

Catalogue no. 11F0019M — No. 469
ISSN 1205-9153
ISBN 978-0-660-48100-5

Analytical Studies Branch Research Paper Series

Paid Employment, Self-employment and Gig Work in Administrative and Survey Data

by Ping Ching Winnie Chan, Sung-Hee Jeon and
Yuri Ostrovsky

Release date: June 6, 2023



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under “Contact us” > “[Standards of service to the public.](#)”

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© His Majesty the King in Right of Canada as represented by the Minister of Industry, 2023

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

Cette publication est aussi disponible en français.

Paid Employment, Self-employment and Gig Work in Administrative and Survey Data

by

Ping Ching Winnie Chan, Sung-Hee Jeon and Yuri Ostrovsky

Social Analysis and Modelling Division
Statistics Canada

11F0019M No.469
2023003
ISSN 1205-9153
ISBN 978-0-660-48100-5

DOI: <https://www.doi.org/10.25318/11f0019m2023003-eng>

June 2023

Analytical Studies Branch Research Paper Series

The Analytical Studies Branch Research Paper Series provides for the circulation of research conducted by Analytical Studies Branch staff and collaborators. The Series is intended to stimulate discussion on a variety of topics, such as labour, immigration, education and skills, income mobility, well-being, aging, firm dynamics, productivity, economic transitions, and economic geography. Readers of the Series are encouraged to contact the authors with their comments and suggestions.

All the papers in the Analytical Studies Branch Research Paper Series go through institutional and peer review to ensure that they conform to Statistics Canada's mandate as a governmental statistical agency and adhere to generally accepted standards of good professional practice.

Table of contents

Abstract	5
Executive summary	6
1 Introduction	7
2 Data	8
3 Results	10
3.1 Mapping the Labour Force Survey labour market activities into tax data income source categories	10
3.2 What can be learned about sole proprietors and gig workers from the linked data from the Labour Force Survey and the Longitudinal Worker File?	13
3.3 Regression analysis	17
4 Conclusions	20
5 Appendix	21
Appendix A: The impact of age on the linkage rate between the Labour Force Survey and the Longitudinal Worker File.....	21
Appendix B: Methodology	23
Appendix C: A comparison between results based on alternative annualization methods ..	24
References	28

Abstract

This study combines survey and administrative data to examine the correspondence between paid-employment and self-employment activities reported in each of these data sources by the same individuals. The study also looks at the role of self-employment as a supplemental income source for individuals whose self-declared main labour market activity is wage employment. It uses tax data information to identify gig workers specifically and to examine possible links between various aspects of the main wage job and participation in gig work activities. The analysis is based on data from the 2016 Labour Force Survey linked to 2016 administrative data from the Longitudinal Worker File.

Keywords: self-employment, incorporated, unincorporated, gig economy, gig work, alternative work arrangements, Longitudinal Worker File, Labour Force Survey

Disclaimer: The definition of gig work used in this study is not an official Statistics Canada definition of gig work.

Executive summary

This study combines survey and administrative data to document both similarities and differences between labour market activities reported by survey respondents and the income sources of the same respondents recorded in tax data. The study is part of a growing literature that examines linked survey and administrative data to create a more complete picture of individual labour market activities than that which can be seen in either of these sources alone. It also contributes to the recent literature on alternative work arrangements or “gig work.”

The analysis is based on two main data sources. The first is the 2016 Labour Force Survey (LFS), which has served as the main source of official labour statistics in Canada since 1945. The second data source is the 2016 Longitudinal Worker File (LWF). The LWF is a database in which data from various administrative sources, such as individual tax returns (T1) and Statements of Remuneration Paid (T4), are linked together using unique individual and business identifiers. Based on the T1 information in the LWF, it is possible to identify five income sources usually associated with unincorporated self-employment: fishing, farming, business, professional and commission incomes. The focus of this study is on self-employed individuals (sole proprietors) with non-zero business, professional and commission income, because these are the labour market activities usually associated with non-traditional work arrangements and gig work. Not all sole proprietors are gig workers. An essential element of gig work is a low expectation of continuity and weak predictability of future earnings. This study identifies gig workers in the same way as Jeon et al. (2021) do. The analytical sample was constructed by linking all 2016 LFS monthly records to the 2016 LWF annual file.

Close to 98% of LFS respondents who said their main activity was wage employment also had wage employment income in tax data, and more than 80% of those whose main labour force status in the LFS was unincorporated self-employed reported non-zero self-employment income on their tax returns. The share of the LFS incorporated self-employed who could be identified as owners of incorporated businesses in the LWF was 74.5%.

Among wage employees in the LFS, individuals who were permanent full-time employees—the largest category—were less likely to be either sole proprietors (5.8%) or gig workers (4.3%) than other wage workers.

The study also shows that 9.6% of individuals who were temporary part-time employees in their main job were identified as gig workers in the administrative data. The regression analysis indicates that, all else being equal, the following employees are more likely than others to be identified as gig workers in administrative data: part-time employees; employees involved in temporary jobs; university degree holders; employees working in small firms; and those employed in educational services, in information and cultural industries, and in arts, entertainment and recreation.

1 Introduction

This study combines Canadian survey data from the Labour Force Survey (LFS) and administrative data from Statistics Canada’s Longitudinal Worker File (LWF) to document both similarities and differences between labour market activities reported by survey respondents and the income sources of the same respondents recorded in tax data. The study is part of a growing literature that examines linked survey and administrative data to create a more complete picture of individual labour market activities than that which can be seen in either of these sources alone. It also contributes to the recent literature on alternative work arrangements or “gig work.” Using the methodology for identifying gig workers in administrative data proposed by Jeon et al. (2021), this study focuses on workers whose main activity in the LFS is paid employment, and it documents the relationship between various characteristics of their jobs and participation in gig work. The study examines how aspects of the precariousness of a worker’s main job, such as part-time status and temporary employment, are correlated with the presence of self-employment and gig income in tax data.

The labour market activities of the same individual can be different in survey and tax data for various reasons. Surveys usually ask only about main and sometimes secondary labour market activities, while tax data record income from all activities, including minor “side” jobs. Respondents in household surveys may not be fully familiar with the labour market activities of other household members and provide incomplete information about their activities. There is also the possibility of a recall or data entry error, or reluctance to mention certain activities to a survey interviewer. Tax data are an attractive alternative to survey data, but they usually reveal little information about the nature of a job, hours of work or hourly wages. Individual tax data are not well suited for identification of certain types of self-employment, particularly incorporated self-employment. In short, both survey and tax data have advantages and drawbacks, and it is important to understand the degree to which the labour market information from these sources overlaps and what can be learned about individuals’ labour market activities by combining information from both sources.¹

Linked survey and tax data also offer new possibilities to researchers trying to measure the size of the gig economy and understand motivations behind participation in gig work. There is still substantial disagreement in the literature regarding the extent of the gig economy, with studies based on survey data usually documenting a substantially higher share of individuals involved in informal or non-traditional work arrangements than studies based on administrative data.² Part of the problem is that terms like “gig work” are difficult to define. Occasionally working as an Uber driver during weekends would be considered a gig activity by most observers; working as an Uber driver every day for several hours as a main income-generating activity may not be recognized as gig work. Some studies apply the term “gig work” only to work activities mediated by online platforms, while other studies define gig work in terms of work attributes, regardless of how the work activity is mediated (Alake-Apata 2021).³

Abraham et al. (2018) introduced a conceptual framework for identifying workers involved in the gig economy based on a typology of work arrangements and a set of characteristics associated

-
1. In a recent study, Abraham et al. (2021) found that only 48.5% of those who were self-employed in survey data also had self-employment recorded in tax data in the United States. They concluded that the magnitude of the differences “reflects the complexity of self-employment activity. There are many different types of self-employment work and a highly heterogeneous set of arrangements under which such work might occur. Neither the household survey data nor the administrative data may be ideally suited to pick up all of that activity” (p. 18).
 2. A study based on the Bank of Canada’s Canadian Survey of Consumer Expectations found that about 30% of the Canadian population were involved in the gig economy (Kostyshyna and Luu 2019), while a study based on administrative data put the number at around 8% to 10% (Jeon et al. 2021). Recent U.S. studies revealed a similar gap: a study based on survey data found that about 28% of respondents were involved in an informal work activity during the previous month (Abraham and Houseman 2019), while the estimates based on administrative data generally fell between 8% and 12% (Collins et al. 2020; Lim et al. 2019).
 3. Koustas (2020) noted that gig workers are part of the “alternative workforce,” a broader category that also includes wage employees doing temporary and contingent jobs.

with each such arrangement. This framework makes a broad distinction between employees and self-employed individuals, and it further categorizes the self-employed into business owners, independent contractors or freelancers, day labourers and on-demand platform workers. The work arrangement characteristics of the last three categories of self-employed individuals—freelancers, day labourers and on-demand workers—are different from all other work arrangements: they are not paid wages or salaries, they work on a task basis and do not have a contract for a continuing relationship, and they do not have a determined work schedule or predictable earnings. Abraham et al. (2018) deemed this category “gig workers” and introduced a methodology that allowed them to link the work arrangement characteristics of gig workers with specific forms and schedules required by the Internal Revenue Service to report these work arrangements to the U.S. tax authorities.

Using the typology of work arrangements introduced by Abraham et al. (2018), Jeon et al. (2021) developed a methodological strategy to identify gig workers in Canada using Canadian administrative data. In addition, they linked individual tax data to the census records of the same individuals to expand the inquiry into the determinants of gig work participation by capturing important human capital characteristics of gig workers unavailable in tax data, such as their highest level of educational attainment and occupation. The LFS–LWF data used in this study offer a further opportunity to examine the role of self-employment and gig work in supplementing income from the main labour market activity.⁴ One of the main questions in the literature related to gig work is why individuals, particularly those who have a wage job and whose main labour market activity is wage employment, engage in gig work. Looking in greater depth at the relationship between main job characteristics and the likelihood of participation in gig work is one of the objectives of the study.

2 Data

This study is based on two main data sources. The first is the LFS, which has served as the main source of official labour statistics in Canada since 1945. The LFS target population is the non-institutionalized population aged 15 years and older (Statistics Canada 2022). Responding to the survey is mandatory, and responses are collected for all household members. The LFS collects information about various aspects of employment and unemployment, including individuals’ labour force status, earnings, work hours, part-time or full-time employment status, industry, occupation, self-employment and employment insurance benefits. The LFS is a monthly survey that uses a rotating panel sample design. Selected individuals remain in the LFS sample for six consecutive months, and one-sixth of the total sample is replaced every month to start a new six-month panel. Each monthly sample consists of about 100,000 individuals from about 56,000 dwellings.

The second key data source is the LWF. The LWF is a database in which data from various administrative sources, such as individual tax returns (T1) and Statements of Remuneration Paid (T4), are linked together using unique individual and business identifiers (Statistics Canada 2021). Individuals receiving wages and salaries can be identified in the LWF using information from annual T4 files. Incorporated self-employed individuals cannot be identified from either T1 or T4 data. However, owners of incorporated businesses can be identified from Schedule 50 (Shareholder Information) data, which are now also part of the LWF. A Schedule 50 lists all owners of a private corporation with shares of 10% or more and is attached to the corporation tax return (T2). Hence, owners of incorporated enterprises can be identified in the LWF based on their presence in Schedule 50 files.

Based on the T1 information in the LWF, it is possible to identify five income sources usually associated with unincorporated self-employment: fishing, farming, business, professional and

4. Jeon et al. (2021) reported that 51.4% of gig workers in 2016 had at least one wage job.

commission incomes.⁵ The focus of this study is on self-employed individuals with non-zero business, professional and commission income, since these are the labour market activities usually associated with non-traditional work arrangements and gig work. The key to the analysis below is the Statement of Business or Professional Activities (Form T2125) used by unincorporated self-employed individuals to report their business, professional and commission income (and business expenses) as part of their individual tax returns (T1). Statistics Canada receives annual T2125 information, along with information from several related files from the Canada Revenue Agency, and aggregates it in financial declaration (FD) files more suitable for research purposes. The 2016 LWF was merged with FD files to identify tax filers reporting T2125 income in 2016. Unincorporated self-employed individuals reporting T2125 income can be either sole proprietors or partners in partnerships (about 8.5% in 2016). Hereafter, all unincorporated self-employed individuals who have non-zero T2125 income and are not partners in partnerships will be referred to simply as “sole proprietors.”

Not all sole proprietors are gig workers. Essential elements of gig work are a low expectation of continuity (relative to wage earners and those who operate a well-established business) and weak predictability of future earnings. The Jeon et al. (2021) strategy of identifying gig workers in administrative data takes into consideration whether a sole proprietor has a business number (BN)—it takes the absence of a BN as a signal of weaker expectations of business continuity and lesser predictability of future earnings, which are the main characteristics of gig work, according to the typology of work arrangements by Abraham et al. (2018). This study identifies gig workers in the same way as Jeon et al. (2021).

The analytical sample is constructed by linking all 2016 LFS monthly records to the 2016 LWF. Because the LFS and administrative data have different unique individual identifiers, the linkage between the two data sources requires probabilistic linkage. The linkage process is based on the classic Fellegi and Sunter (1969) theory of record linkage. In total, 292,100 unique “linkable” LFS respondents aged 15 and older were identified in 2016.⁶ About 246,100 of them could be linked to LWF records, resulting in an 84.2% linkage rate.

An important question is whether there are any systematic differences between linked and non-linked LFS individuals; this issue is investigated in Appendix A. The primary takeaway from the analysis in Appendix A is that the main underrepresented category in the linked data is individuals younger than 20, while those aged 65 and older are somewhat overrepresented. For this reason, the main analysis in this study is restricted to individuals from 20 to 64 years of age.⁷

The main methodological challenge in this study is related to linking **monthly** LFS records to **annual** LWF data. A possible annualization strategy of converting higher frequency (monthly) into lower frequency (annual) LFS records could involve some sort of averaging or aggregation across all monthly records for each individual. However, a considerable drawback of this approach is that it would necessarily cause at least a partial loss of information, which is especially undesirable in the context of this study. For instance, a transformation of the person-month-level data into person-level data would require either assigning a single “annual” labour force status to individuals whose labour force status changes from month to month, or trying to capture such transitions with additional variables, which can be quite messy.

5. The T1 information is drawn from the T1 personal master file (T1PMF), which does not include data on those who submitted their tax returns after the Canada Revenue Agency deadline (usually April 15 of each calendar year). Messacar (2014) compared key demographic and labour market characteristics of tax filers in the T1PMF and T1 historical files, which include late filers, and concluded that “late filers tend to represent a sufficiently small group of people that biases resulting from their exclusion are negligible” (p. 12).

6. All counts are rounded to the nearest 100.

7. The downside of excluding younger and older individuals from the analysis is that the results are not representative of the whole LFS population. To address this issue, a set of alternative results was obtained for LFS respondents aged 15 and older, and the results based on the broader sample were compared with the main results. The differences between the two sets of results were minor. (The results for the “15 and older” sample are available upon request.)

Instead of aggregating monthly records into a single annual person-level record in an ad hoc way, the approach taken in this study is to keep all individual monthly records but divide each individual weight by 12 (the number of months in a year) to maintain the representativeness of the LFS sample at the annual level. Technical aspects of this annualization strategy are discussed in Appendix B. The main advantage of this approach, compared with the averaging alternatives, is that it preserves all the information available in the LFS. This is especially important in the case of labour market activities that can change from month to month. Temporary jobs last for only a short period of time, and a substantial number of them are likely to be aggregated out if individual monthly data are combined into a single person-level annual record. On the downside, this strategy treats multiple observations for the same individuals in the same way as it does single observations for multiple individuals. To deal with this issue in the regression analysis in Section 5, standard errors will be clustered on the individual.

Although the preferred annualization strategy implemented in this study is to retain all person-month observations, the sensitivity of the main results to the choice of annualization method was also investigated. For the sensitivity analysis, LFS respondents who were observed for less than six months were dropped from the main sample, and the resulting subsample was analyzed using the preferred annualization strategy that retains all person-month records (person-month-level analysis) and an alternative strategy that aggregates all personal monthly records into a single individual record (person-level analysis). Details of the sensitivity analysis and its results are discussed in Appendix C.

3 Results

3.1 Mapping the Labour Force Survey labour market activities into tax data income source categories

The LFS asks respondents about their primary and secondary labour market activities. The first question examined in this study is the following: what are the T1-based income sources of survey respondents who report various paid-employment and self-employment activities in the LFS? As the first step in answering this question, the primary activities of currently employed individuals were broken down by secondary activity to better understand the patterns of multiple activities in the LFS (Table 1). Among currently employed LFS respondents, 94.3% reported not having any secondary activity. The percentage was slightly higher for employees (94.5%) but lower for those whose primary activities were incorporated (94.0%) and unincorporated (92.3%) self-employment. Therefore, according to the LFS data, a large majority of labour market participants were involved in only one labour market activity at the time of the survey.

Table 1
Labour Force Survey employment and self-employment status of multiple job holders

Primary activity	Secondary activity			
	No	Yes		
		Wage employee	Incorporated self-employed	Unincorporated self-employed
		percent		
Currently employed	94.3	64.8	9.8	24.6
Wage employee	94.5	70.5	6.7	22.2
Incorporated self-employed	94.0	26.1	52.5	19.6
Unincorporated self-employed	92.3	46.9	5.0	46.6

Notes: Labour Force Survey respondents aged 20 to 64. The numbers in the "Yes" columns represent row percentages of those who reported a secondary activity (unpaid work excluded).

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

For individuals who did report a secondary activity, it was often the same as their primary activity: 70.5% of those whose primary activity was wage employment, 52.5% of those whose primary activity was incorporated self-employment and 46.6% of those whose primary activity was unincorporated self-employment were in the same category in their secondary activity (Table 1). Notably, conditional on having a secondary activity, 46.9% of those whose main labour market activity was unincorporated self-employment reported wage employment as their secondary activity. Primarily unincorporated self-employed respondents were also considerably more likely to be wage-employed (in addition to being self-employed) than the primarily incorporated self-employed (26.1%). The relatively low percentage of the incorporated self-employed whose secondary activity was wage employment is consistent with the emerging evidence of differences between the characteristics of incorporated and unincorporated self-employed individuals, and also the possible differences in the degree of commitment that these two types of self-employment require (Levine and Rubinstein 2017).

Based on income information for the same individuals, Table 2 directly maps LFS main activities into three LWF labour market categories—wage employment, incorporated self-employment and unincorporated self-employment. An important feature of Table 2 is that individuals can have multiple sources of income in tax data, so the row percentages can exceed 100%. In the first three rows, the diagonal numbers are quite high for all LFS categories. About 97.6% of LFS respondents who said their main activity was wage employment also had wage employment income in tax data. One of the salient results in Table 2 is that 80.7% of those whose main labour force status in the LFS was unincorporated self-employed indicated non-zero self-employment income on their tax returns. This number implies that 19.3% of the unincorporated self-employed in the LFS reported no self-employment income in the tax data.⁸

Table 2
Mapping Labour Force Survey employment status and main labour market activities to
Longitudinal Worker File employment status and income sources

LFS primary activity	LWF status			Total (weighted counts)
	Wage employee (T4 income)	Incorporated self-employed	All unincorporated self-employed	
	row percent			count
Currently employed				
Wage employee	97.6	4.4	7.8	11,550,900
Incorporated self-employed	68.4	74.5	24.7	897,900
Unincorporated self-employed	29.2	10.4	80.7	1,036,300
Not currently employed				
Unemployed	70.8	4.0	8.9	876,100
Not in labour force	28.3	4.3	6.9	3,284,400

Notes: LWF = Longitudinal Worker File; LFS = Labour Force Survey. LFS respondents aged 20 to 64. Row percentages may exceed 100% because LFS respondents can have income from multiple sources in the LWF. All counts are rounded to the nearest 100.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

There may be several reasons why there are unincorporated self-employed individuals in the LFS who have no self-employment income in the tax data. Boeri et al. (2020) note that “[i]n survey data, workers are often confused about the nature of their employment relationship” (p. 174). One possibility is that some of the self-employed erroneously report their self-employment income as

8. To put this result in perspective, this is a far lower percentage than the 51.5% of unincorporated self-employed survey respondents who reported no self-employment income in tax data documented in a U.S. study by Abraham et al. (2021).

employment income or “other income.”⁹ Another possibility is that some non-working LFS respondents think about their perceived status rather than current income sources, when asked about their main labour market activity. For instance, individuals may still think of themselves as being mainly self-employed, even if they are retired or did not earn any self-employment income that year. Some individuals may mistakenly report being unincorporated self-employed, even though they own an incorporated business. As always, there is also a possibility of a recording or processing error.¹⁰

Table 1 shows that less than 4% of those whose main labour market activity was unincorporated self-employment mentioned wage employment as their secondary activity. However, the share of mainly unincorporated self-employed in the LFS who had wage employment earnings was considerably higher, at 29.2% (Table 2). Some unincorporated self-employment activities last only a short period of time, so it is likely that some of those who were unincorporated self-employed at the time of a monthly interview were wage-employed at a different point during the same calendar year. It is also possible that some of the unincorporated self-employed neglected to mention occasional wages as their secondary activity.

Table 2 also shows that the share of the LFS incorporated self-employed who could be identified as owners of incorporated businesses in the LWF was 74.5%, which is comparable to the corresponding number for unincorporated self-employment (80.7%). Some of the reasons not all incorporated self-employed individuals in the LFS are identified as such in the LWF are similar to those mentioned for the unincorporated self-employed. Additionally, only those who own 10% or more shares of a private enterprise are required to be listed in a Schedule 50, so smaller shareholders are not identified as incorporated business owners in the LWF. Many owners of incorporated businesses pay themselves a salary, so it is not surprising that 68.4% of the incorporated self-employed in the LFS received a T4 in the LWF.

A considerable number of individuals who were not employed during the LFS reference week either had wage earnings or were self-employed in 2016. Among LFS respondents who said they were currently unemployed, 70.8% received T4 earnings at some point during 2016. Some of those who were “not in the labour force” also received T4 earnings (28.3%), and smaller but substantial shares either owned an incorporated business (4.3%) or had self-employment income (6.9%).

When an LFS interviewer contacts a household, a single household member usually provides information for all household members. One possible concern related to the results above is that they may be influenced by proxy response. In particular, individuals responding to interviewers’ questions may not always be aware of the self-employment status of other household members and may conflate it with wage employment. To investigate this possibility, the main sample was split into two subsamples—direct and proxy respondents—and Table 2 was replicated for each of the subsamples.¹¹ Although the results were generally similar for both subsamples, there was one notable difference: the percentage of the unincorporated self-employed in the LFS who had wage earnings was lower in the direct response subsample (25.3%) than in the proxy response subsample (32.3%), and the percentage of the unincorporated self-employed in the LFS who also had self-employment income in the LWF was higher in the direct response subsample (85.1%), compared with the proxy response subsample (77.2%). The results appear to give some credence to the view that household members interviewed for the LFS may not always be fully aware of the self-employment status of other household members, and when the self-employed have a chance to provide direct responses regarding their self-employment status, the information they provide may better correspond with their actual labour force status. However, what is important in the

9. “Other income” is a tax category that refers to taxable income not listed elsewhere in the tax return, such as project grants, loans, certain annuity payments and training allowances. In the past, it also included limited partnership income.

10. Abraham et al. (2021) also mentioned the possibility of survey respondents thinking about their “off the books” work, which they may not report to the tax authorities.

11. These results are available upon request.

context of this study is that the magnitude of the differences does not appear to alter the main conclusions drawn from the results in Table 2, which hold for both subsamples: a large majority of the unincorporated self-employed in the LFS have self-employment earnings in the LWF, and the size of this majority is generally similar, whether the proxy response is included or excluded.¹²

3.2 What can be learned about sole proprietors and gig workers from the linked data from the Labour Force Survey and the Longitudinal Worker File?

Previous studies have suggested that many unincorporated self-employed individuals, and especially gig workers, use their self-employment earnings to supplement earnings from their main labour market activity. The linked LFS–LWF data offer an opportunity to take a closer look at the degree of involvement in unincorporated self-employment and gig work among LFS respondents, using detailed characteristics of their main activities. The analysis in this subsection focuses on sole proprietors and excludes the unincorporated self-employed who report income from partnerships.¹³ Table 3 shows the shares of sole proprietors for each LFS primary activity broken down by full time vs. part time and permanent vs. temporary status. LFS respondents who identified themselves as part-time unincorporated self-employed in their main job were most likely to be sole proprietors (73.2%). They were also most likely to be identified as gig workers based on their tax records (52.2%). Among wage employees, individuals who were permanent full-time employees—the largest category—were less likely to be either sole proprietors (5.8%) or gig workers (4.3%) than other wage workers. Permanent full-time and temporary part-time wage workers are essentially opposites in terms of the precariousness of their employment, and this appears to be reflected in their propensity to be engaged in self-employment and gig work. The percentage of sole proprietors among currently unemployed LFS respondents was substantial (7.8%) and similar to the percentage of sole proprietors among wage employees who worked full time but were temporarily employed (8.2%), and substantially lower than the percentage of sole proprietors among temporarily employed wage workers who worked part time (11.0%). By contrast, 5.3% of LFS respondents who said that they were not currently in the labour force were sole proprietors.

12. For a comparison of the results in Table 2 with results obtained using alternative aggregation strategies, see Appendix C.

13. Unincorporated self-employed individuals reporting partnership income represented about 7.6% of all individuals with T2125 income in 2016.

Table 3**Prevalence of sole proprietorship and gig work, by Labour Force Survey primary activity, permanent vs. temporary and full-time vs. part-time status**

LFS primary activity	All sole proprietors	Gig workers percent	Share of gig workers among all
			sole proprietors
Not currently employed, unemployed	7.8	6.3	80.6
Not currently employed, not in labour force	5.3	4.5	85.6
Currently employed, wage employee	6.5	5.1	77.3
Permanently employed full time	5.8	4.3	74.7
Permanently employed part time	9.0	7.5	84.0
Temporarily employed full time	8.2	6.5	79.7
Temporarily employed part time	11.0	9.6	87.2
Currently employed, incorporated self-employed	18.8	11.6	61.6
Full time	17.7	10.7	60.1
Part time	26.8	18.5	68.8
Currently employed, unincorporated self-employed	70.7	38.9	55.0
Full time	70.0	33.4	47.9
Part time	73.2	52.2	71.4

Notes: LFS = Labour Force Survey. LFS respondents aged 20 to 64.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

The last column in Table 3 shows that the share of gig workers among sole proprietors whose primary labour market activity was wage employment was the lowest for permanent full-time employees (0.747) and the highest for temporary part-time workers (0.872).¹⁴ These numbers strongly suggest that sole proprietors whose main labour market activities are associated with precarious employment are particularly likely to be gig workers. Although it is not entirely clear whether those who hold precarious jobs feel it is necessary to take on gig work to supplement their income, or whether they choose precarious jobs intentionally to have more time for involvement in gig activities, the fact that they highlight wage employment as their main labour market activity increases the likelihood that gig work is done to supplement wage employment earnings.

Another interesting result in the last column of Table 3 is related to the share of gig workers among sole proprietors whose main labour market activity in the LFS was unincorporated self-employment. First, this share was lower (55.0%) for the unincorporated self-employed in the LFS than for any other category. Second, this share was considerably smaller for those who reported being full-time self-employed workers than for those who reported being only part-time self-employed. These results highlight the conceptual difference between being a gig worker and owning a stable (unincorporated) business with some expectation of stability, continuity and predictability of future income. Those who report being self-employed full time are likely to belong to the latter category, so it is not surprising that the share of gig workers among all sole proprietors for this LFS category was the lowest among all categories.¹⁵

Further insights into the relative importance of self-employment income and informal work can be gained by considering how a decision to become self-employed or engage in gig work by one spouse or partner is related to the main labour market activity of the other spouse or partner. This

14. These findings echo the results of Abraham and Houseman (2019), who found that part-time workers are considerably more likely to be engaged in informal work than full-time workers. Similar to this study, they also found high prevalence of informal work among unemployed individuals and low prevalence of informal work among those not in the labour force.

15. To investigate possible gender differences in the prevalence of unincorporated self-employment and gig work, Table 3 was redone separately for men and women. The results were generally similar, although the prevalence of gig work was higher for women, particularly those whose main labour market activity was unincorporated self-employment. These results are available upon request.

relationship is determined by numerous factors, and it is difficult to say a priori whether, for example, having a self-employed spouse would increase or decrease the rate of participation in self-employment and gig work among individuals whose main activity is wage employment. The linked data make it possible to break down the prevalence of sole proprietorship and gig work, not just by the main labour market activities of LFS respondents, but also by the labour market activities of their spouses or partners. These activities include not working, being a wage employee and being self-employed. Table 4 shows that having a self-employed spouse was associated with the highest prevalence of being a sole proprietor for both men and women whose main labour market activity in the LFS was not self-employment. However, the opposite was true for respondents whose main activity was either incorporated or unincorporated self-employment. For instance, 55.0% of unincorporated self-employed men with self-employed spouses were sole proprietors, compared with 70.3% of unincorporated self-employed men whose spouses were wage-employed. For unincorporated self-employed women, the corresponding numbers were 60.3% and 83.0%. Table 4 also shows similar patterns of involvement in gig work, although related gender differences were even more notable: whereas the shares of gig workers among unincorporated self-employed men with wage-employed and self-employed spouses were similar (25.7% and 24.3%), the share of gig workers was substantially higher for unincorporated self-employed women with wage-employed spouses than for those with self-employed ones (58.1% and 38.2%).

Table 4
Prevalence of self-employment and gig work, by respondent's and spouse's Labour Force Survey main labour market activity

Primary activity in LFS	Not in a couple		Self-employed spouse		Wage-employed spouse		Non-working spouse	
	Sole proprietors	Gig workers	Sole proprietors	Gig workers	Sole proprietors	Gig workers	Sole proprietors	Gig workers
percent								
Men								
Not currently employed								
Unemployed	6.0	4.7	13.9	9.9	10.3	7.6	6.7	4.1
Not in labour force	3.9	3.3	11.7	9.8	5.9	4.2	6.0	4.6
Currently employed								
Wage employee	5.3	4.1	12.0	7.8	6.5	4.4	6.8	4.2
Incorporated self-employed	21.3	13.8	13.9	9.4	16.2	8.4	18.3	9.6
Unincorporated self-employed	67.1	33.6	55.0	24.3	70.3	25.7	65.2	26.8
Women								
Not currently employed								
Unemployed	7.5	6.9	12.1	11.0	6.9	6.1	5.8	5.0
Not in labour force	4.2	3.9	10.8	9.9	6.3	5.5	3.4	2.9
Currently employed								
Wage employee	6.5	5.6	9.6	8.2	6.2	5.3	5.9	4.9
Incorporated self-employed	29.7	20.2	15.7	22.8	28.3	18.9	24.5	15.7
Unincorporated self-employed	77.7	52.1	60.3	38.2	83.0	58.1	78.2	53.9

Notes: LFS = Labour Force Survey. LFS respondents aged 20 to 64.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

These results underscore the complexity of the relationship that may exist between the primary labour market activities of spouses and their involvement in self-employment. The relatively high percentage of wage-employed and non-employed respondents with self-employed spouses may be related to income splitting opportunities for self-employed spouses that do not exist for wage-employed spouses in Canada (Lloyd 2020). A non-working spouse may report self-employment income from a family business to reduce the tax burden on the spouse who runs the business. Income splitting, however, does not seem to explain a particularly high percentage of sole proprietors among self-employed LFS respondents with wage-employed spouses.

So far, the analysis has focused mainly on the primary LFS activity. Table 5 shows that those who reported a secondary activity were considerably more likely to be sole proprietors: 30.1% of LFS respondents with a secondary activity were sole proprietors, while this was the case for 11.2% of those with no secondary activity. Similar disparities were observed in the percentage of gig workers: 22.0% of those with a secondary activity were identified as gig workers, while the corresponding number for those without a secondary activity was only 7.3%. Two-thirds of LFS respondents who identified their secondary activity as unincorporated self-employment were sole proprietors in the tax data, and almost half of them (48.3%) were gig workers. About 27.7% of those whose secondary activity was incorporated self-employment and 16.2% of those whose secondary activity was wage employment were sole proprietors, according to the tax data.

Table 5
Prevalence of gig and non-gig sole proprietors, by Labour Force Survey secondary activity

LFS secondary activity	All sole proprietors	Gig workers	Other sole proprietors (non-gig workers)
	percent		
No secondary activity	11.2	7.3	4.0
Secondary activity	30.1	22.0	8.1
Wage employee	16.2	12.7	3.5
Incorporated self-employed	27.7	18.0	9.8
Unincorporated self-employed	67.7	48.3	19.5

Notes: LFS = Labour Force Survey. LFS respondents aged 20 to 64.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

Overall, LFS respondents who engaged in more than one labour market activity were likelier to be sole proprietors in the LWF, even if their secondary activity in the LFS was not unincorporated self-employment. Self-employment income may come from the main or secondary LFS activities if either of these is unincorporated self-employment, but it may also come from additional activities not reported in the LFS. The results in Table 6 also suggest that the self-employment income for those who reported a secondary activity was much more likely to be associated with gig work (22.0%) than with running an established and stable business (i.e., with being a sole proprietor who is not a gig worker) (8.1%).

Table 6
Average hourly wages and percentages of all sole proprietors and gig workers, by hourly wage quintile

LFS hourly wages	Average wage	All sole proprietors	Gig workers
	dollars	percent	
Lowest quintile	12.6	6.0	5.0
Second quintile	18.0	7.0	5.6
Middle quintile	23.5	6.3	4.8
Fourth quintile	31.7	6.6	4.9
Upper quintile	48.5	6.8	4.9
All quintiles	26.8	6.5	5.0
		hours	
Average hours of work	...	34.8	34.0

... not applicable

Notes: LFS = Labour Force Survey. LFS respondents aged 20 to 64.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

Abraham and Houseman (2019) made an important observation that “[t]he value of informal work to the household engaging in it could be considerable even if the aggregate amount of income it

generates is modest” (p. 111). They found that 38.4% of those who found it “difficult to get by” financially were involved in some kind of informal work, compared with 24.4% of those “living comfortably.” Moreover, 19.0% of individuals in the “difficult to get by” category were engaged in two or more informal work activities. Jeon and Ostrovsky (2020) showed that the degree of dependence on self-employment income is very different for gig workers who also have T4 earnings and for those who do not. Those with T4 earnings generally derive a much smaller share of their total annual income from self-employment income, regardless of whether their total income is low or high. However, skilled professionals who earn high hourly wages do not necessarily have high T4 annual earnings—these depend not just on the hourly wage but also on whether the individual works part time or full time, and whether they were continuously employed throughout the calendar year, or experienced employment interruptions. Hourly wages are usually not available in administrative data, so administrative data alone are insufficient to determine whether low- or high-wage workers are more likely to do gig work.

The LFS–LWF data make it possible to look directly at the hourly wages of workers and classify them according to their place in the wage distribution. To gauge the relative importance of gig work and unincorporated self-employment activities for high- and low-wage earners, the hourly wage distribution was constructed for individuals whose primary LFS activity was wage employment. Based on this distribution, individuals were sorted into five hourly wage quintiles, from lowest to highest. The first column in Table 6 shows the average wage in each hourly wage quintile: from \$12.6 in the lowest quintile to \$48.5 in the highest quintile. The lowest wage quintile also had the lowest prevalence of sole proprietors (6.0%). However, the percentages of gig workers were about the same in the lowest (5.0%) and highest (4.9%) quintiles. The highest prevalence of sole proprietors (7.0%) and gig workers (5.6%) was observed in the second quintile. Above the second quintile, the percentage of sole proprietors appears to increase with hourly wage, but there were no discernible across-quintile differences in the prevalence of gig work.

3.3 Regression analysis

Jeon et al. (2021) found that about half of individuals identified as gig workers in administrative data also had earnings from wage employment. Individuals whose main labour market activity in the LFS was wage employment were subject to a more formal investigation into how different individual and work characteristics of wage workers are associated with being an unincorporated sole proprietor and, more specifically, a gig worker. The regression analysis focused on the LFS variables that appeared to be closely associated with participation in self-employment and gig work activities in the analysis above. The following probit model was estimated:

$$P(Y_{it} = 1 | S_{it}^m, H_{it}^l, \log W_{it}, X_{it}) = F \left(\alpha + \sum_m \gamma_m S_{it}^m + \sum_l \gamma_l H_{it}^l + \gamma_w \log W_{it} + \theta X_{it} \right), \quad (3)$$

where Y_{it} is the outcome for individual i at time t ; S_{it}^m is a set of dummy variables for the highest levels of educational attainment m ; H_{it}^l is a set of dummies for different categories of work status (e.g., temporary and seasonal work); $\log W_{it}$ is the log of hourly wages; and X_{it} are controls that include age, sex, family status, region of residence, industry of the main job and a dummy variable for part-time status.¹⁶ All the explanatory variables are LFS-based, and the categorical variables are categorized as shown in Table 7. The model was estimated on the subsample of LFS respondents aged 20 to 64 whose main labour market activity was wage employment, the only category for which hourly wages were available. Standard errors were clustered at the individual

16. The hourly wage variable enters the model in its log form because, moving from lower to higher values in the wage distribution, the impact of a \$1 hourly wage increase on the probability of a positive outcome is likely to be diminishing.

level because, as mentioned in Section 3, most LFS respondents contributed multiple observations.

Table 7
Estimation results for probability of Longitudinal Worker File labour market status as a function of Labour Force Survey variables

LFS variables	LWF: All sole proprietors		LWF: Sole proprietors, gig workers		LWF: Sole proprietors, non-gig workers		LWF: Incorporated self-employed	
	partial effects	standard errors	partial effects	standard errors	partial effects	standard errors	partial effects	standard errors
Female	-0.010000 ***	0.0023	-0.00056	0.0021	-0.00940 ***	0.0012	-0.0096 ***	0.0020
Age (reference category: 45 to 49)								
20 to 24	-0.027000 ***	0.0043	-0.01800 ***	0.0038	-0.00970 ***	0.0020	-0.0270 ***	0.0038
25 to 29	-0.011000 **	0.0042	-0.00570	0.0037	-0.00610 **	0.0020	-0.0230 ***	0.0034
30 to 34	0.001400	0.0043	0.00500	0.0039	-0.00340	0.0019	-0.0170 ***	0.0032
35 to 39	0.004000	0.0043	0.00330	0.0039	0.00071	0.0020	-0.0099 **	0.0034
40 to 44	0.007600	0.0044	0.00640	0.0040	0.00140	0.0021	-0.0051	0.0034
50 to 54	0.000027	0.0042	-0.00180	0.0038	0.00170	0.0021	0.0027	0.0035
55 to 59	-0.001000	0.0044	-0.00290	0.0039	0.00200	0.0022	0.0065	0.0039
60 to 64	-0.007600	0.0049	-0.00970 *	0.0041	0.00210	0.0028	0.0099 *	0.0047
Family status (reference category: couple with children)								
Single with children	-0.002700	0.0059	-0.00340	0.0051	-0.00024	0.0033	0.0260 ***	0.0037
Single, no children	-0.006200	0.0060	-0.00270	0.0052	-0.00410	0.0034	0.0180 ***	0.0038
Couple, no children	-0.010000	0.0060	-0.00380	0.0053	-0.00690 *	0.0033	-0.0015	0.0037
Region (reference category: Ontario)								
Atlantic	-0.018000 ***	0.0026	-0.01200 ***	0.0023	-0.00560 ***	0.0013	-0.0110 ***	0.0021
Quebec	-0.005100	0.0028	-0.00089	0.0025	-0.00410 **	0.0014	-0.0016	0.0023
Manitoba and Saskatchewan	0.002200	0.0027	0.00400	0.0024	-0.00170	0.0013	0.0062 **	0.0022
Alberta	-0.007500 *	0.0033	-0.00120	0.0030	-0.00620 ***	0.0014	0.0190 ***	0.0030
British Columbia	0.008900 *	0.0035	0.00850 **	0.0031	0.00029	0.0017	0.0041	0.0028
Education (reference category: less than high school diploma)								
High school diploma or equivalent	-0.002000	0.0041	0.00220	0.0036	-0.00350	0.0019	0.0092 **	0.0028
Some postsecondary education	0.011000 **	0.0039	0.01000 **	0.0034	0.00053	0.0019	0.0110 ***	0.0026
University degree	0.030000 ***	0.0043	0.02700 ***	0.0038	0.00340	0.0022	0.0380 ***	0.0031
Log hourly wages	-0.002700	0.0025	-0.00580 **	0.0022	0.00260 *	0.0012	0.0190 ***	0.0021
Union membership	-0.005800 *	0.0024	-0.00420	0.0022	-0.00150	0.0011	-0.0100 ***	0.0019
Firm size (reference category: fewer than 20 employees)								
20 to 99 employees	-0.023000 ***	0.0036	-0.01600 ***	0.0032	-0.00780 ***	0.0018	-0.0260 ***	0.0033
100 to 500 employees	-0.032000 ***	0.0037	-0.02400 ***	0.0032	-0.00810 ***	0.0020	-0.0390 ***	0.0033
More than 500 employees	-0.038000 ***	0.0033	-0.02800 ***	0.0029	-0.01000 ***	0.0018	-0.0420 ***	0.0030
Job status (reference category: permanent)								
Seasonal	0.006400	0.0063	0.00870	0.0060	-0.00220	0.0022	-0.0063	0.0043
Temporary, term or contract	0.026000 ***	0.0043	0.02100 ***	0.0038	0.00490 *	0.0021	0.0120 **	0.0039
Casual	0.019000 **	0.0061	0.01400 **	0.0052	0.00490	0.0035	-0.0037	0.0046
Other	0.045000	0.0400	0.04500	0.0370	-0.00520	0.0089	0.0091	0.0220
Part-time job	0.033000 ***	0.0032	0.02800 ***	0.0028	0.00450 **	0.0017	0.0140 ***	0.0027

* significantly different from reference category ($p < 0.05$)

** significantly different from reference category ($p < 0.01$)

*** significantly different from reference category ($p < 0.001$)

Notes: LWF = Longitudinal Worker File; LFS = Labour Force Survey. LFS respondents aged 20 to 64. All counts are rounded to the nearest 100.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

Table 7
Estimation results for probability of Longitudinal Worker File labour market status as a function of Labour Force Survey variables
(continued)

LFS variables	LWF: All sole proprietors		LWF: Sole proprietors, gig workers		LWF: Sole proprietors, non-gig workers		LWF: Incorporated self-employed	
	partial effects	standard errors	partial effects	standard errors	partial effects	standard errors	partial effects	standard errors
Industry (reference category: agriculture, forestry, fishing and hunting)								
Mining, quarrying, and oil and gas extraction	-0.02000 *	0.0091	-0.00670	0.0082	-0.01300 **	0.0044	-0.0240 **	0.0088
Utilities	-0.021000 *	0.0099	-0.00470	0.0092	-0.01600 ***	0.0043	-0.0320 **	0.0110
Construction	0.002100	0.0078	0.00420	0.0066	-0.00310	0.0042	-0.0150	0.0077
Manufacturing	-0.019000 *	0.0074	-0.00630	0.0064	-0.01200 **	0.0040	-0.0320 ***	0.0075
Wholesale trade	-0.010000	0.0075	-0.00031	0.0064	-0.01000 *	0.0041	-0.0260 ***	0.0077
Retail trade	-0.012000	0.0082	-0.00037	0.0071	-0.01100 *	0.0043	-0.0250 **	0.0081
Transportation and warehousing	0.016000	0.0085	0.01200	0.0071	0.00310	0.0046	-0.0120	0.0084
Information and cultural industries	0.041000 ***	0.0110	0.04000 ***	0.0100	0.00064	0.0061	-0.0310 ***	0.0093
Finance and insurance	-0.001000	0.0085	0.01200	0.0074	-0.01300 **	0.0042	-0.0210 *	0.0085
Real estate and rental and leasing	0.066000 ***	0.0140	0.02400 *	0.0100	0.04100 ***	0.0097	-0.0100	0.0110
Professional, scientific and technical services	0.018000 *	0.0084	0.01700 *	0.0071	0.00053	0.0047	-0.0094	0.0081
Administrative and support, waste management and remediation services	0.012000	0.0090	0.01300	0.0075	-0.00086	0.0052	-0.0380 ***	0.0081
Educational services	0.024000 **	0.0084	0.03300 ***	0.0073	-0.00920 *	0.0042	-0.0370 ***	0.0077
Health care and social assistance	0.011000	0.0077	0.02000 **	0.0066	-0.01000 *	0.0041	-0.0310 ***	0.0075
Arts, entertainment and recreation	0.039000 ***	0.0110	0.03900 ***	0.0098	-0.00110	0.0059	-0.0380 ***	0.0087
Accommodation and food services	-0.013000	0.0080	-0.00073	0.0068	-0.01300 **	0.0042	-0.0240 **	0.0081
Other services (except public administration)	0.020000 *	0.0088	0.02600 ***	0.0077	-0.00630	0.0044	-0.0430 ***	0.0076
Public administration	-0.001300	0.0082	0.00900	0.0070	-0.00980 *	0.0043	-0.0360 ***	0.0078
Number of observations	461,100		461,100		461,100		461,100	

* significantly different from reference category ($p < 0.05$)

** significantly different from reference category ($p < 0.01$)

*** significantly different from reference category ($p < 0.001$)

Notes: LWF = Longitudinal Worker File; LFS = Labour Force Survey. LFS respondents aged 20 to 64. All counts are rounded to the nearest 100.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

Four outcome variables were considered: (i) an indicator for being a sole proprietor, (ii) an indicator for participating in gig work, (iii) an indicator for being a sole proprietor but not a gig worker and (iv) an indicator variable for being an owner of an incorporated enterprise. The latter outcome was added to the analysis with the idea that the comparison of the results for incorporated and unincorporated self-employment may provide additional clues regarding individuals' motivation for supplementing their wage earnings with earnings from unincorporated self-employment and gig work. It should be stressed that the objective of the regression analysis was not to establish a causal relationship between any right-hand-side variables and the outcomes, but only to gauge the degree of association between them, adjusting for a rich set of observed covariates.

The estimated partial effects from the probit model defined by (3) are shown in Table 7. The regression analysis confirms several descriptive results discussed earlier. University degree holders were significantly more likely to be gig workers (0.027) or unincorporated sole proprietors (0.030) than individuals with the lowest level of educational attainment. A somewhat higher prevalence of gig work among university degree holders has been noted in several previous studies (Abraham and Houseman 2019; Collins et al. 2020; Jeon et al. 2021). However, the results in Table 7 reveal another interesting dynamic related to participation in gig work: while higher levels of education are associated with higher probabilities of being a gig worker, the opposite is true for hourly wages. Controlling for education and several essential individual and main job characteristics, a 1% increase in hourly wages is associated with about a 0.6 percentage point decline in the probability of being a gig worker. By contrast, the probability of being an unincorporated self-employed business owner who is not a gig worker is positively related to

hourly wages, although the relationship appears weak and statistically significant only at the 95% level.¹⁷

Other notable results concern two important aspects of work status. LFS respondents working part time at their main jobs were considerably more likely to be unincorporated sole proprietors (0.033) and gig workers (0.028) (Table 7). Similarly, a higher probability of being a gig worker was strongly associated with the temporary work status, especially with being a temporary, term or contract employee (0.021). By contrast, the correlation between the temporary work status variables and the probability of being a sole proprietor other than a gig worker was much weaker, and only one of these variables (temporary, term or contract job) was weakly significant at the 95% level (0.005).¹⁸

As mentioned above, an additional model was estimated for the probability of being an incorporated self-employed worker. The results in the last two columns of Table 7 show that the wage and education gradients are clearer and stronger for the ownership of incorporated firms, and the probability of owning an incorporated firm increases with both hourly wages and education level. Based on the results in Table 7, it seems likely that the characteristics of individuals who do gig work, and their motivations for becoming gig workers, are very different than the characteristics and motivations of those who become owners of incorporated firms, in line with the argument of Levine and Rubinstein (2017).¹⁹

4 Conclusions

This study had two main objectives. First, it looked at the correspondence between labour market activities reported in the LFS and the income sources of the same individuals in the tax data. It was estimated that the vast majority of LFS respondents whose main labour market activity was wage employment also had earnings from wage employment in tax data (97.6%). Also, more than 80% of LFS respondents whose main labour market activity was unincorporated self-employment reported self-employment income in the tax data.

The second objective was to learn more about the role of self-employment as a supplemental source of income. The study examined the likelihood of having self-employment income or being a gig worker for individuals who reported wage employment as their main labour market activity. Wage-employed university graduates appeared more likely to be gig workers or sole proprietors than less educated wage-employed individuals. Temporary work status and part-time employment were strongly associated with being a gig worker. The descriptive analysis showed that the prevalence of gig work activities among temporarily employed part-time wage employees was considerably higher (9.6%) than the prevalence of gig work activities among permanently employed full-time wage employees (4.3%). The prevalence of gig work was also higher among wage-employed individuals whose spouses were self-employed, compared with those whose spouses were also wage-employed or not working. More than two-thirds of LFS respondents whose secondary activity was unincorporated self-employment were sole proprietors, and 48.3% were identified as gig workers.

These results should not be interpreted as evidence of a causal impact of education, hourly wages or part-time work status on participation in gig work. Also, because of the difference between the frequencies of the LFS and LWF data, wage jobs and gig work were not necessarily done concurrently—some self-employment and gig work probably followed (or preceded) wage

17. The self-employment and gig work rates for wage employees in Table 4 (6.5% and 5.1%, respectively) are the estimates for the overall probabilities of being self-employed or participating in gig work.

18. See also Appendix C for regression results based on alternative aggregation strategies

19. The effects of the main variables of interest were similar when regression analyses were performed separately for men and women. These results are available upon request. As an additional sensitivity check, the model in (3) was re-estimated without the hourly wage variable. The level of education, firm size, part-time employment and temporary work remained strongly associated with being a sole proprietor and a gig worker.

employment. Nevertheless, documenting the strong correlational links described above is an important step toward gaining a better understanding of the role played by self-employment and gig work as supplemental income sources.

An important question for future research is whether the findings reported in this study were affected by labour market changes after March 2020. Only pre-2017 administrative data on gig work were available when the study began. As new administrative data become available, the issue of how individuals report their income in administrative and survey data is likely to be revisited.

5 Appendix

Appendix A: The impact of age on the linkage rate between the Labour Force Survey and the Longitudinal Worker File

Although the linkage rate between the Labour Force Survey (LFS) and the Longitudinal Worker File (LWF) discussed in Section 2 is high (84.2%), an important question is whether there are any systematic links between the characteristics of LFS respondents and the probability of being linked to the LWF. A look at the age distribution among individuals not linked to the LWF suggests that the non-linkage rate among LFS respondents younger than 20 years is very high, but it falls quickly with every additional year of age: 81.1% among respondents aged 15, 63.2% for those aged 16, 37.6% for those aged 17 and 21.4% for those aged 18 to 19.²⁰ The non-linkage rate continues to decline after age 20 (down to 11.3% for individuals aged 50 and older), but at a very slow pace. For LFS respondents aged 25 to 54 (“prime working age”), the overall linkage rate was 84.8%, very similar to the overall linkage rate for all age groups (84.2%).

Regression analysis was used for a more comprehensive assessment of the linkage patterns. An indicator of whether an LFS respondent could be linked to the LWF data was regressed on a set of the following demographic, human capital and geographic variables that may potentially influence the linkage rate: sex; age, divided into five-year age categories from 15 to 65 years (e.g., 15 to 19, 20 to 24, 25 to 29) and a separate category for individuals aged 65 and older; education level, divided into four categories (less than a high school diploma, a high school diploma or equivalent, some postsecondary education and a university degree); family status, divided into four categories (single [neither married nor cohabiting] without children aged 18 years and younger, single with children aged 18 and younger, married or cohabiting without children aged 18 and younger, and married or cohabiting with children aged 18 and younger); and geographic region, divided into six categories (the Atlantic provinces, Quebec, Ontario, Manitoba and Saskatchewan, Alberta and British Columbia). The results from the linear probability model are shown in Appendix Table A1. The largest, by far, and most dramatic drop in the probability of LFS records being linked to the LWF used in the study is associated with being aged 15 to 19 years (-0.184; the reference category is LFS respondents aged 45 to 49). One way to assess the magnitude of this drop is to compare it with the benchmark probability of being linked to the LWF given by the constant term (0.846):²¹ an 18.4 percentage point drop represents a 21.7% decline, relative to the benchmark probability. LFS respondents aged 65 and older have the highest probability of being linked (0.106). Other age coefficients—although statistically significant—do not appear to suggest large deviations from the reference category. Not having children is associated with a somewhat lower probability of being linked to the LWF for individuals with any family status. A possible reason for this is that individuals with young children have greater incentives to file tax returns than other individuals because of child benefits. Respondents from

20. A table with the full set of these results is available upon request from the authors.

21. This is almost the same probability as the overall linkage rate, 84.2%.

Quebec are more likely to be linked to the LWF than those in other parts of Canada; this may be related to higher tax-filing rates in the province.

Appendix Table A1

Regression results for linkage between Labour Force Survey and Longitudinal Worker File

	Outcome: linked to the LWF	
	coefficients	standard errors
Women	0.017	0.001 ***
Age group (reference group: 45 to 50)		
15 to 19	-0.184	0.004 ***
20 to 24	0.025	0.003 ***
25 to 29	0.009	0.003 ***
30 to 34	0.000	0.003
35 to 39	-0.006	0.003 *
40 to 44	-0.015	0.003 ***
50 to 54	0.020	0.003 ***
55 to 59	0.050	0.003 ***
60 to 64	0.082	0.003 ***
65 or older	0.106	0.003 ***
Education group (reference group: university degree)		
Less than high school diploma	-0.08	0.002 ***
High school diploma or equivalent	-0.03	0.002 ***
Some postsecondary education	0.00	0.002
Family status (reference group: couple with children aged 18 or younger)		
Single with children aged 18 or younger	0.010	0.004 **
Single without children aged 18 or younger	-0.045	0.004 ***
Couple without children aged 18 or younger	-0.104	0.004 ***
Province group (reference: Ontario)		
Atlantic provinces	0.074	0.002 ***
Quebec	0.122	0.002 ***
Prairies	0.050	0.002 ***
Alberta	0.021	0.002 ***
British Columbia	-0.017	0.002 ***
Constant	0.846	0.005 ***
Total number of LFS individuals	292,100	...
		percent
Number of individuals linked to LWF	246,100	84.24

... not applicable

* significantly different from reference category ($p < 0.05$)

** significantly different from reference category ($p < 0.01$)

*** significantly different from reference category ($p < 0.001$)

Notes: LFS = Labour Force Survey; LWF = Longitudinal Worker File. LFS records of individuals deemed non-eligible for linkage by the Social Data Linkage Environment and those of individuals younger than 15 were excluded. All counts shown are rounded to the nearest 100.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

Based on the results of this assessment, the main analysis is restricted to LFS respondents aged 20 to 64.

Appendix B: Methodology

The Labour Force Survey (LFS) data contain individual sampling weights determined by several weighting factors, the most important being the inverse of the probability of selection. The weights are cross-sectional, and weighted sample counts reflect the population counts at the chosen level of geography. The panel structure of the LFS implies that each individual selected in the LFS sample can have up to six monthly records in a given year. For instance, those selected in May 2016 could be observed for up to six months, from May to October.²² Those selected in November can be observed for only November and December of 2016. In any given month m ,

$$P_m = \sum_s \sum_i w_{ism}, \quad (1)$$

where P_m is the total population count and w_{is} is the weight corresponding to individual i in stratum s .²³ In other words, the total population count in month m can be obtained by summing all weights within each stratum and then across all strata.

To fix the idea, suppose the population of a country, P_m , is 240,000 individuals, which remains stable during the whole year. The country is divided into 100 strata ($s = 100$) of equal size (2,400 individuals), and 40 individuals are selected from each stratum for the panel that starts in January and ends in June; 40 different individuals are selected for the panel that starts in July and ends in December.²⁴ For simplicity, it is assumed that there is no non-response, so the total sample contains 48,000 records for 8,000 individuals. Since 40 individuals in each stratum represent 2,400 individuals living in that stratum, each selected individual is assigned a weight of 60 ($w_{ism} = 60$). Using formula (1), $P_{January} = P_{February} = \dots = P_{December} = 60 \times 40 \times 100 = 240,000$. As mentioned in Section 2, the approach taken in this study is to keep all individual monthly records, but divide each individual weight by 12 (the number of months in a year) to maintain the representativeness of the LFS sample at the annual level:

$$P_a = \sum_s \sum_i \sum_m \tilde{w}_{ism}, \quad (2)$$

where $\tilde{w}_{ism} = w_{ism} / 12$. In the example above, each weight is divided by 12, so that each individual's weight is now $\tilde{w}_{ism} = 60 / 12 = 5$. The total population count can be obtained using (2) as $P_a = 5 \times 6 \times [2 \times 40] \times 100 = 240,000$. The result reflects the population size in year a , and it is the same as the cross-sectional monthly population in the previous formula. The number in the square brackets is the total number of unique individuals in year a (two panels multiplied by 40 individuals), and the second number (6) is the number of months the individual is observed in the panel. In the example, all individuals are assumed to be observed for six months. However, m will be different for different individuals in equation (2), depending on how many observations are available for that individual in year a .²⁵ Intuitively, the contribution of an individual observed in the LFS for six months is equal to the contribution of six individuals observed for one month. Also, the combined contribution of two individuals observed for six months represents the contribution

22. Not all individuals would be present in the survey for all six months because of non-response.

23. The LFS sampling framework is based on a stratified multi-stage design. Each province is divided into large geographic areas (strata), further subdivided into smaller geographic areas (clusters), from which dwellings are selected.

24. To simplify the example, the subdivision of strata into clusters is ignored.

25. In 2016, 31.0% of all unique individuals in the analysis sample were observed for six months, and 41.4% were observed for three months or less. The percentages were close in the linked and pre-linked LFS data, so linking the LFS to Longitudinal Worker File records did not alter the "months present" distribution in a significant way. Close to 80% of those observed in the LFS for six months reported the same main labour market activity in all six months.

of a single hypothetical individual in the annualized cross-sectional LFS counts. If the population remained unchanged for all 12 months of year a , then the monthly and annual cross-sectional counts would be the same, and $P_a = P_m$ for any m in a .

It is important to note that the main objective of the empirical analysis below is to understand how the labour market activities of 2016 LFS respondents translate into their work status and income sources based on tax data. The main annualization method used in this study allows for a fairly straightforward interpretation of the findings. For instance, the study may find that $X\%$ of all LFS respondents whose main labour market activity is wage employment also have T4 earnings in the Longitudinal Worker File (LWF). Because the LFS data are monthly and the LWF data are annual, the T4 information applies to all LFS records. By contrast, if the denominator is the number of people in the LWF who received a T4 in 2016 and the numerator is the number of LFS respondents whose main labour market activity in **a particular month** was wage employment, the results are more difficult to interpret. For this reason, the analysis is benchmarked to the LFS population.

Appendix C: A comparison between results based on alternative annualization methods

To better understand how the choice of an annualization strategy impacts the results, an alternative aggregation-based annualization strategy centred on aggregating all individuals' monthly observations into a single annual observation was also implemented. Because Labour Force Survey (LFS) respondents in the main sample could be observed for any number of months between one and six, the sample was restricted only to individuals with six interviews in 2016 to make the aggregation consistent for all individuals in the sample. For comparison, one set of results was obtained using the same person-month approach as in the main analysis, and another set of results was obtained using the person-level aggregate data. To create a person-level sample, a respondent's main labour market activities were collapsed into a single person-level annual record by choosing the most frequent main labour market activity (i.e., the mode activity) and averaging the LFS monthly weights of that individual into a single annual value.²⁶

The upper panel of Appendix Table A2 shows the results for the restricted sample using the same methodological approach as in the main analysis (Table 2). The percentage of the unincorporated self-employed in the LFS who have self-employment income in the Longitudinal Worker File (LWF) is slightly higher in the restricted sample (82.4%) than in the main sample (80.7%), but, overall, the percentages in the upper panel of Appendix Table A2 are quite similar to those in Table 2. For comparison, the results in the lower panel of Appendix Table A2 show that 88.8% of those whose aggregated main labour market LFS activity was unincorporated self-employment also had self-employment earnings in the LWF. Hence, the person-level aggregate approach produces an even stronger correspondence between unincorporated self-employment in the LFS and tax data than the preferred person-month approach.²⁷

26. Individuals with multiple modes of the main labour market activity were excluded, causing the loss of about 7% of the restricted sample.

27. Intuitively, those who are self-employed in the LFS in just one month will receive much less weight in the calculations than those whose mode main labour market activity over a six-month period is self-employment. This is probably one of the explanations for the results. Also, those whose mode main labour market activity is self-employment may be more likely to report self-employment income in their tax returns than those who are self-employed in the LFS in just one month.

Appendix Table A2

Mapping Labour Force Survey employment status and main labour market activities to Longitudinal Worker File employment status and income sources; sample restricted to those with six interviews

LFS primary activity	LWF status				Total (weighted counts)
	Wage employee (T4 income)	Incorporated self-employed	All unincorporated self-employed	Unincorporated self-employed with T2125 income	
	row percent				count
Person month sample					
Currently employed					
Wage employee	97.7	4.5	7.8	6.9	5,607,800
Incorporated self-employed	67.8	75.3	24.5	20.9	447,900
Unincorporated self-employed	27.6	9.8	82.4	77.0	518,800
Not currently employed					
Unemployed	70.7	4.6	9.0	7.9	393,000
Not in labour force	26.5	4.3	6.7	5.6	1,536,100
Aggregate sample, most common activity					
Currently employed					
Wage employee	98.1	4.3	7.5	6.7	944,300
Incorporated self-employed	68.7	82.7	21.7	18.1	69,900
Unincorporated self-employed	22.0	7.8	88.8	83.0	82,300
Not currently employed					
Unemployed	65.8	4.8	9.1	7.7	44,400
Not in labour force	22.4	4.3	6.4	5.3	242,200

Notes: LWF = Longitudinal Worker File; LFS = Labour Force Survey. LFS respondents aged 20 to 64 with six interviews available in 2016. T2125 income is self-employment income from business and professional activities (including commissions). Row percentages may exceed 100% because LFS respondents can have income from multiple sources in the LWF. All counts are rounded to the nearest 100.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

To test the robustness of the results in Table 7, two additional models based on the same specification were estimated on the restricted sample discussed in Section 4. The first model (first two columns in Appendix Table A3) was based on the person-month sample restricted to those with six interviews in the 2016 LFS. The second model (last two columns) was estimated on the aggregate person-level sample also restricted to those with six interviews in the 2016 LFS.²⁸ The standard errors in both sets of results are substantially higher than in Table 7 because of the smaller sample size.

28. The explanatory variables were represented by the modes of monthly values (e.g., mode hourly wage for all six interviews). In the case of multiple modes, one of the modes was chosen randomly.

Appendix Table A3

Partial effects for probit models estimated on restricted sample (six interviews): person month and aggregate records

LFS variables	Person month-level records				Person-level records			
	LWF: All sole proprietors		LWF: Sole proprietors, gig workers		LWF: All sole proprietors		LWF: Sole proprietors, gig workers	
	partial effects	standard errors	partial effects	standard errors	partial effects	standard errors	partial effects	standard errors
Female	-0.00870 *	0.0037	-0.00059	0.0033	-0.0096 *	0.0039	-0.00230	0.0035
Age (omitted: 45 to 49)								
20 to 24	-0.02800 ***	0.0069	-0.01500 *	0.0060	-0.0280 ***	0.0077	-0.01500 *	0.0067
25 to 29	-0.01500 *	0.0067	-0.00740	0.0057	-0.0150 *	0.0074	-0.00800	0.0062
30 to 34	-0.00048	0.0068	0.00810	0.0060	-0.0013	0.0073	0.00880	0.0065
35 to 39	0.00360	0.0069	0.00880	0.0061	0.0021	0.0074	0.00720	0.0066
40 to 44	0.01100	0.0072	0.01100	0.0062	0.0069	0.0077	0.00900	0.0067
50 to 54	-0.00030	0.0066	0.00220	0.0057	-0.0023	0.0072	0.00096	0.0062
55 to 59	-0.00290	0.0068	0.00210	0.0060	-0.0024	0.0075	0.00250	0.0066
60 to 64	-0.01400	0.0071	-0.01100	0.0058	-0.0200 **	0.0076	-0.01600 **	0.0061
Family status (omitted: couple with children)								
Single with children	-0.00990	0.0098	-0.00840	0.0085	-0.0110	0.0100	-0.00990	0.0093
Single, no children	-0.01100	0.0100	-0.00700	0.0088	-0.0120	0.0110	-0.00950	0.0096
Couple, no children	-0.01500	0.0100	-0.00910	0.0089	-0.0160	0.0110	-0.01200	0.0098
Region (omitted: Ontario)								
Atlantic	-0.01800 ***	0.0042	-0.01500 ***	0.0037	-0.0180 ***	0.0044	-0.01600 ***	0.0038
Quebec	-0.00870	0.0045	-0.00580	0.0039	-0.0086	0.0047	-0.00680	0.0042
Manitoba and Saskatchewan	0.00240	0.0044	0.00290	0.0039	0.0030	0.0046	0.00300	0.0041
Alberta	-0.00820	0.0053	-0.00360	0.0048	-0.0073	0.0055	-0.00320	0.0050
British Columbia	0.00680	0.0056	0.00530	0.0050	0.0056	0.0059	0.00350	0.0052
Education (omitted: less than high school diploma)								
High school diploma or equivalent	0.00450	0.0062	0.00460	0.0057	0.0033	0.0069	0.00210	0.0064
Some postsecondary education	0.01500 *	0.0059	0.01300 *	0.0054	0.0130 *	0.0066	0.00980	0.0062
University	0.03600 ***	0.0067	0.02900 ***	0.0060	0.0360 ***	0.0075	0.02700 ***	0.0069
Log hourly wages	-0.00067	0.0038	-0.00450	0.0034	0.0018	0.0052	-0.00270	0.0046
Union membership	-0.00440	0.0038	-0.00460	0.0034	-0.0070	0.0041	-0.00620	0.0036
Firm size (omitted: fewer than 20 employees)								
20 to 99 employees	-0.01600 **	0.0056	-0.00830	0.0049	-0.0220 **	0.0068	-0.01100	0.0060
100 to 500 employees	-0.02600 ***	0.0057	-0.01900 ***	0.0048	-0.0340 ***	0.0067	-0.02400 ***	0.0057
More than 500 employees	-0.03200 ***	0.0050	-0.02000 ***	0.0043	-0.0360 ***	0.0060	-0.02200 ***	0.0052
Job status (omitted: permanent)								
Seasonal	0.01500	0.0100	0.02000	0.0100	0.0180	0.0130	0.02500 *	0.0130
Temporary, term or contract	0.02000 **	0.0067	0.01500 *	0.0058	0.0140	0.0076	0.01300	0.0067
Casual	0.02300 *	0.0095	0.01800 *	0.0084	0.0220	0.0130	0.01300	0.0110
Other	0.03700	0.0630	0.04200	0.0570	0.0270	0.0600	0.03200	0.0550
Part-time job	0.04200 ***	0.0052	0.03400 ***	0.0045	0.0390 ***	0.0063	0.03300 ***	0.0056
Industry	yes		yes		yes		yes	
Number of observations	224,500		224,500		37,700		37,700	

* significantly different from reference category (p < 0.05)

** significantly different from reference category (p < 0.01)

*** significantly different from reference category (p < 0.001)

Notes: LWF = Longitudinal Worker File; LFS = Labour Force Survey. All counts are rounded to the nearest 100. "Yes" = variables were included as controls in the model.

Source: Statistics Canada, Labour Force Survey and Longitudinal Worker File linkage, authors' calculations.

The key result is that most estimated partial effects are similar for the person-month and person-level models, and also similar to the estimated partial effects in Table 7. The estimated partial effects for the probability of being a gig worker are 0.034 in the person-month and 0.033 in the person-level models (0.028 in Table 7). The estimated partial effects for university education are also very similar (0.029 and 0.027, respectively, compared with 0.027 in Table 7). The person-level model produces somewhat weaker results for temporary or seasonal work (0.012) than the person-month model (0.021). A possible reason for this is that temporary jobs last for only a short

period of time, and a substantial number of them are aggregated out in the person-level approach. For example, if someone reports a temporary or seasonal job only in June and July and has a permanent job in other months, the aggregated job status for that person is “permanent.” As mentioned in Section 3, loss of information is one of the pitfalls of the person-level aggregation approach, and its effect is not straightforward to assess. This is why the person-month approach was deemed preferable in the main analysis.

References

- Abraham, K., Haltiwanger, J., Hou, C., Sandusky, K., & Spletzer, J. (2021). Reconciling survey and administrative measures of self-employment. *Journal of Labor Economics*, 39(4).
- Abraham, K., Haltiwanger, J., Sandusky, K., & Spletzer, J. (2018). *Measuring the gig economy: Current knowledge and open issues*. Cambridge, MA: National Bureau of Economic Research, Working Paper 24950.
- Abraham, K., & Houseman, S. (2019). Making ends meet: the role of informal work in supplementing Americans' income. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(5), 110–131.
- Alake-Apata, B. (2021). *Making sense of gig work*. LMI Insight Report no. 45, Labour Market Information Council, Ottawa.
- Boeri, T., Giupponi, G., Krueger, A., & Machin, S. (2020). Solo self-employment and alternative work arrangements: A cross-country perspective on the changing composition of jobs. *Journal of Economic Perspectives*, 34(1), 170–195.
- Collins, B., Garin, A., Jackson, E., Koustas, D., & Payne, M. (2020). *Has the gig economy replaced traditional jobs over the last two decades? Evidence from tax returns*. Unpublished manuscript.
- Fellegi, I. P., & Sunter, A. B. (1969). A theory of record linkage. *Journal of the American Statistical Association*, 64(328), 1183–1210.
- Jeon, S.-H., Liu, H., & Ostrovsky, Y. (2021). Measuring the gig economy in Canada using administrative data. *Canadian Journal of Economics*, 54(4).
- Jeon, S.-H., & Ostrovsky, Y. (2020). *Self-employed individuals, self-employment income and the post-COVID-19 financial strain*. Ottawa: Statistics Canada.
- Kostyshyna, O., & Luu, C. (2019). *The size and characteristics of informal (“gig”) work in Canada*. Ottawa: Bank of Canada.
- Koustas, D. (2020). *Insights from new tax-based measures of gig work in the United States*. Munich: Ifo Institute - Leibniz Institute for Economic Research at the University of Munich.
- Levine, R., & Rubinstein, Y. (2017). Smart and illicit: who becomes an entrepreneur and do they earn more? *The Quarterly Journal of Economics*, 132(2), 963–1018.
- Lim, K., Miller, A., Risch, M., & Wilking, E. (2019). *Independent contractors in the U.S.: New trends from 15 years of administrative tax data*. Internal Revenue Service. Internal Revenue Service, no. 45. Retrieved from: <https://www.irs.gov/pub/irs-soi/19rpindcontractorinus.pdf>.
- Lloyd, N. (2020). *Strategic self-employment and family formation*. UBC: Unpublished report.
- Messacar, D. (2014). *Report on the Comparison of the T1 Personal Master File*. Ottawa: Statistics Canada, Social Analysis and Modelling Division.
- Statistics Canada (2022). *Labour Force Survey (LFS)*. Retrieved from: <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3701#a1>.
- Statistics Canada (2021). *Longitudinal Worker File (LWF)*. Retrieved from: <https://www.statcan.gc.ca/en/microdata/data-centres/data/lwf>.