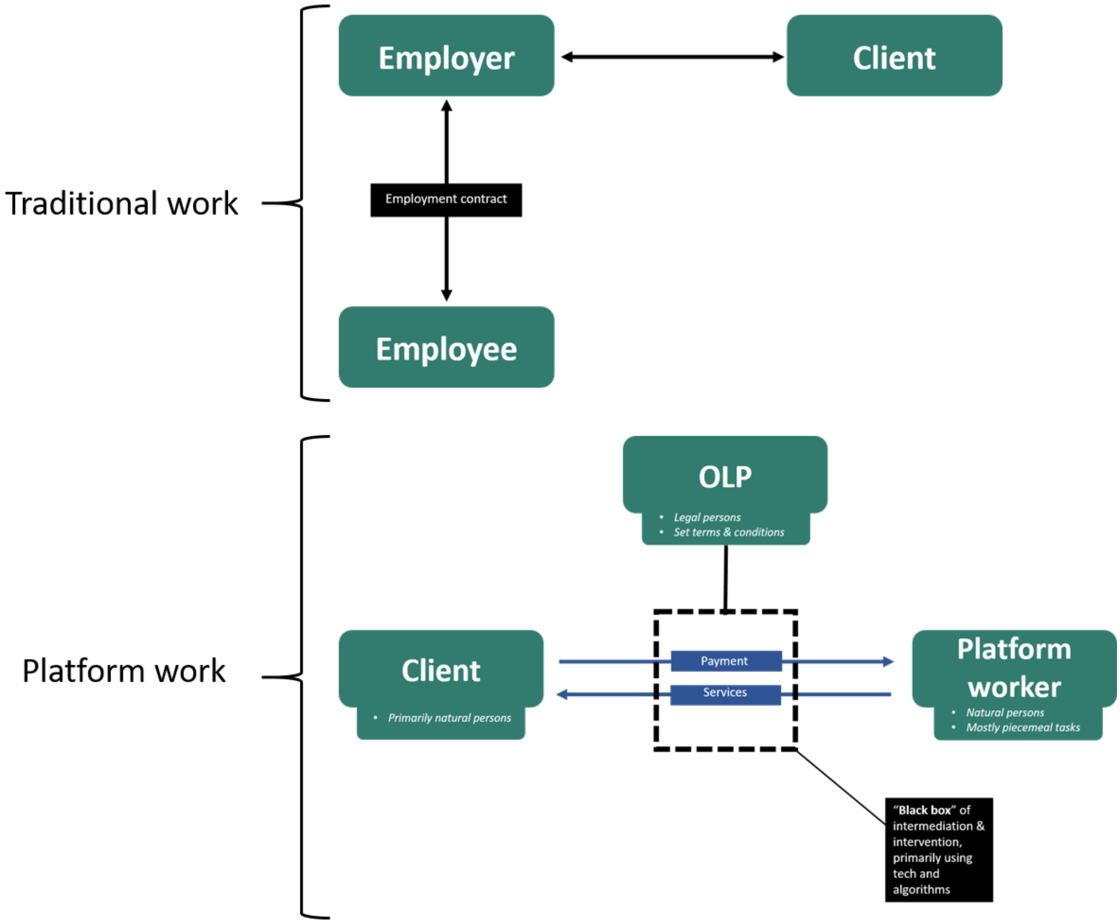
4 Adapted from Eurofound (2018).

5 The term OLP is taken from Oxford Internet Institute publications such as Kässi and Lehdonvirta (2018). It avoids awkward constructions like 'platform work platforms'.

6 While the term worker has different legal meanings, such as in EU treaties and UK labour law, I do not assign any legal significance to the term. The word 'worker' in the term 'platform worker' simply means anyone who performs paid work.

7 Standard employment or standard work refers to a full-time, open-ended employment contract with one employer. Employees perform their services under the subordination of the employer, but have no direct contractual relationship with the client (Kilhoffer *et al.*, 2020).

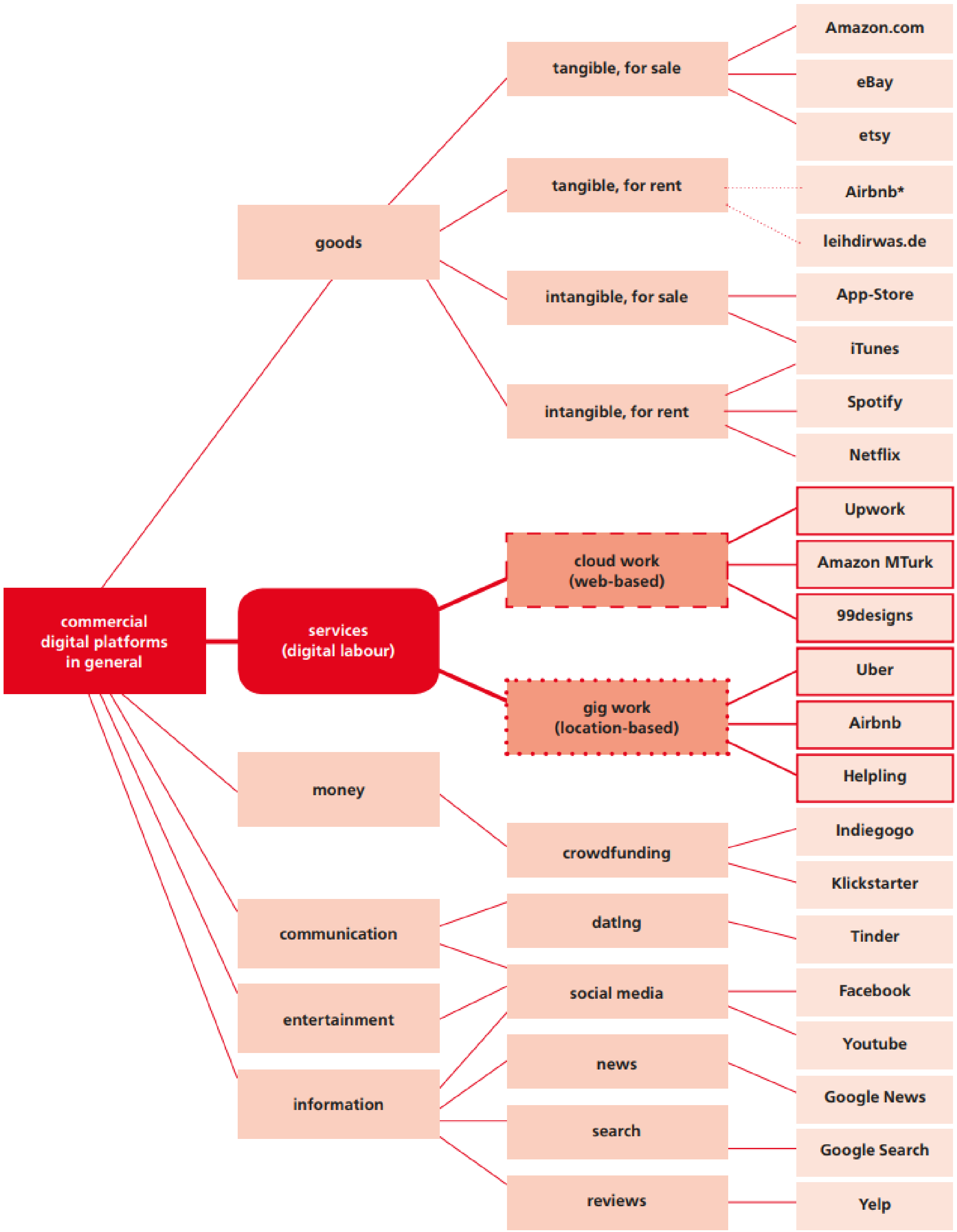
Figure 1. Platform work and standard work conceptualisations



Note: A platform worker may have an employment contract (as in certain platforms like Hilfr and Foodora), but overwhelmingly do not, and instead act as independent contractors, solo self-employed, or similar arrangements depending on national legal context (Kilhoffer *et al.*, 2020).
Source Author's own elaboration, adapted from Kilhoffer *et al.* (2020)

With that, we can explore the messier question of **what does and does not qualify as platform work**. In principle, any type of labour can be intermediated through an online platform. However, certain tasks have taken off more quickly and at larger scale. These include a range of **tasks performed on-location**, such as leisure and hospitality services, personal transport (taxi) services, delivery of goods, delivery of food from restaurants, and a variety of services in the home, such as housekeeping, 'handyman' services, and many more. While most platforms specialise in one or a few types of services, some are much broader. For example, ListMinut offers massage, tutoring, and beauty services, alongside cleaning, furniture assembly, gardening, etc.

Figure 2. Schmidt's categorisation of digital labour markets



Important factors across all platform types are the emergence of monopolies, network effects, biased terms of service, lack of transparency, permanent tracking and rating of user behaviour, and lack of data protection – all of which have problematic consequences for digital labour platforms in particular.

* Because of its many structural similarities Airbnb is treated here as part of gig work, even though the role of labour is secondary on this particular platform.

Source Schmidt (2017: p. 6)

The **tasks performed online** are similarly diverse. These include clerical tasks (customer services, data entry, transcription), online professional services (accounting, legal, project management), creative and multimedia work (animation, graphic design, photo editing), sales and marketing support (lead generation, social media management, search engine optimisation), software development and technology work (data science, game development, app development), writing and translation (article writing, copywriting, proofreading, translation), micro tasks (object classification, tagging, content review, website feedback), and interactive services (lessons and consultations).[8] Some platforms are specialised towards a subset of these tasks, such as 99designs for creative multimedia work, and Mighty AI for training autonomous vehicles through micro tasks. Other platforms offer a huge range. For example, Upwork includes listing for platform workers offering essentially any online task.

Many authors have created **typologies of platform work**. Some are primarily concerned with the types of task performed, such as the eleven types of tasks identified by the COLLEEM survey (Brancati *et al.*, 2019; Pesole *et al.*, 2018). Others use factors like where the task is performed (on-location or online), and the way clients select a platform worker (a specific worker they select, a 'cloud' of potential workers who make offers), and other criteria (Schmidt, 2017; Eurofound, 2018). While these typologies are useful exercises and have some importance in policy debates, they are not particularly important for the purpose of this report. What matters for us is not categorising types of platform work, rather determining if a given type of work counts as platform work. We have already mentioned a number of OLPs above, and Appendix 1 contains a table with all in this report.

Capital versus labour

A few cases at the margins are also worth specific consideration. Taking AirBnB for example, some studies consider AirBnB hosts to be platform workers (Schmidt, 2017), whereas others consider AirBnB a *capital* and not *labour* platform, and therefore outside the realm of platform work (Bogliacino *et al.*, 2020). Those arguing hosts are not platform workers highlight that clients are primarily paying to rent a piece of property, rather than receive a service.[9] I find this unconvincing for several reasons. First, AirBnB hosts must work to procure and prepare property for rental, or oversee those they pay to do so. Second, hosts either perform (or hire out) services like cleaning required to maintain their property. Third, hosts often interact with their guests, providing advice about their stay, and even guiding them in tours that AirBnB styles as 'experiences'.[10]

To summarise this argument, the requirement for capital investment does not impact whether or not services are involved. The prototypical OLP, Uber, certainly requires a large capital investment, but few would argue that it does not intermediate labour.[11] Similarly, if managing a hotel or rental properties qualifies as labour, then certainly AirBnB hosts also perform labour, and clients on AirBnB are purchasing services as well as renting a piece of real estate. Thus, **if services are an integral part of what a platform offers, those providing the services ought to be considered platform workers, even if the platform does not clearly qualify as an OLP**.

Goods versus services

Consider also people who sell goods on e-commerce marketplaces like eBay, Etsy, and Amazon. Generally, such people would not be considered platform workers because they are selling goods rather than labour. However, the distinction is not as simple as it might first appear. Consider a digital artist selling custom designs on Upwork, compared with a carpenter selling custom furniture on Etsy. The former would almost always be considered a platform worker because they seem to be selling a service rather than a tangible good. However, the digital artist can take time to make a design (performing labour) without being paid; the client is actually interested in paying for a completed good, which is the digital image.

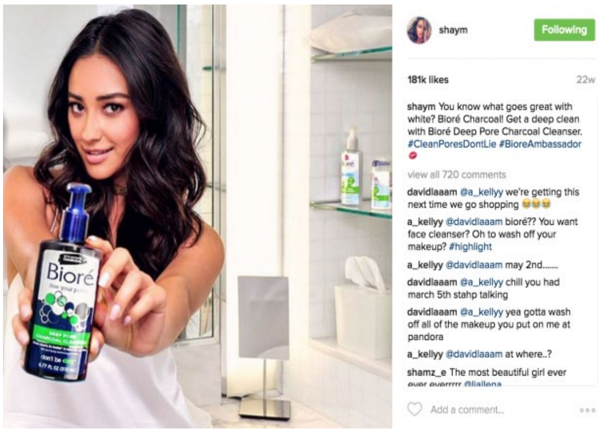
The primary difference between the digital artist on Upwork and the carpenter on Etsy is that the former produces a digital good, while the latter produces a physical good. Otherwise, both individuals use their labour to produce goods and sell them on a digital platform. Both use a platform to

8 Adapted from Brancati *et al.* (2019).
9 This is the argument in Bogliacino *et al.*, for example (2020).
10 AirBnB bills these as 'one-of-a-kind activities hosted by experts', and can include essentially any type of leisure activity.
11 Although Uber has consistently argued that it merely provides digital intermediation, courts often see otherwise (Kilhoffer *et al.*, 2020).

communicate with a client, understand their needs, and work to achieve some specific outcome in exchange for payment. The conclusion must be either: (1) the digital artist on Upwork is not a platform worker, because their product is a mixture of service and good rather than uniquely a service; or (2) the carpenter on Etsy is a platform worker because the good they sell is intrinsically tied to the provision of service. I find the first conclusion unsatisfactory because it would imply that many people using OLPs are not platform workers because they are producing digital goods rather than digital services. If we knew there were 1,000,000 active Upworkers completing an average of one contract per week, we would be no closer to saying anything about platform work. In reality, the distinction between goods and services is difficult to determine – especially in digital marketplaces. I therefore propose that, **although Amazon, Etsy, and similar platforms may be better called e-commerce platforms than OLPs, people earning money through them can rightly be considered platform workers**.

Content creators and influencers

One internet phenomenon has received little attention in the platform work literature: content production on platforms like YouTube, Twitch, and Instagram. On YouTube, content creators produce videos and/or stream from a live feed. While the majority of accounts on YouTube do not earn money, some YouTubers have found a great deal of success earning money from Google Adsense, affiliate links, merchandising, and donations and subscriptions from third party memberships like Patreon. With Twitch, the focus is on livestreaming content; usually a streamer plays videogames while chatting live with their viewers. Twitch streamers earn money through sponsorships, and from their viewers via one-off donations and various subscription tiers. On social media, the term *influencer* refers to internet celebrities who have amassed a large following and often monetised their online presence. Influencers on Instagram do not earn any money from the platform or through the platform directly. However, through sponsored postings and other marketing strategies, some influencers earn a successful living completely from posting on the platform.



Content creators and influencers further illustrate the issue with a sharp delineation between goods and services, platforms and OLPs. While YouTube, Twitch, and Instagram do not primarily function as OLPs, they can clearly be used as such. Content creators and influencers perform labour for pay using platforms, and as such can be considered a type of platform worker. Furthermore, initiatives like the YouTubers Union illustrate that many of the concerns over payment and working conditions are shared for platform workers and content creators.[12]

Teleworkers, homeworkers, and internal crowdsourcing

The final consideration in this section is more traditional workers who may use digital platforms in the course of their work. Teleworking, homeworking, and internal crowdsourcing are not examples of platform work. However, they do illustrate how platform work is influencing traditional work.

12 IG Metall, Europe's largest trade union, has been involved in efforts to unionise platform workers as well as YouTubers. While the YouTubers Union is not technically a trade union, it is a grassroots campaign to achieve more fair conditions for YouTubers. The group is demanding greater transparency, a more stable and equitable system of monetisation, and the ability to negotiate with the platform on decisions that impact their livelihoods (Grey Ellis, 2019).

Telework refers to (own emphasis):[13]

> *a form of organising and/or performing work, using information technology,* **in the context of an employment contract/relationship***, where work, which could also be performed at the employer's premises, is carried out away from those premises on a regular basis.*

Telework cannot be a form of platform work because: (1) it is a form of standard work, and; (2) does not imply any algorithmic intermediation (as per Figure 1).

The legal definition of homework is not consistent across the EU countries, but ILO Convention 177 defines it as (own emphasis):[14]

> *work carried out by a person (i) in his or her home or in other premises of his or her choice, other than the workplace of the employer; (ii) for remuneration; (iii) which results in a product or service as specified by the employer, irrespective of who provides the equipment, materials or other inputs used,* **unless this person has the degree of autonomy and of economic independence necessary to be considered an independent worker** *under national laws, regulations or court decisions.*

Similar to telework, homework requires a traditional employer with a fixed premises. Furthermore, it implies direction from an employer, rather than algorithmic intermediation between a worker and client.

Lastly, internal crowdsourcing is an interesting case. Normally, crowdsourcing means that a client posts an open call on the internet, where a large group of people can contribute to some goal. It usually occurs in an *external* context, meaning companies (or individuals) outsource certain tasks to independent contractors via an OLP (Leimeister *et al.*, 2009).

IBM and other companies used a digital mechanism to distribute work within their organisation. A company would post a task, and employees could compete amongst each other, offering different solutions. The principle is exactly the same as crowdsourcing, except that the crowd consists of employees within a company rather than independent contractors. As such, internal crowdsourcing is a novel way to distribute work within the context of a standard employment relationship, but does not constitute platform work.

Concluding thoughts on terminology

It is not always clear where platform work begins and ends. This is to be expected, given that the platform economy is new and quickly changing. The lines will continue to blur as platform work models influence traditional work, and algorithms, ratings, and reviews become a more normalised part of standard employment.

Because of this, prudence suggests that researchers and policymakers ought not take an overly narrow view of platform work. The determining factor in many cases is whether an intermediary, typically an algorithm, makes decisions to connect the supply and demand for paid work. When this is the case, many foundational concerns related to fairness and working conditions are essentially the same, whether the platform is Uber, Upwork, YouTube, or anything else.

13 Article 2 of the European Framework Agreement on Telework of 2002, cited from Eurofound.

14 Cited from Eurofound.

3. Overview of data types

Most labour market data come from national labour surveys and administrative reporting. However, data on platform work are not or cannot be gathered in most of the ways that other forms of work can be (Kilhoffer *et al.*, 2020). Official labour market statistics are therefore inadequate to say very much about platform work (Riso, 2019).[15]

In spite of difficulties in access, many researchers have found ways to gather data and statistics on the platform economy and used them in novel ways. Many have compensated for data shortcomings with mixed method approaches.[16] This section presents a brief overview of the most-used data types in the platform work literature, as well as considerations for researchers.

3.1 Survey data

Survey data is gathered from a sample of respondents who take a survey, typically through phone or internet. Official labour market data are collected periodically through national or EU-level labour force surveys. Because surveys are the traditional source of high-quality labour market data, researchers have extensively deployed surveys in an attempt to understand platform work.

However, **using survey data to assess platform work presents a number of challenges**. First, platform workers are a relatively small demographic, and reaching sufficient numbers for a representative sample can be difficult. For example, a number of surveys on platform work in Germany have relatively small sample sizes, which are unlikely to provide representative results (Serfling, 2018). One common way of increasing sample size is to pay participants, but this creates a coverage bias, as: (1) people completing surveys online have access to and familiarity with computers and the internet more than the general population, and are therefore more likely to do platform work; and (2) completing surveys for pay is itself a form of platform work.

Second, **'platform worker' is an ill-defined term** to begin with, and surveys may struggle to communicate the idea to participants. Rather than asking participants questions like 'Are you a platform worker?', or 'Have you done work via digital platforms?', which are likely to be misunderstood, many surveys identify certain dimensions of platform work and allow cross-tabulation of variables (Huws *et al.*, 2017).[17] Still, this introduces new challenges in how to cross-tabulate data for a given research question (i.e. how many platform workers are there?), and measurement bias can be a significant problem. The mode of the survey impacts the accuracy of responses, which is especially important when surveys are complex, and for potentially comparing different survey results (Riso, 2019).

Third, and similar to the former point, **studies use different conceptualisations of platform work** than the present report. Some focus on only a few times of labour, such as micro-tasking (Berg, 2016), a specific platform like Uber (Hall & Krueger, 2018; ORB International, 2017), while others look to any type of work with electronic mediation (US Bureau of Labor Statistics, 2018). Again, the conceptual complexity of platform work impacts the data quality.

15 However, note that relevant ministries in number of countries including the Canada, Denmark, Finland, France, the UK, and the US have added questions relevant to platform work to labour force surveys, or released dedicated ad-hoc modules.

16 Mixed methods use a combination of other methodologies to help verify findings and mitigate shortcomings. In studying platform work, researchers might combine quantitative methods with surveys, administrative data, and big data, and qualitative methodologies such as focus groups and in-depth interviews.

17 The variables identify dimensions key to a typology.

Fourth, many studies provide little information about the extent to which they cover the target population (Riso, 2019). This makes it difficult or impossible to know if the results can be generalised to the total population, and if studies are comparable to one another.

Due to these difficulties, the methodologies of surveys require extra careful consideration. Methods such as verifying online results with phone surveys, or triangulating methodologies, are a few potential options, but by no means guarantee a solution to coverage and measurement biases (Riso, 2019).

3.1.1 Official economic statistics

Official economic statistics are understood as those collected directly by national statistical offices or government departments (Riso, 2019). They tend to be easier for researchers to access than other sources.

At EU level, Eurostat has increased efforts to gather data on the platform economy, such as through the 2017 Community survey on ICT usage in households and by individuals. While not aimed at platform work per se, it did collect data on the share of people in the EU arranging accommodation and personal transport online via websites or apps from another individual.[18]

At national level, a handful of European countries have used surveys to gather data relevant to platform work. For example, in 2017 the UK introduced a module to its Opinions and Lifestyle Survey, which is conducted using computer-assisted telephone interviews. The module includes questions about the use of intermediary websites or apps to arrange accommodation or transport, and using the internet for social networking and online shopping (Office for National Statistics, 2017). Statistics Netherlands also began gathering data, asking individuals about ICT-usage including AirBnB and Uber-like platforms (Heerschap *et al.*, 2018). Statistics Netherlands also surveys enterprises on the perceived impact of online platforms on their turnover.

In Finland and Denmark, the national statistical agencies sampled nationals in 2017 to estimate how many people earn money from digital platforms. For both countries, the modules asked respondents about specific platforms, and therefore do not represent all platform workers (Riso, 2019). In France, the labour force survey of 2017 asked self-employed respondents whether they accessed clients exclusively or not through a digital intermediary or more traditional business intermediary (Arnault *et al.*, 2018). However, the question was formulated such that 'digital platform' is open to interpretation.

3.1.2 Unofficial economic statistics

Unofficial statistics are those commissioned to third parties or carried out by and on the initiative of private organisations. Absent official statistics on platform work, more of these are available at EU level and for individual Member States, but the data is not necessarily available for public use.

A notable early example is the Eurobarometer Flash Survey 438, which investigated the (now dated terms) 'collaborative economy' and 'collaborative platforms'. The survey asked about renting accommodation, car sharing, and small household jobs, alongside usage of search engine, social networks, and online marketplaces (European Commission, 2016). Additional surveys covering the EU or subsets for different types of platform work include that by VVA, Milieu and GFK for DG Just (Hausemer *et al.*, 2017), PWC (2018), and McKinsey Global Institute (2016). The ILO further commissioned a survey of platform workers who use CrowdFlower and Amazon Mechanical Turk (Berg, 2016). Finally, Huws *et al.* (2017) and Pesole *et al.* (2018) both conducted ambitious surveys spanning a number of EU countries. The latter two reports are discussed in more detail in the section Divergence in size estimates of platform economy.

18 See here.

At national level, the UK, France, Norway and Germany have all commissioned research towards understanding platform work, though the UK and Germany seem to be the most active due to higher scrutiny about labour platforms.

Looking to Norway, an initial internet survey sampling 1,525 Norwegian adults found that some 10% of respondents indicated they had done platform work at some point, and 2% on a weekly basis (Jesnes *et al.*, 2016). However, this study had significant error margins. A subsequent survey carried out by phone, with more carefully worded questions, resulted in more conservative estimates (Alsos *et al.*, 2017). This indicates further evidence that we should treat internet surveys with caution, and the wording of questions has large implications for estimations.

Community initiatives have also contributed to our understanding of platform work, primarily in to gauge working conditions and fairness in OLPs. For example, FairCrowdWork is a sort of watchdog organisation run by German trade union IG Metall, a number of social partners, and partner platforms. Among other activities, FairCrowdWork collected and compiled data from platform workers about the OLPs they have used. It then created a rating overview of OLPs based on the fairness and desirability of working through them. Additionally, FairCrowdWork conducted a 2016 survey of workers on six German platforms, carried out in collaboration with the platform operators, which revealed that workers found fair payment by far the most important principle. Therefore, this principle was 'clarified and strengthened as much as possible given the platforms' current business models and economic circumstances' in the current Code of Conduct, which a number of platforms abide by.[19]

Recommended readings using survey data include:
- Huws *et al.* (2017);
- Pesole *et al.* (2018);
- Brancati *et al.* (2019).

3.2 Administrative and big data

Administrative and big data represent a diverse assortment of data sources. They have been important resources for platform work research due to the challenges with surveys and traditional labour data. Sources that have been used in the context of platform work include bank transactions (Harris & Krueger, 2015), API calls or webscraping the platforms themselves (Kässi & Lehdonvirta, 2018), and Google trends data (De Groen *et al.*, 2017).

Administrative data are collected by governments or other organisations, primarily not for research purposes, but to provide overviews on registration, transactions and record keeping. Administrative data are very diverse, accounting for usership data from platforms, bank transfer data, tax records, and much more.

Administrative data may have certain advantages, such as avoiding measurement biases found in many surveys. On the other hand, administrative data on platform work is often lacking. Sources like tax statements may not accurately reflect earnings from platform work, which often go unreported. Furthermore, administrative data are rarely available to the general public, limiting their use to those with insider access. Platforms regard proprietary data as an important part of their competitive advantage and tend to keep them confidential (Fabo *et al.*, 2017).

Big data refer to extremely large data sets whose size precludes traditional processing, but can be mined for information. In many cases this requires newly-developed techniques in machine learning. Due to the digital nature of platform work, big data is a promising resource to understand the

19 See here.

platform economy. At present, national statistical offices in Canada, the UK, and Italy are using or considering big data to study platform work (Riso, 2019).

However, big data can suffer from a number of shortcomings. Crawford cautions that big data have intrinsic biases because 'data and data sets are not objective; they are creations of human design' (2013: p. 1). Big data are often low quality, necessitating extensive quality checks prior to serious analysis (Dasgupta, 2013). Furthermore, and similar to surveys, big data frequently suffer from poor representativeness, which makes it difficult for researchers to generalise findings (Lenaerts *et al.*, 2016). Big data are not necessarily reliable and comparable over time, and lastly, big data is subject to regulatory constraints (Boyd & Crawford, 2012). The extent of these challenges very much depends on the particular data, and what we are asking of the data.

Big data can be acquired in several ways, primarily via application programming interfaces (APIs), webscraping, webcrawling, and voluntary data sharing. **APIs** are common on many types of websites and facilitate functions like retrieving data and posting data. APIs allow structured requests for data, which is usually returned in a format like JSON for easy analysis (Dewi *et al.*, 2019). While some APIs do not require registering an account, most public APIs require an access key acquired after submitting an application and agreeing to abide by the website's terms of services (TOS).[20] This means that, provided researchers are approved and follow the TOS, they do not risk legal challenges with the website. Sometimes these conditions are problematic for researchers, as APIs may explicitly forbid, for example, creating a database of retrieved data. While some APIs are completely free, websites also monetise them by charging for API access. Often the free versions of APIs have restrictive rate limiting, which means retrieving a large volume of data requires spacing out requests over time. This can be problematic for researchers interested in acquiring comprehensive data or analysing trends over time. Depending on the volume of data needed, Russel and Klassen recommend considering third party data vendors (2019). These can be expensive, but may be the only practical way to acquire certain bulk data (Kilhoffer, 2020). Key benefits with using APIs include ease of use and the possibility of acquiring real-time data.

Webscraping is a means of gathering data from websites, typically making use of automated tools developed with programming languages to extract data from webpages (Dewi *et al.*, 2019). Some websites are very open to be scraped, as it can help them increase their traffic. For example, search engines rely on webscraping or webcrawling to index the internet and show relevant results for search queries. Many websites, however, explicitly forbid webscraping. This creates ethical (e.g. privacy) and legal obstacles for researchers (vanden Broucke & Baesens, 2018). In principle, anything that can be viewed online can be webscraped using a variety of techniques. Unstructured text, structured data, images, and anything else contained on a website can be retrieved (Kilhoffer, 2020). Webscraping can be considered something of a last resort to gather data when other methods are not feasible.

Webcrawling is similar to webscraping, however it systematically browses one or more websites, typically for the purpose of indexing. Whereas webscraping tends to retrieve pre-defined data from a specific website, a webcrawler will recursively browse a site, any hyperlinks it finds, and index (store in a database) all that it finds. Karanovic *et al.* used webcrawlers to collect over 120,000 blog posts from uberpeople.net - the most popular forum for Uber drivers - and assess the content of the unstructured text with natural language processing techniques (2017). Numerous studies have used webcrawlers to study AirBnB rental markets.[21]

Voluntary data sharing is when internet services collaborate with external researchers, providing data on a limited basis, free of charge, for specific research projects (Kilhoffer, 2020). This type of

20 Essentially a string functioning as a password.

21 See for example Edelman and Luca (2014), Teubner *et al.* (2017), and Zervas *et al.* (2015), which webcrawled AirBnB and analysed price, rating, etc.

data acquisition is generally unproblematic legally, though it is limited to what internet services are willing to share and for what purposes.

3.2.1 Government sources

Governments have access to certain administrative data, and a number of efforts to gather and analyse administrative data are underway. These are often closely linked to tax records, as governments have been concerned about the potential for undeclared earnings from platform work (Lenaerts *et al.*, 2017).

In France, since the introduction of the Finance Bill 2016, all platforms are required to provide an annual earnings statement to platform workers in order to facilitate their tax returns. In Estonia, the tax authorities are working with OLPs to develop a system whereby tax is withheld directly via the platform to facilitate income tax collection. For example, Uber drivers are allowed to opt-in to a system where Uber sends drivers' income data to the tax office, so it is automatically added to their tax return. Furthermore, since 2016 Belgium has had a favourable tax regime for platform workers, but only for selected platforms. In order for platform workers to benefit from this tax regime, platforms must require certain basic information data (i.e. proof of no bankruptcy) (Eurofound, 2018).

In the UK, the ONS is considering analysis on tax returns (Beck *et al.*, 2017), while in Norway, the Fafo Institute for Applied Social Science collected tax data from the Norwegian Tax Administration about UberPop drivers (Alsos *et al.*, 2018). Finally, in Italy, the Italian National Institute for Public Policy Analysis INAPP collected economic and employment information on platform businesses from business registers and administrative data sources (Guarascio & Sacchi, 2018).

One takeaway is that governments can cooperate with OLPs to collect data on taxable earnings. This can help mitigate undeclared earnings from platform work, ease administrative burdens for platform workers, and generate data of use to social scientists. Depending on the arrangement, cooperation between governments and OLPs may also allow external researchers access to useful data, but no examples of this were found.

3.2.2 Platform sources

Platforms often release certain administrative data describing the number of services provided in a certain timeframe, workers signed up, etc. This is often for promotional purposes. Additionally, publicly listed platforms are typically required to release detailed information on revenues, profits, and so on in their country of registration.

Platforms are certainly well placed to provide data on their users, however, in many cases these data must be treated with scepticism. As noted in Fabo *et al.* (2017), platforms have an incentive to misrepresent the number of platform workers to appear as large and relevant as possible, thereby attracting more workers, customers, and investment. Furthermore, platforms often provide figures of limited value, such as the total number of individuals subscribed as service providers, which is much larger and much less relevant than the number of individuals consistently working through a platform.

In a number of cases, platforms have published or commissioned studies based on their internal administrative data and surveys. For example, Etsy released an economic impact study built on an online interview of 2,658 Etsy sellers from 24 April to 14 May 2018, primarily answering how the platform's sellers contribute to the US economy. Etsy indicated that sellers contribute $4.7 billion annually to the US economy, including $1 billion for themselves, and $850 million in wages and income for US workers in other segments of the economy.[22] Additional findings confirmed Etsy's

22 See Guta (2018). The link to Etsy's actual report is broken as of 28 October 2020.

outsized role for women, who represented some 86% of sellers on the platform in 2015 (Fortune, 2015). Unfortunately, the report seems to have been taken offline, so it is difficult to say more about the methodology and findings.

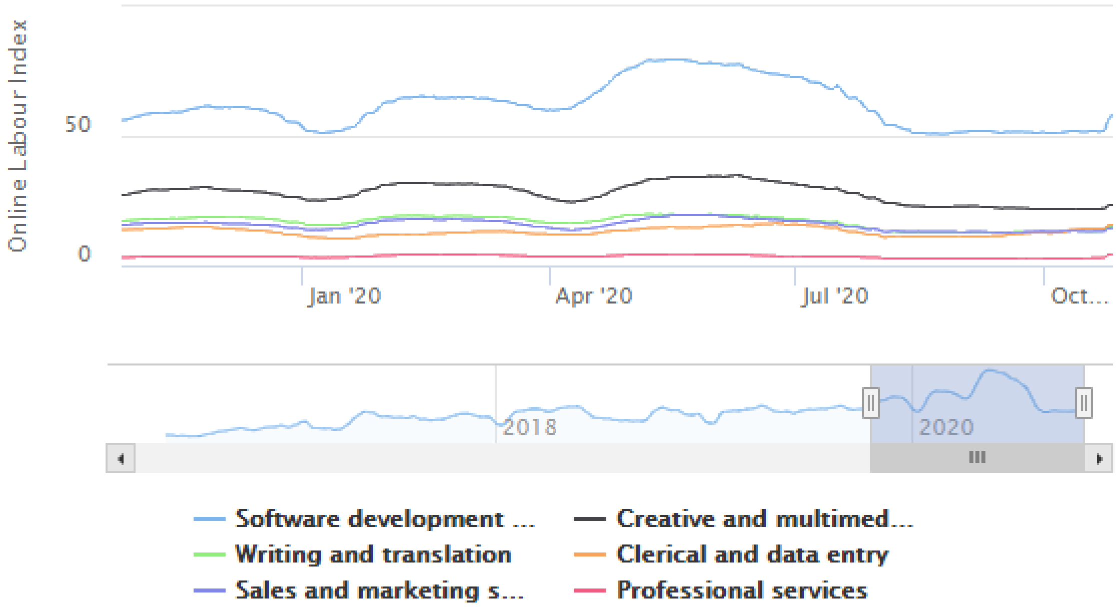
One influential paper is based on a survey commissioned by Uber combined with administrative data (Hall & Krueger, 2018), which has appeared in US Congressional committee hearings, a Federal Trade Commission's workshop on the sharing economy, and other policy venues (Berg & Johnston, 2019). The report paints Uber in a flattering light, finding that 81% of Uber drivers are very satisfied or somewhat satisfied, and earn from $18.75-19.41 per hour: far higher than the average for taxi drivers and chauffeurs. It also estimates that Uber drivers should safely earn more than taxi drivers and chauffeurs even accounting for expenses. However, Berg and Johnston have strongly criticised the results (2019). Problems included non-response bias (10% response rate), missing questions (average number of hours driven for Uber per week), posing leading questions, improper scaling of job satisfaction questions, misinterpretation of job satisfaction results, and so on.

Other platforms, such as BlaBlaCar and AirBnB, have published reports with different research focuses from revenues to environmental impacts. However, methodologies are often unclear, and most of the time data is not available for reproduction (Riso, 2019). The takeaway should be that reports commissioned by the platforms themselves are often undertaken for promotional purposes and should be viewed with scepticism.

One particularly good use of data from platform sources is from the Oxford Internet Institute, which set up the Online Labour Index (OLI). The OLI tracks all tasks posted to the five largest English-language OLPs,[23] representing at least 60% of the market by traffic. The Oxford Internet Institute has reached an agreement with these platforms to gather data through API calls or periodically webcrawling vacancies. The result is an easy-to-use tool to view supply of tasks by country, time period, type of occupation, and growth trends (Kässi & Lehdonvirta, 2018). For example, the OLI generally indicates overall growth around 25% annually, however Figure 3 shows a slump in demand associated with the COVID-19 pandemic. The OLI illustrates the value of big data provided in near real-time.

23 As indicated by unique monthly visitor estimates from Alexa.

Figure 3. Screenshot from OLI - Demand for online labour by occupation



Note: 28-day moving average used. Dates shown are 25 October 2019 – 26 October 2020. Screenshot made 26 October 2020.
Source OLI

3.2.3 Other administrative and big data sources

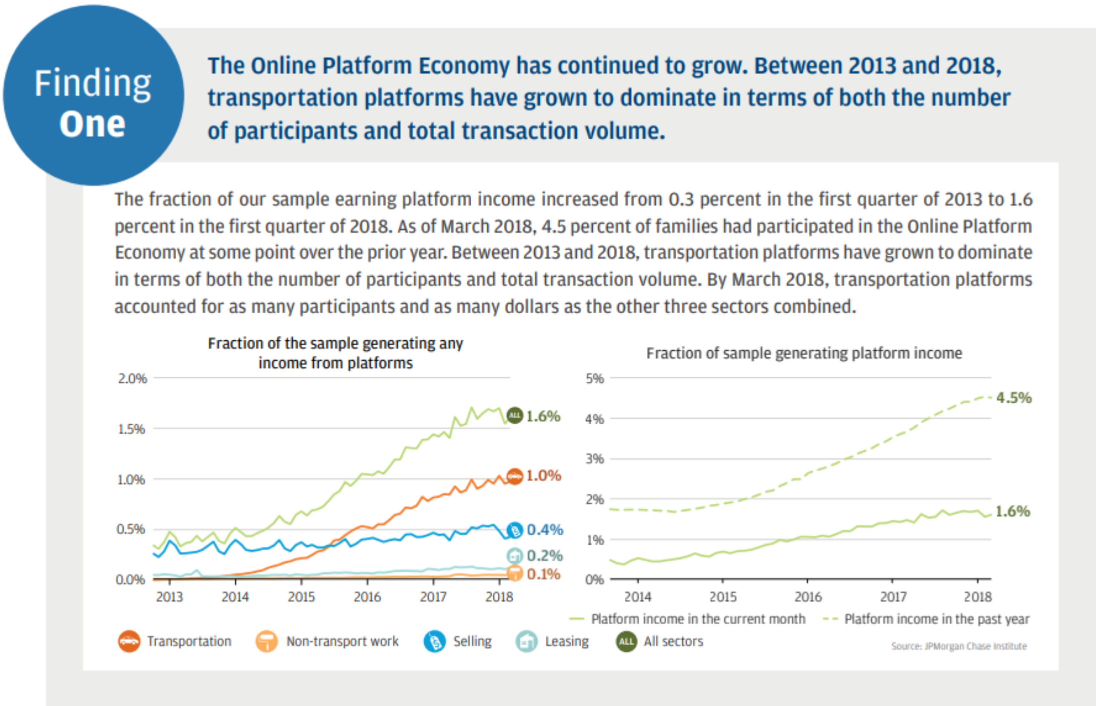
Additional administrative and big data comes from sources like financial institutions. For example, a few notable papers used proprietary data on JP Morgan Chase's American customers' bank account transactions. These studies attempt to estimate participation in the platform economy, defined as labour platforms (like Uber) and capital platforms (like AirBnB). The earlier of these reports (Farrell and Greig, 2016) estimates participation rates based on only 30 platforms, and the data are skewed towards older bank account holders. In a follow-up, Farrell *et al.* (2018) used a new sampling method[24] on 39 million anonymised bank accounts over a 66-month period from October 2012 to March 2018, and 128 platforms in further disaggregated sectors.[25] Riso considers these results robust, reliable, and repeatable,[26] but notes that such analysis is only possible for banks which own the data (2019).

24 In the earlier study (Farrell and Greig, 2016), checking accounts were only analysed if they existed for all 36 months between October 2012 and September 2015, and had at least five outflow transactions per month. This dropped the sample from 28 million checking account customers to 6 million. In the latter study (Farrell *et al.*, 2018), the unit of analysis was the 'account-month', where the only inclusion criterion was at least five outflows from October 2012 to March 2018. This allowed analysis of more accounts, which is important because customers open, close, and modify the accounts they use.

25 The four sectors were transportation, non-transport work (essentially on-location platform work excluding transport), selling (essentially e-commerce), and leasing (renting homes, parking spaces, or other assets).

26 For example, Farell *et al.*'s (2018) estimates are extremely close to those from the US Bureau of Labor Statistics (National Academies of Sciences and Medicine, 2020).

Figure 4. Growth findings from big data (39 million US bank accounts)



Source Farrell *et al.* (2018: p. 3)

A different approach was used in De Groen *et al.* (2017), combining the JRC dataset on platforms, which contains gross revenues and the number of active workers for some platforms. The 173 platforms meeting their definition of 'crowd employment' were used for analysis, and these data were interpolated to scale revenues from the global to the EU level, assuming that platforms generate the same gross revenues per unique visitor across the globe. The authors then enriched the JRC database with data on the number of unique visitors to a website and location of visitors from Amazon's Alexa, which served as a proxy for the amount of activity on a platform. The results estimated 12.8 million active platform workers in the EU for 2016, or 5.9% of total employment (ages 15-62). This is similar to Eurobarometer (European Commission, 2016), but still likely inflated because of the reliance on self-reporting by platforms. For example, one platform reported nearly 3 million registered platform workers, which is incredible given the site had fewer than 1 million unique visitors for the surveyed month.

A final initiative to gather big data comes from Hara *et al.* (2018), who developed a browser plugin for platform workers on AMT ('Turkers') to use on an a voluntary, opt-in basis. The plugin was designed to gather descriptive data[27] at the task level, allowing analysis of task ('HIT') duration, HIT reward, and effective hourly wages Turkers earned. This resulted in logged data of 2,676 Turkers who performed 3.8 million tasks. The findings highlight the issue of low pay in platform work, as only 4% of Turkers earn more than $7.25 per hour, while the median hourly wage is just $1.77.

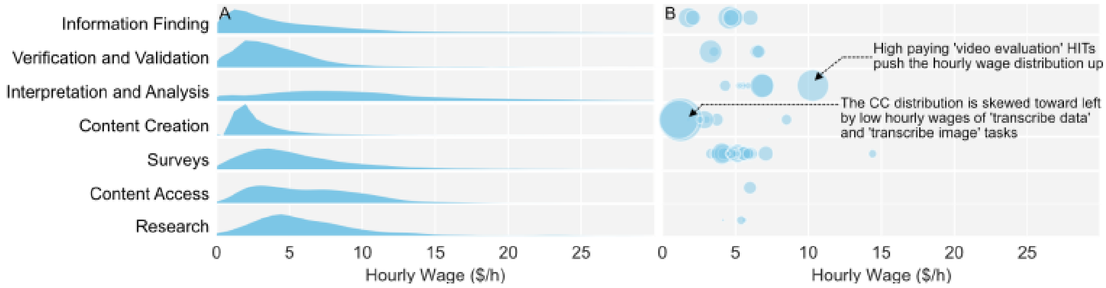
While this plugin strategy is very appealing, it still has potential weaknesses. First, the authors concede that it is difficult to define unpaid work given that Turkers (and other platform workers) often spend much of their time searching for paid tasks, which itself is not compensated. In fact, Hara *et al.* (2018) identify three types of unpaid work: searching for tasks, working on tasks that are rejected, and working on tasks that are not submitted. Second, the sample may suffer from self-selection bias, as the most productive 'super turkers' may not have been interested in using the plugin.

27 For example, when workers accept, submit, and return a task; reward (payment); and other metadata about HITs.

Third, accessing such data is not straightforward. Researchers must develop a plugin, then ensure a large enough number of platform workers actually use it.

Nevertheless, Hara *et al.*'s (Ibid.) report allows for a more accurate and granular analysis of working time than any other methods discussed in the present report. For example, the authors calculated the expected earnings from different types of tasks that Turkers commonly perform, as shown in Figure 5.

Figure 5. Hourly wage distribution of tasks



Note: A represents hourly wage distributions of seven HIT categories provided by Gadiraju *et al.* (2014) with an additional category Research, while B shows strip plots with the median hourly wages of HITS.

Source Hara *et al.* (2018)

Recommended readings using administrative and big data include:

- Farell *et al.* (2018);
- Hara *et al.* (2018);
- Kässi and Lehdonvirta (2018).

4. Divergence in size estimates of platform economy

A number of efforts have attempted to assess the number of total or active platform workers, amount of revenue earned by platform workers, the economic value added of platform work, or some combination of these. These estimations have mostly relied on one or a combination of three sources: surveys, administrative data, and big data.

Available literature shows broad consensus about certain ideas: (1) the platform economy represents a fairly small portion of the overall economy, both in terms of revenues and workers; (2) the platform economy is less developed in Europe than in the US; (3) the platform economy is rapidly growing; and (4) a few 'giants' such as Uber and Airbnb comprise a very large portion of the platform economy (De Groen *et al.*, 2017).

Beyond these similarities, size estimates vary greatly due to theoretical and methodological challenges particular to platform work. One of the most contested and policy-relevant questions is the share of platform workers out of the population or labour force. Table 1 presents an overview of estimates of the share of platform workers in select EU Member States.

Table 1. Share of platform workers in selected EU Member States

Country	%	Scope	Source
Croatia, Finland, France, Germany, Hungary, Italy, Lithuania, Netherlands, Portugal, Romania, Slovenia, Spain, Sweden, UK	2%	European working-age population (aged 16–74) engaged in platform work as a main job	Pesole et al, 2018
	6%	European working-age population (aged 16–74) for whom platform work generates a significant income (25% of average wage for a standard working week of 40 hours)	
	8%	European working-age population (aged 16–74) who perform tasks through digital platforms at least once a month	
Austria	19%	Population who has ever done platform work	Huws et al, 2017
Denmark	1%	Platform workers earning money at least once over the last 12 months	Ilsøe and Madsen, 2017
Finland	0.3%	Population aged 15–74 who had earned more than 25% of their income from work-related and non-work-related platform activities during the previous 12 months	Statistics Finland, 2017
Germany	1.8%	Population aged 15 and over registered as platform workers in January 2017	Mrass and Peters, 2017
	12%	Population who has ever done platform work	Huws et al, 2017
Italy	22%	Population who has ever done platform work	Huws et al, 2017
Netherlands	9%	Population who has ever done platform work	Huws et al, 2017
Sweden	2.5%	Working-age population who performed platform work	SOU, 2017
	4.5%	Working-age population who looked for work on platforms	
	10%	Population who has ever done platform work	Huws et al, 2017
United Kingdom	4%	Gig economy, performing tasks online, providing transport or delivering food or other goods at least once over the past 12 months	CIPD, 2017
	4.4%	Gig economy, involving exchange of labour for money between individuals or companies via digital platforms that actively facilitate matching between providers and customers, on a short-term and payment-by-task basis	BEIS, 2018
	9%	Population who has ever done platform work	Huws et al, 2017

Note: The estimates shown do not necessarily reflect this report's understanding of platform work.
Source Eurofound (2018: p. 13)

As shown in the first grouping of countries, platform work as a main job is fairly rare, at 2% of the working age population. However, Table 1 shows great variation even among recent and widely-cited studies.

To highlight the differences, take Spain for example. These figures in Table 2 come from the follow-up survey Huws *et al.* (2019), which extended the same methodology of Huws *et al.* (2017) to more countries, and the COLLEEM survey data.

Table 2. Spanish platform work estimates as part of working age population

	Reported Figure	Source
17.0%	At least weekly platform work	Huws *et al.* (2019)
10.5%	Less than weekly platform work	
20.6%	Seeking but not undertaking platform work	
12.5%	Ever platform work	COLLEEM (Pesole *et al.*, 2018; Brancati *et al.*, 2019)
6.3%	Platform work 0-10 hours weekly and <25% income	
2.6%	Platform work 10-20 hours per week and/or >25% income	
2.7%	Platform work 20+ hours per week and/or >50% of income	

Source Author's elaboration based on Huws *et al.* (2019) and COLLEEM data (Pesole *et al.*, 2018; Brancati *et al.*, 2019)

Huws *et al.* found that 17% of the Spanish working-age population earn money from platform work at least weekly, and 10.5% less than weekly. An additional 20.4% were seeking but not currently undertaking platform work. From the COLLEEM data, 12.5% of Spanish working age have ever performed platform work, and only 6.3% at least weekly. Obviously, the COLLEEM estimation is far more modest.

With such similar timeframes, and both studies relying on internet surveys, it is worth discussing the reason for such divergence. Consider the following differences in how the researchers approached the problem.

First, Huws *et al.* used a random sample of the entire working age population by adding extra questions to a standard online omnibus survey,[28] while COLLEEM sampled working age *internet users* in an online panel survey.[29] In the first report publishing the COLLEEM results, Pesole *et al.* argue, 'this is not a bug but a feature of the COLLEEM sample: it would be absurd to sample non-internet users for a study of work on internet platforms' (2018: p. 16). Pesole *et al.* further state that this approach required scaling estimates of internet usage to the general population, for which they use Eurostat ICT Survey data. However, Huws *et al.* do no such scaling. This is interesting because, while Huws *et al.* reported figures for the working age population, it would have been much more accurate to report figures for internet users given the survey mode.

This sampling bias should not be understated, and it is not just that only internet users, by definition, can be platform workers. One common type of platform work is filling out online surveys for payment, which means surveys for pay inherently draw a higher proportion of platform workers than the general population, or even the population with internet. Huws *et al.* (2017) attempted to check for this bias by conducting the survey via phone and face-to-face in two of the seven countries (UK and Switzerland) covered by the online survey. In the UK, the online survey found that 9.3% undertake any crowd work, while the offline survey was lower at 7.38%. The difference is not trivial; the online estimate is 26% larger. The authors also found that 'these survey mode effects may be severe in the Swiss tele- phone-based survey' (Huws *et al.*, 2017: p. 20). Furthermore, we know that the UK has relatively high internet penetration compared to some other sampled countries; in 2014, for example, 90% of UK households had internet access versus 74% in Spain.[30] This suggests that the discrepancy between online and offline surveys may have been small in the UK compared to some other countries. Nevertheless, **Huws *et al.* presented all findings without any adjustments to the (online) survey mode**. This is likely to have biased results significantly upwards, as noted by De Groen *et al.* (2017) and others.

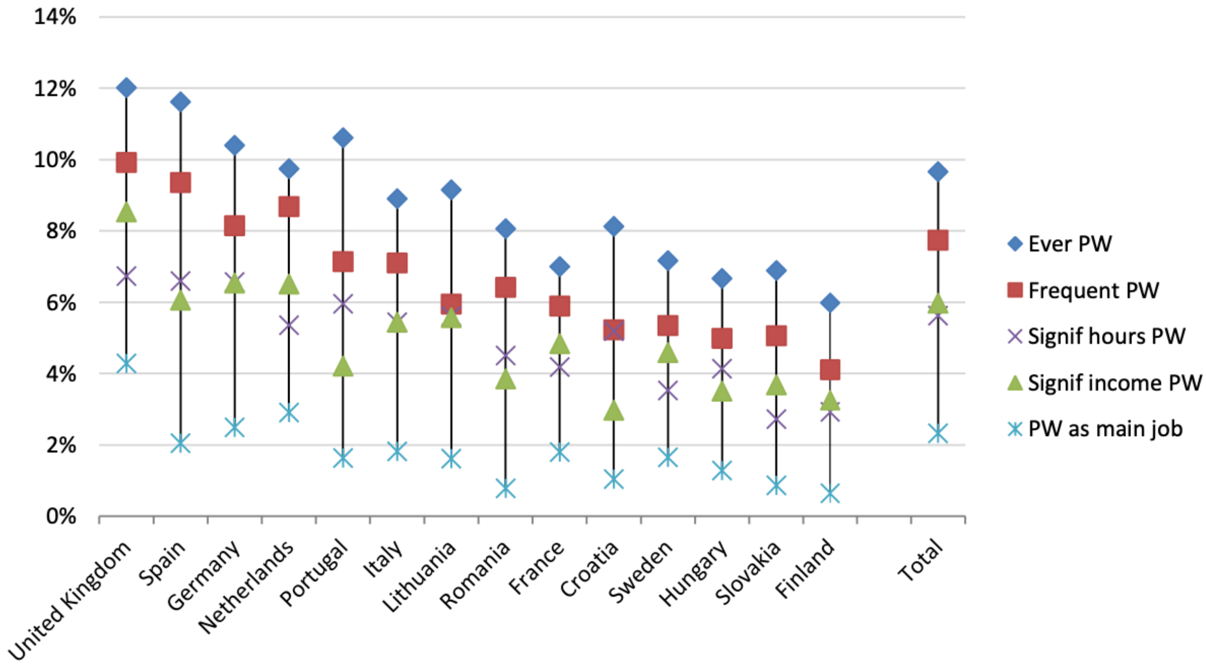
Second, the two studies differently delineate sporadic, occasional, and full-time platform workers. There is no objectively correct way to do this, but it must be done to highlight differences in platform work frequency. Taking all who have performed platform work ever is too broad a measure of platform workers, and not as relevant from a policy standpoint. Huws *et al.* focused on percentage of earnings (>50% or not), whereas Pesole *et al.* asked more questions allowing them to further delineate between frequency, periodicity, and earnings.

28 Specified in Huws *et al.* (2017).

29 The sampling frame was a commercially available list of internet users, with non-probability quota sampling of respondents by gender and age group (Pesole *et al.*, 2018).

30 See Eurostat data here.

Figure 6. Different estimates of platform work (PW) using COLLEEM data



Source Pesole *et al.* (2018: p. 19)

Third, the surveys had subtle but important differences in types of platform work and wording of questions, which impact how respondents qualify or not as platform workers. Huws *et al.* use 'crowd work' defined as paid work via an online platform (2017: p. 16). This includes three categories of platform workers with survey questions worded as follows (UK version) – those who:

1. look for work you can carry out from your own home on a website such as Upwork, Freelancer, Clickworker;
2. look for work you can carry out for different customers somewhere outside your home on a website such as Handy, Taskrabbit or Mybuilder;
3. offer to drive someone to a location for a fee using an app or website such as Uber or Blablacar.

In the first category, some who carries out work (1) *from your own home* and (2) *on a website* would be considered a crowd worker. However, a teleworking web designer (employed by a company) could meet both criteria, though they should not be understood as a platform worker. The question relies on familiarity with the example OLPs given. One nuance of the wording is that platform workers should not be understood to be doing work *for a website* (the wording in the report) or *on a website* (the wording in the survey), but rather *matching with customers through a website*. This is better reflected in the second and third categories. However, the ambiguity is still such that the survey probably resulted in false positives.

Pesole *et al.* attempted to limit this problem by being more specific and using ten categories of tasks. One advantage of this is more granularity in results, since numerous types of tasks can be undertaken on a platform like Upwork. On the other hand, Pesole *et al.* only provide two types of on-location platform work.

These are listed as follows:

1. transportation and delivery services (e.g. driving, food delivery, moving services and similar);
2. on-location services (e.g. housekeeping, beauty services, on-location photography services and similar).

Certain types of platform work performed on location, such as pet-sitting, retail intelligence, and others would seem to fit into the second category, but require more interpretation than the examples explicitly listed. For that reason, on-location platform work may be under-estimated. Indeed, COLLEEM is notable for estimating that online platform workers form the large majority of platform workers, going against the conventional wisdom that Uber drivers are most common (Harris & Krueger, 2015; De Groen *et al.*, 2017).

Final thoughts

Both the survey in Huws *et al.* (2017) and COLLEM (Pesole *et al.*, 2018) are important sources of data available on platform work in Europe. However, the former significantly overestimates the preponderance of platform workers overall compared to other reputable surveys. The latter is much better in this regard for its handling of bias from online surveys. Overall, the COLLEEM survey is the most reliable source of data on platform work in the EU (Kilhoffer *et al.*, 2020).

Comparing the COLLEEM survey (Pesole *et al.*, 2018; Brancati *et al.*, 2019) to Huws *et al.* (2017; 2019) illustrates the difficulty of surveys on platform work. Even when the survey modes are properly handled, the wording of questions remains an important challenge. If the academics rarely agree on what constitutes a platform worker, it is no wonder that survey questions - designed to be brief - are unable to adequately communicate the idea. For this reason, subsequent COLLEEM surveys will continue to refine their method (Brancati *et al.*, 2019). This should result in more robust results, but changing methodologies also make time-series analysis more complex.

5. Conclusions

This report has argued that platform work is increasing in relevance. Given the rate of change in the platform economy, policymakers and researchers ought to take a broad view of what constitutes platform work. Failing to do so will result in new forms of platform work being overlooked, and more risks of precarious work falling in a regulatory grey zone.

Many examples of data and statistics on working conditions are available, covering a broad range of platform work types. However, these data are often publicly unavailable and challenging for researchers to access. Overall, the data suggest both advantages and disadvantages of platform work, such as the benefit of more flexible working times and the risk of inadequate social protection, while showing that platform work remains a small but growing fraction of the labour force.

Measuring platform work remains very difficult, especially when the goal is to generalise about the entire population. Surveys, administrative and big data all struggle in this regard. The best data sources acknowledge this shortcoming and attempt to handle it with methods like triangulation, weighting by population, and others.

Future efforts should continue to leverage surveys, which are especially well-suited for understanding worker sentiment. This would continue to deepen our understanding of the circumstances under which platform work is most advantageous - or precarious - for workers. However, extra care is required to avoid sampling and response bias when relying on online surveys – particularly when respondents are paid.

The most promising initiatives involve collaboration. For example, the Online Labour Index (OLI) is a collaboration between academics and platforms, resulting in real-time insights into the demand for platform work. Using big and administrative data from platforms, whether from voluntary sharing or obligatory disclosures, is already being pursued by a number of governments both for research purposes and to ensure proper taxation. Without a complete overview of all Online Labour Platforms (OLPs), these efforts cannot result in a comprehensive view of platform work, but they can account for the majority of platform work given the dominance of relatively few platforms.

At EU-level and for individual Member States, policymakers should consider requiring platforms to provide administrative data, in order to receive preferential tax treatment (as in Belgium), or to operate at all (as with AirBnB in Amsterdam). This can help ensure conformity with regulations and better socio-economic insights.

appendix 1

Table a1. Mentioned platforms

99designs
AirBnB
Amazon
Amazon Mechanical Turk (Mturk/AMT)
BlaBlaCar
Clickworker
eBay
Etsy
Freelancer
Handy
Hilfr
Instagram
ListMinut
Mighty AI
Mybuilder
Taskrabbit
Twitch
Uber
Upwork
YouTube

Table a2. Selected data and statistics on platform work

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2014	Airbnb, 2014	UK	Airbnb	Administrative and survey data	n.a.	Economic impact and job creation	Nov 2012 - Oct 2013
2014	Hawksworth and Vaughan, 2014 (PwC)	US	Crowdfunding and P2P lending and accommodation, online staffing, car sharing, streaming, (video/music)	Administrative data	Forecasting method	Platforms' revenues in the five sectors	n.a.
2014	Owyang *et al.*, 2014	UK, US, Canada	Peer-to-peer platforms (covering 5 broad categories of collaboration: goods, services, transportation, space and money including money lending and crowdfunding)	Survey data	Two survey rounds. First round part of a general omnibus survey (N=90,112), and follow-up survey (N=2,550)	Participation in the sharing economy (as consumer) and motivations	Oct 2013 - Jan 2014
2014	Nielsen, 2014	World (60 countries throughout Asia-Pacific, Europe, Latin America, the Middle East, Africa and North America)	For profit good and service platforms	Survey data	Online survey (N=30,000 internet users). Non-probability sampling	Willingness in participating (as consumer) in sharing economy activities	14 Aug - 6 Sept 2013
2014	Stokes *et al.*, 2014 (Nesta)	UK	Internet-enabled collaborative activities across a selection of sectors (transport, holidays, off jobs and tasks, technologies and electronics, clothing and accessories, media, children's equipment and toys, households goods and appliances)	Survey data	Online survey (N=2,000 adults 16 and older). No information on sampling technique used	Participation as consumer or provider	May 2014
2015	Burston-Marsteller, the Aspen Institute and TIME, 2015	US	Ride sharing, accommodation, food delivery platforms and other services platforms	Survey data	Online survey (N= 3,000 US adults). No information on sampling technique used	Participation as consumer or provider in sharing activities	Nov 2015

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2015	DGE, 2015	FR	Peer-to-peer transactions involving a wide range of goods and services. Also transactions without monetary exchange covered.	Survey data	Online consumers survey (N=2,006 adults aged 18 and over). Non-probability sampling	Types of transactions, frequency, spending, offers, purchase	15-22 Oct 2014
2015	Harris and Krueger, 2015	US	Labour platforms	Big data	Collection and analysis of google trends data	Size of the workforce engaging in the gig economy	Jan – Nov 2015
2015	ING International, 2015	AU, AT, CZ, BE, DE, ES, FR, IT, LU, NL, PO, RO, TR, UK, US	Capital platforms	Survey data	Online survey (N= 14,829 adults aged 18 and older). No information on sampling technique	Awareness, participation, earned income and attitudes towards sharing	16 Jan - 2 Feb 2015
2015	Kuek *et al.*, 2015	World	Microwork and online freelancing platforms	Administrative data	Forecasting method	Market size and number of registered workers	2013 (projections to 2016)
2015	Maselli and Fabo, 2015 (CEPS)	World	CoContest (design work platform)	Big data	Data collected from Google searches and web crawling	Number of submissions per designer, level of earnings (compared to local wages)	Sept 2015
2015	Nesta, 2015	UK	Selling, lending, giving or leasing own assets or skills on the internet	Survey data	Part of a face-to-face omnibus survey (N=2,010 adults aged 15 and over). No information on sampling technique	Participation as provider and earnings, Estimation of monetary value of transactions	Feb 2015

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2015	PwC, 2015	US	• Hospitality and Dining (CouchSurfing, Airbnb, Feastly, LeftoverS wap) • Automotive and Transportation (RelayRides, Hitch, Uber, Lyft, Getaround, Sidecar) • Retail and Consumer Goods (Neighborgoods, SnapGoods, Poshmark, Tradesy) • Media and Entertainment (Amazon Family Library, Wix, Spotify, SoundCloud, Earbits)	Survey data	Online survey of consumer panellists (N=1,000)	Familiarity and engagement, benefits, concerns	17-22 Dec 2014
2016	Berg, 2016	World (CrowdFlower), US and India (AMT)	Micro tasks platforms (CrowdFlower and AMT)	Survey data	Online survey of CrowdFlower (N=67.7) and AMT workers (N=1,167)	Demographics, work experience and work history	Nov - Dec 2015
2016	Collaboriamo and Trailab, 2016a	IT	Capital and labour platforms	Administrative data	Mapping exercise drawing from existing literature and information provided directly by platforms through an online questionnaire (64 out of 138 identified platforms responded)	Number of active platforms, sector distribution, demographics of platform owners and workforce	Oct 2016
2016	Collaboriamo and Trailab, 2016b	IT	Crowdfunding platforms (divided into donation reward, DIY, equity and lending platforms)	Administrative data	Mapping exercise drawing from existing literature and information provided directly by platforms through an online questionnaire (41 out of 70 identified platforms responded)	Number of active platforms, amounts raised for each platform type, demographics of workforce	Oct 2016

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2016	De Groen *et al.*, 2016 (CEPS)	BE	ListMinut (local personal services platform)	Big data	Web crawling; data supplemented with Belgian administrative data	Types of tasks posted/provided and hourly remuneration	23 Dec 2013 - 22 Dec 2015
2016	De Groen and Maselli, 2016 (CEPS)	EU28	Uber ride-hailing platform	Big data	Collection and analysis of Google search data	Number of active workers	End of 2015
2016	European Commission, 2016a (Flash Eurobarometer 438)	EU28	Online service platforms (renting accommodation and car sharing to small household jobs)	Survey data	Telephone-based survey (N=14,050, EU residents aged 15 years and over). Multi-stage, random (probability) sampling	Awareness and frequency of use of 'collaborative platforms'	March 2016
2016	European Commission, 2016b (Special Eurobarometer 447)	EU28	Search engines, online social networks, online marketplaces	Survey data	Face to face survey (N=27,969, EU residents aged 15 and over). Multi-stage, random (probability) sampling	Frequency of use and attitudes towards online platforms	April 2016
2016	Evans and Gawer, 2016	World (five world regions and 22 countries)	Transaction platforms; innovation platforms; integrated platforms; investment platforms	Administrative data	Data collected using different search tools and databases (Quid Web Intelligence tool, CB insights, Thomson Reuters Eikon financial database), analysed and compiled in a database	Geographic and sector distribution, ownership structure	2015
2016	Farrell and Greig, 2016 (JP Morgan Chase and Co. Institute)	US	Capital and labour platforms (30 in total)	Big data	Analysis of American JP Morgan Chase customers' bank account transactions	Income from platforms	Oct 2012 - Sept 2015
2016	Freelancers Union and Upwork, 2016	US	Social media, online freelance marketplaces and sharing economy sites	Survey data	Online panel survey (N=6,002 of US adults). No information on sampling technique used	Use of online social media and online platforms to find work	2016

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2016	Hall and Krueger, 2016	US	Uber ride-hailing platform	Survey and administrative data	Analysis of data from two online surveys of Uber drivers (N= 601 in 2014; N= 833 in 2015). Survey data supplemented by administrative data on Uber drivers' driving histories, schedules and earnings between 2012 and 2014	Demographics of Uber drivers, income situation and motivations	Dec 2014, Nov 2015
2016	Jesnes *et al.*, 2016	NO	Capital and labour platforms	Survey data	Online survey (N=1,525 Norwegian adults aged 18 and over)	Engagement in online platforms and frequency of use	2016
2016	Katz and Krueger, 2016	US	Labour platforms	Survey data	RAND-Princeton Contingent Work Survey (RPCWS), a version of the CWS, as part of the RAND American Life Panel (N=3,850). Sample recruited through a variety of means (including a group recruited for the University of Michigan internet panel, a random digit dial sample, and a snowball sample)	Size of workforce in platform work	Oct - Nov 2015
2016	McKinsey Global Institute, 2016	DE, ES, FR, SE, UK, US	Digital platforms for independent work, comprising platforms for people to sell goods or lease assets or provide labour services	Survey data	Online panel survey. Sampling working age population (N=8,131)	Engagement in independent work and digital platform; motivations; and incomes earned from digital platforms	June - July 2016
2016	Robles and McGee, 2016	US	Online labour platforms and selling sites	Survey data	Online survey (N= 2,483 qualifying respondents out of a total sample of 6,898 US adults aged 18 and over). Probability-based online sampling	Engagement in online platforms	Oct - Nov 2015

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2016	Vaughan and Daverio, 2016 (PWC)	BE, DE, ES, FR, IT, the NL, PO, SE, UK.	Peer-to-peer accommodation; peer-to-peer transportation; on- demand household services; on- demand professional services; collaborative finance	Administrative data	Secondary data sources used, enabling 'data translation and triangulation exercise'	Size of the platform economy in terms of value of transactions and platforms' revenue	2013-2015
2016	Smith, 2016a	US	Labour platforms, capital platforms, and crowdfunding sites	Survey data	Panel survey (N= 4,787 US adults). Probability sampling	Attitudes, awareness and use (as clients) of online platforms	Nov-Dec 2015
2016	Smith, 2016b	US	Capital and labour platforms	Survey data	Panel survey (N= 4,579 US adults). Probability sampling	Use (as provider) of online platforms	July-Aug 2016
2016	Kässi and Lehdonvirta, 2016	World	Five prominent English language online labour platforms intermediating digital services	Big data	API access and web scraping. Tracking projects and tasks posted across major English- language online labour platforms	Utilisation of online labour across countries and occupations, projects and tasks posted	May 2016 - October 2016
2016	BMAS, 2016	DE	Two 'crowdworking' platforms	Survey data	Survey (N=408)	Socio-economic background, employment status and motivations	Feb 2015
2017	Alsos *et al.*, 2017	NO	Labour platforms and Airbnb	Survey data	Telephone survey (N=1,000 Norwegians aged 18-65 years)	Size of workforce engaging in platform work	Sept 2016-Oct 2017
2017	Bonin and Rinne 2017	DE	Labour platforms	Survey data	Omnibus telephone survey (N=10,017, aged 18+)	Size of the workforce engaging in platform work	April-June 2017
2017	Balaram *et al.*, 2017	UK	Labour platforms	Survey data	Face-to-face omnibus survey (N=7,656 UK residents aged 15 and older)	Engagement in platform work, motivation, working time, work-life balance	11 Nov 2016-10 Jan 2017

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2017	CIPD, 2017	UK	Labour platforms	Survey data	Online survey (N=5,019 adults aged 18-70). Non-probability sampling	Size of workforce engaging in platform work, motivations and level of income	Dec 2016
2017	De Groen *et al.*, 2017	EU28	Labour platforms	Administrative data	Secondary data sources used to calculate estimates (data extrapolations to estimate missing data). Clustering technique used to categorise online platforms	Size of work- related platform economy in terms of gross revenues and number of active workers	2016
2017	European Commission, 2017	BG DE, DK, ES, FR, IT, NL, PO, SL, UK	Peer-to-peer online platform in five sectors of activity: (re)sale of goods; sharing/ renting accommodation; sharing/renting goods; odd jobs; and ride sharing/hiring.	Survey data	Online survey (N=10,019 internet users)	Participation in peer-to-peer online market as consumer or provider or both	May 2016
2017	Eurostat, 2017	EU28	Peer-to-peer accommodation and transport services platforms	Survey data	General population / household survey (N=200,000 EU residents aged 16-74). Telephone/ face-to-face/ web interviews. Stratified, random (probability) sampling	Share of people arranging accommodation and transport services online via websites or apps	In most countries, second quarter of 2017

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2017	Fabo *et al.*, 2017	EU28	Transportation (of people and goods) platforms; platforms trading online services (for example design, IT services); and platforms trading offline, local services (for example delivery or housework)	Administrative data	Mapping exercise drawing from existing literature, media articles and information provided directly by platforms	Number of platforms active in the EU, platform size and turnover, work assignment method and business models, required skill level of workers, number of employees	
2017	Huws *et al.*, 2017	AT, CH, DE, IT, NL, SE, UK	Work platform (delivery of tasks online and on- location)	Survey data	**Online surveys:** AT, N=1,969, 18-65 years CH, N=2,001, 16-70 years DE, N=2,180, 18-65 years IT, N=2,199, 16-70 years NL, N=2,126, 16-70 years SE, N=2,146, 16-65 years UK, N=2,238, 16-75 years **Companion surveys:** Telephone based survey CH, N=1,205, 15-79 years **Face-to-face survey:** UK, N=1,794, 16-75 years	Size of workforce engaged in platform work, frequency of work, income, employment status	22-26 Jan 2016 (UK, online) 24 March-4 April 2017 (UK, offline) 26 Feb -7 March 2016 (SE) 1-4 April 2016 (DE) 1-4 April 2016 (AT) 22-27 April 2016 (NL) 31 March-5 April 2017 (IT) 3-14 April 2017 (CH, online) 27 March-7 April 2017 (CH, offline)
2017	Ilsøe and Madsen, 2017 (Denmark LFS)	DK	Labour platforms and capital platforms	Survey data	Ad-hoc module of the Danish LFS (N=18,043 Danes aged 15-74). Random sampling	Size of workforce engaging in online platforms and earning an income	Jan-March 2017
2017	Jackson *et al.*, 2017	US	'Gig economy' platforms identified in tax returns data (specific words and phrases such as ride share or ridesharing, or names of specific platform providers)	Administrative data	Analysis of tax returns. 109,700 individuals filing a return reporting income from online platform	Number of workers filing self- employment income and reporting income from an online intermediary.	2014

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2017	Karanovic, 2017	World	Uber	Big data	Webscraping and natural language processing	Sentiment analysis	April 2014 - January 2017
2017	ORB International, 2017	UK	Uber ride-hailing platform	Survey data	Telephone survey (N=1,002 Uber drivers)	Income, working time, work-life balance, motivation and employment status	8-17 Sept 2017
2017	Statens Offentliga Utredningar, 2017 (SOU)	SE	Peer-to-peer assets- based and services platforms	Survey data	Online survey (N=7,069 adults aged 16-64)	Size of workforce using online platforms	Sep 2016
2017	Statistics Canada, 2017 (LFS)	Canada	Peer-to-peer rental platforms and ride services platforms	Survey data	Telephone-based survey. (N=100,000 adults aged 18 and over). Multi-stage, random (probability) sampling	Total expenditure and use of online platforms as both provider and consumer	Oct 2016
2017	Statistics Finland, 2017 (LFS)	FI	Airbnb, Uber, Tori.fi / Huuto.net, Solved (and others specified by respondents)	Survey data	Telephone-based survey (N=43,0005 aged 15-74 residents in Finland). Stratified, random sample	Income from work and non-work- related platforms	2017
2017	Zervas *et al.*, 2017	US	Airbnb	Survey and administrative data	Data collected directly from Airbnb website, And supplemented with other data sources (Texas Comptroller, county demographics from US Census Bureau, airport passenger counts from US Bureau of Transportation Statistics (BTS), Current Population Survey (CPS) from the US BLS, and hotel reviews from TripAdvisor. Difference in difference technique for data analysis	Economic impact of Airbnb on hotel industry (in revenue terms)	Jan 2003 –Aug 2014

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2018	BEIS, 2018	Great Britain	Labour platforms	Survey data	NatCen Panel, a probability-based online survey (N=2,184, aged 18 and over). YouGov Omnibus, non- probability online panel survey (N=11,354, aged 18 and over)	Size of the workforce engaging om platform work and characteristics of platform work	July-Aug 2017
2018	Bureau of Labour Statistics, 2018 (BLS)	US	Electronically- mediated work, online and in person	Survey data	Contingent Worker Survey (CWS) is a supplement to the monthly Current Population Survey (CPS). Data collected via telephone and face-to-face (N=46, 000, aged 16 and over). Probability sampling	Size of the workforce engaging in electronically-mediated work	May 2017
2018	European Commission, 2018	EU28	For profit and not- for profit peer-to- peer and peer-to-business online platforms in four sectors of economic activity (transport, accommodation, finance, and online skills including on- demand household services, on-demand professional services)	Administrative and big data	Data collected through online web questionnaire sent to 1,012 identified platforms (64 full responses and 108 partial responses). Supplemented with secondary data obtained from web searches and web scrapping. Different data sources enabled data triangulation and validation	Size of the collaborative economy in terms of revenues and employment	July - Oct 2017
2018	Farrell *et al.*, 2018	US	Capital and labour platforms (128 in total)	Big data	Analysis of American JP Morgan Chase customers' bank account transactions	Income from online platforms	Oct 2012 -March 2018

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2018	Guarascio and Sacchi, 2018	IT	Capital platforms for intermediation services for real estate, accommodation and classified ads (Subito.it, Casa.it and Booking), labour platforms providing food-delivery (Deliveroo, Just-Eat, Foodora) and pet care services (Petme), and Italian branches of three global platforms (Amazon, Facebook and Google)	Administrative data	Descriptive analysis of data drawn from business and administrative data sources	Economic and employment characteristics of most prominent online platforms operating in Italy	2012-2016
2018	Insee, 2018 (French LFS)	FR	Intermediaries (including digital platforms). Types of platform unspecified.	Survey data	Ad-hoc module of the French LFS (Enquête Emploi) (N= 3,103.000 self-employed with and without employees). Probability sampling	Access to clients through an intermediary (including a digital platform)	2017
2018	Kässi and Lehdonvirta, 2018	World	Five prominent English language online labour platforms intermediating digital services	Big data	API access and web scraping. Tracking projects and tasks posted across major English-language online labour platforms	Supply and demand of online freelance labour over time and across countries and occupations. Collected data was used to construct an online labour index.	Jul 2016 Feb 2017 Jan 2018
2018	MBO partners, 2018	US	Online job platforms	Survey data	Online survey (N=3,584, US residents aged 21 and older). Non-probability sampling	Size of the independent workforce and motivations. Use of digital platforms to find work.	March 2018

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2018	ORB International, 2018	UK	Uber ride-hailing platform	Survey data	Telephone survey (N=1,001 Uber drivers)	Socio- demographics, income, working time, motivation, subjective well-being	18-28 March 2017
2018	Pesole *et al.*, 2018 (European Commission's JRC)	DE, ES, FI, FR, NL, HR, HU, IT, LT, PT, RO, SE, SL, UK	Labour platforms	Survey data	Online survey (N=32,409 Internet users aged 16- 74). A commercially available list of internet users in the selected countries (CINT) used as sampling frame, with non-probability quota sampling of respondents by gender and age groups	Size of workforce engaging in platform work, their characteristics, motivations and working conditions	June 2017
2018	PwC, 2018	AT, BE, CH, DE, NL and TR,	For profit and not- for profit peer-to- peer and business- 'For profit and not- for profit peer-to- peer and business-to-peer online platforms in selected industry segments (media and entertainment, hotels and accommodation, automotive and transport, retail and consumer goods, services, finance, and machinery)'	Survey data	Online survey (N=4,500). No information on sampling technique	Size and acceptance of platform economy in the selected sectors.	June - Aug 2017
2018	Statistics Canada, 2018	CA	Selling sites (for example Etsy and eBay); freelance services platforms, peer-to-peer ride, delivery and accommodation services platforms.	Survey data	Online / telephone survey (N=12,000 Canadians aged 18 and older). Two-stage, random (probability) sampling	Income from selected platform activities	June - July 2018
2018	Weel *et al.*, 2018	NL	On-location work platform	Survey data	Online survey. No information on sampling technique	Size of workforce engaging in platform work	n.a.

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2018	Nezhyvenko, 2018	UA	'IT Freelance' platforms such as Upwork, AMT, and Ukrainian versions	Survey, interview, focus group data	Online survey (N=1,000) on three leading platforms, InPoll, snowball technique	Activity, income, other working conditions characteristics	2017
2018	Berg *et al.*, 2018	75 countries	Five major microtask platforms	Survey data, interviews, surveys	Survey (N=3,500)	Motivations and work characteristics	2015 and 2017
2019	Huws *et al.*, 2019	AT, CZ, EE, FI, FR, DE, IT, NL, SI, ES, SE, CH, UK	Online and on-location platforms, online marketplaces for selling goods	Survey data	Online survey	Detailed demographic data and usage of various platform types	January 2016-May 2019
2019	Joen *et al.*, 2019	CA	Identifying gig workers based on characteristics of work and how reported in tax data.	Administrative data and survey data (census microdata)	Combine national survey and administrative data	Proportion of workers considered gig workers	2016
2019	Katz and Krueger, 2019	US	AMT	Survey data	Online survey (N=2,291 AMT workers, aged 18 and older). Sample was not chosen to be representative instead selected to include a large number of workers who worked on multiple jobs, often on a casual basis, and determine the extent to which multiple job holders neglect to report that they worked on multiple jobs based on the standard BLS Current Population Survey (CPS) question	Multiple job holding using CPS (BLS)-like question	March 2015
2019	Le Ludex *et al.*, 2019	FR	How many French microworkers, and how active.	Survey data	Online survey (N=2,792)	Count and activity levels of microworkers.	2018

Year	Publication (authors)	Geographic focus	Research focus (platform types)	Type of data	Method for data collection and/or analysis	Measurement(s)	Reference period
2019	Piasna and Drahokoupil, 2019	BG, HU, LV, PL, SK	Online forms of platform work	Survey data	Online survey (ETUI Internet and Platform Work Survey), random sampling	Prevalence of platform work, working conditions	2019-2019
2019	Serfling, 2019	DE	Labour platforms	Survey data	Open-access web panel survey (N=494,970). Non-probability sampling	Size of crowd work work-force, socio-demographics, remunera-tion, task duration, moti-vations and satisfaction	July 2017 - Oct 2018
2020	Brancati *et al.*, 2020	ES, UK, PT, DE, LT, NL, IT, RO, HR, FR, SE, SK, HU, FI	10 diverse types of platform work	Survey data	Survey (N=38,022)	Proportion active as plat-form workers, amount of activity, demographics, working conditions	2018

Note: This table builds on the summary found in Riso's excellent work 'Mapping the contours of the platform economy' (2019). The descriptions Riso used are essentially unchanged. The author extends warm thanks to Eurofound for permission to reproduce this work.

References

Airbnb (2014), 'New Study: Airbnb Community Generates $824 Million in Economic Activity in the UK', available at https://www.airbnb.ie/press/news/new-study-airbnbcommunity-generates-824-million-in-economic-activity-in-the-uk

Alsos, K. *et al.* (2017), 'Nar sjefen er en app', When the boss is an app). Fafo-report, Vol. 41.

Alsos, K. *et al.* (2018), 'Nar sjefen er en app–delingsøkonomi i et arbeidsperspektiv', Praktisk økonomi & finans, Vol. 34, No. 02, pp. 101–110.

Arnault, S. *et al.* (2018), 'Emploi, chômage, revenus du travail. Edition 2018. Dossier 3. Les agents de l'Etat en 2014: leurs profils, leurs carrières'.

Balaram, B., Warden, J. and Wallace-Stephens, F. (2017), 'Good gigs: A fairer future for the UK's gig economy', RSA (Action and Research Centre), UK.

Beck, P. *et al.* (2017), 'The feasibility of measuring the sharing economy: November 2017 progress update', United Kingdom Office for National Statistics, https://www.ons.gov.uk/economy/economicoutputandproductivity/output/articles/thefeasibilityofmeasuringthesharingeconomy/november2017progressupdate.

BEIS (2018), 'The characteristics of those in the gig economy', Research Paper: 2018 no. 2. Department for Business, Energy and Industrial Strategy (BEIS), UK Government.

Berg, J. *et al.* (2018), 'Digital labour platforms and the future of work | Towards decent work in the online world', ILO (https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_645337.pdf).

Berg, J. (2016), 'Income Security in the On-Demand Economy: Findings and Policy Lessons from a Survey of Crowdworkers', SSRN Scholarly Paper ID 2740940, Social Science Research Network, Rochester, NY, March (https://papers.ssrn.com/abstract=2740940).

Berg, J. and H. Johnston (2019), 'Too Good to Be True? A Comment on Hall and Krueger's Analysis of the Labor Market for Uber's Driver-Partners', ILR Review, Vol. 72, No. 1, pp. 39–68.

BLS (US Bureau of Labor Statistics) (2018), 'Electronically Mediated Work: New Questions in the Contingent Worker Supplement'. Monthly Labor Review, US Bureau of Labor Statistics Washington DC.

BMAS (Bundesministerium für Arbeit und Soziales) (2016), 'Chancen und Risiken der Digitalisierung der Arbeitswelt für die Beschäftigung von Menschen mit Behinderung', Forschungsbericht 467.

Bogliacino, F. *et al.* (2020), 'Quantity and Quality of Work in the Platform Economy', in K.F. Zimmermann (ed.), Handbook of Labor, Human Resources and Population Economics, Cham: Springer International Publishing (https://doi.org/10.1007/978-3-319-57365-6_18-1).

Bonin, H. and Rinne, U. (2017), Omnibusbefragung zur Verbesserung der Datenlage neuer Beschäftigungsformen, Kurzexpertise im Auftrag des Bundesministeriums für Arbeit und Soziales, IZA Research Report No. 80, IZA, Institute of Labor Economics, Bonn, Germany

Boyd, D. and K. Crawford (2012), 'Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon', Information, communication & society, Vol. 15, No. 5, pp. 662–679.

Brancati, C.U. *et al.* (2019), 'Digital Labour Platforms in Europe: Numbers, Profiles, and Employment Status of Platform Workers', Text, July (https://ec.europa.eu/jrc/en/publication/digital-labour-platforms-europe-numbers-profiles-and-employment-status-platform-workers).

Burston-Marsteller, Aspen Institute and TIME (2015) The Collaborative Economy Survey, available at www.burson-marsteller.com/ondemand-survey/

CIPD (2017), 'To gig or not to gig? Stories from the modern economy', Chartered Institute of Personnel and Development, UK.

Collaboriamo and Trailab (2016a), Sharing economy: La mappatura delle piattaforme italiane 2016.

Collaboriamo and Trailab (2016b), Il crowdfunding in Italia, Report 2016.

Crawford, K. (2013), 'The hidden biases in big data', Harvard Business Review, Vol. 1, p. 1.

Dasgupta, A. (2013), 'Big data: The future is in analytics', Geospatial World, Vol. 3, No. 9, pp. 28–36.

De Groen, W.P. *et al.* (2017), 'The Impact of the Platform Economy on Job Creation', Intereconomics, Vol. 52, No. 6, pp. 345–351.

De Groen, W.P., Maselli, I. and Fabo, B. (2016), 'The Digital Market for Local Services: A one nightstand for workers?', CEPS Special Report No. 133, CEPS, Brussels, Belgium.

Dewi, L.C. *et al.* (2019), 'Social Media Web Scraping using Social Media Developers API and Regex', Procedia Computer Science, Vol. 157, pp. 444–449.

DGE (Direction Générale dés Entreprises) (2015), 'Enjeux et perspectives de la consummation collaborative', Annexe – Enquête auprès des consommateurs, Etudes Economiques, July.

Edelman, B.G. and M. Luca (2014), 'Digital discrimination: The case of Airbnb. com', Harvard Business School NOM Unit Working Paper, No. 14–054.

Eurofound (2018), 'Employment and working conditions of selected types of platform work', Digital Age, Eurofound, Luxembourg (https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef18001en.pdf).

European Commission (2016a), 'Flash Eurobarometer 438: The use of collaborative platforms', Directorate-General for Communication, Brussels, July.

European Commission (2016b), 'Special Eurobarometer 447: Online platforms', Brussels, Belgium

European Commission (2017), 'Exploratory study of consumer issues in online peer-to-peer platform markets', prepared VVA, Milieu, and GfK for the European Commission (DG Just), Brussels, Belgium

European Commission (2018), 'Study to Monitor the Economic Development of the Collaborative Economy at sector level in the 28 EU Member States', prepared by Technopolis, Trinomics, VVA Consulting for the European Commission (DG Grow), Brussels, Belgium.

Eurostat (2017), 'Online peer-to-peer accommodation services used by 1 in 6 EU citizens', news release, 199, 20 December.

Evans P. C. and Gawer, A. (2016), 'The Rise of the Platform Enterprise: A Global Survey', The Emerging Platform Economy Series No. 1, New York, US.

Fabo, B. *et al.* (2017), 'An Overview of European Platforms: Scope and Business Models'.

Farrell, D. and F. Greig (2016), 'Paychecks, paydays, and the online platform economy'.

Farrell, D. *et al.* (2018), 'The online platform economy in 2018: Drivers, workers, sellers, and lessors', JPMorgan Chase Institute.

Fortune (2015), 'A Huge Percentage Of Etsy Sellers Are Women', July (https://fortune.com/2015/07/23/etsy-sellers-women/).

Freelancers Union and Upwork (2016), 'Freelancing in America: 2016' (https://www.upwork.com/i/freelancing-in-america/2016/).

Gadiraju, U. *et al.* (2014), 'A taxonomy of microtasks on the web'.

Grey Ellis, E. (2019), 'YouTubers Must Unionize, No Matter What Google Says', Wired (https://www.wired.com/story/youtube-union/).

Guarascio, D. and S. Sacchi (2018), 'Digital platforms in Italy; an analysis of economic and employment trends', INAPP policy brief, Vol. 8.

Guta, M. (2018), 'Etsy Study Says Small Businesses Selling on the Platform Contribute $4.7B to the US Economy', November (https://smallbiztrends.com/2018/11/2018-etsy-statistics.html).

Hall, J.V. and A.B. Krueger (2018), 'An analysis of the labor market for Uber's driver-partners in the United States', ILR Review, Vol. 71, No. 3, pp. 705–732.

Hall, J.V. and Kreuger, A.B. (2016), 'An Analysis of the Labor Market for Uber's DriverPartners in the United States', NBER Working Paper No. 22843, National Bureau of Economic Research, Washington.

Hara, K. *et al.* (2018), 'A data-driven analysis of workers' earnings on Amazon Mechanical Turk'.

Harris, S.D. and A.B. Krueger (2015), A Proposal for Modernizing Labor Laws for Twenty-First-Century Work: The' Independent Worker', Hamilton Project, Brookings (https://www.shrm.org/ResourcesAndTools/legal-and-compliance/employment-law/Documents/modernizing_labor_laws_for_twenty_first_century_work_krueger_harris.pdf).

Hausemer, P. *et al.* (2017), 'Exploratory Study of consumer issues in peer-to-peer platform markets'.

Hawksworth, J. and Vaughan, R. (2014), 'The Sharing Economy - Sizing the Revenue Opportunity', PwC, London, UK.

Heerschap, N. *et al.* (2018), 'Measuring online platforms', Universiteit van Amsterdam, Amsterdam, December.

Huws, U. (2014), Labor in the global digital economy: The cybertariat comes of age, NYU Press.

Huws, U. *et al.* (2017), 'Work in the European gig economy: Research results from the UK, Sweden, Germany, Austria, the Netherlands, Switzerland and Italy'.

Huws, U. *et al.* (2019), 'The Platformisation Of Work In Europe: Highlights from research in 13 European countries', Foundation for European Progressive Studies, July (http://researchprofiles.herts.ac.uk/portal/en/publications/the-platformisation-of-work-in-europe-highlights-from-research-in-13-european-countries(6cb3297c-861e-4d75-962e-5bc9ff031f53).html).

Ilsøe, A. and Madsen, L. W. (2017), 'Digitalization of work and digital platforms in Denmark', Employment Relations Survey Centre (FAOS), University of Copenhagen.

ING International (2015), 'The European sharing economy set to grow by a third in the next 12 months', ING.com.

Insee (Institut national de la statistique et des études économiques) (2018), 'Emploi, chômage, revenus du travail', Insee.

Jackson, E., Looney, A., and Ramnath, S. (2017), 'The Rise of alternative work arrangements: Evidence and implications for tax filing and benefit coverage', Department of Treasury, Office of Tax Analysis, US.

Jesnes, K. *et al.* (2016), 'Aktører og arbeid i delingsøkonomien', Delrapport. Fafo-notat 2016, Vol. 23.

Joen *et al.*, 2019

Karanovic, J. *et al.* (2017), ''Is Platform Capitalism Doomed'? Ask the Workers'.

Kässi, O. (2016), 'How is online work classified in the OLI?', The iLabour Project, web page, available at http://ilabour.oii.ox.ac.uk/how-is-online-work-classified.

Kässi, O. and V. Lehdonvirta (2018), 'Online labour index: Measuring the online gig economy for policy and research', Technological Forecasting and Social Change (https://linkinghub.elsevier.com/retrieve/pii/S0040162518301331).

Katz, L. F. and Krueger, A. B. (2016), 'The rise and nature of alternative work arrangements in the United States, 1995-2015', Working paper 22667, National Bureau of Economic Research.

Kilhoffer, Z. (2020), ''BEST PRACTICE' REPORT ON HOW SOCIAL MEDIA CAN BE USED TO UNDERSTAND THE LABOUR MARKET', Milestone 112, Leuven.

Kilhoffer, Z. *et al.* (2020), 'Study to gather evidence on the working conditions of platform workers', DG EMPL.

Kuek, S.-C. Paradi-Guilford, C., Fayomi, T., Imaizumi, S., Ipeirotis, P. (2015), 'The global opportunity in online outsourcing', World Bank Group, Washington DC, US.

Leimeister, J.M. *et al.* (2009), 'Leveraging crowdsourcing: activation-supporting components for IT-based ideas competition', Journal of management information systems, Vol. 26, No. 1, pp. 197–224.

Lenaerts, K. *et al.* (2016), 'Prospects for utilisation of non-vacancy Internet data in labour market analysis - an overview', IZA Journal of Labor Economics, Vol. 5, No. 1, p. 1.

Lenaerts, K. *et al.* (2017), 'Government Responses to the Platform Economy: Where do we stand?', Policy Insights 2017–30, Brussels, July (https://www.ceps.eu/system/files/PI2017-30_Government%20Responses%20to%20the%20Platform%20Economy.pdf).

Maselli, I. and Fabo, B. (2015), 'Digital workers by design? An example from the collaborative economy', CEPS Working Document 414, CEPS, Brussels.

MBO partners (2018, 'The State of independence in America, 2018: the new normal', available online at https://www.mbopartners.com/uploads/files/state-of-independencereports/State_of_Independence_2018.pdf.

McKinsey Global Institute (2016), 'Choice, Necessity, and the Gig Economy', San Francisco. McKinsey Global Institute.

National Academies of Sciences and Medicine (2020), Measuring Alternative Work Arrangements for Research and Policy, National Academies Press.

Nesta (2015), 'Towards an Index of the Collaborative Economy', Nesta, London.

Nezhyvenko, O. (2018), 'Methodology Behind the Measurement of the Platform Economy (Survey on the Platform Economy in Ukraine)', presentation for Expert workshop 'Measuring the platform economy: size and impact', CEPS, Brussels (http://www.inclusivegrowth.eu/files/Call-5/INGRID2_ExpertWorkshop_Session1_Oksana-Nezhyvenko.pdf).

Nielsen (2014), 'Global share community report: Is sharing the new buying?', Nielsen.com.

Office for National Statistics (2017), 'Internet access—households and individuals', Statistical Bull.

ORB International (2017), 'Uber poll 2017 - Flexibility key for drivers', ORB International, September (http://www.orb-international.com/2017/09/26/uber-poll-flexibility-key-drivers).

ORB International (2018), 'New ORB International poll for Uber and Oxford Martin School – Uber happy? Work and well-being in the 'gig economy'', 2 October, available at www.orbinternational.com/2018/10/02/uberhappy.

Owyang, J., Samuel, A. and Grenville, A. (2014), 'Sharing is the new buying, how to win in the collaborative economy', Vision Critical and Crowd companies.

Pesole, A. *et al.* (2018), 'Platform Workers in Europe Evidence from the COLLEEM Survey', EUR - Scientific and Technical Research Reports, Publications Office of the European Union (http://publications.jrc.ec.europa.eu/repository/handle/111111111/52393).

PwC (2015), 'The sharing economy', Consumer Intelligence Series, PricewaterhouseCoopers.

PWC (2018), 'Share Economy 2017: The New Business Model', PricewaterhouseCoopers.

Riso, S. (2019), 'Mapping the contours of the platform economy', Working Paper, Eurofound, Dublin.

Robles, B.J. and M. McGee (2016), 'Exploring Online and Offline Informal Work: Findings from the Enterprising and Informal Work Activities (EIWA) Survey', SSRN Scholarly Paper ID 2875204, Social Science Research Network, Rochester, NY, October (https://papers.ssrn.com/abstract=2875204).

Russel, M. and M. Klassen (2019), Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More, O'Reilly Media, Inc.

Schmidt, F.A. (2017), 'Digital Labour Markets in the Platform Economy: Mapping the Political Challenges of Crowd Work and Gig Work'.

Serfling, O. (2018), 'Crowdworking Monitor Nr. 1', Für das Verbundprojekt „Crowdworking Monitor'. CIVEY, Hochschule Rhein-Waal, gefördert durch Bundesministerium für Arbeit und Soziales. Abgerufen von https://www.bmas.de/DE/Presse/Meldungen/2018/ausmass-von-plattformarbeit-in-deutschland-hoeher-als-erwartet. html.

Silberman, M.S. (2017), 'Fifteen criteria for a fairer gig economy', democratization, Vol. 61, No. 4, pp. 589–622.

Smith, A. (2016a), 'Shared, collaborative and on demand: The new digital economy', Pew Research Center.

Smith, A. (2016b), 'Gig Work, Online Selling and Home Sharing', Pew Research Center.

Statens Offentliga Utredningar (SOU) (2017), 'Ett arbetsliv I förändring – hur påverkas ansvaret för arbetsmiljön?', Stockholm.

Statistics Canada (2017), 'The sharing economy in Canada', Press release, 28 February, available at https://www150.statcan.gc.ca/n1/daily-quotidien/170228/dq170228b-eng.htm

Statistics Canada (2018), 'Digital economy, July 2017 to June 2018', Press release, 29 August, available at https://www.statcan.gc.ca/eng/about/smr09/smr09_094

Statistics Finland (2017), 'Labour force survey 2017: platform jobs', available at http://www.stat.fi/til/tyti/2017/14/tyti_2017_14_2018-04-17_en.pdf.

Stokes, K., Clarence E., Anderson, L. and Rinne, A. (2014), 'Making Sense of the UK Collaborative Economy', Nesta, London.

Teubner, T. *et al.* (2017), 'Price determinants on AirBnB: How reputation pays off in the sharing economy.', Journal of Self-Governance & Management Economics, Vol. 5, No. 4, pp. 53–80.

US Bureau of Labor Statistics (2018), Electronically Mediated Work: New Questions in the Contingent Worker Supplement. Monthly Labor Review, US Bureau of Labor Statistics Washington DC.

vanden Broucke, S. and B. Baesens (2018), Practical Web Scraping for Data Science Best Practices and Examples with Python, KU Leuven.

Vaughan, R. and Daverio, R. (2016), 'Assessing the size and presence of the collaborative economy in Europe', PwC, London, UK.

Weel, B., Werff, van der, S., Bennaars, H., Scholte, R., Fijnje, J., Westerveld, M. and Mertens, T. (2018), 'De opkomst en groei van de kluseconomie in Nederland', SEO (SEO Economisch Onderzoek), Amsterdam.

Zervas, G. *et al.* (2015), 'A first look at online reputation on Airbnb, where every stay is above average', Where Every Stay is Above Average (January 28, 2015).

Supporting expertise in inclusive growth

www.inclusivegrowth.eu

Working paper

PLATFORM WORKERS AND LABOUR MARKET RISK

Who is vulnerable?

Leonie Westhoff

February 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 730998

Abstract

Platform work is an emerging labour market phenomenon that is increasingly drawing the attention of policymakers and researchers. This exploratory study contributes to the growing body of evidence on platform work and its connection to labour market risks. In particular, the study explores in how far labour market risk is stratified within the population of platform workers due to platform workers' individual characteristics and the types of platform work they work in. Drawing on a new data set on platform workers in Europe, the 2018 COLLEEM survey, three types of labour market risk are investigated: low income, inferior working conditions and educational mismatch. The results of the analysis demonstrate that there is significant heterogeneity in platform worker vulnerability to labour market risk. While low education significantly increases the likelihood of all types of labour market risks for platform workers, the influence of demographic characteristics is less systematic and depends on the type of labour market risk. In addition, the influence of types of platform work tasks on labour market risks is complex. While low-skilled platform work tasks are associated with a higher likelihood of low income, workers in higher-skilled types of platform work are more likely to report experiencing inferior working conditions. The results of the study suggest that the relationship between platform worker characteristics and labour market risk is nuanced and warrants further examination in future research.

Acknowledgements

I would like to thank Sara Baiocco, Zachary Kilhoffer and Stephanie Steinmetz for their useful comments on previous drafts of this work. I would also like to thank the Joint Research Centre of the European Commission for providing access to the COLLEEM data set.

48

Contents

1. Introduction

Recent decades have seen the transformation of labour markets through processes such as digitalisation, globalisation and automation. Significantly, platform work is an emerging phenomenon in labour markets, which is increasingly drawing the attention of academic research. Against this background, examining the characteristics of platform work and its economic and social consequences is of critical importance. However, given the relative newness of the phenomenon and lack of high-quality data, existing research remains limited.

In this exploratory study, platform work is understood as all labour provided through, on or mediated by online platforms, where platforms operate in a wide range of sectors, work can be of varied forms and is provided in exchange for payment (Kilhoffer *et al.*, 2020). A platform worker is then a natural person providing platform work (Ibid.). There are several reasons to expect platform workers to be vulnerable to 'old' labour market risks, such as unemployment and occupational health and safety risks, but also 'new' risks such as non-payment and dependency on ratings for receiving work. In addition they might also experience exacerbated risks compared to more traditional forms of work, including increased job and income instability, low income, a lack of training and skills development opportunities, and skills or educational mismatch (Allaire *et al.*, 2019). Nevertheless, platform work is extremely heterogeneous, both in the composition of the workforce and the nature of platform work tasks (Schor *et al.*, 2020). As such, the individual and job characteristics of platform workers may influence their vulnerability to labour market risk. This association is examined in this study.

The empirical analysis draws on a survey of platform workers, the Collaborative Economy and Employment (COLLEEM) data set (Pesole *et al.*, 2018; Urzì Brancati, Pesole and Fernandez Macias, 2020), covering 16 EU countries in 2018. This survey is considered to be the most reliable data source on platform work in the EU (Kilhoffer *et al.*, 2020). The study contributes to the growing literature on the platform economy in several ways. Most significantly, it provides a systematic analysis of the influence of different individual characteristics and platform work types on platform worker vulnerability to labour market risk. Furthermore, a comparative analysis of three types of labour market risks is advanced: low income, inferior working conditions and educational mismatch. These are 'old' labour market risks, which all workers may face, but which could be exacerbated by platform work. Indeed, the three labour market risks examined have been identified as particularly relevant to platform workers in previous studies (Eurofound, 2018; Kilhoffer *et al.*, 2020).

Hence, the following research questions are examined:

1. Is there an association between the individual characteristics and platform work tasks of platform workers and their vulnerability to labour market risk?
2. If so, does this association differ for different types of labour market risk?

2. Which platform workers are most vulnerable to labour market risk?

The nature of platform work is often vulnerable, as platform workers may be susceptible to inferior working conditions and lack access to social, labour and health and safety protections (Garben, 2019). This section reviews literature on three types of labour market risk - low income, inferior working conditions and educational mismatch - and assesses the vulnerability of platform workers to these risks. In this sense, vulnerability is understood as both a product of one's inherent individual characteristics as well as relational and context specific, being dependent on (among other things) social and economic circumstances (Mackenzie, Rogers & Dodds, 2014).

The main interest of the study is to examine which platform workers are most vulnerable to labour market risk *within* the population of platform workers. Given the relative lack of existing analysis on this issue, this study is explicitly exploratory and descriptive in nature. The theoretical discussion refers to literature on the working conditions of platform workers wherever possible; however, it also draws on findings from the general literature on labour market stratification and working conditions where appropriate. This literature provides a useful reference point to examine whether patterns observed in the general labour market also apply for platform work. Types of labour market risk are examined separately to analyse potential differences in effects.

Platform work is a diverse field of work that covers a multiplicity of tasks. These can be broadly distinguished based on two main factors: whether they are conducted online or offline, and whether they are high-skill or low-skill (De Groen & Maselli, 2016). Following Pesole *et al.* (2018), a fine-grained typology of ten distinct types of platform work is examined: online clerical and data-entry tasks, online professional services, online creative and multimedia work, online sales and marketing support work, online software development and technology, online writing and translation work, online microtasks, interactive services, transportation and delivery services, on-location services. This is assumed to encompass the main types of platform work.

2.1 Risk of low income

Previous literature has identified low pay and pay irregularity as an issue for many platform workers, exacerbated by a struggle to get a sufficient amount of work (Berg, 2016; Forde *et al.*, 2017; Eurofound, 2018). Platform workers face economic and legal insecurity, have little bargaining power and are often not unionised or covered by minimum wages (Risak, 2018; Kilhoffer *et al.*, 2020), leaving them with low labour market standing that can be related to lower wages. Moreover, for online platform work, global competition means that platform workers in high-income countries compete with workers who can undercut their wages. Workers may also have to cover costs associated with platform work, and there may be issues with non-payment and unpaid time, for instance due to complaints or poor ratings (Risak, 2018). Where platform work is worker-initiated, a significant proportion of time may be spent looking for tasks, researching requesters and communicating with clients, and thus not be remunerated (Berg *et al.*, 2018). While many platform workers use platform work to supplement income from other sources (Eurofound, 2018), there are also some that count it as their main source of income (Kilhoffer *et al.*, 2020) and others that would prefer to work more, often combining platform work with other poorly-paid work (Fabo, Karanovic & Dukova, 2017).

Many studies have examined the influence of individual and job-level characteristics on individual incomes in the overall labour market. This literature has shown that wage returns are structured by the socio-economic characteristics of workers, with higher wage returns expected for older (Fournier & Koske, 2012), more educated (Card, 1999) and male (Blau & Kahn, 2017) workers, among others. On the one hand, the traditional dividing lines of factors such as age and gender may be replicated in the platform economy. On the other hand, due to the additional flexibility, ease of work substitutability and transparency it offers, the platform economy may lower entry barriers for specific groups traditionally disadvantaged in labour markets (Urzì Brancati, Pesole & Fernandez Macias, 2020). If so, lower entry barriers may be associated with reduced pay gaps based on socio-economic characteristics. For instance, the higher flexibility of the platform economy may enable women to increase their labour supply within the context of persistent gender gaps in domestic work (Foong *et al.*, 2018). However, evidence on the influence of socio-economic characteristics of platform workers on their income is generally scarce, and existing evidence indicates that pay gaps do persist within the platform economy. Indeed, studies examining labour platforms in the United States (Foong *et al.*, 2018; Cook *et al.*, 2020) and in Ukraine (Aleksynska, Bastrakova & Kharchenko, 2019) show evidence of a gender pay gap in platform work, putting in doubt whether this emerging form of work could close gender differences in labour markets.

In addition to socio-economic characteristics, types of platform work tasks will likely influence platform workers' pay. While there are some highly-qualified platform tasks, such as online professional services, that are highly rewarded, platform work also includes tasks that tend to be low-paid, such as translation, social media and administrative tasks (Gomez-Herrera, Martens & Mueller-Langer, 2018; Schwellnus *et al.*, 2019). Generally, virtual services with general skills will be least rewarded (De Groen and Maselli, 2016; Fabo, Karanovic & Dukova, 2017). In particular, online microwork, where work is broken up into extremely small, low-skilled tasks, is an area where extreme low pay is an issue (Berg, 2016; Webster, 2016; Berg *et al.*, 2018; Cantarella & Strozzi, 2019). In contrast, income from on-location platform work is usually relatively stable, if modest (Eurofound, 2018).[31] As such, platform workers engaging in low-skilled platform work such as microtasks are expected to be at higher risk of low income, while the location of platform work as such is not expected to affect the risk of low income.

2.2 Risk of inferior working conditions

Stress

Work-related stress is associated with adverse impacts on individual health, safety at work and well-being (Jain & Leka, 2019). Due to the nature of their work, stress may be common among platform workers. Many forms of platform work are characterised by a high level of competition between workers and the dependency of workers on rating mechanisms (Garben, 2019), which may encourage a rapid pace of work without breaks. Indeed, platform workers may experience pressure to constantly be online to be available for work and attempt to take on as many tasks as possible, especially given low rates of pay for some tasks, as described above (Forde *et al.*, 2017). In addition, platform workers often have to coordinate multiple tasks at the same time, sometimes jointly with regular employment, and have to cope with deadlines posed by clients (Ibid.). Such high levels of work intensity and pressure are likely associated with stress and anxiety (Kilhoffer *et al.*, 2020).

However, the stress platform workers experience at work may be mediated by other factors. In general, socio-demographic factors matter for work-related stress. Across European countries, studies focused on the general population have shown that higher levels of stress are observed among women and lower-educated workers (De Smet *et al.*, 2005). Marriage can also be a stress factor, given

31 For on-location platform work where platform workers set their own prices, pay may be more variable.

the higher financial pressure associated with family responsibilities and the potential for work-home conflict (Marinaccio *et al.*, 2013). The relationship between stress and age is u-shaped, with both younger workers and older workers experiencing higher levels of work-related stress (De Smet *et al.*, 2005; Marinaccio *et al.*, 2013). In principle, socio-economic factors are expected to affect platform workers in a similar manner as workers in the overall labour market.

Moreover, the type of platform work may have an impact on levels of stress associated with it. While tight deadlines are common across types of platform work, the threat of replaceability will be higher among workers that perform tasks requiring a low level of skills (Kilhoffer *et al.*, 2020). For online work, microwork has been found to be associated with high levels of stress for workers (Webster, 2016; Eurofound, 2018). As such, low-skilled platform work is expected to be associated with higher levels of stress, while the location of platform work is not expected to play a significant role.

Long working hours

In addition to work-related stress, working long hours, equally associated with psychosocial and health and safety risks (Jain & Leka, 2019), can be a concern for some platform workers. While many platform workers only devote a few hours to platform work, often using it as a supplement to regular employment, anecdotal evidence suggests that there is a proportion that work very long hours, though there are no precise estimates of how large this proportion is (Kilhoffer *et al.*, 2020). Characteristics of platform work that encourage long hours are the dependency on supply and demand, the pressure to remain available and work overload, as elaborated on above (Ibid.). The potential for low pay and the need to spend time waiting or searching for new tasks can further exacerbate this.

There is scarce evidence on the factors that may increase the likelihood of working long hours among the population of platform workers. Within the general labour market, men are more likely to work longer hours than women, particularly when they have children (Kodz *et al.*, 2003; Lee, McCann & Messenger, 2007; Anxo & Karlsson, 2019). Workers in prime age are also most likely to work particularly long hours (Kodz *et al.*, 2003; Lee, McCann & Messenger, 2007). In addition, migrants may be more likely to work particularly long hours, while the role of education is disputed (Anxo & Karlsson, 2019). The empirical analysis will explore whether these patterns also obtain within the population of platform workers. However, given the lack of existing evidence on why platform workers work long hours, concrete expectations cannot be formed. As regards the role of platform work tasks, expectations are also unclear. On the one hand, it may be the case that workers are more likely to work long hours in low-paid tasks such as online microwork (Berg *et al.*, 2018). On the other hand, similar to the regular labour market, workers in higher-level occupations may be more likely to work longer hours (Kodz *et al.*, 2003).

Health and safety

Finally, platform workers may be exposed to health and safety risks through their work. Traditional health and safety risks at work can be exacerbated by the nature of platform work (EU OSHA, 2017). Protective occupational health and safety regulations are not necessarily guaranteed for platform workers given their unclear employment status, though this will depend on country-specific regulation and the type of platform work. Workers often do not receive health and safety training and are usually responsible for providing their own protective equipment and tools. In addition, platform workers often do not have the right to paid sick leave.

Within the general population, research has shown that sociodemographic characteristics of workers can have an influence on workplace health and safety (Schulte *et al.*, 2012). Low education may be a determinant of the risk of work injury, as education can provide better knowledge in health and safety related issues or help workers advance to safer positions (Piha *et al.*, 2013). Migrant workers are also more likely to be exposed to inferior working conditions, in turn related to occupational health and safety risk (Pérez *et al.*, 2012). Men, younger workers and lower-skilled workers are at higher risk of accidents at work, as laid out by Eurostat (2010). In contrast, women are at higher risk of mental health problems, as are mid-age workers and higher-educated workers. Female workers are also at higher risk of occupational diseases, the incidence of which strongly increases with age. In principle, these mechanisms may apply in similar ways within the platform economy, though this cannot be said with certainty given the lack of existing evidence.

In addition, the type of platform work undertaken is expected to be associated with differences in exposure to health and safety risk. In the first place, the types of risks differ substantially between online and offline work, but it is not clear which type of platform work is associated with higher health and safety risk. While the risks associated with on-location platform work include accidents, chemical exposure and harassment, online work exposes workers to postural disorders, eye strain and psychosocial risks due to the lack of an office environment (EU OSHA, 2017; Eurofound, 2018; Kilhoffer *et al.*, 2020). Moreover, certain types of platform work, particularly low-skilled platform work, are more likely to be associated with time pressure, stress and long working hours, as elaborated on above, which may also exacerbate health and safety risks associated with this work (EU OSHA, 2017). As such, low-skilled platform work, is expected to be related to a higher health and safety risk relative to high-skilled platform work.

2.3 Risk of educational mismatch

Finally, platform workers are at risk of educational mismatch. Overeducation is defined as an individual working in a job for which a lower level of education than their own is required, whereas undereducation, conversely, is defined as working in a job for which a higher level is required (Allen & van der Velden, 2001). Educational mismatch is a form of skills mismatch, that is, a poor fit between a worker's qualifications and skills and those required by their job (OECD, 2013). Such mismatch can have negative consequences for workers, both in the form of wages and job satisfaction (Quintini, 2011a).

Many providers of platform work engage in activities that have traditionally been blue or pink collar, despite being highly educated (Schor, 2017). As such, most platform workers are overqualified for the type of work they perform through platforms (Kilhoffer *et al.*, 2020). While the types of skills applied on platforms vary strongly (Eurofound, 2018), for many workers, tasks are frequently short and repetitive, despite workers' high level of education (Berg *et al.*, 2018), resulting in educational and skills mismatch. In addition, access to training is significantly lower than for other forms of work, and most workers indicate that there are few opportunities to learn or develop skills (Eurofound, 2018; Kilhoffer *et al.*, 2020). However, some workers highlight the opportunity to learn new things such as navigating work in an online environment or improving their English (Berg *et al.*, 2018).

Existing research on the overall labour market has shown that socio-demographic characteristics influence the propensity of workers to be over- and under-qualified (for a thorough review see Quintini, 2011a, 2011b). In particular, young workers and immigrants are substantially more likely to be over-qualified for their work, which can be traced back to their lack of work experience and difficulty in transferring qualifications to the domestic labour market. Some studies also find that women are more likely to be overeducated as well as undereducated, though the evidence is more mixed in this case. Equally, the role of marital status and children in influencing educational mismatch is unclear.

It is an interesting question to what extent these traditional labour market dynamics will translate to platform work. On the one hand, similar patterns as in the general labour market may be observed for platform workers. If demands for qualifications, work experience and language skills on platforms are similar to other jobs, workers that are traditionally more likely to face educational mismatch may face similar risks in the platform economy and be trapped in lower-skilled platform work that does not match their qualifications. On the other hand, it could also be the case that workers traditionally at the margins of labour markets and struggling to find jobs matching their qualifications or skills in

54

the traditional economy are more able to do so in the platform economy. If this is the case, vulnerability to educational mismatch may not be stratified by socio-economic characteristics in the platform economy.

3. Data, variables and methods

One of the challenges in conducting research on the platform economy is the general lack of high-quality data. This study is able to draw on perhaps the highest-quality survey of platform workers available to date, the COLLEEM survey (Pesole *et al.*, 2018; Urzì Brancati, Pesole & Fernandez Macias, 2020). COLLEEM is an online panel survey of internet users aged 16-74, that contains detailed information on respondents' socio-demographic characteristics and activity in the platform economy. COLLEEM data is collected by the Joint Research Centre of the European Commission in partnership with the Directorate General for Employment, Social Affairs and Inclusion. The first pilot wave of the survey was carried out in 2017, with fieldwork for a second survey carried out in autumn of 2018. The data was collected through questionnaires administered online.

The data used originates from the second COLLEEM survey. This data includes 38,787 responses, of which 5,489 are platform workers, from 16 European countries.[32] The survey data is a representative sample of internet users aged 16-74 based on a non-probability quota-based sampling approach. The fact that only internet users are sampled in the COLLEEM survey is intentional, as the authors state that non-internet users should not be sampled for a study on work on internet platforms (Pesole *et al.*, 2018). Nevertheless, there may be some bias in the sample and the results cannot necessarily be generalised to the whole population of platform workers.

A variety of dependent variables are defined. First, a dummy variable on low income is created. Ideally, the indicator of income would directly measure the income individuals make in the platform economy. However, while COLLEEM includes a measure of hourly wages from work in the platform economy, this indicator cannot be used in the analysis. First, it measures wages from the main platform individuals work on, not wages generally received in the platform economy. Second, preliminary examination of the variable showed many missing values, a large variation in outcomes and improbably high values for many respondents, suggesting that the measure is not reliable. Instead, the indicator of low income is coded as a dummy equalling one when individuals place themselves as receiving an income corresponding to the lowest quartile of their country's income distribution. However, this measure also has limitations. Besides relying on individuals accurately reporting their income, a particular issue is that the measure refers to all income, not only income in the platform economy. This is problematic, particularly as many individuals obtain income from platform work as a supplement to their regular income. To mitigate potential error resulting from this, the analysis is limited to individuals that make at least half of their income in the platform economy. While this should make the results more reliable in linking individual characteristics to income from the platform economy, they should nevertheless be interpreted with substantial caution. Indeed, as a substantial number of cases are lost due to the restriction (see table below) and the link between independent variables and income from platform work remains somewhat uncertain, the results of the regression on income should be understood as exploratory and not generalisable.

Second, a variety of variables are used to examine working conditions. All these variables rely on workers' subjective evaluation of their working conditions. Stress is measured as a dummy variable equalling one if a respondent indicates that they agree/strongly agree that their work on platforms is stressful. Long hours are coded as a dummy variable equalling one if a respondent indicates that they often/sometimes work more than ten hours a day on platforms. Health risk is a dummy variable

32 These include the UK, Germany, France, Italy, Spain, Finland, the Netherlands, Sweden, Hungary, Slovakia, Romania, Croatia, Lithuania, Ireland, the Czech Republic and Portugal.

equalling one if a respondent indicates that they agree/strongly agree that their work on platforms is a risk to their health. In addition to these three variables, a joint dummy variable is created, equalling one if an individual is subject to all three inferior working conditions. This adds value, given that the theoretical discussion indicated that the three types of inferior working conditions may be interconnected. Individuals who are exposed to all three types of inferior working conditions can be regarded as particularly vulnerable in the labour market.

Third, educational mismatch is measured in the form of two dummy variables. Overeducation is measured as an individual's level of education being higher than the level of education required for the main task they carry out on platforms, while undereducation is defined as an individual's level of education being lower than the level of education required for the main task they carry out on platform. Education levels were assigned to platform work tasks by matching them to ISCO codes and corresponding skill levels (ILO, 2012). The assigned education levels are given in the Appendix, Table a1.

However, it should be pointed out that this can only be a rough approximation, given the diversity of platform work even within task categories, and since it is unclear in how far platform work tasks map onto the ILO occupational classification.

A range of independent variables are included to examine associations between platform worker characteristics and the measures of labour market risk. Socio-economic characteristics include gender (a dummy equalling one if the respondent is female), age (categories 16-25, 26-40, 41-55, 56-64), education (categories primary and lower secondary education (reference), upper secondary and post-secondary non-tertiary education, tertiary education), a dummy variable for being in a couple, a dummy variable for dependent children being present in the household and a dummy for migration (not being born in the country of residence). In addition, the type of platform work is included in the model, based on the main task respondents indicate they perform on platforms. This is done based on the COLLEEM classification of platform work into ten categories: online clerical and data-entry tasks, online professional services, online creative and multimedia work, online sales and marketing support work, online software development and technology, online writing and translation work, online microtasks, interactive services, transportation and delivery services and on-location services.

The regression analysis focuses on the working age population, defined as respondents aged 16-64. In addition, to focus on platform workers who do a substantial amount of platform work, the sample is limited to individuals who do platform work at least once a week. As robustness checks, the analysis was also performed for a sample including platform workers who do platform work at least monthly, rather than weekly (Appendix, Table a4.) and platform workers of any age (Appendix, Table a5.). This did not substantially change results. Given the restrictions indicated above, the sample size of the analysis is reduced, as summarised in Table 2. Most of the regression analysis focus on weekly platform workers. However, as explained above, the regression on income focuses on weekly platform workers who make at least 50% of income from platform work, which substantially reduces the size of the sample. While the sample size is judged to be large enough for regression analysis (Green, 1991), substantial caution should be exercised when interpreting the results.

Table 1. Summary of sample sizes (individuals 16-64)

Sample	Number of responses
Full COLLEEM Sample	35,035
Platform workers	5,009
Monthly platform workers	3,825
Weekly platform workers	2,871
Weekly platform workers who make at least 50% of their income from platform work	444

Note: Observations with missing values on independent variables included.

Logit regressions are run on the dependent variables to determine in how far the different variables affect the likelihood of platform workers to experience the different types of labour market risk. In addition, as a robustness check, linear probability models (LPM) were also run (Appendix, Table a6. - a8.), as they can be an alternative for models with binary dependent variables where the sign and significance of the relation are the main interest, rather than non-linearity of the relation per se (Mood, 2010). The results of the LPM were virtually identical to the average marginal effects (AME) in the logit regression. Regressions are run in a stepwise manner, first including only the socio-economic characteristics of respondents and then adding their main task on platforms. In this way, it can be determined in how far the main tasks of platform workers account for the effect of their socio-economic characteristics. Given the small sample size for individual countries, the regressions are run on the pooled sample of European countries. To account for potential country-level effects on the dependent variable, country fixed effects are included in all models. All data are weighted to account for sampling design.

4. Results: summary and discussion

4.1 Descriptive statistics: the population of platform workers across European countries

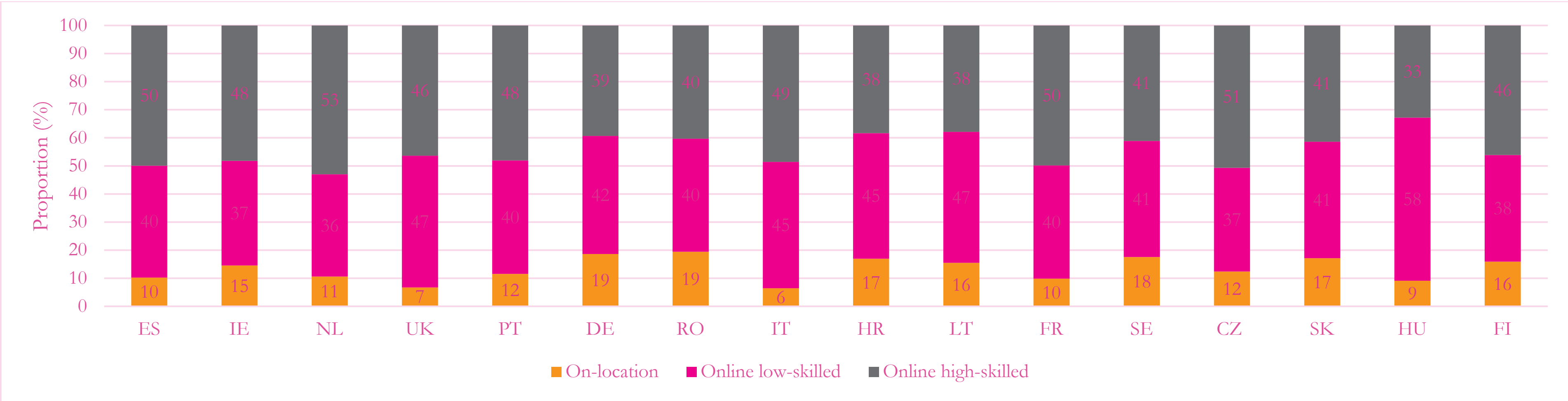
Figure 1 illustrates the proportion of platform workers that engage in platform work at least weekly in the sample, within the working age population. There is strong variation in the size of the platform economy across countries. Estimates range from 12% in Spain to 4.2% in the Czech Republic. Figure 2 shows the relative proportions of high-skilled online work, low-skilled online work and on-location work in each country. In all countries, the vast majority of respondents engage in online platform work. However, this may be a function of the sampling design, given that data was collected through an online survey. The relative proportions of high-skilled and low-skilled online work differ across countries, though in most, the split is relatively even.

The appendix (Table a2.) contains further descriptive statistics on the composition of the sample with regard to the independent and dependent variables. About a quarter (25.29%) of platform workers indicate that their income is located in the lowest quartile of the income distribution of their respective country. Inferior working conditions are more common than low income among platform workers. More than half of platform workers surveyed indicate that they experience high stress (53.55%), frequently work more than 10 hours per day (61.85%) or perceive their work to be associated with a high health risk (52.31%). Around a third of workers (31.81%) experience all three. As regards educational mismatch, 26.14% of platform workers in the sample are undereducated for their work, while a larger proportion (39.54%) are overeducated. A cross tabulation of the dependent variables with the independent variables is shown in the Appendix (Table a3.).

Figure 1. Proportion of weekly platform workers by country

Note: Estimates are adjusted by internet use using data from Eurostat, ICT Survey.

Figure 2. Distribution of types of platform work by country



4.2 Regression results: the relationship between platform worker characteristics and vulnerability to labour market risk

Table 2 shows the results of the logistic regressions on low income including only socio-economic characteristics (Model 1) and adding the main task on platforms (Model 2). All results are presented as Average Marginal Effects (AMEs) to facilitate comparison across models (Mood, 2010). Starting with Model 1, relative to platform workers aged 16-25, prime-age workers' probability of low income is decreased by 26.7%, whereas there is no significant difference between younger and older workers (41 or older). In addition, the risk of low income decreases with the level of education. Being in a couple and having children are also associated with a decrease in the probability of low income. As such, age, education and family characteristics appear to influence the risk of low income to platform workers. In contrast, gender and immigration do not have a significant effect on the probability of being in the low-income group for platform workers. This is contrary to what has been found by studies on the general labour market and in existing studies examining gender pay gaps in the platform economy, as laid out in the theoretical section. This could provide some first evidence that work in the platform economy could indeed help to overcome income discrimination. However, these results should be interpreted with caution, given the issues with linking income to work in the platform economy and the small sample size of the regression, as elaborated in the previous section.

In a second step, information on the main task workers perform on platforms is added to the regressions (Model 2). While the significance of socio-economic characteristics does not change, the size of the AME of education and family characteristics on low income decreases, indicating that this effect can be partly be explained by the different tasks higher-educated workers and workers with family engage in on platforms. Nevertheless, education and family characteristics continue to have a substantial effect in the full model. As regards the role of platform tasks, relative to online professional services, aligning with expectation, workers performing online microtasks are at significantly higher probability (33.2%) of low income. In addition, doing online writing and translation work (33.6%) and online software development and technology (34%) are also associated with a higher probability of low income. Effects are also found for interactive services (29.6%) and on-location services (20.5%). While these are only marginally significant, at significance level 10%, these effects are nevertheless interesting, given the small sample size of the regression.

Table 2. Results of logistic regression on low income, average marginal effects, sample of 16 EU countries, platform workers who do platform work at least weekly, 16-64

	Y = Low income	
	Model 1	**Model 2**
Female	-0.044	-0.021
	(0.055)	(0.059)
Age (ref = 16-25)		
26-40	-0.267***	-0.263***
	(0.072)	(0.068)
41-55	-0.119	-0.107
	(0.105)	(0.085)
56-64	-0.249	-0.192
	(0.159)	(0.194)
Education (ref = primary and lower secondary)		
Upper secondary and post-secondary non-tertiary	-0.284**	-0.230*
	(0.140)	(0.126)
Tertiary	-0.284**	-0.247**
	(0.138)	(0.122)
Couple	-0.125**	-0.115**
	(0.057)	(0.051)
Dependent children in household	-0.240***	-0.210***
	(0.056)	(0.055)
Foreign-born	0.008	0.055
	(0.074)	(0.058)
Main task on platform (ref = online professional services)		
Online clerical and data entry tasks		0.060
		(0.070)
Online creative and multimedia work		0.139
		(0.131)
Online sales and marketing support work		0.114
		(0.094)
Online software development and technology		0.340***
		(0.095)
Online writing and translation work		0.336***
		(0.098)
Online microtasks		0.322***
		(0.104)
Interactive services		0.296*
		(0.165)
Transportation and delivery services		-0.005
		(0.093)
On-location services		0.205*
		(0.105)
Observations	417	417
Country FE	Yes	Yes

Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Regression results from logistic regression on low income.
All regression results weighted.
Regressions on low income exclude workers who gain less than 50% of their income in the platform economy.

Next, the results for working conditions are examined focusing on indicators for stress, long hours and health and safety risk, as well as on the cumulative indicator for experiencing all three simultaneously. The results are shown in Table 3, for both a model including only socio-economic characteristics (Model 1) and the full model (Model 2). In most cases, gender and immigration do not play a role in influencing working conditions. However, women are more likely to work long hours on platforms than men, in contrast to previous studies on the regular economy. It is possible that such differences are observed due to the additional flexibility of platform hours compared to regular work, so that women can combine this work with caretaking responsibilities, unlike other forms of employment, where they are more likely to work lower hours. However, this exploratory analysis cannot definitely claim to identify such pathways.

There are also age patterns, in that prime-age workers are more likely to experience health and safety risks at work, as well as work long hours, as expected. In contrast, relative to workers aged 16-25, somewhat older workers (41-55) are less likely to report feeling stressed or experiencing health risks at work, while there is no significant difference to older workers. With regard to all three indicators of inferior working conditions, as expected, higher levels of education decrease the probability of experiencing them. Hence, the platform economy reflects patterns found in the overall labour market, where skills, and educational qualifications which function as signals of skills, are a significant determinant of labour market risk in the form of working conditions. Finally, family characteristics play a role as expected, in that being in a couple or having children in the household increases the risk of inferior working conditions. As regards the cumulative measure of working conditions, the probability of experiencing all three inferior working conditions increases for prime-age workers (11.7%) and those with children (9.8%) but decreases significantly with higher levels of education. Overall, therefore, the patterns found in the general labour market, elaborated on in the theoretical section, appear to be reflected in the platform economy, workers with lower education or with family are at higher risk of inferior working conditions.

Once the task content of platform work is added to the model, as seen in Model 2, most effects of socio-economic characteristics persist, and the size of effects does not change substantially. However, the influence of being in a couple on perceived health and safety risk becomes insignificant, suggesting that this effect is driven by individuals in a couple working in different types of tasks than those who are single. As regards the role of types of platform tasks as such, interesting patterns are found. Contrary to expectation, it does not appear that low-skilled platform work is associated with inferior working conditions. Rather, relative to online professional services, online microtasks and online writing translation work are associated with a lower probability of perceived inferior working conditions, on all measures. In addition, online creative and multimedia work is associated with a lower probability of working long hours (14.2%) compared to online professional services, while transportation and delivery services are associated with a lower probability of health and safety risk (19.8%). Hence, it appears that the inferior working conditions associated with platform work are not only experienced by workers performing low-skilled tasks. However, it should be borne in mind that these results are based on the subjective evaluation of workers in these tasks. It may be the case that workers in higher-skilled tasks evaluate working conditions differently to those in lower-skilled tasks, which could influence results, for instance due to different relative benchmarks in the classic labour market. In addition, examination of the source data revealed that in the categories found to be significantly different from online professional services, comparatively lower proportions of workers derive at least 50% of their income from platform work or do platform work daily.[33] The subjective evaluation of platform work as more dangerous and stressful could thus be a result of greater dependency on it.

33 The shares of these types of platform workers by type of platform work task are shown in the Appendix, Table a9.

Table 3. Results of logistic regression on different measures of working conditions, average marginal effects, sample of 16 EU countries, platform workers who do platform work at least weekly, 16-64

	Y = Stress		Y = Long hours		Y = Health risk		Y = Working conditions (cumulative)	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Female	0.010	0.017	0.055*	0.062**	-0.023	-0.016	0.027	0.033
	(0.032)	(0.031)	(0.030)	(0.029)	(0.031)	(0.031)	(0.030)	(0.030)
Age (ref = 16-25)								
26-40	-0.005	-0.001	0.133***	0.134***	0.096**	0.099***	0.117***	0.117***
	(0.039)	(0.038)	(0.037)	(0.036)	(0.038)	(0.037)	(0.036)	(0.036)
41-55	-0.183***	-0.168***	-0.014	-0.004	-0.108**	-0.099**	-0.045	-0.039
	(0.045)	(0.044)	(0.045)	(0.044)	(0.044)	(0.043)	(0.038)	(0.038)
56-64	-0.104	-0.100	-0.026	-0.038	-0.059	-0.064	-0.011	-0.016
	(0.075)	(0.072)	(0.073)	(0.069)	(0.072)	(0.069)	(0.067)	(0.064)
Education (ref = primary and lower secondary)								
Upper secondary and post-secondary non-tertiary	-0.097**	-0.093*	-0.077*	-0.077*	-0.098**	-0.096**	-0.125**	-0.127**
	(0.049)	(0.049)	(0.046)	(0.045)	(0.049)	(0.048)	(0.049)	(0.050)
Tertiary	-0.128***	-0.131***	-0.117***	-0.121***	-0.155***	-0.165***	-0.166***	-0.174***
	(0.047)	(0.047)	(0.045)	(0.044)	(0.047)	(0.047)	(0.046)	(0.048)
Couple	0.058*	0.039	0.064**	0.034	0.084**	0.057*	0.063*	0.046
	(0.034)	(0.034)	(0.033)	(0.032)	(0.033)	(0.032)	(0.032)	(0.032)
Dependent children in household	0.081**	0.078**	0.119***	0.107***	0.092***	0.088***	0.098***	0.094***
	(0.035)	(0.035)	(0.032)	(0.031)	(0.034)	(0.034)	(0.033)	(0.034)
Foreign-born	-0.027	-0.040	0.005	-0.014	0.060	0.044	0.046	0.039
	(0.046)	(0.047)	(0.044)	(0.044)	(0.046)	(0.046)	(0.040)	(0.041)

Table 3. **Results of logistic regression on different measures of working conditions, average marginal effects, sample of 16 EU countries, platform workers who do platform work at least weekly, 16-64 (continued)**

	Y = Stress		Y = Long hours		Y = Health risk		Y = Working conditions (cumulative)	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Main task on platform (ref = online professional services)								
Online clerical and data entry tasks		-0.071		-0.045		-0.028		-0.043
		(0.057)		(0.051)		(0.054)		(0.057)
Online creative and multimedia work		0.002		-0.128**		-0.032		-0.071
		(0.059)		(0.057)		(0.060)		(0.061)
Online sales and marketing support work		-0.060		-0.099*		-0.101*		-0.078
		(0.061)		(0.057)		(0.060)		(0.061)
Online software development and technology		-0.061		-0.039		-0.032		-0.087
		(0.068)		(0.065)		(0.064)		(0.066)
Online writing and translation work		-0.119*		-0.127**		-0.107*		-0.137**
		(0.063)		(0.060)		(0.061)		(0.064)
Online microtasks		-0.244***		-0.324***		-0.266***		-0.230***
		(0.061)		(0.055)		(0.057)		(0.057)
Interactive services		-0.137*		-0.042		0.001		-0.098
		(0.075)		(0.065)		(0.072)		(0.073)
Transportation and delivery services		-0.088		-0.089		-0.198***		-0.126*
		(0.078)		(0.070)		(0.076)		(0.076)
On-location services		-0.011		-0.048		-0.036		-0.050
		(0.077)		(0.071)		(0.077)		(0.082)
Observations	2,722	2,722	2,722	2,722	2,722	2,722	2,722	2,722
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Regression results from logistic regression stress, long hours, health risk and a cumulative measure of working conditions.
All regression results weighted.

Finally, the results for regressions on over- and undereducation of platform workers are evaluated (Table 4). These regressions do not include variables on education or main platform task, as the dependent variable is constructed from these two indicators. Hence, the second step of the regressions is not conducted for educational mismatch. The results show that socio-economic characteristics do have an influence on educational mismatch. Matching the results of most general labour economics literature, female platform workers are more likely to be overeducated than male platform workers (6.7%) and, conversely, are less likely to be undereducated (7.2%). With regard to age patterns, it is evident that the likelihood of undereducation decreases with age, that is, younger workers are more likely to be undereducated. Conversely, the group of mid-age workers (41-55) is most likely to be overeducated, rather than the reference group of workers aged 16-25. Similarly, contrary to findings for the general labour market, platform workers with an immigrant background are less likely to be overeducated (16.5%), and more likely to be undereducated (10.5%). This could suggest that workers such as the young and migrants, who may be struggling to access the regular labour market due to a lack of official qualifications or experience, or difficulties in transferring these to the domestic labour market, can apply these skills in the platform economy, potentially providing a stepping stone to the regular labour market eventually. Finally, family patterns are unsystematic - individuals in couples are less likely to be undereducated while those with children are more likely to be overeducated.

Table 4. Results of logistic regression on platform work potential risk factors, average marginal effects, sample of 16 EU countries, platform workers who do platform work at least weekly, 16-64

VARIABLES	Overeducated	Undereducated
Female	0.067**	-0.072**
	(0.030)	(0.030)
Age (ref = 16-25)		
26-40	0.031	-0.074*
	(0.035)	(0.038)
41-55	0.086**	-0.109**
	(0.042)	(0.043)
56-64	0.038	-0.137**
	(0.066)	(0.059)
Education (ref = primary and lower secondary)		
Upper secondary and post-secondary non-tertiary		
Tertiary		
Couple	0.017	-0.076**
	(0.031)	(0.031)
Dependent children in household	-0.066**	-0.023
	(0.032)	(0.033)
Foreign-born	-0.165***	0.107***
	(0.044)	(0.039)
Observations	2,853	2,853
Country FE	Yes	Yes

Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Regression results from logistic regression overeducated and undereducated.
All regression results weighted.

5. Conclusion

Platform work is a labour market phenomenon that is increasingly attracting the attention of researchers. To add to this growing body of work, this exploratory study has sought to examine the factors which influence the vulnerability of platform workers to labour market risk. There are reasons to believe that platform workers are indeed vulnerable to labour market risk, including low income, inferior working conditions - in form of stress, long working hours and health and safety risk - and educational mismatch. However, the analysis using a high-quality data set on the platform economy, the 2018 COLLEEM survey, showed that not all platform workers are equally vulnerable to such risks. This is not a surprising finding, given the large heterogeneity in both the composition of the platform workforce and the nature of tasks carried out on platforms.

While lower levels of education appear to significantly increase the likelihood of all types of labour market risk, the influence of socio-demographic characteristics is more heterogeneous. Though non-prime-age and single platform workers are most at risk of low income, it is prime-age platform workers as well as platform workers in couples or with children that are most likely to report experiencing inferior working conditions on platform. In contrast, gender and immigration play a limited role. This is a contrast to findings of literature on the stratification of labour market risk within the overall working population, which tends to find that these characteristics are significant in influencing labour market divides. Hence, factors that typically stratify labour markets appear to matter less in platform work than in the general labour market, or even work in the opposite direction in some cases. However, this exploratory study cannot definitely make such a claim, particularly given the limitations of the data available.

Equally, the influence of types of platform work tasks on labour market risk is complex. Lower-skilled platform work tasks, such as online microtasks, are associated with increased risk of low income relative to higher-skilled online tasks. However, it is in fact workers in higher-skilled types of platform work, such as online professional services, that appear particularly vulnerable to experiencing inferior working conditions in platform work. While one should be careful to draw definitive conclusions based on this first explorative study on this topic, this suggests that the relation between types of platform work and labour market risk is more nuanced than might be thought initially.

While a significant contribution to existing research on the platform economy, this study has several limitations. First, there are significant issues with online panel surveys such as COLLEEM, as laid out by Pesole *et al.* (2018), including the potential lack of generalisability of results given the sampling strategy. Second, the small sample size of the data set does not allow for a disaggregated analysis by country, which would add an important dimension. Third, as detailed in the methodological section, given the nature of the data set, several approximations have to be used in the measurement of labour market risk.

As such, the results should be interpreted with caution and further research is needed to supplement the evidence of this study. However, this would require providing higher-quality data on platform workers. While the COLLEEM data set is a very welcome step in this direction, large-scale representative surveys on platform workers are still lacking. In particular, it would be desirable to add items on platform work to existing data sets, such as the European Labour Force Survey. In addition, longitudinal data is needed to examine moves into and out of platform work and their implications. This would enable researchers, for instance, to examine whether platform work can serve as a stepping stone to regular employment.

Ultimately, this study underlines, similar to previous research, the large heterogeneity in platform work. This extends not only to the types of platform work, but also to working conditions and other labour market circumstances associated with it. For future research, but also for policy decisions on the platform economy, it is essential that this heterogeneity is adequately taken into account.

appendix 1

Table a1. Skill level allocated to platform work task groups

Task	Corresponding education level
Online clerical and data-entry tasks (e.g. customer services, data entry, transcription and similar)	Primary and lower secondary education
Online professional services (e.g. accounting, legal, project management and similar)	Tertiary education
Online creative and multimedia work (e.g. animation, graphic design, photo editing and similar)	Tertiary education
Online sales and marketing support work (e.g. lead generation, posting ads, social media management, search engine optimisation and similar)	Tertiary education
Online software development and technology work (e.g. data science, game development, mobile development and similar)	Tertiary education
Online writing and translation work (e.g. article writing, copywriting, proofreading, translation and similar)	Upper secondary and post-secondary non-tertiary education
Online microtasks (e.g. object classification, tagging content review, website feedback and similar)	Primary and lower secondary education
Interactive services (e.g. language teaching, interactive online lessons, interactive consultations and similar)	Upper secondary and post-secondary non-tertiary education
Transportation and delivery services (e.g. driving, food delivery, moving services and similar)	Primary and lower secondary education
On-location services (e.g. housekeeping, beauty services, on-location services and similar)	Primary and lower secondary education

Table a2. Descriptive statistics on the sample composition, weekly platform workers 16-64, weighted sample

Variable	Proportion (in %)
Female	34.80
Age	
16-25	26.63
26-40	48.20
41-55	19.57
56-74	5.61
Education	
Primary	17.17
Upper secondary/Post-secondary non-tertiary	39.15
Tertiary	43.69
In a couple	65.79
Dependent children in household	72.25
Not resident in country of birth	15.03
Main platform work task	
Online clerical and data entry	15.42
Online professional services	10.75
Online creative and multimedia work	12.43
Online sales and marketing support	13.02
Online software development and technological support	9.85
Online writing and translation	9.39
Online microtasks	12.05
Interactive services	5.48
Transportation and delivery	6.22
On-location services	5.40
Lowest quartile of income distribution	25.29
High stress through work	53.55
High health risk	52.31
Work > 10h per day	61.85
Working conditions - cumulative	31.81
Educational mismatch	
No mismatch	34.31
Undereducated	26.14
Overeducated	39.54

Note: Data on low income includes only respondents who gain at least 50% of their income from platform work.

Table a3. Crosstabulation, dependent variables and independent variables, weekly platform workers 16-64, weighted sample

Variable	Low income (%)	High stress (%)	Long hours (%)	Health risk (%)	Working conditions – cumulative	Undereducated (%)	Overeducated (%)
Gender							
Men	26.15	53.7	60.4	53.14	31.16	28.51	37.39
Female	23.79	53.27	64.57	50.73	33.03	21.76	43.53
Age							
16-25	41.84	56.42	55.02	49.73	26.45	34.24	36.08
26-40	12.5	57.9	69.7	60.07	39.63	24.51	38.94
41-55	26.17	41.18	55.42	40	22.29	21.11	45.23
56-74	11.66	45.53	49.24	40.5	22.82	19.18	41.26
Education							
Primary	56.25	60.98	66.79	61.22	42.48	60.25	0
Upper secondary/Post-secondary non-tertiary	26.59	51.86	59.92	50.76	29.24	40.36	43.9
Tertiary	17.39	51.23	60.82	49.24	28.55	0	51.17
Family characteristics							
Not in a couple	41.36	52.19	57.2	47.88	28.21	33.92	37.29
In a couple	16.84	54.25	64.27	54.58	33.66	22.19	40.69
No dependent children in household	53.83	46.99	51.29	42.98	23.23	28.28	45.26
Dependent children in household	16.22	56.05	65.9	55.89	35.09	25.32	37.34
Residency status							
Resident in country of birth	22.87	52.91	61.01	50.46	29.87	24.39	42
Not resident in country of birth	36.97	57.14	66.57	62.74	42.75	36.39	25.13

Table a3. Crosstabulation, dependent variables and independent variables, weekly platform workers 16-64, weighted sample (continued)

Variable	Low income (%)	High stress (%)	Long hours (%)	Health risk (%)	Working conditions – cumulative	Undereducated (%)	Overeducated (%)
Main platform work task							
Online clerical and data entry	16.54	54.64	68.24	57.37	36.71	0	86.35
Online professional services	11.31	60.95	72.14	60	40.03	41.48	0
Online creative and multimedia work	27.37	62.23	59.4	58.45	35.97	54.68	0
Online sales and marketing support	12.9	54.75	63.66	51.48	33.17	61.38	0
Online software development and technological support	59.24	57.03	68.43	59.38	33.59	54.57	0
Online writing and translation	37.21	46.45	56.58	47.06	24.33	6.34	53.65
Online microtasks	43.42	36.71	37.48	31.78	16.91	0	82.59
Interactive services	24.26	50.29	70.96	63.01	33.52	18.5	38.32
Transportation and delivery	7.78	53.64	63.21	42.25	28.89	0	74.75
On-location services	32.5	60.05	65.1	53.45	33.33	0	80.88

Note: Data on low income includes only respondents who gain at least 50% of their income from platform work.

Table a4. Results of logistic regression on platform work potential risk factors, average marginal effects, sample of 16 EU countries, platform workers who do platform work at least monthly, 16-64

Variables	Low income	Stress	Long hours	Health risk	Working conditions (cumulative)	Overeducated	Undereducated
Female	-0.029	0.027	0.047*	-0.010	0.031	0.071***	-0.069***
	(0.055)	(0.027)	(0.026)	(0.026)	(0.025)	(0.025)	(0.025)
Age (ref = 16-25)							
26-40	-0.293***	-0.003	0.133***	0.105***	0.112***	0.060**	-0.105***
	(0.063)	(0.033)	(0.032)	(0.032)	(0.030)	(0.029)	(0.032)
41-55	-0.123	-0.146***	0.004	-0.083**	-0.040	0.111***	-0.122***
	(0.083)	(0.038)	(0.038)	(0.038)	(0.032)	(0.036)	(0.037)
56-64	-0.210	-0.143**	-0.012	-0.082	-0.039	0.092	-0.143***
	(0.197)	(0.063)	(0.060)	(0.059)	(0.052)	(0.058)	(0.051)
Education (ref = primary and lower secondary)							
Upper secondary and post-secondary non-tertiary	-0.270**	-0.105**	-0.077*	-0.109**	-0.139***		
	(0.113)	(0.043)	(0.041)	(0.042)	(0.043)		
Tertiary	-0.294***	-0.147***	-0.115***	-0.166***	-0.175***		
	(0.109)	(0.041)	(0.040)	(0.041)	(0.041)		
Couple	-0.120***	0.036	0.030	0.059**	0.041	0.015	-0.051*
	(0.045)	(0.028)	(0.028)	(0.028)	(0.027)	(0.027)	(0.027)
Dependent children in household	-0.196***	0.069**	0.102***	0.101***	0.090***	-0.074***	-0.010
	(0.052)	(0.029)	(0.028)	(0.029)	(0.028)	(0.027)	(0.027)
Foreign-born	0.076	-0.005	0.033	0.078**	0.057*	-0.138***	0.102***
	(0.054)	(0.040)	(0.038)	(0.039)	(0.034)	(0.038)	(0.033)

Table a4. Results of logistic regression on platform work potential risk factors, average marginal effects, sample of 16 EU countries, platform workers who do platform work at least monthly, 16-64 (continued)

Variables	Low income	Stress	Long hours	Health risk	Working conditions (cumulative)	Overeducated	Undereducated
Main task on platform (ref = online professional services)							
Online clerical and data entry tasks	0.084	-0.114**	-0.061	-0.049	-0.084*		
	(0.069)	(0.049)	(0.047)	(0.048)	(0.049)		
Online creative and multimedia work	0.149	-0.025	-0.114**	-0.066	-0.096*		
	(0.124)	(0.051)	(0.050)	(0.052)	(0.051)		
Online sales and marketing support work	0.144*	-0.061	-0.073	-0.064	-0.069		
	(0.083)	(0.053)	(0.050)	(0.052)	(0.053)		
Online software development and technology	0.343***	-0.094	-0.034	-0.090	-0.118**		
	(0.086)	(0.058)	(0.058)	(0.057)	(0.055)		
Online writing and translation work	0.310***	-0.172***	-0.184***	-0.131**	-0.190***		
	(0.095)	(0.052)	(0.052)	(0.053)	(0.052)		
Online microtasks	0.296***	-0.281***	-0.313***	-0.231***	-0.244***		
	(0.085)	(0.051)	(0.050)	(0.051)	(0.048)		
Interactive services	0.329**	-0.136**	-0.060	0.002	-0.116*		
	(0.161)	(0.064)	(0.062)	(0.061)	(0.061)		
Transportation and delivery services	0.113	-0.048	-0.085	-0.140**	-0.097		
	(0.098)	(0.066)	(0.063)	(0.066)	(0.066)		
On-location services	0.232**	-0.059	-0.039	-0.024	-0.061		
	(0.091)	(0.065)	(0.061)	(0.065)	(0.069)		
Observations	475	3,628	3,628	3,628	3,628	3,800	3,800
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Regression results from logistic regression stress, long hours, health risk and a cumulative measure of working conditions.
All regression results weighted.
Regressions on low income only include respondents who gain at least 50% of their income from platform work.

Table a5. Results of logistic regression on platform work potential risk factors, average marginal effects, platform workers who do platform work at least weekly

Variables	Low income	Stress	Long hours	Health risk	Working conditions (cumulative)	Overeducated	Undereducated
Female	-0.015	0.028	0.048*	-0.016	-0.010	0.069**	-0.078***
	(0.059)	(0.026)	(0.025)	(0.030)	(0.026)	(0.030)	(0.030)
Age (ref = 16-25)							
26-40	-0.263***	-0.004	0.132***	0.098***	0.105***	0.031	-0.075**
	(0.068)	(0.033)	(0.032)	(0.037)	(0.032)	(0.035)	(0.038)
41-55	-0.109	-0.148***	0.003	-0.094**	-0.084**	0.087**	-0.110**
	(0.085)	(0.038)	(0.038)	(0.043)	(0.038)	(0.042)	(0.043)
56-74	-0.265	-0.162***	-0.016	-0.082	-0.071	0.075	-0.124**
	(0.174)	(0.057)	(0.054)	(0.062)	(0.054)	(0.060)	(0.054)
Education (ref = primary and lower secondary)							
Upper secondary and post-secondary non-tertiary	-0.228*	-0.103**	-0.082**	-0.110**	-0.116***		
	(0.125)	(0.043)	(0.040)	(0.047)	(0.042)		
Tertiary	-0.251**	-0.144***	-0.123***	-0.176***	-0.173***		
	(0.119)	(0.041)	(0.039)	(0.046)	(0.041)		
Couple	-0.108**	0.037	0.035	0.059*	0.058**	0.013	-0.072**
	(0.049)	(0.028)	(0.027)	(0.032)	(0.028)	(0.031)	(0.031)
Dependent children in household	-0.197***	0.073**	0.100***	0.092***	0.101***	-0.061*	-0.032
	(0.053)	(0.029)	(0.027)	(0.033)	(0.029)	(0.032)	(0.032)
Foreign-born	0.058	-0.005	0.035	0.036	0.075*	-0.163***	0.107***
	(0.056)	(0.040)	(0.038)	(0.045)	(0.039)	(0.043)	(0.038)

Table a5. Results of logistic regression on platform work potential risk factors, average marginal effects, platform workers who do platform work at least weekly (continued)

Variables	Low income	Stress	Long hours	Health risk	Working conditions (cumulative)	Overeducated	Undereducated
Main task on platform (ref = online professional services)							
Online clerical and data entry tasks	0.059	-0.112**	-0.067	-0.028	-0.049		
	(0.072)	(0.049)	(0.046)	(0.053)	(0.048)		
Online creative and multimedia work	0.138	-0.026	-0.109**	-0.029	-0.061		
	(0.135)	(0.051)	(0.050)	(0.059)	(0.052)		
Online sales and marketing support work	0.109	-0.063	-0.078	-0.105*	-0.065		
	(0.091)	(0.052)	(0.050)	(0.058)	(0.052)		
Online software development and technology	0.336***	-0.095	-0.034	-0.033	-0.086		
	(0.098)	(0.058)	(0.057)	(0.064)	(0.057)		
Online writing and translation work	0.336***	-0.166***	-0.190***	-0.097	-0.129**		
	(0.098)	(0.052)	(0.051)	(0.060)	(0.052)		
Online microtasks	0.323***	-0.280***	-0.317***	-0.264***	-0.232***		
	(0.105)	(0.051)	(0.049)	(0.056)	(0.051)		
Interactive services	0.209	-0.141**	-0.073	-0.001	0.014		
	(0.163)	(0.063)	(0.060)	(0.071)	(0.061)		
Transportation and delivery services	-0.001	-0.043	-0.088	-0.192**	-0.137**		
	(0.093)	(0.065)	(0.062)	(0.075)	(0.065)		
On-location services	0.201*	-0.052	-0.040	-0.036	-0.024		
	(0.106)	(0.065)	(0.061)	(0.076)	(0.065)		
Observations	419	2,769	2,769	2,769	2,769	2,904	2,904
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Regression results from logistic regression stress, long hours, health risk and a cumulative measure of working conditions.
All regression results weighted.
Regressions on low income only include respondents who gain at least 50% of their income from platform work.

Table a6. Results of linear probability model, regression on low income, sample of 16 EU countries, platform workers who do platform work at least weekly, 16-64

	Y = Low income	
	Model 1	**Model 2**
Female	-0.026	-0.005
	(0.050)	(0.051)
Age (ref = 16-25)		
26-40	-0.229***	-0.223***
	(0.065)	(0.061)
41-55	-0.118	-0.117
	(0.098)	(0.087)
56-64	-0.243*	-0.192
	(0.137)	(0.137)
Education (ref = primary and lower secondary)		
Upper secondary and post-secondary non-tertiary	-0.268**	-0.249**
	(0.115)	(0.107)
Tertiary	-0.267**	-0.259**
	(0.116)	(0.107)
Couple	-0.142**	-0.112*
	(0.066)	(0.057)
Dependent children in household	-0.303***	-0.268***
	(0.072)	(0.068)
Foreign-born	0.018	0.041
	(0.077)	(0.064)
Main task on platform (ref = online professional services)		
Online clerical and data entry tasks		0.042
		(0.064)
Online creative and multimedia work		0.105
		(0.096)
Online sales and marketing support work		0.076
		(0.077)
Online software development and technology		0.336***
		(0.098)
Online writing and translation work		0.308***
		(0.112)
Online microtasks		0.299***
		(0.108)
Interactive services		0.220
		(0.138)
Transportation and delivery services		-0.020
		(0.097)
On-location services		0.161
		(0.122)
Observations	417	417
Country FE	Yes	Yes

Note: Standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.
Regression results from logistic regression on low income.
All regression results weighted.
Regressions on low income exclude workers who gain less than 50% of their income in the platform economy.

Table a7. Results of linear probability model, regression on different measures of working conditions, sample of 16 EU countries, platform workers who do platform work at least weekly, 16-64

	Y = Stress		Y = Long hours		Y = Health risk		Y = Working conditions (cumulative)	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Female	-0.001	0.005	0.051*	0.057**	-0.023	-0.017	0.029	0.033
	(0.032)	(0.032)	(0.030)	(0.029)	(0.031)	(0.031)	(0.031)	(0.031)
Age (ref = 16-25)								
26-40	-0.007	-0.004	0.129***	0.128***	0.096**	0.097***	0.117***	0.118***
	(0.038)	(0.038)	(0.036)	(0.035)	(0.038)	(0.037)	(0.037)	(0.037)
41-55	-0.171***	-0.160***	-0.006	-0.001	-0.105**	-0.095**	-0.049	-0.041
	(0.045)	(0.044)	(0.045)	(0.043)	(0.045)	(0.043)	(0.040)	(0.040)
56-64	-0.095	-0.094	-0.037	-0.052	-0.059	-0.066	-0.011	-0.019
	(0.074)	(0.073)	(0.072)	(0.069)	(0.071)	(0.069)	(0.066)	(0.063)
Education (ref = primary and lower secondary)								
Upper secondary and post-secondary non-tertiary	-0.094*	-0.087*	-0.061	-0.059	-0.103**	-0.103**	-0.126**	-0.125**
	(0.050)	(0.050)	(0.048)	(0.047)	(0.050)	(0.049)	(0.050)	(0.050)
Tertiary	-0.123***	-0.122**	-0.097**	-0.100**	-0.159***	-0.170***	-0.167***	-0.174***
	(0.048)	(0.048)	(0.046)	(0.046)	(0.048)	(0.047)	(0.047)	(0.048)
Couple	0.046	0.026	0.057*	0.028	0.084**	0.059*	0.063*	0.045
	(0.034)	(0.034)	(0.033)	(0.032)	(0.033)	(0.033)	(0.033)	(0.033)
Dependent children in household	0.080**	0.074**	0.125***	0.110***	0.094***	0.090***	0.094***	0.087***
	(0.035)	(0.035)	(0.033)	(0.033)	(0.035)	(0.035)	(0.032)	(0.032)
Foreign-born	-0.027	-0.037	0.005	-0.009	0.055	0.041	0.053	0.045
	(0.046)	(0.046)	(0.042)	(0.042)	(0.045)	(0.045)	(0.046)	(0.041)

Table a7. Results of linear probability model, regression on different measures of working conditions, sample of 16 EU countries, platform workers who do platform work at least weekly, 16-64 (continued)

	Y = Stress		Y = Long hours		Y = Health risk		Y = Working conditions (cumulative)	
	Model 1	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Main task on platform (ref = online professional services)								
Online clerical and data entry tasks		-0.063		-0.044		-0.030		-0.042
		(0.057)		(0.050)		(0.055)		(0.059)
Online creative and multimedia work		-0.017		-0.143**		-0.037		-0.072
		(0.060)		(0.058)		(0.060)		(0.062)
Online sales and marketing support work		-0.073		-0.095*		-0.105*		-0.080
		(0.061)		(0.055)		(0.060)		(0.063)
Online software development and technology		-0.054		-0.046		-0.036		-0.087
		(0.067)		(0.063)		(0.064)		(0.068)
Online writing and translation work		-0.124**		-0.136**		-0.104*		-0.136**
		(0.063)		(0.060)		(0.061)		(0.063)
Online microtasks		-0.244***		-0.337***		-0.266***		-0.227***
		(0.061)		(0.055)		(0.057)		(0.058)
Interactive services		-0.119		-0.036		-0.004		-0.095
		(0.074)		(0.062)		(0.071)		(0.076)
Transportation and delivery services		-0.084		-0.092		-0.200**		-0.128
		(0.079)		(0.070)		(0.078)		(0.079)
On-location services		0.014		-0.054		-0.039		-0.054
		(0.077)		(0.071)		(0.078)		(0.082)
Observations	2,722	2,722	2,722	2,722	2,722	2,722	2,722	2,722
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Regression results from logistic regression stress, long hours, health risk and a cumulative measure of working conditions.
All regression results weighted.

Table a8. Results of linear probability model, regression on educational mismatch, sample of 16 EU countries, platform workers who do platform work at least weekly, 16-64

Variables	Overeducated	Undereducated
Female	0.068**	-0.071**
	(0.031)	(0.029)
Age (ref = 16-25)		
26-40	0.031	-0.075**
	(0.035)	(0.038)
41-55	0.087**	-0.108**
	(0.042)	(0.043)
56-64	0.040	-0.139**
	(0.066)	(0.059)
Education (ref = primary and lower secondary)		
Upper secondary and post-secondary non-tertiary		
Tertiary		
Couple	0.016	-0.081**
	(0.031)	(0.033)
Dependent children in household	-0.068**	-0.026
	(0.033)	(0.033)
Foreign-born	-0.153***	0.116**
	(0.038)	(0.045)
Observations	2,853	2,853
Country FE	Yes	Yes

Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Regression results from logistic regression overeducated and undereducated.
All regression results weighted.

Table a9. Share of platform workers gaining at least 50% of income from platform work and working on platform daily or almost daily by type of platform work, employees 16-64, platform workers doing platform work at least weekly, weighted sample

Type of platform work	Proportion gaining at least 50% of income from platform work	Proportion working on platforms daily or almost daily
Online clerical and data-entry tasks (e.g. customer services, data entry, transcription and similar)	17.00	52.06
Online professional services (e.g. accounting, legal, project management and similar)	21.54	52.87
Online creative and multimedia work (e.g. animation, graphic design, photo editing and similar)	12.99	46.49
Online sales and marketing support work (e.g. lead generation, posting ads, social media management, search engine optimisation and similar)	12.8	44.48
Online software development and technology work (e.g. data science, game development, mobile development and similar)	12.09	48.06
Online writing and translation work (e.g. article writing, copywriting, proofreading, translation and similar)	12.20	42.46
Online microtasks (e.g. object classification, tagging content review, website feedback and similar)	8.24	45.52
Interactive services (e.g. language teaching, interactive online lessons, interactive consultations and similar)	4.17	41.57
Transportation and delivery services (e.g. driving, food delivery, moving services and similar)	8.83	36.06
On-location services (e.g. housekeeping, beauty services, on-location services and similar)	13.32	38.81

References

Aleksynska, M., Bastrakova, A., & Kharchenko, N. (2019). *Working Conditions on Digital Labour Platforms: Evidence from a Leading Labour Supply Economy*. IZA Discussion Paper No. 12245.

Allaire, N., *et al.* (2019). *Covering risks for platform workers in the digital age*. Available at: https://www.sciencespo.fr/public/chaire-numerique/wp-content/uploads/2019/05/covering-risks-plateform-workers-digital-age.pdf.

Allen, J., & van der Velden, R. (2001). Educational mismatches versus skill mismatches: Effects on wages, job satisfaction, and on-the-job search', *Oxford Economic Papers*, *53*(3), 434–452. doi: 10.1093/oep/53.3.434.

Anxo, D., & Karlsson, M. (2019). *Overtime work: A review of literature and initial empirical analysis*. Conditions of Work and Employment Series No. 104. Geneva: International Labour Office.

Berg, J. (2016). Income security in the on-demand economy: findings and policy lessons from a survey of crowdworkers, *Comparative labor law and policy journal*, *37*(3), 506–543.

Berg, J. M., *et al.* (2018). *Digital labour platforms and the future of work: Towards decent work in the online world.* Geneva: International Labour Office.

Blau, F.D., & Kahn, L.M. (2017). The Gender Wage Gap : Extent, trends, and Explanations, *Journal of Economic Literature*, *55*(3), 789–865.

Cantarella, M., & Strozzi, C. (2019). *Workers in the Crowd: The Labour Market Impact of the Online Platform Economy.* IZA Discussion Paper No. 12327.

Card, D. (1999). The causal effect of education on earnings, in Ashenfelter, O. & Card, D. (eds) *Handbook of Labor Economics, Volume 3*. Amsterdam: Elsevier, pp. 1801–1863. doi: 10.1016/S1573-4463(99)03011-4.

Cook, C. *et al.* (2020). The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers. NBER Working Paper 24732. doi: 10.3386/w24732.

EU OSHA (2017). *Protecting workers in the online platform economy: An overview of regulatory and policy developments in the EU.* Luxembourg: Publications Office of the European Union. doi: 10.2802/918187.

Eurofound (2018). *Employment and working conditions of selected types of platform work*. Luxembourg: Publications Office of the European Union. doi: 10.2806/42948.

Eurostat (2010). *Health and safety at work in Europe*. Luxembourg: Publications Office of the European Union. doi: 10.2785/38630.

Fabo, B., Karanovic, J., & Dukova, K. (2017). In search of an adequate European policy response to the platform economy, *Transfer*, 23(2), 163–175. doi: 10.1177/1024258916688861.

Foong, E., *et al.* (2018). Women (Still) Ask For Less, *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW). doi: 10.1145/3274322.

Forde, C., *et al.* (2017). *The Social Protection of Workers in the Platform Economy*. Luxembourg: Publications Office of the European Union.

Fournier, J.-M., & Koske, I. (2012). The Determinants of Earnings Inequality: Evidence From Quantile Regressions', *OECD Journal: Economic Studies*, 2012(1), 5–36.

Garben, S. (2019). *Tackling social disruption in the online platform economy*. FEPS Policy Paper. Available at: https://www.feps-europe.eu/attachments/publications/feps paper - garben - clean final.pdf.

Gomez-Herrera, E., Martens, B., & Mueller-Langer, F. (2018). *Trade, Competition and Welfare in Global Online Labour Markets: A 'Gig Economy' Case Study*. JRC Digital Economy Working Paper 2017-05. Luxembourg: Publications Office of the European Union. doi: 10.2139/ssrn.3090929.

Green, S. B. (1991). How many subjects does it take to do a regression analysis?, *Multivariate Behavioral Research*, 26(3), 499–510.

De Groen, W. P., & Maselli, I. (2016). *The Impact of the Collaborative Economy on the Labour Market*, CEPS Special Report No. 138. Available at: http://aei.pitt.edu/76467/.

ILO (2012). *International Standard Classification of Occupations: Structure, Group Definitions and Correspondence Tables*. Geneva: International Labour Office.

Jain, A., & Leka, S. (2019). Occupation and Health, in Llewellyn, C. D. *et al.* (eds) *Cambridge Handbook of Psychology, Health and Medicine*. Cambridge: Cambridge University Press, pp. 28–31.

Kilhoffer, Z., *et al.* (2020.) *Study to gather evidence on the working conditions of platform workers*. Luxembourg: Publications Office of the European Union. doi: 10.2767/26582.

Kodz, J., *et al.* (2003). *Working long hours: a review of the evidence. Volume 1—Main report*. Employment Relations Research Series 16. doi: 10.13140/2.1.4808.3527.

Lee, S., McCann, D., & Messenger, J.C. (2007). *Working time around the world: Trends in working hours, laws and policies in a global comparative perspective*. Geneva: International Labour Organization. doi: 10.4324/9780203945216.

Mackenzie, C., Rogers, W. & Dodds, S. (2014). Introduction: What is Vulnerability and Why Does It Matter for Moral Theory?', in Mackenzie, C., Rogers, W., and Dodds, S. (eds) *Vulnerability: New essays in ethics and feminist philosophy*. Oxford: Oxford University Press.

Marinaccio, A., *et al.* (2013). 'The relevance of socio-demographic and occupational variables for the assessment of work-related stress risk', *BMC Public Health*, *13*(1). doi: 10.1186/1471-2458-13-1157.

Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it, *European Sociological Review*, 26(1), 67–82. doi: 10.1093/esr/jcp006.

OECD (2013). *Skills Outlook 2013: First Results from the Survey of Adult Skills.* Paris: OECD Publishing.

Pérez, E.R., *et al.* (2012). Differences in working conditions and employment arrangements among migrant and non-migrant workers in Europe', *Ethnicity and Health*, *17*(6), 563–577. doi: 10.1080/13557858.2012.730606.

Pesole, A., *et al.* (2018). *Platform Workers in Europe: Evidence from the COLLEEM Survey*. Luxembourg: Publications Office of the European Union. doi: 10.2760/742789.

Piha, K., *et al.* (2013). Socio-economic and occupational determinants of work injury absence, *European Journal of Public Health*, *23*(4), 693–698. doi: 10.1093/eurpub/cks162.

Quintini, G. (2011a). *Over-Qualified or Under-Skilled : A Review of Existing Literature*. OECD Social, Employment and Migration Working Papers No. 121. Paris: OECD Publishing.

Quintini, G. (2011b). *Right for the Job : Over-Qualified or Under-Skilled ?* OECD Social, Employment and Migration Working Papers No. 120. Paris: OECD Publishing.

Risak, M. (2018) 'Fair working conditions for platform workers : Possible regulatory approaches at the EU level'. Available at: https://library.fes.de/pdf-files/id/ipa/14055.pdf.

Schor, J.B. (2017). Does the sharing economy increase inequality within the eighty percent?: Findings from a qualitative study of platform providers', *Cambridge Journal of Regions, Economy and Society*, *10*(2), 263-279. doi: 10.1093/cjres/rsw047.

Schor, J.B., *et al.* (2020). Dependence and precarity in the platform economy, *Theory and Society*. [online]. doi: 10.1007/s11186-020-09408-y.

Schulte, P.A., *et al.* (2012). Interaction of occupational and personal risk factors in workforce health and safety, *American Journal of Public Health*, *102*(3), 434–448. doi: 10.2105/AJPH.2011.300249.

Schwellnus, C., *et al.* (2019). *Gig economy platforms: Boon or Bane ?* OECD Economics Department Working Papers No. 1550. Paris: OECD Publishing. doi: 10.1787/fdb0570b-en.

De Smet, P., *et al.* (2005). Gender and regional differences in perceived job stress across Europe, *European Journal of Public Health*, *15*(5), 536–545. doi: 10.1093/eurpub/cki028.

Urzì Brancati, C., Pesole, A., & Fernandez Macias, E. (2020). *New evidence on platform workers in Europe: Results from the second COLLEEM survey*. Luxembourg: Publications Office of the European Union. doi: 10.2760/459278.

Webster, J. (2016). Microworkers of the Gig Economy', *New Labor Forum*, *25*(3), 56–64. doi: 10.1177/1095796016661511.

InGRID

Supporting expertise in inclusive growth

www.inclusivegrowth.eu

Deliverable 12.3

WHAT DETERMINES PRICE IN PLATFORM WORK?

An empirical study on the impact of ratings, experience, and duration

Zachary Kilhoffer

April 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 730998

Abstract

Platform work is a fast-growing form of non-standard work characterised by an online platform that intermediates paid services. Among the main concerns discussed in the platform work literature at present are fair pay, ratings, and algorithmic management. However, very little empirical evidence has attempted to shed light on these issues.

This report builds on a data set of platform worker profiles (n=1,420) retrieved by webscraping a German labour platform for cleaning services. This platform allows workers to set their hourly price, allowing analysis to better understand what determines price for platform workers.

Ratings are found to be a statistically significant, but economically insignificant, predictor of price. The most likely reason for this is that ratings are very inflated; essentially all platform workers are rated from 4-5. Experience is found to be statistically significant predictor of price; for every 100 tasks completed, hourly wages rise 2-4%. Duration on the platform is a significant and stronger predictor of price, as people who have been on the platform for longer earn more. This may be related to the importance of perceived trustworthiness for platform workers, whom clients invite into their homes to perform services. Gender and unemployment are also found to be significant, as men earn 3-4% less hourly than women, and an additional 1% unemployment reduces hourly earnings about 5%.

The findings suggest that ratings and reputation mechanics are an important topic for platform workers' earnings. The interface and design choices of platforms have real-world impacts on platform workers, which should elevate them as consequential issues in policy discussions.

Acknowledgements

I would like to thank, in no particular order, the following individuals for their invaluable feedback and discussions over the course of writing this paper: Hannah Johnston, Janine Berg, Six Silberman, Sara Baiocco, Mattia Di Salvo, Leonie Westhoff, Andreas Haupt, Daniel Gros, Karolien Lenaerts, and Stephanie Steinmetz.

I would further like to thank Annarosa Pesole and Uma Rani for their approval to use these data, which I gathered on the basis of an earlier project in cooperation with the Joint Research Centre. The data will also appear in their forthcoming European Commission Working Paper 'How many online workers? Estimated based on selected online web-based and location-based platforms'.

Contents

1. Introduction

Data on platform work are essential to understand the phenomenon, but good data are hard to come by for a number of technical and theoretical reasons.[34] This shortcoming is all too evident in efforts to define and measure the platform economy (e.g. the number of platform workers), which scholars have been attempting for years. Conceptual challenges in this task include definitional complexity and a lack of standardised terminology. On the technical side, data on platform work is not or cannot be gathered in most of the ways that data on other forms of work can be, such as national labour surveys and administrative reporting (Riso, 2019; Kilhoffer *et al.*, 2020).

This focus on the whole landscape of platforms is understandable; after all, the entire platform economy is more impactful than any single platform. Still, data on the platform economy are composed of data on individual platforms. While understanding the forest remains the ultimate goal, we can certainly gain from examining individual trees. This has a number of advantages, foremost of which is feasibility. Even though platforms are extremely diverse, better understanding a single platform can give important clues to how platform business models impact the workers.

This study attempts to shed light on certain factors about a particular platform – particularly, the relationship between a platform worker's characteristics (rating, gender, experience, etc.) and their earnings. This is relevant for decent working conditions, algorithmic management, and fairness in platform design.

Box 1: Selected platform description

The selected platform offers cleaning services, primarily in clients' homes. While based in and originating in Germany, it has expanded to other EU countries.

The selection process begins when a client enters a location (zip code), date, and time of day, then sees a list of candidate platform workers available at that time and place. The client sees each candidate's name, gender, average rating, experience (number of cleanings), duration (time since registering on the platform), and price per hour (which the workers themselves choose).

The client can then select a platform worker, who accepts or rejects the client's work offer. Upon acceptance, client and platform worker are put in contact to arrange logistics, and the platform worker meets the client and cleans for a specified number of hours. The client can rate the worker if they wish, and the worker's profile is updated to reflect their new average rating and experience. The service can be one-off or recurring.

Rather than relying on administrative data, surveys, or other more common methodologies, this case study uses data web-scraped in January and February 2020 from a German **cleaning platform**. This platform falls into a category of platform work sometimes called 'on-location' or 'location-dependent' (Eurofound, 2018; Brancati *et al.*, 2019).

A primary reason for selecting this platform is the pricing model. Unlike many on-location platforms, such as the big players offering personal transportation (Uber, Taxify) or food delivery (UberEats, Deliveroo, TakeAway), the selected platform allows platform workers to set their own hourly prices. Furthermore, many 'handyman' or multipurpose platforms intermediate an array of services, which makes it more difficult to compare similar work. By contrast, the selected platform only offers cleaning services.

In short, the characteristics of the selected platform allow us to gather data on and empirically explore policy-relevant research questions concerning prices and earnings on platform work. This

34 See discussion in Kilhoffer (2020a).

report aims to understand how platform workers set their price, and how signals visible to clients and platform workers impact this decision.

2. Theory

The analysis builds on **signalling theory**, which focuses on information asymmetries between at least two different market sides during the initiation of transactions. Platforms are multi-sided markets, making them an appropriate subject to apply the theory.

Sellers of goods and services know more about what they sell than clients, creating an information asymmetry. Traditional business transactions occur face-to-face, at least initially, which helps clients get a clearer idea of what they are purchasing, while establishing rapport between parties. This is usually not the case in platform work, and is not possible in many transactions occurring exclusively online (e.g. hiring a freelancer to design a graphic on 99Designs or UpWork). This creates an asymmetric information problem that platforms and platform workers must address.

Sellers can help to mitigate the information asymmetry and assure prospective clients by signalling the quality of their product service with descriptions, guarantees, warranties, or branding (Spence, 2002). Signals can be *conventional* (e.g. self-descriptions, promises), *handicap* (a product guarantee or well-crafted CV), or *index*. Index signals rely on some form of confirmation through an independent third party that used or experienced the product or service (Teubner *et al.*, 2017), and are generally considered the strongest type of signal (Aiken & Boush, 2006). The data scraped from the selected platform include several index signals:

1. average client rating;
2. number of jobs completed (experience);
3. duration on platform.

The first topic of interest is **client ratings** for platform workers. Ratings, typically shown as a number of stars between 1 and 5, are a critical component of platform work and online interactions more generally, and have become a **ubiquitous factor in addressing information asymmetry** (Belleflamme & Peitz, 2018). Van Doorn (2017: p. 903) writes:

> … *ratings have become a major decentralised and scalable management technique that outsources quality control to customers of on-demand platforms, creating a generalised audit culture in which service providers are continually pushed to self-optimise and cater to the customer's every whim.*

Ratings are essential to understanding how Uber created a business model where clients willingly enter a complete stranger's car. Caretaking, 'handyman' and cleaning platforms must go a step further, as clients invite strangers into their homes with or without supervision.[35]

For the service providers, **ratings are extremely consequential**. Platform workers' ratings can serve as a threshold for satisfactory work; Uber previously 'deactivated' drivers who failed to maintain a 4.6 out of 5 (Cook, 2015). Ratings attract or deter prospective clients, who primarily rely on reviews of earlier work to choose a platform worker (De Groen & Kilhoffer, 2019). Thus, ratings matter a great deal for platform workers' income and future job prospects (Huws *et al.*, 2019; Kilhoffer *et al.*,

35 Note that ratings are not the only risk mitigating factor for most platforms. For example, the selected platform offers clients insurance for property damaged during the cleaning.

2020). They also play a role in discrimination. For example, a worker's ethnicity might negatively impact their ratings (Ye *et al.*, 2017), or improve trust to counteract biases (Cui *et al.*, 2017).

Rating systems also provide an important source of **network effects**. The more platform workers on a platform, and the more ratings they have, the better informed clients are in their purchasing decisions. Rating systems are therefore part of self-reinforcing mechanisms that cause successful platforms to become more successful at the expense of smaller rivals (Belleflamme and Peitz, 2018). Similarly, within a single platform, a few platform workers with many reviews and high ratings are likely to edge out others, thereby attracting ever more clients in the future.

In spite of their importance, **rating systems are imperfect**. For example, Zervas et. al describe **rating inflation**, finding that 95% of AirBnB properties have an average rating between 4.5 and 5 stars, while virtually none have 3.5 or lower (2015). Wired Magazine described a situation familiar to many Uber users – a white-knuckle, harrowing ride through the city, which they rated a five star experience (Kane, 2015). Online marketplaces contain reviews that may be sincere, malicious, or fraudulent (Kaghazgaran *et al.*, 2017).

In platform work, five star reviews seem to be the default, while only the most unsatisfied clients leave one star reviews.[36] In previous interviews, workers of the selected platform suggested that below a four star rating, clients would never hire them (Eurofound, 2018). It is therefore unsurprising that platform workers and their advocates continue to call for more transparent and contestable review systems (Berg *et al.*, 2018).

Because ratings are important for both the platform (versus other platforms) and the platform workers themselves, I investigate **the relationship between average ratings and hourly price**. For the reasons explained above, and empirical evidence on how various signals translate to price (Edelman & Luca, 2014), **higher ratings are expected to increase the hourly price** a platform worker requests.[37]

H1: *Higher average ratings are positively correlated with hourly price.*
An additional index signal is **the number of completed tasks**. This serves to demonstrate a platform worker's experience, signalling trustworthiness and capacity to perform services. Some measure of experience is likely to be very important for platform workers, whom (on this platform) clients cannot select on the basis of more detailed job histories provided in a CV or similar.

H2: Higher experience is positively correlated with hourly price.
Next, **longer duration on the platform** is likely to signal that a platform worker is an established member of the community, which could impact the price a platform worker can successfully charge. For example, a platform worker who began yesterday is probably more likely to be fraudulent, as very little time has passed allowing fraud or malign intentions to be uncovered. Moreover, longer duration on the platform signals higher social capital, which may increase clients' satisfaction (Huang *et al.*, 2017; Teubner *et al.*, 2017).

H3: Longer duration on the platform is positively correlated with hourly price.
Finally, I expect that the number of completed tasks impacts how clients perceive the value of ratings, and thus explore an interaction effect between the average rating score and the total number of ratings. Consider two platform workers, each with an average five-star rating, but one having completed a single task, and one having completed a hundred. The latter would signal a consistent record of success and experience, whereas the former has too small a sample size to draw any strong con-

36 Two additional points on this note. First, some interviewed platform workers of the selected platform indicated that satisfied clients sometimes leave a four-star review, either not understanding or not caring that this greatly harms the worker. Some interviewees further described receiving reviews from elderly clients who misunderstood the star system. The written review was a glowing recommendation, but the client left a single star, apparently not understanding that this is the worst possible review.

37 On the other hand, platform workers, particularly those with more general skills (e.g. cleaning), have limited ability to set rates (De Groen & Kilhoffer, 2019). This would indicate that the size of the effect is not expected to be very large.

clusions.[38] With this theoretical grounding and empirical support (Gutt & Herrmann, 2015), I propose:

H4: The positive wage return to ratings increases with the number of tasks a platform worker has completed.

The expected relationships are summarised in Table 1.

Table 1. Summary of expectations for independent variables

Variable	Hypothesis	Models	Impact on Price
Rating	1	1, 9, 13	+
Experience	2	2, 10, 13	+
Duration	3	3, 11, 13	+
Interaction (rating*experience)	4	4, 12	+

* + indicates I expect a positive correlation between the independent and dependent variables.

I rely on four control variables: gender, unemployment, GDP per capita, and location (NUTS-3 region). Most general and platform work literature suggests that men earn more than women, which is also my expectation (Cook *et al.*, 2018; Aleksynska *et al.*, 2019). I also expect that areas of Germany with lower unemployment, higher GDP per capita, and higher living costs are associated with higher price per hour.[39] The main tool of analysis is ordinary least squares (OLS) linear regression.

2.1 Assumptions

A number of assumptions underlie this analysis. While those listed below are not exhaustive, they are useful to simplify the complex interactions between platform workers and clients in this online marketplace.

1. The platform operates on a hedonic price model, meaning that marketable features will be reflected in market prices (Rosen, 1974).
2. Clients and platform workers are aware that rating, experience, and duration signal the quality of services offered.
3. Platform workers charge the highest price per hour that allows them to continue finding work.
4. Clients seek to maximise the quality of services received and minimise price paid.

38 Or as Germans would say, *'Einmal ist keinmal'* – approximately 'Once is never'.

39 Unemployment data are from Eurostat representing NUTS-2 region in 2019. GDP per capita data are from Eurostat representing NUTS-3 region in 2019. Location is a categorical variable for NUTS-3 region, derived from the zip-code (*Postleitzahl*) used during webscraping.

3. Methodology

3.1 Data collection

To gather the data, I used Python to develop a webscraping programme.

Box 3: Webscraping description

Webscraping is a means of gathering data from websites. Webscraping typically makes use of automated tools developed with programming languages to extract data from webpages. In principle, anything that can be viewed online can be webscraped using a variety of techniques. Unstructured text, structured data, images, and anything else contained on a website can be retrieved.

Source: Kilhoffer (2020b: p. 8).

Webscraping has a number of pros and cons,[40] so researchers must be wary of ethical, practical, and legal restrictions on the practice. For example, webscraped data can be noisy, and not all websites are possible to scrape due to CAPTCHA, rate limiting, log-in requirements, etc. Nevertheless, webscraping can help equalise what researchers, policy-makers, and the platforms themselves know about their business models.

Typically, webscraping with Python uses the libraries Requests and BeautifulSoup, which is relatively quick and easy. However, the selected platform was unsuited for this method. More specifically, the platform's website does not list workers unless one specifies a date, time, and location. At this point, the website displays a list of workers who have made themselves available at this time and place. In testing, I was unable to recreate the request headers necessary to retrieve worker data using the website's API,[41] so I relied on a slower webscraping method.

With the goal of attaining a comprehensive dataset of workers on the platform, I used the Python library Selenium. Selenium is a tool for automating web browsers, primarily used for testing web applications, but it can also be used for webscraping. The main downside is that it is significantly slower and more computationally intensive than alternative methods.

I used Selenium to automate a web browser, simulating a human requesting available platform workers at various locations, dates, and times, and writing the data to CSV files. The process required a few weeks and took place in late January and early February 2020. I observed polite scraping protocol, placing delays between requests so as to not spam the platform's servers too frequently. I also ensured that the data were scraped and handle in accordance with the European Parliament's GDPR guidelines on webscraping.[42]

The webscraping programme searched all German zip codes at different dates and times. This aimed to ensure that platform workers would be found even if they are only sporadically available, such as only on weekends, or at certain times of day. Even so, only platform workers who made themselves available to work on the searched days and times could be retrieved. If they were registered on the platform but selected no dates to be available, they would not be found. Similarly, if a platform worker was taking a long break from the platform, they could not be retrieved.

40 See discussion in Kilhoffer (2020b).

41 Application programming interface.

42 The data gathered cannot be linked to a natural person; they do not contain last names, addresses, birthdays, emails, or anything else typically classified as protected, personal data. See European Parliament resolution (2016/2225(INI)) (European Parliament, 2017).

3.2 Pre-processing

The data needed to represent a single platform worker per row for analysis. Thus, we needed to merge and de-duplicate all of the CSV files created during webscraping. In some cases, the same platform worker was available in a number of zip codes, all days of the week, and virtually all times of day. Other platform workers were only sporadically available.

Exactly de-duplicating could not be performed using all columns of data, as certain aliases turned out to be common, and the scrape took place over a few weeks. I deliberately retrieved no worker ID number or other data that could definitively be linked to a person. In practice, this meant that some platform workers would have multiple rows where they were potentially listed with different numbers of completed cleanings, ratings, or durations. In exceptional cases, I had to manually check that workers with the same name and locations were not the same person. I could not be certain of six cases, where a row may or may not represent a duplicate of another person. These were removed from the dataset. Next, I added the NUTS-3 region for each platform worker based on zip-code searched.

3.3 Describing the data

The data consist of 1,420 observations, each representing a unique platform worker. Each worker has an alias, average rating, experience (count of completed assignments), duration on platform, hourly price, gender, and zip code.[43]

The average rating is a value from 1-5 to a single decimal place, or a blank if no client had yet left a rating. Experience was given as a whole number greater than five, except for those who had not yet completed five assignments; these were listed as '<5'.

The duration on the platform is listed in one of five categories: new, 1-6 months, 6-12 months, 1-3 years, or greater than 3 years. Hourly price is given in Euros and cents, while gender was given as male or female[44] based on the avatar displayed.

Table 2 Descriptive overview of continuous variables

	Count	Mean	Std.	Min.	25%	50%	75%	Max.
Price (€)	1,420	18.3	3.7	10.0	16.2	16.6	19.9	45
Rating (1-5)	1,065	4.7	0.4	1.0	4.6	4.8	4.9	5
Experience (# of jobs)	947	152.3	266.3	5	20	53	154	2,456

We see that hourly price is known for all 1,420 platform workers in the data. Average rating was known for 1,065, while the website displayed a blank for the remainder. In 947 cases, experience (the number of completed tasks) was listed as an integer. I discuss the treatment of these missing data below.

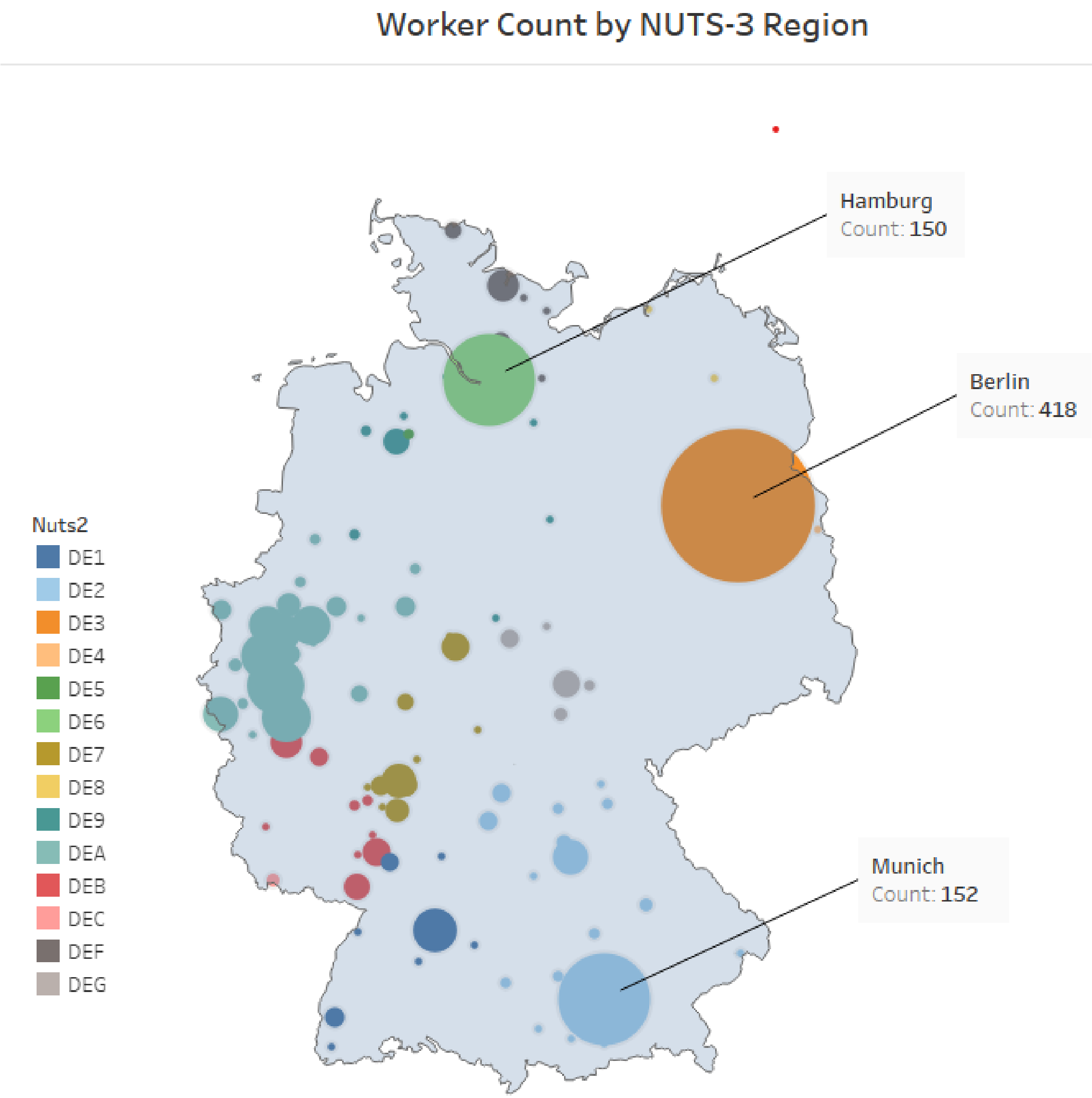
3.4 Analysis

Location data reveal platform workers are concentrated in the most urban areas of Germany (Figure 1). Berlin alone has almost 30% of all platform workers. The tendency for more activity in urban areas is unsurprising; other forms of on-location platform work are thought to be similar (Kilhoffer *et al.*, 2020).

43 From the search query used in the web-scraping programme.

44 Later in 2020, the platform began displaying profile pictures. This was not the case at the time of the scrape; all platform workers had either a generic female or male avatar.

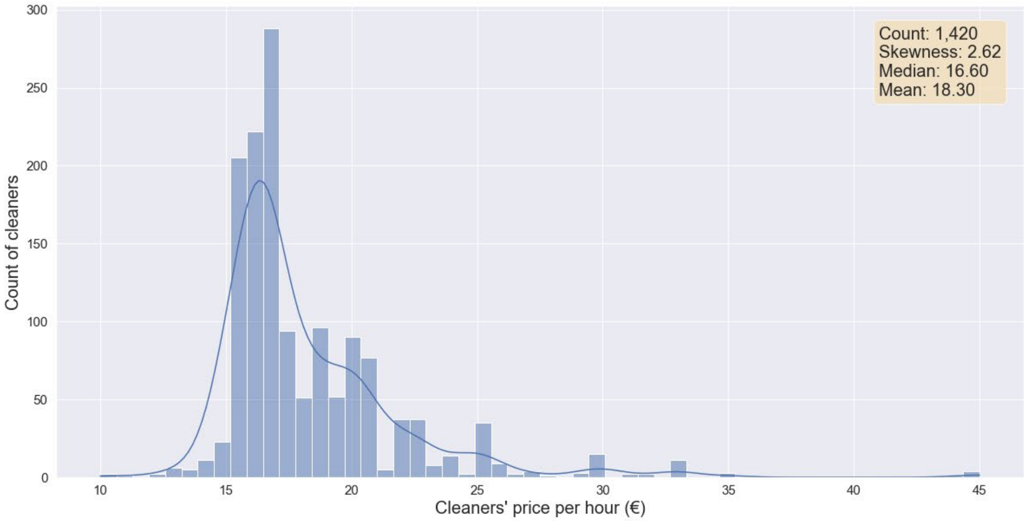
Figure 1. Distribution of platform workers in Germany



Note: The colour of each circle represents the NUTS2 region. Each NUTS-3 region with at least 2 platform workers has a circle shown. Larger circles indicate more platform workers.

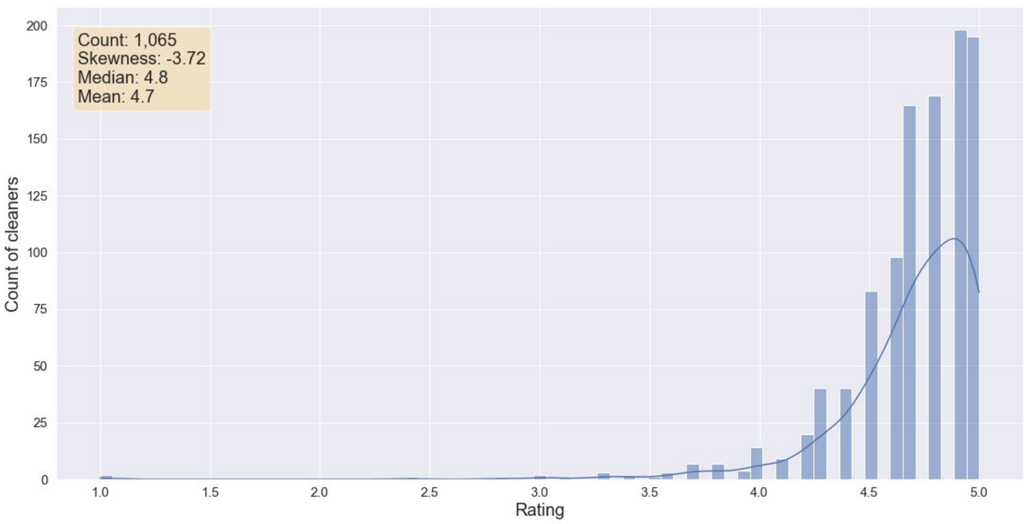
As shown in Figure 2, the distribution of prices has a long right tail with high skewness. A few outliers are notable here. While the mean price is 18.30 EUR per hour, and the 99th percentile is 33 EUR per hour, five workers charge over 44.00 EUR. These are not expected to be illustrative of larger trends; in fact, two of these individuals have virtually no experience and may not be setting realistic prices. To remove outliers, I calculate the z-score for price, then excluding observations with z-scores greater than 3.

Figure 2. Price histogram



The first explanatory variable is rating. Figure 3 shows a long left tail, indicating that lower ratings are very unusual, as discussed in the Theory section. Only 3.2% of the observations (34/1066) have a rating below 4, and the first percentile is 3.4. Exactly two people (0.14%) have a one-star rating. I therefore trim these outliers below the first percentile and leave NAs[45] in place.

Figure 3. Rating histogram



As discussed above, I assume that rating is related to price in two ways: having any rating, and the magnitude of the rating. In other words, there is a price penalty for having no rating, and there is a price reward for having a higher rating. In order to test both with the same variable, and to avoid excluding NAs from the analysis, I create an interaction variable.

First, I use a dummy variable, which is 1 if a cleaner has no rating, and 0 otherwise. Second, I replace NAs in the rating column with 0.5.[46] Third, I calculate the rating interaction as follows:

45 Not Applicable, or simply missing data.

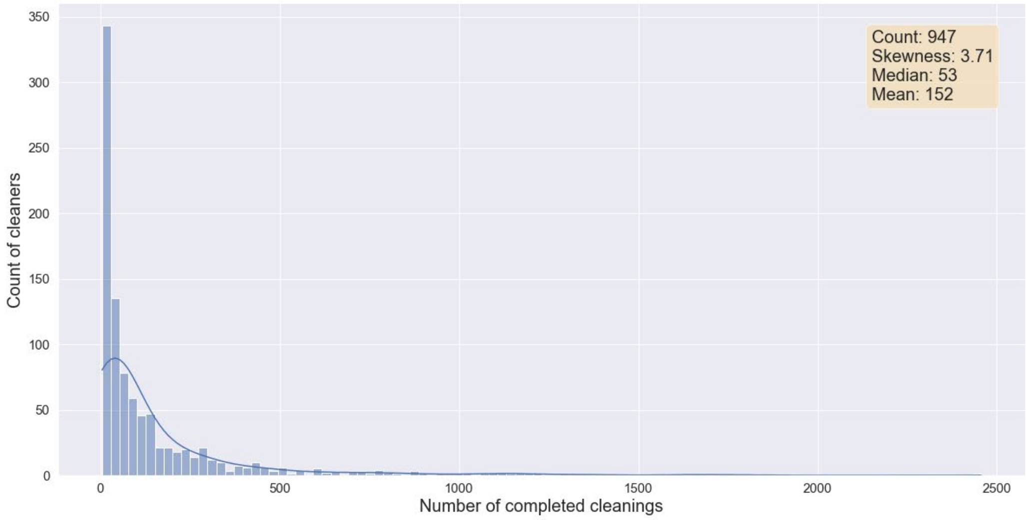
46 The value chosen to replace NA with does not matter, as it will be multiplied by 0.

Rating interaction = rating dummy + (1 – rating dummy) * (rating + 0.5)

In effect, this assigns a 1-star rating to those without a rating, while giving workers with a rating a 'bonus' of 0.5.

Moving to experience, Figure 4 shows the distribution, dropping all observations where the number of completed cleanings is '<5'.

Figure 4. Experience histogram



Before proceeding, I must handle the cleaner experience data shown as '<5'. For the moment, I replace '<5' with 2, being the mean of possible values: 0, 1, 2, 3, and 4. The new distribution is shown in Figure 5.

Figure 5. Experience histogram ('<5' replaced with 2)

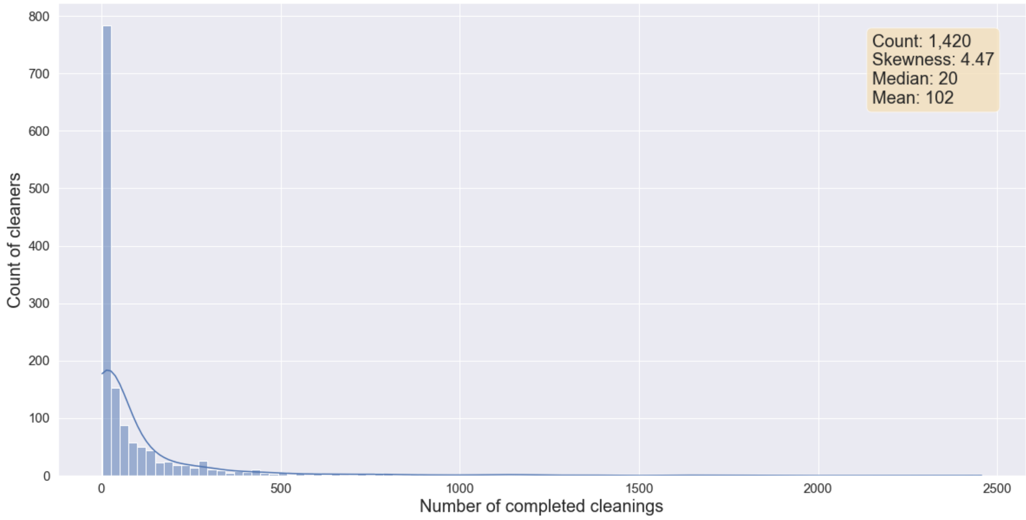


Figure 5 again shows an extremely long right tail and high skewness. The most experienced worker is an individual who has completed 2,456 tasks – far exceeding the second most experienced worker at 1,823. I thus remove outliers by excluding rows where the z-score for experience is greater than 3.

Similar to the rating variable, I assume that experience is related to price in two ways: having any (not '<5') experience, and the magnitude of experience.

I treat the experience variable similarly to rating. First, I create a dummy variable, which is 1 if a cleaner's experience is '<5', and 0 otherwise. Second, I replace '<5' in the experience column with 2. Third, I calculate the experience interaction as follows:

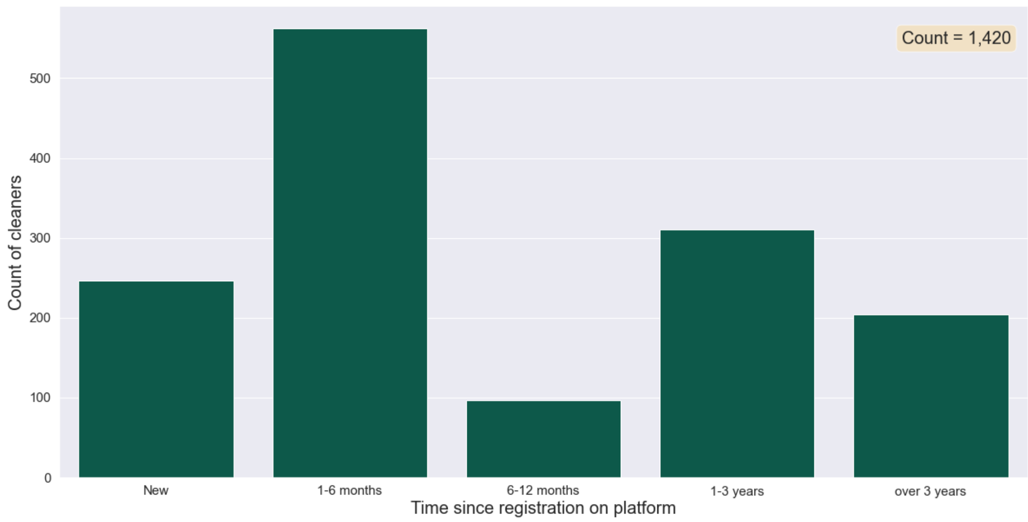
Experience interaction = experience dummy + (1 – experience dummy) * (experience + 5)

In effect, this assigns an experience of 1 to those who have less than five cleanings, while giving those with five or more cleanings a 'bonus' of 5.

While this strategy is not perfect, it does allow us to avoid excluding a large part of the dataset, while accounting for both the *presence of the information* (on rating and experience) and the *magnitude of the information*.[47]

As shown in Figure 6, duration is given in five categories. In order to compare how increasing levels of duration affect price in the regression, I set the reference level of the duration variable to 'New'.

Figure 6. Duration box plot

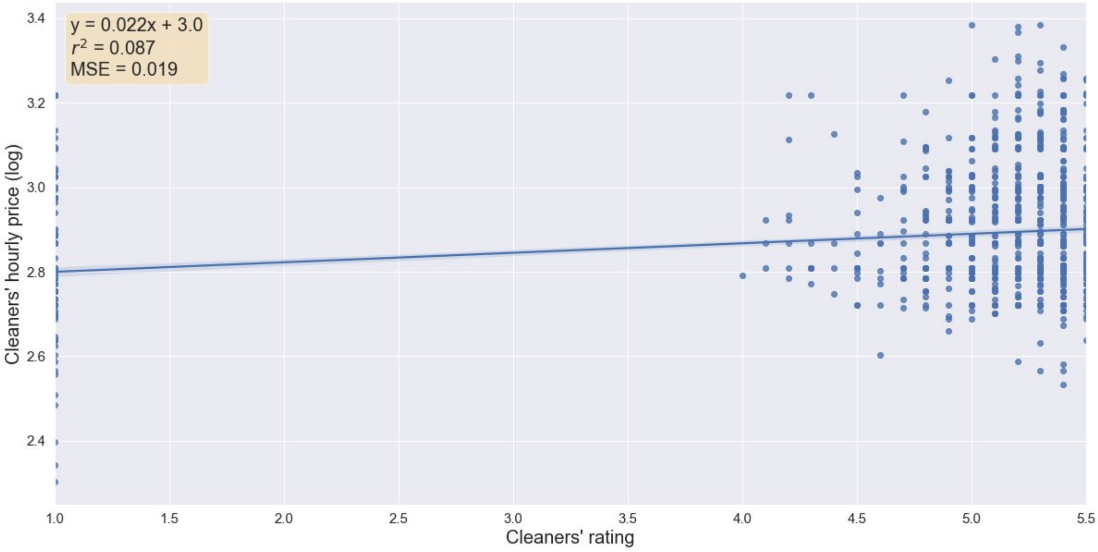


3.4.1 Bivariate analysis

Before running the full models, it is useful to check bivariate correlations between the independent and dependent variables. The first is rating, as shown in Figure 7, and corresponding to Hypothesis 1.

47 Alternative strategies would result in a smaller dataset or require imputing missing values. On the latter, attempts to impute missing values proved difficult and resulted in less precise models. Unfortunately, the dataset is not of a sufficient size, and lacks columns that may help with the imputation.

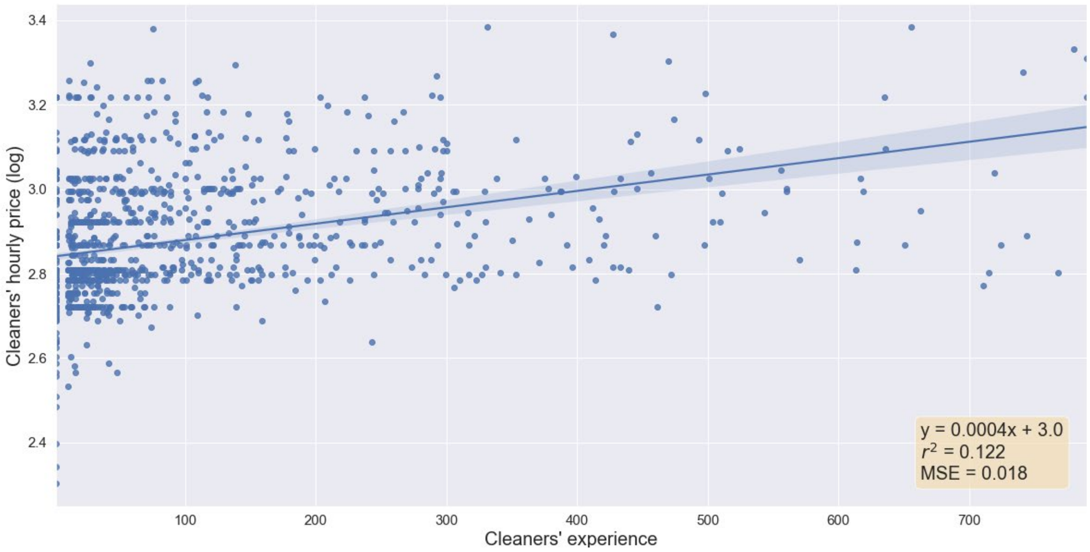
Figure 7. Cleaners' rating and price



Note: Cleaners' rating refers to the modified variable described above.

It does appear that people with higher ratings tend to have a higher price. Next, I plot cleaners' experience and price, as per Hypothesis 2.

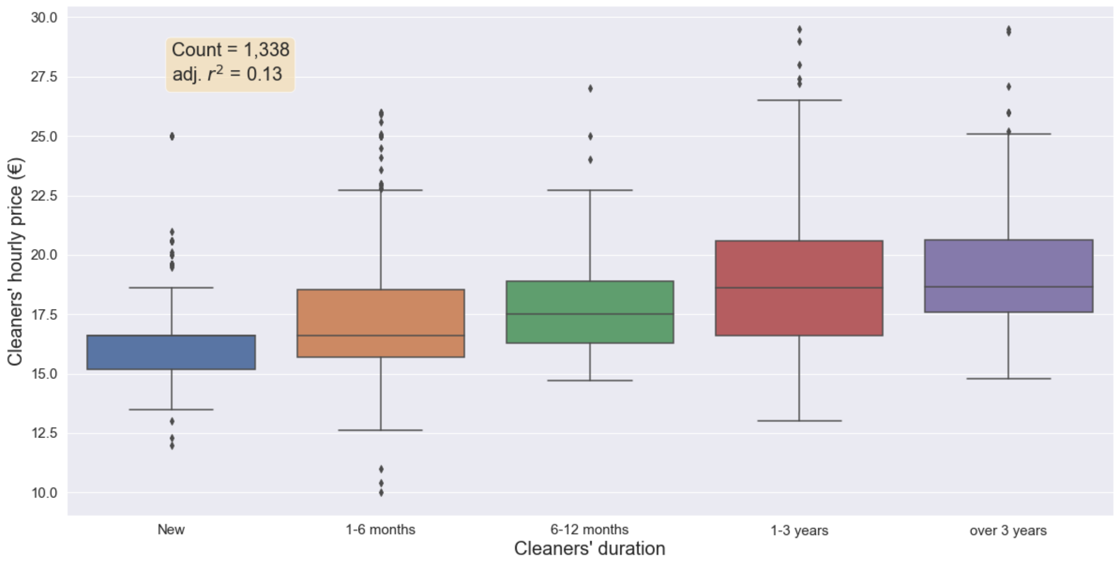
Figure 8. Cleaners' experience and price



Note: Cleaners' experience refers to the modified variable described above.
Source

Cleaners with more experience do seem to have a higher price. Next, I plot cleaners' duration and price, as per Hypothesis 3.

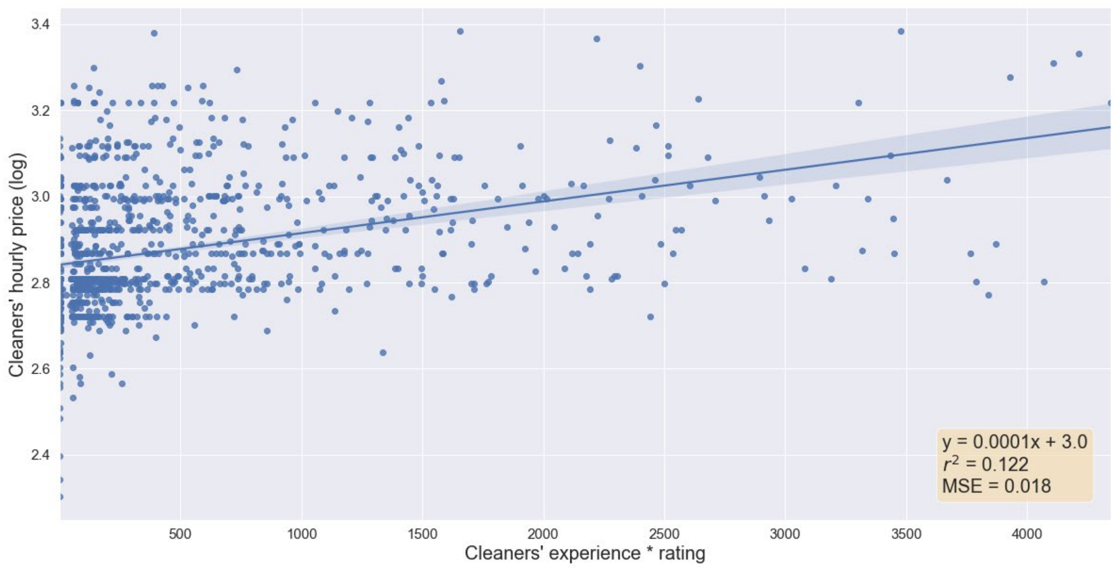
Figure 9. Cleaners' duration and price



As expected, higher duration on the platform seems related to higher price. Each successive category shows this relationship.

Corresponding with Hypothesis 4, I next create an interaction variable, rating times experience, and plot it against price.

Figure10. Cleaners' rating * experience and price

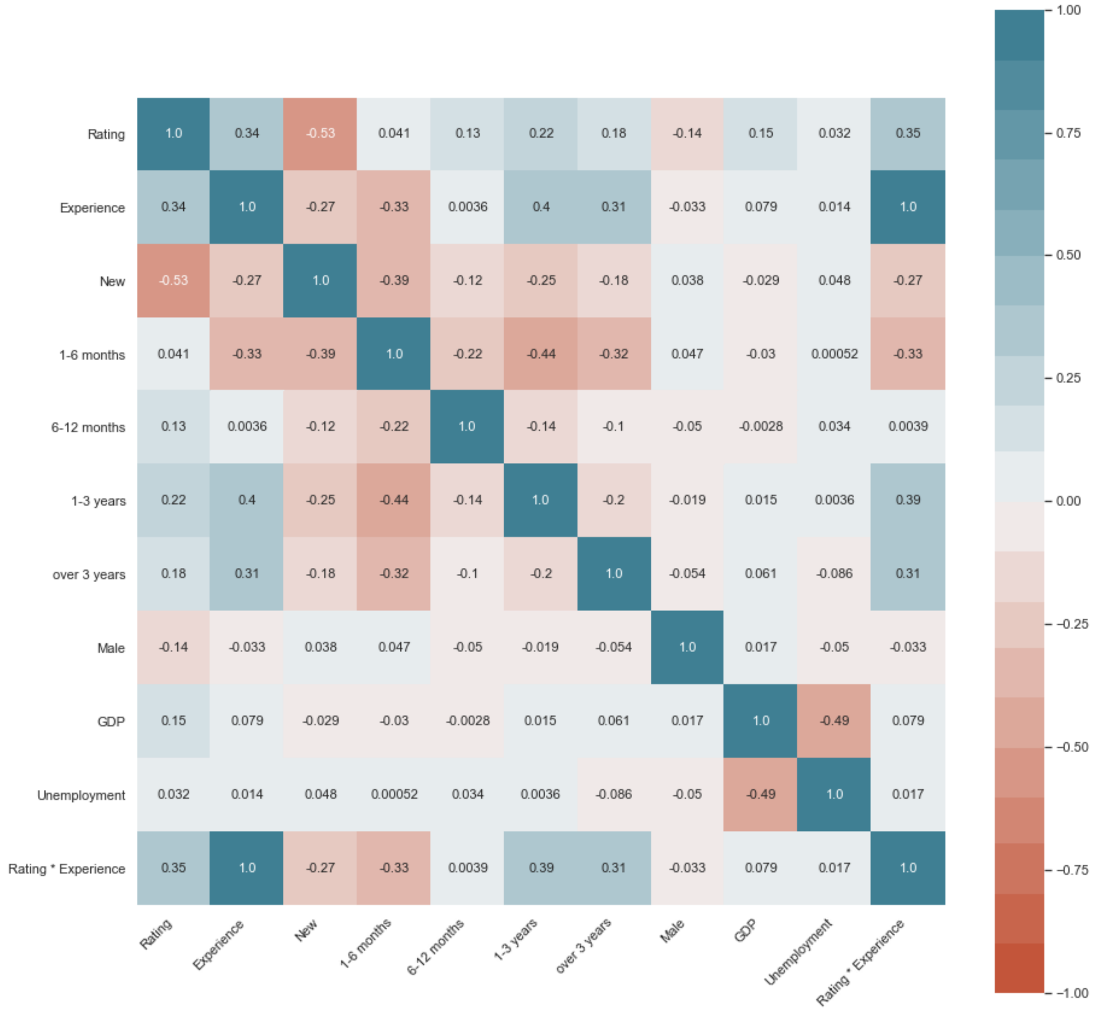


Note: The modified variables for rating and experience, as described above, were multiplied for this interaction variable.

It is not immediately clear if the interaction variable is much different than the experience variable alone, but this should become clearer in the regressions.

Next, I plot the independent and control variables against one another to understand their relationship. I suspect that experience and duration are quite closely related, as well as experience and the interaction variable of experience times rating. For the control variables, unemployment and GDP per capita may be very correlated as well. While I do not show all the variables for location (NUTS-3 region), these would also be very correlated with unemployment and GDP per capita.

Figure 11. Correlation matrix



Note: For the duration variable, each category was made into a Boolean.

A few of the variables are quite correlated. First, the 'New' cleaners have lower rating and experience. Correspondingly, the higher values of duration are correlated with higher experience, and to some extent, rating. Second, the control variables GDP per capita and unemployment are related, with higher GDP related to lower unemployment. Third, the interaction variable Rating * Experience is almost perfectly correlated with Experience. Upon further inspection, this seemed to be the case because there is very little variation in Rating (generally around 4.7), whereas Experience has a large range.

These results suggest that to avoid multicollinearity problems, models should not use location-based variables (Unemployment, GDP per capita, location) simultaneously, or use Experience and the interaction variable Rating * Experience simultaneously.

3.4.2 OLS

I next run linear regressions with ordinary least squares.[48] I use the data trimmed as described above, which reduces the number of observations from 1,420 to 1,338.

48 An additional option is quantile regression, which is sometimes used for pricing models.

The regressions proceed step-wise as outlined in Box 3, starting with independent variables alone, control variables alone, then each independent variable with control variables, before the final model, which is an attempt at best fit using selected independent and control variables.

Box 5: Regression model summaries

Independent variables alone

M1: log(Price) ~ Rating
M2: log(Price) ~ Experience
M3: log(Price) ~ Duration
M4: log(Price) ~ (Rating * experience interaction)

Control variables alone

M5: log(Price) ~ Gender (male)
M6: log(Price) ~ Unemployment
M7: log(Price) ~ GDP per capita
M8: log(Price) ~ NUTS-3

Independent and control variables

M9: log(Price) ~ Rating + Gender + Unemployment
M10: log(Price) ~ Experience + Gender + Unemployment
M11: log(Price) ~ Duration + Gender + Unemployment
M12: log(Price) ~ (Rating * experience interaction) + Gender + Unemployment
M13: log(Price) ~ Rating + Experience + Duration + Gender + Unemployment

4. Results

4.1 Models 1-4

Table 3. OLS regression results part 1

	Dependent variable: Cleaner price (log)			
	Model 1	**Model 2**	**Model 3**	**Model 4**
Rating	0.02251*** (0.00200)			
Experience		0.00039*** (0.00003)		
Rating*experience interaction				0.00007*** (0.00001)
Duration (1-6 months)			0.03746*** (0.01018)	
Duration (6-12 months)			0.07874*** (0.01648)	
Duration (1-3 years)			0.13129*** (0.01149)	
Duration (over 3 years)			0.14966*** (0.01315)	
Intercept	2.777*** (0.009)	2.841*** (0.004)	2.802*** (0.008)	2.841*** (0.004)
Observations	1,338	1,338	1,338	1,338
R^2	0.087	0.122	0.145	0.122
Adjusted R^2	0.086	0.121	0.143	0.121
Residual Std. Error	0.136	0.133	0.132	0.134
F Statistic	126.911***	185.871***	56.743***	184.945***

Note: *p<0.1; **p<0.05; ***p<0.01.
Five significant digits shown for coefficients, three significant digits shown elsewhere.

Models 1-4 mostly turned out as expected.

Model 1 shows that rating is positively and significantly correlated with price, though it is a rather weak predictor. Rating alone explains less than 9% of the variation in price, and moreover, an additional star rating (i.e. from 4 to 5) only results in a 2.25% rise in price. This is hardly impressive considering that almost all ratings fall between 4 and 5.

In Model 2, experience is positively and significantly correlated with price. However, quite a bit of experience is necessary to see meaningful results; every 100 additional cleanings results in about a 4% rise in hourly price. Experience also explains about 12% of the variation in price, making it a stronger predictor than rating.

In Model 3, all categories of duration are in relation to 'New'. As expected, each of them are significant, positive, and successively stronger. Compared to a new platform worker, one with 1 month of experience would earn 3.7% more hourly; after six months, 7.8% more; after one year, 13% more;

and those with over 3 years of experience earn about 15% more. Note that the difference between those with 1-3 years and over 3 years of experience is not nearly as large of a jump as with previous categories. The duration variables explain about 15% of the variance in price.

In Model 4,[49] the interaction term is positive and statistically significant, but like experience, very weak. An additional one star in rating, and an additional 100 cleanings, would result in a 4.5% higher hourly price. If rating were held constant, an additional 100 cleanings would result in a 3.1% higher price. While this does not intuitively seem like much, recall the way that the rating and experience variables were transformed. A person with no rating and no experience has had both values set to 1. As such, the model predicts that a person who performed just ten cleanings, and went from no rating to the median of 4.7, would have an hourly price 3.6% higher than someone just starting out. The interaction term therefore helps show that there is a price penalty associated with starting new as a platform worker. On the other hand, the adjusted R^2 for Model 4 is not better than that for Model 2 (experience alone). In short, the interaction term seems to have the predicted effect, but the impact is almost identical to experience alone. This may be due to the lack of variation in ratings.

4.2 Models 5-8

Table 4. OLS regression results part 2

	Dependent variable: Cleaner price (log)			
	Model 5	**Model 6**	**Model 7**	**Model 8**
Male	-0.035*** (0.008)			
Unemployment		-0.047*** (0.002)		
GDP per capita (EUR 1000s)			0.003*** (0.000)	
DE712				-0.188*** (0.032)
Intercept	2.891*** (0.006)	3.068*** (0.011)	2.710*** (0.010)	3.026*** (0.020)
Observations	1,338	1,338	1,338	1,338
R^2	0.014	0.219	0.186	0.436
Adjusted R^2	0.014	0.219	0.185	0.388
Residual Std. Error	0.141	0.126	0.129	0.111
F Statistic	19.513***	375.464***	304.534***	9.153***

Note: *p<0.1; **p<0.05; ***p<0.01
All figures shown to three significant digits.
DE712 is the first of the NUTS-3 regions, others shown in Appendix. See discussion below.

Models 5-8 show just the control variables.

Model 5 shows that gender is a significant indicator, but its predictive power is quite weak, explaining 1.4% of the variance in price. Unexpectedly, men earn around 3.5% less per hour than women.

49 Note that due to the high level of collinearity between Experience and the interaction term Rating * Experience, the effects on the dependent variable cannot be disentangled if run together in a regression. When running rating, experience, and the interaction variable together, the interaction variable loses its significance and the coefficient becomes slightly negative. Rating stays very significant but its coefficient weakens, while experience becomes slightly less significant, but its coefficient becomes much stronger. Interpreting these results is tricky, as the coefficient of individual variables would hold when the other variables have a value of 0, which cannot be the case in the data.

It may be that clients prefer to hire women to enter and clean their home, and therefore men offer slightly lower rates.

In Model 6, unemployment turns out to be quite a strong predictor. For areas with 1% higher unemployment, cleaners earn 4.7% less per hour. Unemployment explains about 22% of the variance in price, making it the best predictor so far.

Model 7 shows that GDP per capita also significantly and positively impacts price, explaining 18.5% of its variance. Cleaners in an area with 10,000 higher GDP per capita earn 3% more hourly. Given that GDP per capita in the data ranges from 20,000 to 180,000, this can make quite a large difference.

Model 8 shows the geographical location variable, NUTS-3 region. Because over 100 NUTS-3 regions are in the data, only one (DE712) is shown above, while the rest are in the Appendix. As shown in the Appendix, most locations prove to be significant determinants of hourly price. This is sensible for at least two reasons. First, wages and cost of living vary throughout Germany. For example, we would expect (and indeed observe)[50] that cleaners in Stuttgart charge more than average. Second, the amount of clients and platform workers in a given location impacts the supply and demand. If relatively few workers are available, and relatively many clients searching for services, then workers may have more margin with which to set their prices. The NUTS-3 regions explain a great deal of the variation in price - around 39% - making it the strongest predictor in the data.

All control variables prove to be significant. However, it is not possible to use unemployment and GDP per capita simultaneously (see discussion by Figure 11), and NUTS-3 regions (Model 8) add over one hundred variables to the model. To achieve the best and most parsimonious model, the next regressions will use only gender and unemployment as control variables.

50 Alphabetically, the first German NUTS-3 region (DE111) is central Stuttgart, which is the reference for other locations and not shown in the regression tables. Stuttgart recently surpassed Munich as the most expensive city in Germany. Almost all other regions show a negative and significant coefficient, meaning prices are significantly lower per hour than in Stuttgart.

4.3 Models 9-13

Table 5. OLS regression results part 3

	Dependent variable: Cleaner price (log)				
	Model 9	**Model 10**	**Model 11**	**Model 12**	**Model 13**
Rating	0.0226*** (0.0017)				0.0132*** (0.0020)
Experience		0.0004*** (0.0000)			0.0002*** (0.0000)
Duration (1-6 months)			0.0305*** (0.0088)		-0.0035 (0.0095)
Duration (6-12 months)			0.0759*** (0.0143)		0.0200 (0.0149)
Duration (1-3 years)			0.1231*** (0.0100)		0.0468*** (0.0118)
Duration (over 3 years)			0.1248*** (0.0115)		0.0455*** (0.0131)
Rating*experience interaction				0.0001*** (0.0000)	
Male	-0.0300*** (0.0066)	-0.0386*** (0.0064)	-0.0346*** (0.0064)	-0.0386*** (0.0064)	-0.0298*** (0.0062)
Unemployment	-0.0489*** (0.0023)	-0.0486*** (0.0022)	-0.0461*** (0.0022)	-0.0487*** (0.0022)	-0.0480*** (0.0021)
Intercept	2.999*** (0.013)	3.067*** (0.011)	3.023*** (0.013)	3.068*** (0.011)	3.002*** (0.013)
Observations	1,338	1,338	1,338		1,338
R^2	0.014	0.219	0.186		0.436
Adjusted R^2	0.014	0.219	0.185		0.388
Residual Std. Error	0.141	0.126	0.129		0.111
F Statistic	19.513***	375.464***	304.534***		9.153***

Note: *p<0.1; **p<0.05; ***p<0.01.
Four significant digits shown for coefficients, three significant digits shown for summary statistics.

Models 9-13 show the independent variables with control variables.

Model 9 is essentially identical to Model 1, showing that rating is a statistically significant, but rather economically insignificant, predictor of price. The control variables are more or less the same as when testing them independently.

Model 10 is a similar story, as experience and the control variables maintain the same relationship as previous regressions showed.

Model 11 is mostly the same as Model 3, except the coefficients are a bit weaker for 1-3 years and over 3 years. One interesting change is in the adjusted R^2. In Models 2 and 3, which had no control variables, duration was quite a bit better at explaining variance in price. After adding gender and unemployment, the models with experience and duration both explain around 36% of the variance in price.

Model 12, like Model 4, shows that the interaction term rating * experience is positively and significantly correlated with price. However, the coefficient became about 43% stronger (from 0.00007 to 0.0001). In this model, a person with a constant rating of 4.7, would see their hourly price rise 5.2% after an additional 100 cleanings. A person with experience of 105 and rating 4.9 would earn 5.4%

more than a person with rating of 4.4 and experience of 5. Again, rating is not very economically meaningful, unless you have no rating at all.

Finally, Model 13 shows the preferred regression with the most information, explaining 41% of the variance in price. Interestingly, the independent variables become a bit weaker when run together. An additional star rating now increases hourly price only 1.3%, and an additional 100 experience increases price by 2%. The story of experience could be more complicated than simply 'more experience, higher wages'. For example, platform workers in need of money may set lower prices to find work more quickly.

The first two categories of duration, 1-6 months and 6-12 months, lose their significance, and the longer durations, 1-3 years and over 3 years, only earn about 5% more than a baseline worker. This may indicate that price is best predicted by a mixture of multiple factors, which refines the story from Models 1-4, which made it seem like duration was a far better predictor than experience. Unemployment is virtually unchanged from previous models, while the coefficient of male stays about the same.

4.4 Summary

Models 1, 9, and 13 suggest that continuous rating is a statistically significant and positive determinant of price, though rating did not turn out to be a very economically significant indicator. Model 9 suggests that for each additional star rating, price increases 2.2%, where Model 13 shows an increase of 1.3%. This is hardly impressive, given the maximum star rating of 5, and that almost all cleaners fall around 4.7 to 4.8 stars. I nevertheless **accept the first hypothesis:** ***Higher average ratings are positively correlated with hourly price***.

Looking to cleaner experience, Models 2, 10, and 13 reveal a significant and positive correlation with price.[51] Model 8 shows that each additional 100 cleanings leads cleaners to request an additional 4% in hourly price. While Model 13 shows a weaker relationship – an additional 100 cleanings leads to an additional 2% in hourly price – the relationship stays otherwise the same. I therefore **accept the second hypothesis:** ***Higher experience is positively correlated with hourly price***.

In Models 3 and 11 for duration, the reference is those who have been on the platform for under one month. These models show that each successive category becomes more positive, while all are significant. Moreover, the effect of duration appears to be stronger than rating or experience. After one month on the platform, cleaners earn 3.5% more. Six months after starting, cleaners earn 7.5% more than their starting wage, and so on. While the predictive power of duration decreases in Model 13, it still shows the same general relationship with price. I therefore **accept the third hypothesis:** ***Longer duration on the platform is positively correlated with hourly price***.

Finally, we look to the interaction effect. Models 4 and 12 indeed show a significant and positive effect of rating * experience on price. This is grounds to accept the **fourth hypothesis:** ***The weight return to ratings increases with the number of tasks a platform worker has completed.*** However, I would note that it is unclear if the interaction effect of rating * experience performs better than experience alone.

51 In Model 5, where the interaction effect is added, the coefficient of experience is the unique effect of experience on price when rating = 0, which cannot exist. Thus, Model 5 does not provide evidence against Hypothesis 2.

5. Discussion

In this case study, I have presented a strategy to webscrape and analyse data from a labour platform. The analysis shows that the data mostly support the hypotheses; rating, experience, and duration are meaningful signals in platform marketplaces that demonstrably impact workers' earnings. The results illustrate the importance of ratings and evaluations in platform work, lending credence to the notion that **fair platform work requires close attention to rating and reputation systems**. This evidence is especially notable due to the limited flexibility cleaners have to set their prices. I expect rating is a stronger predictor for other types of platform work such as programming or consulting.

The results suggest that rating is a statistically significant, but economically insignificant, determinant of price. This may indicate that rating is more of a pass-fail mechanic, whereby clients will not hire a cleaner below a certain rating, but otherwise do not care very much. Alternatively, workers with low ratings may voluntarily drop out of the platform, or be removed, as in the case of Uber drivers.

A related idea that future researchers could explore is a longitudinal one; how likely are platform workers with low ratings to find future work, or drop off the platform? I suspect that the answer is quite a bit less likely to find work, and quite a bit more likely to drop off voluntarily or otherwise. Moreover, such effects may be more pronounced than that on price discussed in the present report. If so, it would further cement the notion that platform workers need transparency in rating and evaluation metrics, as well as the ability to contest unfair reviews.

I also note the importance of the region for price, as NUTS-3 region turns out to predict more of the variation in price than any other variables. Clearly areas with a higher cost of living would be expected to have higher labour costs. However, when reviewing the data and focusing on any specific NUTS-3 region, prices among platform workers are extremely similar. It may therefore be a sort of price collusion, where cleaners simply match their peers' prices and hope to be selected based on their availability or other factors.

It is also interesting that duration is such a strong indicator of price compared to ratings, at least when considered alone. In this particular type of platform work, one possibility is that maximising the quality of cleaning service is not the main priority of clients. Instead, they may simply want someone who can do a 'good enough' job of cleaning, while being trustworthy. Having a longer track record, rather than a stellar rating, may be more important in this respect. Alternatively, it could simply be that ratings are too inflated to be a good signal of quality.

On the other hand, it is rather strange to think that a cleaner who opened an account a year ago, then was inactive for a year, then began working suddenly, should merit a higher price. This is also interesting with regards to interface design. The platform determines which signals to show clients, and in the case of duration, even sets the value. Based on the results of this study, duration certainly seems to be a stronger indicator than clients' ratings, which are the only direct measure of service quality available.

While this case study does not focus on gender, it revealed an unexpected finding on this particular platform – on average, women earn 3% to 4% more than men. This result is very significant and consistent across all models. I propose two possibilities for the difference: first, clients generally see domestic work as women's work; and second, clients are less comfortable with inviting an unknown man into their home.[52] This may indicate that gender biases present in the general labour market

52 See, for example, van Walsum's research in the Netherlands: All of my informants agreed that Dutch employers generally saw domestic work as women's work and tended to prefer women (2011: p. 153).

persist in platform work, while not necessarily meaning that women's earnings always suffer as a result.

In conclusion, web-scraped data is a valuable tool to examine how labour platforms function, and investigate earnings for platform workers. It is also an interesting strategy to better understand the impacts of interface design. Researchers could further build on this line of inquiry by gathering longitudinal data on different platforms in different locations, and implementing more sophisticated tools than OLS. For example, Lasso (least absolute shrinkage and selection operator) may be more useful to sort out which variables best predict price, while improving prediction accuracy and enhancing interpretability (Brownlee, 2020). It would also be interesting to use machine learning techniques to analyse whether aliases/names or profile pictures impact prices. In this platform, for example, platform workers who look German or have a German-sounding name may earn more or less than their peers.

appendix 1

Table a1. OLS regression results part 2 (all regional variables)

	Dependent variable: Cleaner price (log)			
	Model 5	**Model 6**	**Model 7**	**Model 8**
Male	-0.035*** (0.008)			
Unemployment		-0.047*** (0.002)		
GDP per capita (EUR 1000s)			0.003*** (0.000)	
DE712				-0.188*** (0.032)
DE254				-0.004 (0.034)
DEA13				-0.157*** (0.030)
DE212				0.032 (0.022)
DE300				-0.240*** (0.021)
DE933				-0.104* (0.054)
DEA17				-0.138*** (0.032)
DE128				-0.133*** (0.050)
DE600				-0.130*** (0.023)
DEA23				-0.126*** (0.025)
DEB11				-0.087* (0.050)
DEA22				-0.105*** (0.027)
DEA11				-0.057** (0.028)
DEF0F DEG01				-0.179*** (0.036) -0.275*** (0.037)
DEF09				-0.118*** (0.039)
DEA52				-0.147*** (0.030)
DE731				-0.270*** (0.037)
DEF02				-0.246*** (0.033)
DEB12				-0.191*** (0.033)

	Dependent variable: Cleaner price (log)			
	Model 5	**Model 6**	**Model 7**	**Model 8**
DEC01				-0.226*** (0.067)
DEA55				-0.164 (0.113)
DEA1B				-0.217*** (0.047)
DEA1C				0.043 (0.081)
DEF0D				-0.112** (0.047)
DE711				-0.162*** (0.041)
DE941				-0.215*** (0.039)
DE713				-0.127*** (0.044)
DE408				-0.192*** (0.034)
DEA27				-0.124 (0.081)
DEB3J				-0.001 (0.081)
DE71A				-0.122*** (0.047)
DEA2D				-0.187*** (0.032)
DE71C				-0.045 (0.081)
DE272				-0.461*** (0.113)
DEB34				-0.104*** (0.036)
DEA15				-0.076 (0.081)
DEA1A				-0.160*** (0.044)
DE142				-0.098 (0.113)
DEB33				-0.059 (0.038)
DEA28				-0.026 (0.113)
DE115				-0.060 (0.081)
DE251				-0.217* (0.113)
DE939				-0.160** (0.081)
DEF0A				-0.241** (0.113)

	Dependent variable: Cleaner price (log)			
	Model 5	**Model 6**	**Model 7**	**Model 8**
DEA5A				-0.241*** (0.054)
DE24B				-0.241** (0.113)
DEA2B				-0.309*** (0.081)
DE724				-0.217*** (0.054)
DE252				-0.100* (0.059)
DEA45				-0.162** (0.081)
DE263				-0.270*** (0.050)
DEA36				-0.133*** (0.041)
DEA51				-0.139*** (0.050)
DEA32				-0.217* (0.113)
DE271				-0.241** (0.113)
DE232				-0.241*** (0.067)
DE131				-0.136*** (0.047)
DEB39				-0.241** (0.113)
DEG07				-0.241** (0.113)
DE40A				-0.263*** (0.067)
DE124				0.087 (0.113)
DEA54				-0.175*** (0.047)
DEA37				-0.058 (0.081)
DEA47				-0.218*** (0.050)
DE936				-0.305*** (0.113)
DEG06				-0.252*** (0.050)
DE242				-0.241*** (0.081)
DE26B				-0.243*** (0.050)
DEA26				-0.241*** (0.081)

	Dependent variable: Cleaner price (log)			
	Model 5	Model 6	Model 7	Model 8
DEG0F				-0.235*** (0.067)
DEG05				-0.241*** (0.081)
DE803				-0.241** (0.113)
DEF01				-0.259*** (0.054)
DEF08				-0.241** (0.113)
DE241				-0.241*** (0.081)
DE139				-0.241** (0.113)
DE717				-0.241** (0.113)
DE501				-0.305*** (0.113)
DE944				-0.273*** (0.081)
DEB31				-0.305*** (0.113)
DE114				-0.087 (0.113)
DE211				-0.241** (0.113)
DE734				-0.285** (0.113)
DE80J				-0.241** (0.113)
DE279				-0.241** (0.113)
DE943				-0.136* (0.081)
DE21N				-0.241** (0.113)
DE253				-0.051 (0.113)
DEA53				-0.217* (0.113)
DEB35				-0.188** (0.081)
DE913				-0.241** (0.113)
DE71E				-0.241** (0.113)
DE91C				-0.241** (0.113)
DE732				-0.241** (0.113)

	Dependent variable: Cleaner price (log)			
	Model 5	**Model 6**	**Model 7**	**Model 8**
DE127				-0.241**
				(0.113)
DE222				-0.241**
				(0.113)
DE403				-0.305***
				(0.113)
DEB3C				-0.305***
				(0.113)
DEA5B				-0.056
				(0.113)
DEA35				0.192**
				(0.081)
DE406				-0.211*
				(0.113)
DEB25				-0.103
				(0.113)
DE935				-0.217*
				(0.113)
Intercept	2.891***	3.068***	2.710***	3.026***
	(0.006)	(0.011)	(0.010)	(0.020)
Observations	1,338	1,338	1,338	1,338
R^2	0.014	0.219	0.186	0.436
Adjusted R^2	0.014	0.219	0.185	0.388
Residual Std. Error	0.141	0.126	0.129	0.111
F Statistic	19.513***	375.464***	304.534***	9.153***

Note: *p<0.1; **p<0.05; ***p<0.01.
All figures shown to three significant digits.

References

Aiken, K.D., & Boush, D.M. (2006). Trustmarks, objective-source ratings, and implied investments in advertising: investigating online trust and the context-specific nature of internet signals', *Journal of the Academy of Marketing Science*, *34*(3), 308–323.

Aleksynska, M. *et al.* (2019). Working Conditions on Digital Labour Platforms: Evidence from a Leading Labour Supply Economy, IZA, March (https://www.iza.org/publications/dp/12245/working-conditions-on-digital-labour-platforms-evidence-from-a-leading-labour-supply-economy).

Belleflamme, P., & Peitz, M. (2018). Inside the Engine Room of Digital Platforms: Reviews, Ratings, and Recommendations, *SSRN Electronic Journal* (https://www.ssrn.com/abstract=3128141).

Berg, J. *et al.* (2018). *Digital labour platforms and the future of work | Towards decent work in the online world*, ILO (https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_645337.pdf).

Brancati, M.C.U. *et al.* (2019). *New evidence on platform workers in Europe, JRC Technical Reports*, Joint Research Centre.

Brownlee, J. (2020). How to Develop LASSO Regression Models in Python (https://machinelearningmastery.com/lasso-regression-with-python/).

Cook, C. *et al.* (2018). *The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers*, w24732, National Bureau of Economic Research, Cambridge, MA, June (http://www.nber.org/papers/w24732.pdf).

Cook, J. (2015). Uber's internal charts show how its driver-rating system actually works, *Business Insider*.

Cui, R. *et al.* (2017), Discrimination with incomplete information in the sharing economy: Evidence from field experiments on Airbnb, *Harvard Business School*, pp. 1–35.

De Groen, W.P., & Kilhoffer, Z. (2019). *Employment and working conditions of selected types of platform work: On-location client-determined moderately skilled work*, Working Paper, Eurofound, September.

Edelman, B.G., & Luca, M. (2014). Digital discrimination: The case of Airbnb. com, *Harvard Business School NOM Unit Working Paper*, No. 14–054.

Eurofound (2018). *Employment and working conditions of selected types of platform work, Digital Age*, Eurofound, Luxembourg (https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef18001en.pdf).

European Parliament (2017). OPINION of the Committee on Employment and Social Affairs for the Committee on the Internal Market and Consumer Protection on a European agenda for the collaborative economy (2017/2003(INI)), Opinion 2017/2003(INI), European Parliament, Brussels, March (http://www.europarl.europa.eu/doceo/document/EMPL-AD-592420_EN.pdf?redirect).

Gutt, D., & Herrmann, P. (2015). *Sharing means caring? Hosts' price reaction to rating visibility*.

Huang, Z. *et al.* (2017). A Study on the Influence of Team Members' Social Media Interaction Perception on Employee Creativity, *American Journal of Industrial and Business Management*, *7*(11), 1209–1228.

Huws, U. *et al.* (2019). The Platformisation Of Work In Europe: Highlights from research in 13 European countries, Foundation for European Progressive Studies, July (http://researchprofiles.herts.ac.uk/portal/en/publications/the-platformisation-of-work-in-europe-highlights-from-research-in-13-european-countries(6cb3297c-861e-4d75-962e-5bc9ff031f53).html).

Kaghazgaran, P. *et al.* (2017). *Behavioral Analysis of Review Fraud: Linking Malicious Crowdsourcing to Amazon and Beyond*.

Kane, K. (2015). The Big Hidden Problem With Uber? Insincere 5-Star Ratings, *Wired*, 19 March (https://www.wired.com/2015/03/bogus-uber-reviews/).

Kilhoffer, Z. (2020a). Data on the platform economy - Working Paper D12.3, InGRID, December.

Kilhoffer, Z. *et al.* (2020). Study to gather evidence on the working conditions of platform workers, DG EMPL.

Kilhoffer, Z. (2020b). Best practice' Report on How social media can be used to understand labour market, Milestone 112, Leuven.

Riso, S. (2019). Mapping the contours of the platform economy, Working Paper, Eurofound, Dublin.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition, *Journal of political economy*, *82*(1), 34-55.

Spence, M. (2002). Signaling in retrospect and the informational structure of markets, *American Economic Review*, *92*(3), 434-459.

Teubner, T. *et al.* (2017). Price determinants on AirBnB: How reputation pays off in the sharing economy, *Journal of Self-Governance & Management Economics*, *5*(4), 53–80.

Van Doorn, N. (2017). Platform labor: on the gendered and racialized exploitation of low-income service work in the 'on-demand'economy, *Information, Communication & Society*, *20*(6), 898–914.

van Walsum, S. (2011), 'Regulating Migrant Domestic Work in the Netherlands: Opportunities and Pitfalls', *Canadian Journal of Women and the Law*, *23*(1), 141–165.

Ye, T. *et al.* (2017), Race and rating on sharing economy platforms: The effect of race similarity and reputation on trust and booking intention in Airbnb.

Zervas, G. *et al.* (2015), A first look at online reputation on Airbnb, where every stay is above average, *Where Every Stay is Above Average (January 28, 2015)*.

InGRID-2
Integrating Research Infrastructure for European expertise on Inclusive Growth from data to policy

Referring to the increasingly challenging EU2020-ambitions of Inclusive Growth, the objectives of the InGRID-2 project are to advance the integration and innovation of distributed social sciences research infrastructures (RI) on 'poverty, living conditions and social policies' as well as on 'working conditions, vulnerability and labour policies'. InGRID-2 will extend transnational on-site and virtual access, organise mutual learning and discussions of innovations, and improve data services and facilities of comparative research. The focus areas are (a) integrated and harmonised data, (b) links between policy and practice, and (c) indicator-building tools.
Lead users are social scientist involved in comparative research to provide new evidence for European policy innovations. Key science actors and their stakeholders are coupled in the consortium to provide expert services to users of comparative research infrastructures by investing in collaborative efforts to better integrate microdata, identify new ways of collecting data, establish and improve harmonised classification tools, extend available policy databases, optimise statistical quality, and set-up micro-simulation environments and indicator-building tools as important means of valorisation. Helping scientists to enhance their expertise from data to policy is the advanced mission of InGRID-2. A new research portal will be the gateway to this European science infrastructure.

This project is supported by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 730998.

More detailed information is available on the website: www.inclusivegrowth.eu

Co-ordinator
Monique Ramioul

KU LEUVEN HIVA
RESEARCH INSTITUTE FOR WORK AND SOCIETY

Partners

TÁRKI Social Research Institute Inc. (HU)
Amsterdam Institute for Advanced Labour Studies – AIAS, University of Amsterdam (NL)
Swedish Institute for Social Research - SOFI, Stockholm University (SE)
Economic and Social Statistics Department, Trier University (DE)
Centre for Demographic Studies – CED, University Autonoma of Barcelona (ES)
Luxembourg Institute of Socio-Economic Research – LISER (LU)
Herman Deleeck Centre for Social Policy – CSB, University of Antwerp (BE)
Institute for Social and Economic Research - ISER, University of Essex (UK)
German Institute for Economic Research – DIW (DE)
Centre for Employment and Work Studies – CEET, National Conservatory of Arts and Crafts (FR)
Centre for European Policy Studies – CEPS (BE)
Department of Economics and Management, University of Pisa (IT)
Department of Social Statistics and Demography – SOTON, University of Southampton (UK)
Luxembourg Income Study – LIS, asbl (LU)
School of Social Sciences, University of Manchester (UK)
Central European Labour Studies Institute – CELSI (SK)
Panteion University of Social and Political Sciences (GR)
Central Institute for Labour Protection – CIOP, National Research Institute (PL)

InGRID-2

Integrating Research Infrastructure for European expertise on Inclusive Growth from data to policy Contract N° 730998

For further information about the InGRID-2 project, please contact
inclusive.growth@kuleuven.be
www.inclusivegrowth.eu
p/a HIVA – Research Institute for Work and Society
Parkstraat 47 box 5300
3000 Leuven
Belgium