Abstract

We examine how mobile big data can be used to measure the size and characteristics of the platform economy. It focuses on the gig economy in Poland between 2018-2020, using data from six gig workers apps, with information on 30 million smartphone users. Mobile big data is subject to coverage and measurement errors, and the paper addresses these issues by discussing the data generation process, identifying possible sources of error, comparing estimates with official statistics, and conducting validation studies.

The results show that mobile big data has good coverage for the foreign-born population, and low measurement error for Polish citizens. The platform economy in Poland is growing, with food delivery and transport services through apps expected to grow significantly between 2018 and 2020. The demographic structure of app users confirms previous studies, showing that most platform workers are young men, but the age structure varies across service categories.

The estimated share of active app users represents approximately 0.5-2% of the working population in the nine largest Polish cities. The study demonstrates the potential of mobile big data to measure the size and characteristics of the platform economy, providing insights for policymaking and strategy development in the gig economy sector.

Statement of significance:

Our contribution is as follows: 1) this is the first study using mobile big data collected through programmatic advertising to study the characteristics of gig workers in Poland; 2) we discuss a new type of mobile big data that has not been discussed before in a statistical context; 3) we assess the coverage and measurement error of this source using an administrative and validation study and show that the quality is quite good for active users and Polish citizens, while for foreigners the available data does not allow to fully assess the quality; 4) we provide new evidence on the gig worker population in Poland at a level that has not been published before.

1 Introduction

The changing world of work is characterised by a growing popularity of labour relationships based on non-standard employment contracts. A typical indefinite contract for a full-time job performed on the employer's premises in no longer the predominant employment model in the contemporary labour market. Digitisation and flexibilization are contributing to the emergence of new forms of employment. One of the phenomena that has been expanding recently is platform-mediated labour. According to the latest report published by the International Labour Organization (ILO), the number of online web-based and location-based (taxi and delivery) platforms increased from 142 in 2010 to over 777 in 2020 (ILO, 2021).

Platform work can be defined as "non-standard work facilitated by online platforms, which use digital technologies to 'mediate' between individual suppliers (platform workers) and buyers of labour" (Hauben et al., 2020, p. 98).

Originally, the platform economy, also known as *the gig economy* was associated with passenger transport services, mainly with the *Uber* company, which started to arrange work via an online platform. Nowadays, the range of platform-based activities is rapidly growing and includes several types of jobs, such as odd jobs (e.g. *TaskRabbit*), cleaning (e.g. *Helpling*), care (e.g. *care.com*), food delivery (e.g. *Deliveroo*) or programming and translating (e.g. *Upwork*) (cf. Koutsimpogiorgos et al., 2020). This explains why the phenomenon of platform work is receiving more and more attention from researchers and international institutions (cf. e.g. recent reports published by Eurofound (2019) or ILO (2021)).

Research on these populations is becoming important for several reasons. Primarily, the "platformisation" of labour relations is likely to increase competitiveness and create new opportunities on the labour market. Moreover, as employment relations become increasingly less standard, working conditions are likely to deteriorate in terms of job security and legal protections, because platform workers are predominantly self-employed and are not covered by institutional arrangements, such as collective pay agreement schemes. On the other hand, platform work may create opportunities for some groups in the labour market by increasing labour market participation (Eurofound, 2019) and enabling greater flexibility (Lehdonvirta, 2018).

While the number of studies about the platform economy is rapidly increasing, the reported evidence is mostly anecdotal and is based on interviews and personal first-hand experiences (De Stefano and Aloisi, 2018). This is because workers in the gig economy, while constantly growing in numbers, constitute a hard-to-reach and hard-to-identify population (see Bohning et al., 2017, Chapter 1). There is no sampling frame or a register that provides a full coverage and members of this group are often indistinguishable from the rest of the population (e.g. they can only be identified by examining the content of their smartphone or laptop). Information about this population is fragmented, which is why it is necessary to resort to modern data sources, such as big data, which can be used for estimating its size and characteristics. Because platform workers tend to operate in urban areas, general population sample surveys do not provide accurate estimates of the reference population (e.g. the economically active population) and administrative data often suffer from over-coverage (e.g. owing to outdated information about the place of residence, delays in reporting, etc.).

Big data sources are used for studying the gig economy from different angles. For instance, change in income at the national level has been measured based on transaction-level datasets, such as those generated by bill-paying applications (cf. Koustas, 2019); detailed behaviour and matching between customers and workers have been analysed using detailed data generated by Uber[1] or Lyft[2]. However, access to such data is restricted and often granted only to researchers from financial institutions or providers of the apps.

In this study, we propose a new way to measure the size and characteristics of the working age (18–65) population involved in providing food delivery or transportation services via applications for workers based on mobile big data obtained from advertisement systems on smartphones (Selectivv, 2021a). For this study, we purchased historical data (for 2018–2020) about users of the following apps: *Uber*, *Bolt Driver*, *Glover*, *Wolt Courier*, *TakeAway Courier* and *Bolt Courier*. For comparison, we also purchased data on *iTaxi* and *FREE NOW* apps, which are only used by licensed taxi drivers and we do not treat them as gig worker apps. Due to project budget constraints, we do not study the use of multiple platforms by the same user. Background characteristics, generated by classification algorithms applied by the data

[1] See: `https://eng.uber.com/research/`.

[2] See `https://eng.lyft.com/tagged/data-science`.

provider, include information about working time, location, country of origin, age and sex as well as other characteristics, such as being a student or having children. We assessed misclassification errors in two ways: based on administrative data and a limited validation sample survey.

Our contribution to the literature is twofold. First, we use big data collected via smartphones to passively measure the gig economy in Poland. This is in line with recent trends in official statistics to use all available information rather than create new ones (e.g. surveys). We highlight problems associated with these sources (e.g. measurement error due to using apps for workers) as well as their strengths (e.g. exact location), which can open a new avenue for measuring these populations. This approach makes it possible not only to access hard-to-reach and hard-to-identify populations, such as foreigners, but also to plan targeted surveys that can be used to collect more detailed information concerning motivations, working conditions or the quality of life. Second, we provide detailed information regarding the number and background of platform workers, which has not previously been available in Poland at any level of spatial aggregation. We believe that similar data sources can be identified in other countries and therefore our experiences and results may prove useful to other researchers and policy makers.

The article has the following structure. Section 2 contains a review of the literature and provides an overview of the main issues encountered when measuring the gig economy using classical (e.g. sample surveys) and modern data sources (e.g. administrative data, the Internet, big data). Sections 3 review available data sources about the gig economy in Poland. Section 4 provides description of mobile phone market in Poland, a description programmatic advertisements and variables obtained from Selectivv. Section 5 presents the main results: assessment of coverage and measurement error, characteristics of gig economy workers and compares them with the reference population from the Labour Force Survey (LFS). Finally, Section 6 includes a discussion of the results and describes future steps in the analysis of the gig economy on the basis of mobile big data.

2 Issues in measuring the gig economy

There are a number of terms to describe the ”platformisation” of labour relations, including the most common ones, such as ”platform economy” or ”gig economy”, as well as ”on-demand economy” or ”collaborative economy”. The gig economy can be defined in a broader and narrower sense. The broad definition covers precarious and casual work involving the use of technological intermediation (Aleksynska et al., 2019), while proponents of the narrower perspective focus either on specific types of platforms or equate the gig economy with digital labour markets, without providing a detailed definition (for an overview see Koutsimpogiorgos et al. (2020)). Based on the definitions proposed by Eurofound, platform work refers to “a form of employment that uses an online platform to enable organisations or individuals to access other organisations or individuals to solve problems or to provide services in exchange for payment” (Eurofound, 2018, p. 9). The authors of the definition also mention some key characteristics of platform work: paid work organised through an online platform; with the involvement of three parties: the online platform, the client and the worker; platform work involves providing specific tasks or solving specific problems; it is outsourced; jobs are broken down into tasks; services are provided on demand (Eurofound, 2018). In short, following the review made by Koutsimpogiorgos et al. (2020), it can be concluded that there is no single definition of the gig economy that is commonly accepted among researchers, policymakers, or practitioners.

In this study we narrow the analysis to *the transportation and food delivery sector and concentrate on the work using platforms in these two sectors*. The choice of definition is purely related to the data availability.

A review of the literature shows an increasing number of studies describing the relatively new phenomenon of the gig economy. What they all have in common is the growing realisation that the long-established idea of a typical job, which is performed under an indefinite contract, on a regular basis and during fixed hours, is becoming outdated and is being replaced by new forms of employment. Besides problems with understanding and defining the platform economy, there is a related concern about how it can be measured. In this section, we provide an overview of past studies aimed at measuring the extent of the platform economy.

Please note that we focus solely on the European labour market and do not include results for other countries, which can be found, for example, in Friedman (2014), Mitea (2018), Riggs et al. (2019) for New Zealand, Koustas (2019), Abraham et al. (2018) for US.

Looking at the previous results compiled in Table A1 in Supplementary Materials one can see that the majority of past estimates are based on survey data. It is important to underline the numerous limitations related to survey-based estimates mentioned in the literature. The main concern is about whether surveyed respondents appropriately understand the concept of platform work. As Bonin and Rinne (2017) point out, respondents may have misclassified other online activities, such using job search websites, search engines or professional networks as forms of platform work.

Moreover, in the case of online surveys, there is a risk of many inconsistencies caused by the poor reliability of the pooling technique, which may result in the over-representation of those who are more likely to be engaged in platform work (Piasna and Drahokoupil, 2019). Examples of survey-based studies include Kaczmarczyk et al. (2022) who used Facebook's advertising system to recruit gig economy workers to survey on job quality survey in Poland, Aleksynska et al. (2019), who present survey results for Ukraine; Piasna and Drahokoupil (2019) presenting estimates for Bulgaria, Hungary, Latvia, Poland and Slovakia; Berg (2016) for the USA and India. Remaining results are based on specific case studies of selected types of platforms, conducted by means of qualitative methods, see among others Meszmann (2018) for Hungary, Lenaerts (2018) for Belgium, Sedláková (2018) for Slovakia, Lee et al. (2018) for Hong Kong. However, these case studies provide important knowledge about the characteristics on platform workers rather than estimates of the extent of the phenomenon. Unlike European studies, which are mainly based on survey data, studies conducted in the USA also make use of transaction (financial) data.

For example, Koustas (2019) used a transaction-level dataset obtained from a large financial aggregator and a bill-paying application to identify changes in spending patterns for platform workers. Similarly, Farrell et al. (2019) report results based on de-identified checking accounts from Chase bank. To the best of our knowledge, there have been no similar studies based on transactional data that relate to European countries.

3 Data sources about the gig economy in Poland

3.1 Probability and non-probability sample surveys

To study the gig economy population in Poland, we reviewed existing probability and non-probability surveys in Poland. We focused on official statistics and large internet panels, which unfortunately do not provide reliable estimates at the national level.

The main source of information on internet users in Poland is the Information and Communication Technology (ICT) survey. Details of the methodology can be found at Statistics Poland (2020b). While the ICT survey covers the general population, between 2017 and 2020 the demand side of the platform economy was investigated in the question on transport services (see B.1 in Suplementary Materials). The results for these two questions are not reported by Statistics Poland because too few respondents answered them positively.

The second major survey is the Labour Force Survey (LFS), which measures the size of the active, inactive and unemployed population according to the ILO definition. The LFS is a quarterly survey with a rotating panel design; a detailed description of its methodology can be found in Statistics Poland (2020c). At the time of writing the article, the LFS did not include questions or special sections related directly to the gig economy. However, there are plans to include a special module in the LFS as part of a coordinated action suggested by the Group of Experts on Measuring Quality of Employment within the United Nations Economic Commission for Europe[3]. In the article, we used the LFS to obtain estimates of the working 18–65 population at municipality levels in Poland as reference figures for the supply side of the gig economy. Information about the working population at city level is not published by Statistics Poland. Therefore, based on available unit-level data from 2018 to the 1st half of 2020, we obtained estimates and their standard errors according to the Polish LFS methodology. For more details, see Supplementary Materials.

Another source that we considered is a non-probability sample of Internet users conducted by the PBI company (the acronym stands for *Polskie Badania Internetu*, which means ‘Polish

[3] See for instance `https://unece.org/statistics/events/group-experts-measuring-quality-employment` and the Handbook on Measuring Quality of Employment (`https://unece.org/statistics/publications/handbook-measuring-quality-employment`).

Internet research'). PBI specialises in measuring online audiences and their characteristics. Between 2016 and 2020 PBI conducted the Megapanel PBI/Gemius survey, which is exclusively focused on Internet users. Participants were were asked to install a special computer application that tracked all their online activity. About 150,000 respondents were recruited to participate in the survey by pop-ups placed on major Polish websites.

In 2020, PBI, together with the Gemius company[4] and the Radio Research Committee[5] launched a new survey called the Mediapanel. It is a combination of the Internet research standard, the radio audience standard (Radio Track) and data from the single-source cross-media Gemius measurement. The Mediapanel includes a passive measurement of Internet, TV and radio consumption. In 2021 the panel had over 280,000 users.

We asked PBI to provide an estimate of the number of users between 2018 and 2020 using at least one of 25 websites devoted to freelancing or app testing, such as `www.mturk.com`, `www.clickworker.com`, `www.fiverr.com` or `www.upwork.com`[6]. As in the previous cases, almost none of the above-mentioned websites met the requirement of the minimum number of active users to provide reliable estimates. PBI set up this threshold based on estimated number of users, i.e. 50k for audited (verified by the PBI staff) and 105k for non-audited media channels (not-verified by the PBI staff).

In summary, there are no sources that provide reliable estimates at low levels of spatial aggregation or for detailed cross-classifications. One way to overcome this problem is to turn to administrative data or internet data sources (big data).

3.2 Administrative data

In view of the growing popularity of applications such as Uber, the situation was remedied by an amendment to the Road Transport Act (*Ustawa z dnia 16 maja 2019 r. o zmianie ustawy o transporcie drogowym oraz niektórych innych ustaw, Dz.U. 2019 poz. 1180*), which came to be known as "Uber lex". It came into effect with a transition period until the end of March,

[4]Gemius is an international research and technology company providing data, as well as advanced tools for digital and traditional marketing activities such as web analytics, online campaigns' management and ad serving. For more see: `https://www.gemius.com/about-us.html`.

[5]More information can be found at `https://badaniaradiowe.pl/`.

[6]For the whole list please consult Suplementary Materials

which was then extended to the end of September and, later, to the end of December owing to the pandemic. The *Uber lex* extended the license obligation to Uber and other app's drivers, but at the same time simplified the licensing procedure. In theory, this would mean that we may have a list of all registered drives, but it is not possible.
A license is valid within the municipality in which it was issued by the head of the municipal government. Databases of licensed drivers are created independently by administrators and there is no common IT system that would enable the creation of a central register. Moreover, in practice, data is often collected in paper form. According to the act, records of persons with a taxi license do not have to include information about the method of accounting used by drivers for recording turnover (cash register / virtual cash register). As a result, it is impossible to identify drivers who provide transport services by means of registrations. Therefore, administrative sources are not suitable for measuring the gig economy in Poland.

3.3 Big data – applications for workers

In the following step, we examined mobile apps that could be used to identify gig workers. Table 1 contains a list of such applications, divided into the following categories: transportation, delivery, services, crowdwork and microtasks. In each case, we identified the provider, i.e. the company that develops a given app, the name of the app for workers and additional information.

Table 1: Mobile applications by category, with an indication of the target group of users (as of December 2020)

Category	Provider	The name of the app for workers	Comments
Transportation	Uber	Uber Driver	one app for Uber;✓
	Bolt	Bolt Driver	✓
	Lyft	Lyft Driver	no or small # users
	FREENOW (Mytaxi)	FREE NOW for drivers	✓
	iTaxi	iTaxi Kierowca K3	✓
	Optitaxi	–	no or small # users
Delivery	Deliveroo	Deliveroo Rider	no or small # users
	Glovo	Glover	✓
	UberEats	Uber Driver	one app for Uber; ✓
	Coopcycle	–	GP 5k+
	Take Eat Easy	–	no or small # users
	Pyszne.pl	Takeaway.com Courier	Same as pyszne.pl; ✓
	Lieferando	Takeaway.com Courier	Same as pyszne.pl; ✓
	Takeaway	Takeaway.com Courier	Same as pyszne.pl; ✓
	Bolt Food	Bolt Courier	✓
	Wolt	Wolt Courier Partner	✓
Services	TaskRabbit	Tasker by TaskRabbit	GP 100k+
	Helpling	Helpling Partner	no or small # users
	Fiverr	–	GP 10 mln+
	Upwork	Upwork for Freelancers	no or small # users
	Freelancer	–	GP 5 mln+
	PeoplePerHour	–	GP 100k+
	Toptal	–	GP 10k+
	Guru	–	GP 5k+
	FlexJobs	–	GP 1k+
	Truelancer	–	GP 500k+
	ClearVoice	–	–
Crowdwork	Amazon Mechanical Turk	Turkdroid	GP 50k+
	Clickworker	–	GP 500k+
	Microworkers	–	GP 5k+
	CrowdFlower	–	–
	Spare5	–	–
Microtasks	App Jobber	–	GP 100k+
	ShopScout	–	GP 50k+
	Streetspotr	–	GP 500k+

Note: Symbol '–' in the column *The name of the app for workers* denotes that there were no separate app for workers (i.e. only one app for clients and workers); GP – the number of downloads from Google Play reports provided by Selectivv; *no or small # users* informs that the there were no users in Selectivv data or this number was very small (i.e. hundred or lower); ✓– app was selected for the study.

This is a crucial part of our analysis. We decided to focus on dedicated apps for workers in order to distinguish between the demand and supply side. We found that the majority of transportation and delivery platforms create apps for workers, while platforms for services, crowdwork or microtasks tend to have only one app for both workers and customers. It is important to point out that Uber uses only one app to provide transportation (Uber) and delivery services (Uber Eats). Programmatic services do not have access to user activity within the app, so we only have information about the time spent in the app.

The *Comments* column in Table 1 specifies whether it was possible to use a given app in the study. For instance, three providers (Pyszne.pl, lieferando and takeaway) use the same app: Takeaway.com Courier. Moreover, information about the number of downloads shown in the Google Play store or the App store was misleading. For example, while TaskRabbit has been downloaded over 100,000 times and Fiverr – over 10 million times, data provided by Selectiv indicated that no user in their databases had installed either of these two apps. This suggests that Google Play reports provided by Selectivv probably present information about the global number of downloads, not just for Poland. Other sources such as Mturk demographics tracker[7] suggest that around 2% of their users are located outside the US and India. There are several commercial services such as *Sensor Tower, the Tool* or but as they are not free of charge we were not able to use their data in this paper.

Finally, taking into account these limitations and our main assumptions, we selected 6 apps for further analysis: *Uber*, *Bolt Driver*, *Glover*, *Wolt Courier*, *Takeaway.com Courier* and *Bolt Courier*. For the purpose of comparison, we used *iTaxi* and *FREE NOW*, which are designed for licensed taxi drivers.

4 Big data from smartphones

4.1 Mobile phone market in Poland

According to Office for Electronic Communication (2020b), the penetration of mobile telephony services in Poland will increase from 134.3% in 2018 to 141.5% in 2020, which means

[7] See `https://demographics.mturk-tracker.com/\#/countries/all`

that 51.6 million SIM cards were active in 2018 and 54.1 million in 2020 (the majority of SIM cards were postpaid, about 72-73%). The Office for Electronic Communications conducts an annual *Consumer Survey of Individual Customers*, which allows to characterise the Polish population aged 15+ in terms of mobile phone use. According to the table 2, over 92% use a mobile phone, of which over 70-75% use a smartphone and a negligible proportion use both a smartphone and a traditional telephone. The majority of Poles aged 15+ use a private mobile phone (over 97%) and only 5-6% use a business mobile phone. This may indicate that the problem of multiple device use may be low in the general population, but for some subpopulations (e.g. the 25-34 age group) the share of business mobile phones is twice as high as reported below.

Table 2: Characteristics of Polish 15+ population on mobile phones use

	2018 [%]	2019 [%]	2020 [%]
Sample size	1600	1600	1600
Use a mobile phone	91.9	93.1	92.9
% of those with a mobile phone			
Q1. How many active phone numbers do you have?			
Sample size	1457	1478	1469
1 private	97.2	97.1	96.9
2+ private	2.4	2.7	2.9
1 business	5.3	5.2	5.8
2+ business	0.2	0.0	0.8
1 private and 1 business	4.8	4.9	5.6
Q2. What type of telephone do you use?			
Sample size	1470	1487	1483
Smartphones	69.8	74.9	75.8
Smartphone and traditional	0.2	0.8	0.5

Source: Based on Office for Electronic Communication (2020a) and earlier reports. Note that in sample size in Q1 and Q2 varies because of non-response, refusals or don't know.

According to the SmartBarometer survey conducted by Kantar Public for Digital Care (2022), only 5% of respondents change their smartphone every year or more often, 12% after one and a half years and the rest (83%) after two or more years, as shown in table 3. This indicates that a given user can be tracked over a longer period of time and thus more information about activity can be collected than with cookies, which have an average age of 16 days and a lifetime of about 70 days (Miller and Skiera, 2023). According to this survey, 63% of respondents use a smartphone that is one or more years old, 72% use only one smartphone and the rest use two or more.

Table 3: How often do Polish smartphone users change their device and how old is their main device?

	< 1 year	to 1.5 years	to 2 years	more than 2 years
Replacement after [%]	5	13	46	36
Device age [%]	37	22	17	24

Source: SmartBarometer

Finally, the Polish ICT survey included questions on green ICT, in particular on unused mobile phones and smartphones, tablets and computers (Statistics Poland, 2022). In 2022, 48% of people aged 16-74 decided to keep them in the household, 19% sold or gave them away and only 12% disposed of them in electronic waste collection/recycling. This suggests that some parts of smartphones are circulating in the economy and may be a source of over-coverage due to multiple users in a given period.

Therefore, as smartphone use is increasing and working in the gig economy is not possible without the use of mobile apps, one could consider using passively collected mobile data to study this population in Poland. In this study, we used data from programmatic advertising systems, which are presented below.

4.2 Programmatic advertisement system

For this study we purchased aggregated data from the Selectivv company[8], which uses a *programmatic* advertisement platform to monitor about 350,000 mobile apps used and over 17 million mobile websites visited by over 30 million smartphone users in Poland.

Programmatic advertising platforms are IT systems that automate the process of displaying advertisements to Internet users based on their individual characteristics, such as past online behavior patterns or location. It is based on real-time bidding systems that in a matter of milliseconds decide whether to show a given ad to a given user (Busch, 2016).

In practice, a programmatic system works as follows. When a mobile device user opens a website or an application that contains ad space with an advertisement displayed by a particular

[8] For more information see `https://selectivv.com/en/`.

company (e.g. Selectivv), the company receives basic information from the mobile device along with an individual number associated with the user account. This number is assigned through two systems: *the Google advertising ID* (GAID) and *Apple's Identifier for Advertisers* (IDFA). The first one works on smartphones with Android, the second one – on those with iOS. The number remains the same as long as the user logs in with their Google/iOS or another account. If the user changes a mobile phone and creates a new account, then they are treated as a new user.

The main difference between data collected through programmatic platforms and systems of mobile network operators is the level of available information. The latter can only provide information about customers, often without a distinction between business and private users, about location, based on signalling, or the type of smartphone. Most of these data are taken from Call Detail Records (CDR), which, according to Polish law, are separated from consumer relationship management systems. Thus, data collected by mobile network operators contain no information about apps installed on a give phone or any background details about a given user.

On the other hand, companies working with programmatic systems serve multiple providers and collect information not only about domestic users but also about foreign ones, who are often missed by surveys or registers. As almost all apps contain advertisements, the amount of information collected is considerable and may include: location (GPS / WiFi), time of use, websites visited, applications installed, device type, operating system and its settings (such as language) and mobile network operator. Programmatic systems do not have access to activities within mobile applications and only provide information about the app name, time spent in the app as well as when and where it was used (depending on smartphone settings).

Based on passively collected data, companies profile users of mobile devices by applying machine learning algorithms or heuristic rules. In our case, Selectivv claims that its algorithms can predict up to 360 variables about an individual user. However, the quality of the algorithms is often unknown, but it is a function of user activity, i.e. as more information is produced and collected, better accuracy is observed. Although the store of collected information is rich, it does not contain any personal information such as name, surname or personal ID, except for user characteristics.

Selectivv also employs algorithms to identify users of multiple devices (private and business ones) based on GPS location and WiFi signals. In 2021Q3 Selectivv database, after deduplication of smartphone identifiers (GAID, IDFA) contained around 30,9 mln Polish 18+ users which is close to about 31,2 mln Polish 18+ nationals in the Population Register (PESEL) and around 1,23 mln Ukrainians (of 1,1 mln in the Population Register). This suggests that the deduplication algorithm may work in some cases incorrectly, especially when a given person changes the smartphone. This issue was raised when discussing with Selectivv however the quality of matching and deduplication algorithms is unknown and thus is the main limitation of the study. In the section we will describe variables used in this study.

4.3 Variables obtained from smartphones

We obtained aggregated data for various socio-demographic characteristics describing smartphones users, which had been specified during a discussion with Selectivv's staff. The type and levels of all variables were limited by the project budget. All variables are the output of Selectivv's proprietary classification or heuristic algorithms, so the level of errors associated with these methods is unknown and validation is required. For the study, we selected the following variables:

- `Sex` (Male, Female) – the classification was made on the basis of a user's activity (e.g. visited websites, installed applications, information provided in the apps) and sample surveys conducted by Selectivv via advertisement systems.
- `Age group` (18–30, 31–50, 51–64) – the classification was made on the basis of a user's activity (e.g. visited websites, installed applications, information provided in the apps) and sample surveys by Selectivv via advertisement systems.
- `Nationality` (Polish, Ukrainian, Other) – Ukrainians are defined as people who have a SIM card provided by a Polish operator, have set the language on their smartphone to Russian or Ukrainian and at least once in the last year have been to Ukraine, where they replaced their SIM card with another one provided by a Ukrainian operator.

- `Residence` (Cities, Functional Urban Areas[9] and Province; see supplementary materials) – this information is derived on the basis of location metadata (e.g. GPS, WiFi) and is defined as the most frequent night location (18:00–8:00) in a given period. We provided a shapefile with borders of these areas.
- `Student` – whether a given person is a student; derived based on a user's location data (e.g. points of interest; POI) and their browsing history.
- `Parent of child aged 0-4` – derived on the basis of POIs, such as kindergartens, visits at offline and online shops.
- `Parent of child aged 5-8` – derived on the basis of POIs, such as primary schools, visits at offline and online shops.
- `Time spent in the app` – the mean and standard deviation of the time spent in the app (in seconds).

For the purposes of our study, we defined an active user as *a person who used a given app for at least one minute* within a given time period (month, half year between 2018 and 2020). First, the pragmatic reason was our limited budget, which was only sufficient to obtain data for long periods, such as months or half-years. Second, we decided not to use thresholds such as one, two or more hours because we did not know before the analysis how long certain users used selected apps.

Thanks to the courtesy of Selectivv, after the project we received overall information on the distribution of monthly activity time, as shown in table 4. This shows that the majority of gig workers used apps for more than 40 hours in January 2018 (over 75%), and are heavy smartphone users. Therefore, the information collected about them through programmatic advertising is likely to be rich, and the resulting classification error of socio-demographic variables is likely to be low.

For this study, we obtained four datasets for a period from 2018 to 2020 for the population of users aged 18–64:

[9] A functional urban area consists of a city and its commuting zone. For details see D in Supplementary Materials.

Table 4: Monthly activity times of selected applications in January 2018 [in %]

App	1 min-40h	[40, 80)h	[80, 120)h	[120, 160)h	160h+
Bolt Courier*	16	18	35	29	2
Bolt Driver	6	67	16	6	4
Glover*	13	22	35	28	2
Takeaway.com Courier	25	23	39	12	0
Uber Driver	12	17	27	38	6
Wolt Courier Partner	27	19	39	14	0
Google Chrome	18	38	35	7	2

Note: Symbol * indicates that the time period was different. For Bolt Courier was April 2020 and Glover was July 2019. Google Chrome is used for comparison purposes.

1. the number of active users by app within a given month,
2. the number of active users by sex, age group and nationality within a given half-year; for each cross-classification we got a share of students, parents with children aged 0–4 and parents with children aged 0–4,
3. the number of active users by city, functional urban area and province within a given half-year,
4. the mean and standard deviation of the time spent in the app for the following periods: Mondays-Thursdays 8:00-18:00, Mondays-Thursdays 18:00–8:00; Fridays-Sundays 8:00-18:00 and Fridays-Sundays 18:00–8:00 within a given half-year.

All datasets used in the study are available for readers in Supplementary materials.

5 Results

5.1 Assessment of the coverage error

One way to assess coverage is to verify the share and changes in the population of foreigners in Selectivv databases. This population is often missed by official statistics owing to the lack of up-to-date sampling frames. In a recent report (Selectivv, 2021b), the number of Ukrainians reported in January 2021 was 1,274 million. According to recent experimental studies conducted by Statistics Poland (2020a,d), there were 1,351 million Ukrainians on 31st December 2019. The latter number was obtained by integrating 9 administrative sources concerning the registered population, in particular the economically active population. It can therefore be concluded that Selectivv data provide an adequate coverage of the Ukrainian population.

Figure 1 presents changes in the population of Selectivv users by country of origin. Between 2018 and 2020 the number of all smartphone users in their database increased by 8% but the change for each sub-population was different. The group of Polish users increased by 6%, which, in absolute numbers, corresponds to about 2 million additional users between 2018 and 2020.

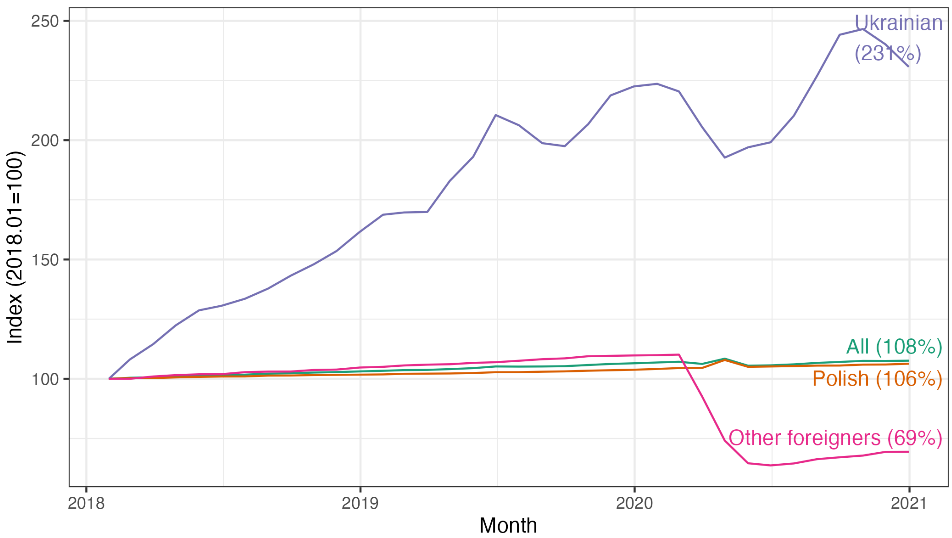

Figure 1: Changes in the number of smartphone users based on monthly data collected by Selectivv between 2018 and 2020 (2018.01 = 100)

The biggest changes can be observed for Ukrainians and other foreigners, whose number fell considerably after the first COVID lockdown in March 2020, but, unlike in the case of Ukrainians, it did not change much over the rest of 2020. In contrast, the number of Ukrainian users rose by over 200% between 2018 and 2020, which is the result of migration to Poland and can also be observed in official sources.

To further verify the degree of coverage of the foreign population, we compared these changes with the index of insured foreigners, calculated on the basis of records maintained by the Social Insurance Institution (ZUS). If a foreigner is legally employed in Poland, they should be listed in the ZUS register. The register includes economically active foreigners and has been consistent with the structure of migration to Poland since 2018.

Figure 2 presents a comparison between monthly and quarterly data from ZUS and Selectivv data for two groups of foreigners: Ukrainians and non-Ukrainians (other). Trend lines for Ukrainians, which are evident in ZUS data, in particular the sharp drop and increase in 2020, are similar to those observed in Selectivv data. The main difference between Selectivv and ZUS regarding this subgroup are the actual index values (January 2019 is used as the baseline). It should be remembered, however, that not all Ukrainians would have applied for a legal job or are insured.

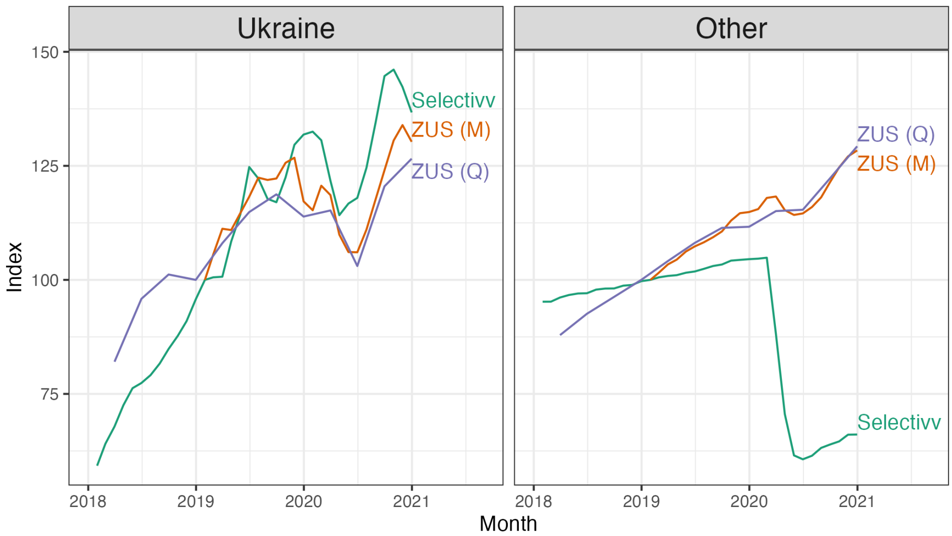

Figure 2: A comparison of changes in the number of users according to Selectivv and the number of insured foreigners according to ZUS (2019.01 = 100). Note: ZUS (M) – monthly statistics; ZUS (Q) – quarterly statistics

In contrast, the drop in the index for non-Ukrainian foreigners is significantly bigger than that observed in ZUS registers. There are three possible explanations for this. First, since we do not know the number of other foreigners in Selectivv databases, the decrease may simply be due to their low initial levels. In ZUS data there are about 200,000 non-Ukrainian foreigners. Second, ZUS covers all foreigners, including those who may have lived in Poland for several years and are already assimilated (e.g. use the Polish language), while Selectivv may be covering short-term migrant workers. Finally, the drop may represent illegal workers who lost their source of income during the COVID pandemic.

After comparing Selectivv data with existing official or administrative sources, it can be concluded with a high degree of confidence that our data source covers the majority of smartphone users in Poland. Certainly, access to more detailed information could improve our assessment but as far as we know there is no other available data source in Poland with this level of coverage. Therefore we are convinced that these data can provide accurate information regarding the number of people engaged in the gig economy in Poland.

5.2 Assessment of the measurement error

To fully assess the quality of the Selectivv data, it is necessary to focus on the measurement error, i.e. the accuracy of the variables they provide. The standard way to assess quality is to carry out a validation survey in a two-stage study design, i.e. a random sample from the dataset under study. This has already been done for administrative data on the labour market or health (Schenkel and Zhang, 2022; Shepherd et al., 2023). In the context of passively collected data, (Neumann et al., 2019) examined the accuracy of classifying cookies by gender and age. The gender accuracy ranged from 25.7% to 62.7% with an overall average of 42.3%, while the age accuracy ranged from 4.3% to 42.5%. It should be noted, however, that the lifetime of cookies is significantly shorter than that of smartphones. Recently, Grow et al. (2022) studied Facebook's advertising platform for research, in particular Facebook's classification of demographic information. They concluded that the classification was correct in over 90% of cases. It should be noted that the authors write that "Facebook's ad delivery algorithms were designed to optimise ad delivery to increase the likelihood that users who were shown an ad would click on it". This means that Facebook may be targeting an active sub-population of its users. In this way, the algorithm selects respondents in proportion to the information it has about them, and the accuracy of its algorithms is therefore higher.

In the paper we approach this problem in two ways. First, we compare the sex and gender distribution of Polish and Ukrainian users with the population register. Second, ideally we would like to have a validation study on users of gig economy apps, but due to budget limitations we were not able to conduct such a study on this sub-population. However, thanks to another research project we were able to conduct a validation survey on a 1,000 sample of active smartphone users.

5.2.1 Overall

In order to assess the measurement error at the database level, we compared the distribution of sex and working age (18-24, 25-29, 30-39, 40-60/64) for Poles and Ukrainians between Selectivv and two registers: Population Register (PESEL) and ZUS[10].

Figure 3 shows the results for Polish citizens for Q2 and Q4 between 2018 and 2020. The reason for choosing these quarters was the limited access to detailed data from the PESEL register. Both sexes in the age groups 18-24 and 25-29 are fairly well reflected in the Selectivv data, while discrepancies are observed for the next age groups: 30-39 and 40-60/64. The reason for this could be twofold: 1) the classification algorithms are not well suited to older groups, 2) once a given user is classified in a certain age group, it is not updated in time. The second case is visible for younger groups, where their proportions decrease over time in the population register, while they are stable in Selectivv.

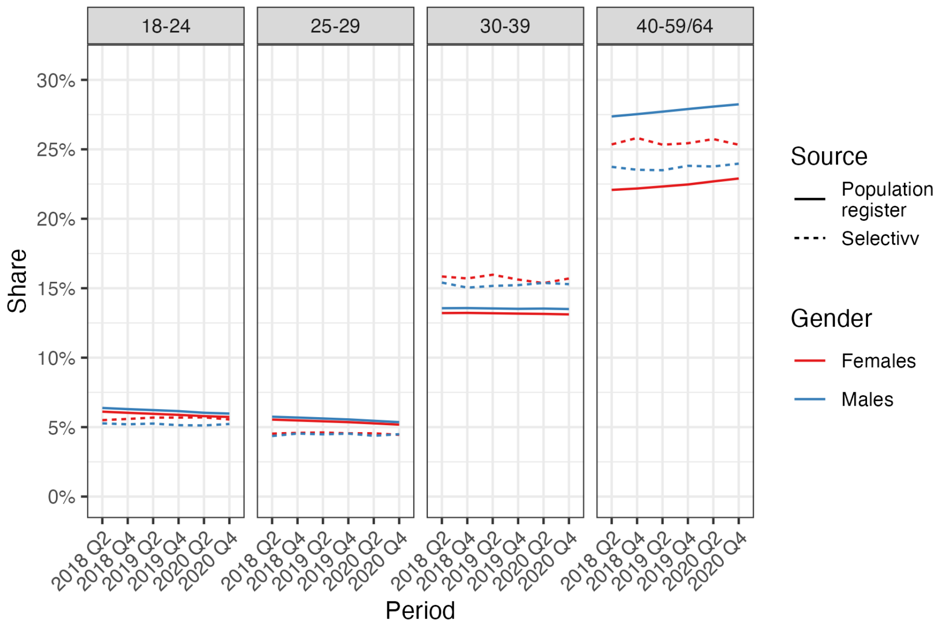

Figure 3: Comparison of gender and age of Polish citizens according to the Population Register (PESEL) and Selectivv

Figure 4 shows a comparison of Selectivv data for Ukrainians with two registers. The reason

[10]Note that the age groups differ from those mentioned in section 4.3 because the Selectivv data we are comparing come from a different project.

for this is the coverage of PESEL and ZUS registers, for example, in 2019Q2, 400k Ukrainians were in PESEL, while over 470k were in ZUS. The reason for this is that not all Ukrainians applied for a PESEL identifier. There are significant differences in the distributions that cannot be attributed to either coverage or measurement error on the basis of this comparison. The youngest group of Ukrainians is underrepresented in the Selectivv data, while those aged 25-29 are over-represented. In addition, the distributions vary between quarters, which may be due to migration patterns. The only group that appears to be similarly represented are men aged 30-39.

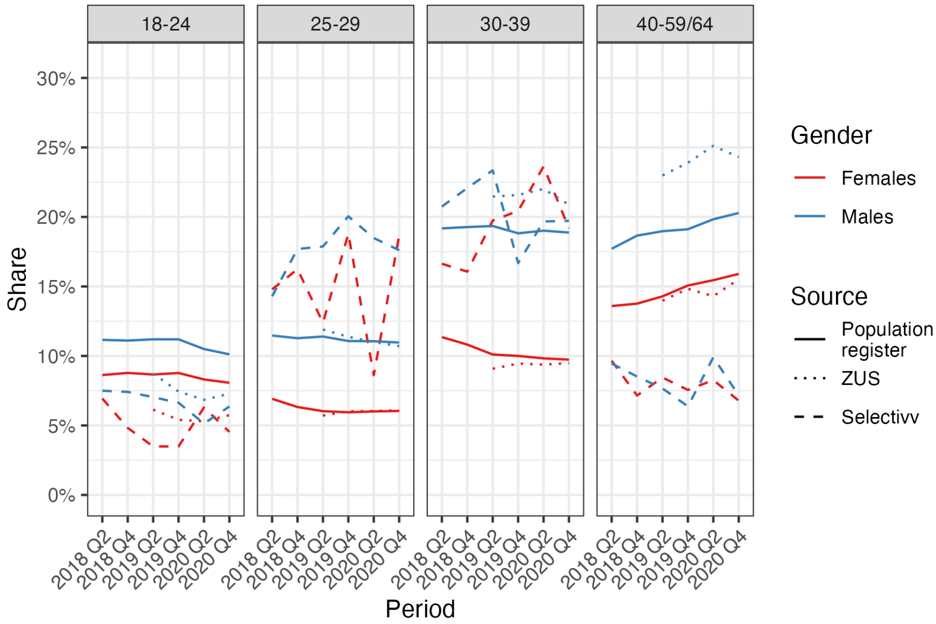

Figure 4: Comparison of gender and age of Ukrainian citizens according to the Population Register (PESEL), Social Insurance Register (ZUS) and Selectivv

It should be noted that the population register or the register of insured persons may reasonably well cover resident foreigners, whereas the smartphone data include both short-term and long-term migrants. Differences in gender and age distributions may be due to different population definitions rather than measurement error per se. Detailed information on Ukrainians from the 2021 Census of Population and Housing had not been published at the time of writing, which limits the possibility of comparison. A validation study is therefore needed to assess the measurement error using unit-level data.

5.2.2 Validation survey

A survey of 1,000 active smartphone users was designed. These users are defined as those for whom Selectivv has a significant amount of information, measured by the time of smartphone use[11]. The quotas for the sub-populations were set in such a way that some groups would be equally represented (gender) and others (age):

- Citizenship: Polish (500), Ukrainian (290), Belarus (210),
- Gender: Males (250), Females (250),
- Age groups: 18-24 (360), 25-29 (320), 30-39 (160), 40+ (160).

The survey was conducted via smartphone pop-up ads, i.e. an invitation was displayed to 49k Polish, 28k Ukrainian and 20k smartphone users (over 96k invitations were sent). The realised sample size was 986 which means that the response rate was around 1%. Tables 5–7 present the results of the classification of nationality, gender and age groups, where the rows refer to the classification by Selectivv and the columns are the self-reported data. In all cases the classification error is very low with an accuracy of 96.9% for all variables together (i.e. whether all variables are correct).

Table 5: Classification table for nationality

Selectivv / Self-reported	Poland	Ukraine	Belarus
Poland	494	0	0
Ukraine	2	281	2
Belarus	3	3	201

Overall classification was: 99,0%.

Table 6: Classification table for gender

Selectivv / Self-reported	Female	Male
Female	491	3
Male	3	489

Overall classification was: 99,4%.

[11] according to the Selectivv representative, this group represents about 10% of all users

Table 7: Classification table for age groups

Selectivv / Self-reported	18-24	25-29	30-39	40+
18-24	339	1	0	0
25-29	5	310	0	0
30-39	1	7	159	0
40+	0	1	1	162

Overall classification was: 98,4%.

The assessment of the classification errors suggests that the whole database may suffer from significant measurement errors, mainly due to the small amount of information on a significant proportion of users. While it may be reliable for a specific sub-population of active users, which are gig workers, it should be further investigated in future projects. Furthermore, we did not validate other variables such as place of residence, being a student or a parent, so the quality of these variables is unknown.

5.3 Characteristics of Polish gig workers

Figure 5 presents changes in the number of active users (in thousands) between January 2018 and December 2020. Note we distinguish Uber as a separate category due to one app for Uber and Uber Eats. Data for Glover and Bolt Courier are only available for a shorter period because these services started operating in Poland in mid 2019 and beginning of 2020. Interestingly, the rate of growth for Glover is significantly higher than for the other apps. This is mainly due to the fact the company offers delivery services not just for meals, including fast-food meals, such as those sold by McDonald's, but also groceries from the biggest chain store in Poland – Biedronka.

In the case of other delivery apps, Takeaway saw the highest absolute increase, from 7,500 to nearly 14,000 users in December 2020. On the other hand, at the end of 2020, Wolt and Bolt Couriers had around 6,000 and 2,000 users, respectively. These disparities have to do with the companies' regional scope of operation. For instance, only 14% of Takeaway users were from Warsaw (Poland's capital city) while the corresponding share for the other apps was between 35% and 40%.

Figure 6 indicates where the increases presented in Figure 5 actually happened. In the case

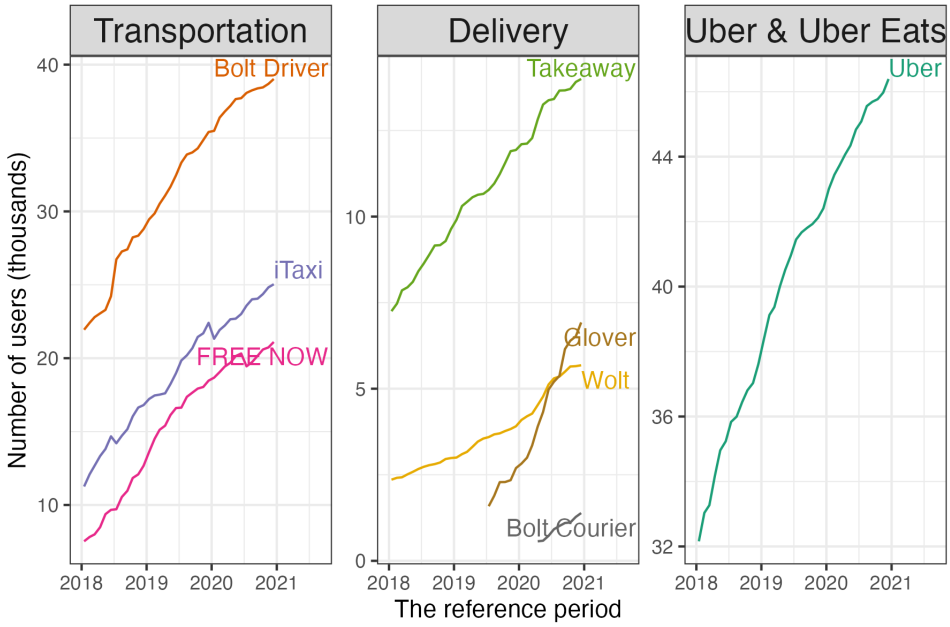

Figure 5: The monthly number of active users in Poland for selected apps by category between 2018 and 2020

of delivery apps, the biggest increases were recorded in cities, which is to be expected given that these services are mainly offered in cities. Changes in functional urban areas (FUAs) and provinces are negligible relative to 2018HY1. In the case of transportation services, changes in cities and FUAs are small and there is even a decrease in the number of Bolt Driver users. The observed increase in the number of users is mainly due to the expansion of these services to areas surrounding cities, which is evident at the province level. For instance, in the case of Tricity (the metropolitan area of Gdansk, Sopot and Gdynia) services offered by Uber actually cover almost the entire Pomorskie province (as of May 2021).

For more details, see Supplementary Materials. With regard to these results, it should be noted that the same users may have been using more than one of these apps, but the level of this overlap is unknown.

Table 8 contains demographic characteristics of active apps users for the last period in our analysis (2020HY2), which may be indicative of the self-selection process. The main difference between transportation and delivery services is the age structure. The majority (95%) of couriers were between 18 and 30, which is mainly related to the mode of transportation – bikes and scooters. There is also a major difference in terms of sex in both categories, with female drivers accounting for only 11-12% of all users, and an even smaller share of female

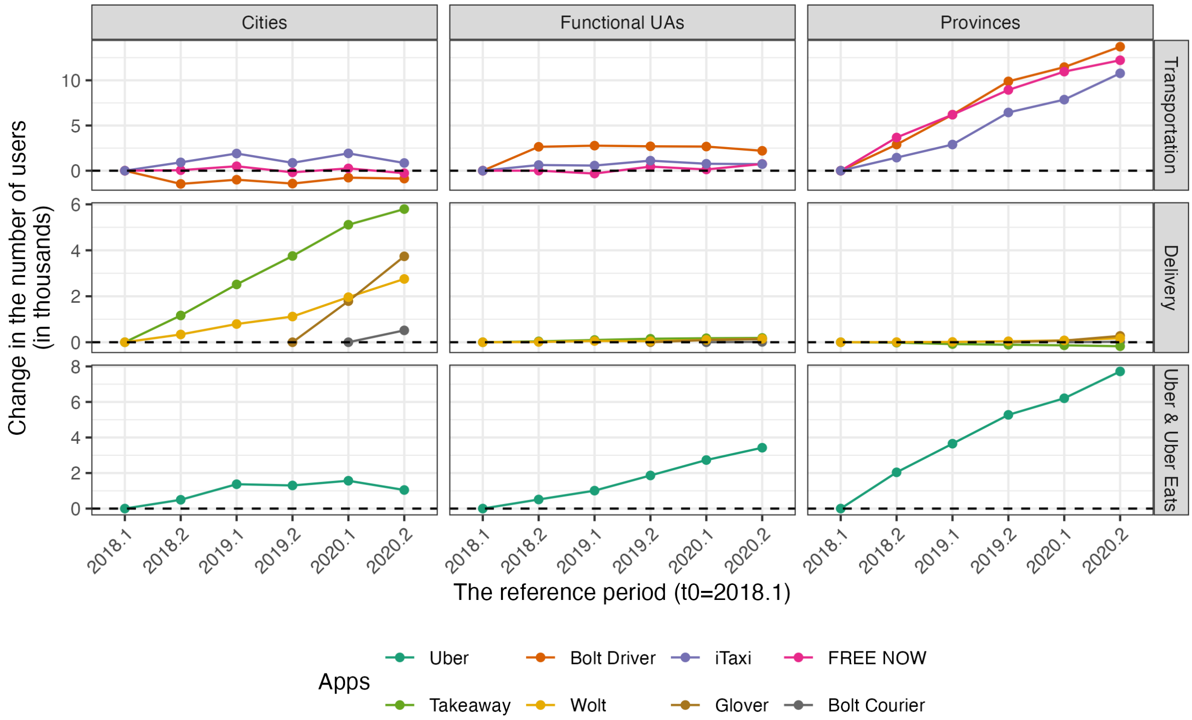

Figure 6: Change in the number of active users in Poland for selected apps by type between 2018 and 2020. Base 2018HY1

couriers (5-10%). This finding is in line with survey results showing that, on average, platform work is mainly performed by younger people (Piasna and Drahokoupil, 2019; Brancati et al., 2020).

Smaller shares of foreigners of other nationalities, students and parents with children aged 0-4 are, most likely, to do with age. The majority of recent migrants in Poland are young people from Ukraine, Belarus, Russia, India, Bangladesh or Nepal.

Figure 7 shows changes in the number of active users by nationality (top), age (middle) and sex (bottom) for different apps[12]. The number of Polish users is significantly higher and the curve is steeper than in the case of Ukrainians and other foreigners. For each app we can observe an increase in the number of active users, which is higher than the overall increase presented in Figure 1. This suggests that existing smartphone users may have installed these apps to start providing services. We can also see a slight impact of COVID-19 on Glover and Wolt, but given the scope of this article we did not investigate this aspect any further.

There are also differences in the number of users of each apps when age is taken into consid-

[12]Plots for other characteristics may be found in F in Supplementary Materials

Table 8: Demographic characteristics of gig app users in 2020HY2

	Sex		Age group			Nationality			Student	Parent of a child	
App	Men	Women	18–30	31–50	51–64	PL	UA	other		aged 0–4	aged 5–10
Uber & Uber Eats											
Uber	88.0	12.0	49.2	46.4	4.4	66.0	24.1	9.8	2.7	3.5	3.1
Transportation											
Bolt	86.3	13.7	56.2	37.8	6.0	65.5	26.1	8.5	3.1	4.4	1.9
FREE NOW	88.6	11.4	40.1	52.4	7.5	76.8	21.7	1.5	0.8	2.0	3.2
iTaxi	88.7	11.3	28.2	58.1	13.7	77.9	20.2	2.0	4.4	0.6	1.1
Delivery											
Takeaway	89.7	10.3	94.6	4.5	0.9	62.1	31.1	6.7	8.1	8.1	1.6
Glover	93.8	6.2	94.1	5.7	0.2	61.6	27.5	10.8	7.8	7.8	1.4
Wolt	92.3	7.7	95.7	2.9	1.4	54.2	28.4	17.5	7.9	8.0	1.5
Bolt Courier	94.7	5.3	100.0	–	–	62.3	27.5	10.2	0.0	0.0	0.0

Note: PL – Polish, UA – Ukrainian, other – other foreigners.

eration. Most users of delivery apps are aged 18–30. The highest increase can be observed for Glover and Takeaway, since these apps have the highest coverage and are the most popular. With regard to transport apps, there is a difference in the age structure of those who use apps for licensed taxi drivers and those who use Uber and Bolt. The plurality of iTaxi users are aged 31-50 and their share grew slowly over the reference period. In contrast, for users of FREE NOW the difference between these two age groups was initially much bigger and rose faster. This may suggest that younger taxi drivers choose FREE NOW, whereas iTaxi is chosen by taxi corporations. Interesting results can be observed for Bolt and Uber. Differences between the age groups may suggest that these apps are used by different groups of users: while the trends are generally similar, the disparity between the number of users aged 18–30 and those aged 31-50 is bigger for Uber. The increase in the number of Bolt users aged 18–30 was bigger than for the other age groups but from 2019HY2 the growth slowed down.

Finally, as can be expected, most users involved in providing transportation and delivery services are male. Previous evidence for other European countries indicates that women account for about 13% of riders, see e.g. a study by Drahokoupil and Piasna (2019) for Belgium. Our results are, therefore, consistent with that proportion. Interestingly, the share of women performing platform work is growing, especially considering the transportation sector. This trend has been recently observed and reported in Europe. Following survey data for European countries provided by Brancati et al. (2020), platform work is becoming a source of income for

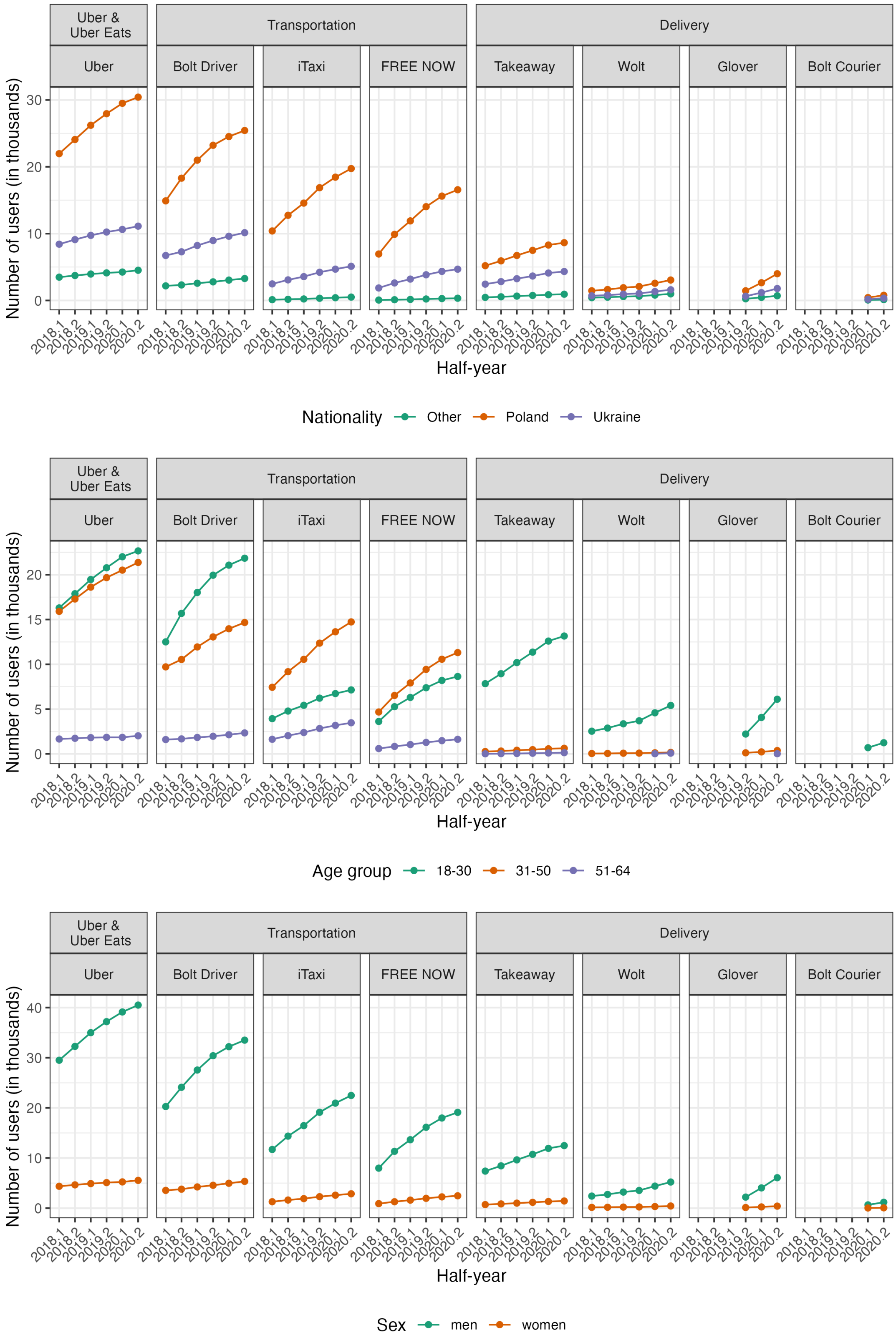

Figure 7: The monthly number of active users of selected apps in Poland by country, age and gender between 2018 and 2020

an increasing proportion of women. Importantly, differences between male and female users are closely associated with the type of tasks performed. Women are more likely to be over-represented in feminised tasks, such as translation or interactive services, while transportation and delivery services tend to be more male-dominated Brancati et al. (2020).

Furthermore, we investigated what share of the working population (based on LFS estimates) in a given city provided services via gig apps. Data for 19 Polish cities are compared, with the exception of the city of Sopot, where sample sizes were too small. As the LFS estimates for individual cities have relatively high variances, we decided to include confidence intervals. Details about the LFS estimates and information about the number of active users are presented Supplementary Materials.

Figure 8 presents point estimates and 95% confidence intervals (CI) of the ratio of the number of active users to the working population aged 16–65 for the last period, i.e. 2020HY2. It was not possible to compare transportation apps for all cities, because, while iTaxi (as well as all delivery apps) are available in all 20 cities of interest, Uber is only available in 9 cities (including Sopot), Bolt – in 10 (including Sopot) and FREE NOW in 8 cities (including Sopot).

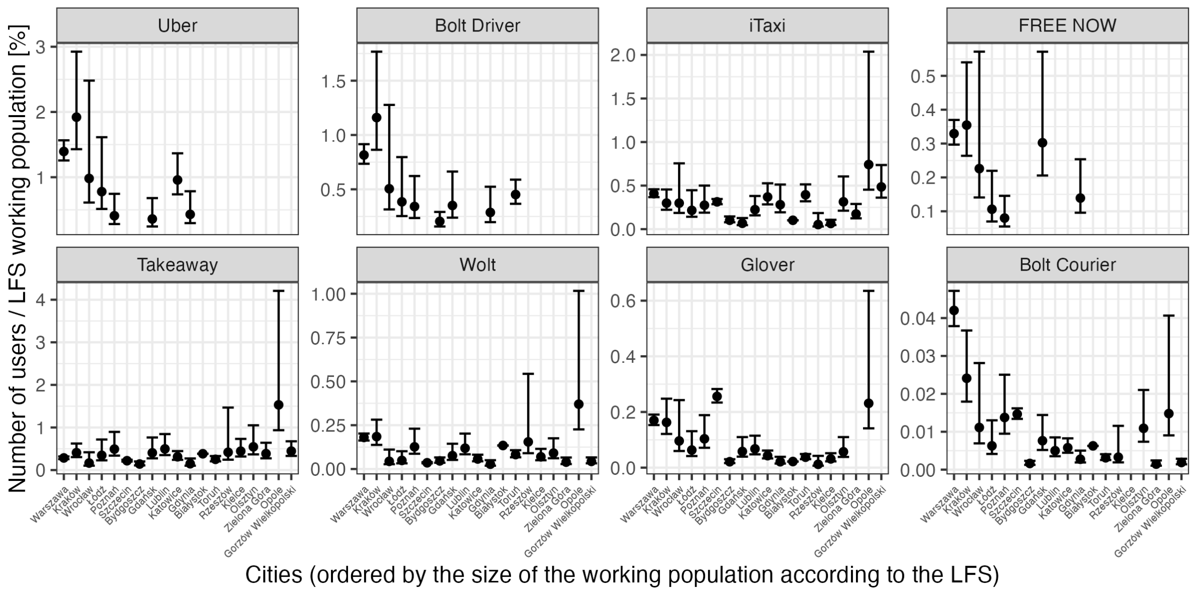

Figure 8: Point estimates and 95% confidence intervals of the ratio of the number of active users to the LFS-based working population aged 16–65 for each app in 2020HY1. Note that the Y axis differs across the apps

In general, the share of active users of transportation apps ranges from 0.5% to 2% of the working population in the selected cities. FREE NOW is the least popular one, with a share ranging from 0.1% to 0.5%. While the share of app users in each city, regardless of the app, is quite similar, it is consistently the highest in Crakow and Warsaw, which could be explained by their specific character: both Crakow and Warsaw ares visited by many foreign tourists have a large student community. The share of iTaxi users is almost the same across the cities, which may have to do with the policy of taxi corporations, which try to adjust the number of taxi drivers to a given market, while Uber or Bolt do not impose limits on the number of users (however, they do control which drivers are matched with particular customers). Additional comparison connected with working times is presented in section H in Supplementry Materials.

Compared to the transportation apps, the shares of delivery app users are generally smaller, even for Takeaway, the most popular app for ordering food in Poland. The highest variability can be observed for Bolt Courier, with the majority of users operating in Warsaw and Crakow. For the other apps, the pattern is similar: the share of the working age population is around 0.5% for Takeaway and 0.1-0.2% for Wolt and Glover. It should be noted that these figures are upper bounds, since the LFS-based working population may be underestimated as a result of non-response among foreign-born respondents and over-coverage is present in Selectivv data at unknown level.

6 Conclusions

Platform work is rapidly affecting the way labour market functions worldwide. In the broad sense, the term covers all job activities conducted by means of digital platforms and applications. Along with technology spillovers, the popularity of web-mediated activities is increasing in all economies, developing as well developed ones. As a result, the literature concerning platform work keeps growing, providing new knowledge about this phenomenon. However, owing to data scarcity, estimates of the size of this segment on the labour market are rather limited.

In the study described in this article we made an attempt to measure the size and characteristics of the platform economy in Poland. To this end, we used big data about 30 million smartphone users collected through an advertisement system. The data enabled us to provide an upper bound of the number of drivers and couriers at very low levels of spatial aggregation. Our results show that the platform economy in Poland is growing. By analysing data about delivery and transportation activities performed via gig applications, we found a growing trend between January 2018 and December 2020. As regards delivery apps, the sharpest increase was observed for Takeaway and Glover, while increases for the transport apps were more or less similar. Taking into account the demographic structure of apps users, we confirmed estimates from past studies: most platform workers are young men. Analysing the sex ratio of app users, it can be seen that the share of women is increasing especially with respect to transport apps. By comparing the number of platform workers with the LFS estimates of the working populations in 9 largest Polish cities, we found that the share of active app users in those cities was at the level of 0.5-2%.

Despite the undeniable advantages of these sources, such as the fact that the data are collected passively and provide an accurate measure of time spent in the app, they are not free from certain shortcomings. While under-coverage may not be a serious problem, over-coverage and mis-classification may be. Based on a validation study, we conclude that big mobile data from programmatic advertising systems cover the dynamics of the foreign-born population reasonably well, while existing administrative sources do not allow for the assessment of measurement errors due to different definitions of the population. On the one hand, information on Polish citizens seems to be measured more accurately and more in line with the population register. A limited validation study suggests that the classification algorithms work well for active smartphone users, i.e. those who are well covered by data collected through advertising systems. The level of over-coverage is unknown at the moment and requires more data to be assessed.

Furthermore, not all platforms provide separate apps for workers and potential customers. This makes it challenging to identify those involved in the supply side of the gig economy. We cannot determine the threshold which should be applied to distinguish those groups. One may consider using the ILO definition of a working person, which is based on the question about whether a given person has worked at least one hour in the last week. But applying this rule to real-time app data may be a challenge as there is no reference date that can be taken into account. One possible solution is to choose an arbitrary date, for example the middle of the month, but this is bound to cause errors and limit comparability with existing official statistics.

Finally, the cost of obtaining big data is non-negligible. Companies like Selectivv collect terabytes of data over short periods of time and use cloud services to store and process them. While the price of these services is decreasing, the cost of obtaining access and the computation time required to compile target data aggregations and statistics are factors that cannot be ignored. Prior to any analyses, researchers need to specify the target population and quantities and negotiate the purchase price with these companies. This limits possibilities of data exploration as access to unit-level data is either forbidden by the company or limited by the way the data are stored.

Another problem is associated with limited access to reliable information about the reference working age population at low levels of spatial aggregation. In general, owing to small sample sizes, population surveys or coverage errors in administrative data (i.e. out-dated information, under-coverage of the foreign-born population) estimates of the size of the gig economy in these areas tend to be characterised by high MSE values. To overcome this problem, official statistics and researchers could take advantage of small area estimation, which relies on multiple sources to provide reliable estimates for areas of interest (see, for example Van den Brakel and Krieg, 2009; Rao and Molina, 2015).

References

Abraham, K. G., Haltiwanger, J. C., Sandusky, K., and Spletzer, J. R. (2018). Measuring the Gig Economy: Current Knowledge and Open Issues.

Aleksynska, M., Bastrakova, A., and Kharchenko, N. (2019). Working Conditions on Digital Labour Platforms: Evidence from a Leading Labour Supply Economy.

Berg, J. (2016). Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. Technical report, International Labour Office, Geneva.

Bohning, D., Van der Heijden, P. G., and Bunge, J. (2017). *Capture-recapture methods for the social and medical sciences*. CRC Press.

Bonin, H. and Rinne, U. (2017). Omnibusbefragung zur Verbesserung der Datenlage neuer Beschäftigungsformen. Technical report, IZA.

Brancati, U., C., Pesole, A., and Fernández-Macías, E. (2020). New evidence on platform workers in Europe. Results from the second COLLEEM survey. Technical report, Publications Office of the European Union, Luxembourg.

Busch, O. (2016). *Programmatic advertising*. Springer, Berlin.

De Stefano, V. and Aloisi, A. (2018). *European Legal Framework for "Digital Labour Platforms"*. European Commission, Luxembourg.

Digital Care (2022). How do poles use their smartphones? Accessed: 28.04.2023.

Drahokoupil, J. and Piasna, A. (2019). Work in the Platform Economy: Deliveroo Riders in Belgium and the SMart Arrangement.

Eurofound (2018). Employment and working conditions of selected types of platform work. Technical report, Publications Office of the European Union, Luxembourg.

Eurofound (2019). Platform work: Maximising the potential while safeguarding standards ? Technical report, Publications Office of the European Union, Luxembourg.

Farrell, B. D., Greig, F., and Hamoudi, A. (2019). The Evolution of the Online Platform Economy: Evidence from Five Years of Banking Data. *AEA Papers and Proceedings*, 109:362–366.

Friedman, G. (2014). Workers without employers : shadow corporations and the rise of the gig economy. *Review of Keynesian Economics*, 2(2):171–188.

Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., Zagheni, E., Flores, R. D., Ventura, I., Weber, I., et al. (2022). Is facebook's advertising data accurate enough for use in social science research? insights from a cross-national online survey. *Journal of the Royal Statistical Society Series A*, 185(S2):S343–S363.

Hauben, H., Lenaerts, K., and Wayaert, W. (2020). The platform economy and precarious work. Publication for the committee on Employment and Social Affairs. Technical report, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg.

Huws, U., Spencer, N. H., Coates, M., and Holts, K. (2019). The platformisation of work in Europe: results from research in 13 European countries. Technical report, Foundation for European Progressive Studies, UNI Europa and University of Hertfordshire, Brussels.

ILO (2021). *World Employment and Social Outlook. The role of digital labour platforms in transforming the world of work.* International Labour Office, Geneva.

Kaczmarczyk, P., Kowalik, Z., and Lewandowski, P. (2022). Job quality gaps between migrant and native gig workers: evidence from poland: evidence from poland. IBS Working Paper WP 09/2022.

Koustas, B. D. K. (2019). What Do Big Data Tell Us about Why People Take Gig Economy Jobs? *AEA Papers and Proceedings*, 109:367–371.

Koutsimpogiorgos, N., Slageren, J. V., Herrmann, A. M., and Frenken, K. (2020). Conceptualizing the Gig Economy and Its Regulatory Problems. *Policy & Internet*, pages 1–21.

Lee, Z. W. Y., Chan, T. K. H., Balaji, M., and Chong, A. Y.-L. (2018). Why people participate in the sharing economy: an empirical investigation of Uber. *Internet Research*, 28(3):829–850.

Lehdonvirta, V. (2018). Flexibility in the gig economy : managing time on three online piecework platforms. *New Technology, Work and Employment*, 33(1):13–30.

Lenaerts, K. (2018). Industrial Relations and Social Dialogue in the Age of Collaborative Economy. Technical report, CEPS.

Meszmann, T. T. (2018). Industrial Relations and Social Dialogue in the Age of Collaborative Economy. National Report Hungary. Technical Report 27, CELSI.

Miller, K. M. and Skiera, B. (2023). Economic consequences of online tracking restrictions. arXiv preprint arXiv:2303.09147.

Mitea, R. D. E. (2018). The Expansion of Digitally Mediated Labor: Platform-Based Economy, Technology-Driven Shifts in Employment, and the Novel Modes of Service Work. *Journal of Self-Governance and Management Economics*, 6(4):7–12.

Neumann, N., Tucker, C. E., and Whitfield, T. (2019). Frontiers: How effective is third-party consumer profiling? evidence from field studies. *Marketing Science*, 38(6):918–926.

Office for Electronic Communication (2020a). Consumer survey 2020 – individual customers. Accessed: 28.04.2023.

Office for Electronic Communication (2020b). Report on the state of the polish telecommunications market in 2020. Accessed: 28.04.2023.

Piasna, A. and Drahokoupil, J. (2019). Digital labour in central and eastern Europe: evidence from the ETUI Internet and Platform Work Survey.

Rao, J. and Molina, I. (2015). *Small Area Estimation*. Wiley Series in Survey Methodology. Wiley, 2 edition.

Riggs, L., Sin, I., and Hyslop, D. (2019). Measuring the "gig" economy: Challenges and options.

Schenkel, J. F. and Zhang, L.-C. (2022). Adjusting misclassification using a second classifier with an external validation sample. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(4):1882–1902.

Sedláková, M. (2018). Industrial Relations and Social Dialogue in the Age of Collaborative Economy. National report: Slovakia. Technical Report 28, CELSI.

Selectivv (2021a). Gig economy in poland. Aggregated data based on smartphones `https://doi.org/10.5281/zenodo.5834791`, doi: 10.5281/zenodo.5834791.

Selectivv (2021b). Ukrainians in Poland 2020 – did they leave, did they come or stayed? https://selectivv.com/en/ukrainians-in-poland-2020-did-they-leave-did-they-come-or-stayed/.

Shao, J. and Tu, D. (2012). *The jackknife and bootstrap*. Springer Science & Business Media.

Shepherd, B. E., Han, K., Chen, T., Bian, A., Pugh, S., Duda, S. N., Lumley, T., Heerman, W. J., and Shaw, P. A. (2023). Multiwave validation sampling for error-prone electronic health records. *Biometrics*, n/a(n/a).

Statistics Poland (2020a). Appendix – The foreign population in Poland during the COVID-19 pandemic. https://stat.gov.pl/en/experimental-statistics/human-capital/appendix-the-foreign-population-in-poland-during-the-covid-19-pandemic,10,2.html.

Statistics Poland (2020b). Information society in Poland in 2020. https://stat.gov.pl/en/topics/science-and-technology/information-society/information-society-in-poland-in-2020,1,7.html.

Statistics Poland (2020c). Labour force survey in Poland IV quarter 2020. https://stat.gov.pl/en/topics/labour-market/working-unemployed-economically-inactive-by-lfs/labour-force-survey-in-poland-iv-quarter-2020,2,40.html.

Statistics Poland (2020d). The foreign population in Poland during the COVID-19 pandemic. https://stat.gov.pl/en/experimental-statistics/human-capital/the-foreign-population-in-poland-during-the-covid-19-pandemic,10,1.html.

Statistics Poland (2022). Information society in Poland in 2022. https://stat.gov.pl/en/topics/science-and-technology/information-society/information-society-in-poland-in-2022,2,12.html.

Van den Brakel, J. A. and Krieg, S. (2009). Estimation of the monthly unemployment rate through structural time series modelling in a rotating panel design. *Survey Methodology*, 35(2):177–190.