Opening the Black Box of Self-Employment: Identifying Alternative Work Arrangements in the United States[1]

Joelle Abramowitz
University of Michigan
3408 ISR-Thompson
426 Thompson Street
Ann Arbor, MI 48109
Phone: (734) 936-0324
Fax: (734) 763-9831
Email: jabramow@umich.edu

Andrew Joung
University of Michigan
238 Lorch Hall
611 Tappan Avenue
Ann Arbor, MI 48109
Phone: (734) 764-2355
Fax: (734) 764-2769
Email: ajoung@umich.edu

Abstract: While 18.4% of workers report engaging in self-employment, there exists a dearth of data identifying heterogeneity in the nature of these work arrangements. To address this gap, this paper uses novel data using machine learning leveraging internal data collected in the 2003-2019 waves of the Panel Study of Income Dynamics on respondents' narrative descriptions of their industry and type of work along with their employer names. The paper uses these data to examine trends in the prevalence and nature of self-employment work arrangements, transitions across these arrangements, and who works in these arrangements. Findings show disparate trends in the share of workers engaging in different types of self-employment work arrangements that would otherwise be masked. Further results suggest that the informally self-employed tend to be less educated, are less likely to be male and non-Hispanic White, have less labor income, and have worse measures of wellbeing.

Keywords: Self-employment, non-traditional work arrangements, independent contractors, work, wellbeing, employment trends

[1] The research reported herein was performed pursuant to a grant from the National Science Foundation Award FW-HTF-P 2128416. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of NSF or any agency of the federal government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States government or any agency thereof. The authors would like to thank Nathan Babcock, Yiran Fang, Akshitha Ginuga, Kyle Kolick, Mitchell Lukas, Sajiv Shah, Lauren Thill, Chuqiao Wang, and Xinyi Wang for excellent research assistance. They would also like to thank conference and seminar participants at the University of Michigan, the University of Minnesota Life Course Center's 2022 Data-Intensive Research Conference, the 2022 Association for Public Policy Analysis and Management Fall Research Conference, the 2023 Outsourcing Research Network Workshop, the 2024 Allied Social Science Associations Annual Meeting, and the Federal Reserve's Uneven Outcomes in the Labor Market Conference for their helpful feedback. For information regarding the data and code, please contact Andrew Joung at ajoung@umich.edu.

Self-employment consists of a wide array of different work arrangements, but is often reported in administrative and survey data as a singular, homogenous category. For example, an individual pursuing self-employment in the transportation sector could choose to innovate a new platform or technology, drive for an app-based ride-sharing service, advertise their own chauffeur services, drive on a contract basis for an established business, or manage their own or someone else's established business. The characteristics of these jobs - the barriers to entry, risks, work stresses, and compensation - are likely to vary considerably. However, in most existing data sources, we would be unable to meaningfully differentiate these jobs.

A substantial share of self-employment consists of alternative work arrangements – work that falls outside of traditional employer-employee relationships. However, as discussed in the National Academies of Sciences, Engineering, and Medicine (2020) report on measuring alternative work arrangements, existing data sources on alternative work arrangements have limitations in their ability to be used to understand the changing nature of work and develop appropriate policy. This is because administrative data do not capture work associated with income not reported to tax authorities, and survey data do not capture activity that survey respondents do not consider to be work. Due to these data limitations, it has been difficult to estimate the size of the workforce engaged in alternative work arrangements, examine how such arrangements are changing over time, and to understand how workers use these jobs to enhance their wellbeing.

Understanding such heterogeneity in self-employment has become all the more important as new technologies such as electronic platforms have introduced new means of engaging in self-employment with potentially more far-reaching effects on the economy. The introduction and growth of the platform gig economy, one type of alternative work arrangement, have raised the question about how big this sector is, as well as the extent to which such jobs crowd out employment in other sectors, and the extent to which such jobs are good for workers' wellbeing. Answering these questions requires identifying such work separately from other types of self-employment.

This paper fills existing gaps by using novel data constructed using machine learning and internal respondent narratives on industry and type of work and employer names collected in the 2003-2019 the Panel Study of Income Dynamics (PSID). These data separately identify wage and salaried employment, business ownership, platform gig work, informal self-employment, and formal self-employment. The paper uses these data to examine: (1) how the prevalence and nature of different self-employment work arrangements has changed over time, (2) how individuals transition across different types of self-employment work arrangements, and (3) the characteristics of individuals working in different types of self-employment work arrangements.

Our findings show that, like Current Population Survey's Annual Social and Economic Supplement (CPS-ASEC), the PSID exhibits a downward trend in the share of self-employed workers among the employed. However, we find, in levels, a far larger share of self-employment in the PSID relative to the CPS-ASEC. We further find that, from 2003 to 2019, formal self-employment shares have fallen, whereas informal self-employment shares have risen. Business ownership experienced large fluctuations, rising following the Great Recession and subsequently returning to pre-Great Recession levels. Additional results show that relative to other self-employed workers, business owners are far more likely to remain business owners across survey waves; on the other hand, informal self-employed workers are nearly twice as likely to enter non-employment relative to other work arrangements. These patterns are stable over time with a rise in the probability of staying in one's work arrangements across survey waves. Finally, we identify differences in the demographics, labor market outcomes, and self-reported wellbeing of individuals working in different types of self-employment work arrangements. Taken together, these findings suggest salient differences in trends, transitions, and work and individual characteristics of the self-employed that would otherwise be masked.

Measuring Self-Employment

Self-employment is notoriously hard to measure. While both administrative data and surveys are able to provide some insight about self-employment, they also each face challenges. As a result of these challenges, discrepancies appear across administrative and survey data sources in identifying trends in self-employment broadly and in specific arrangements such as contingent work and gig employment (Abraham et al., 2018, 2021a; Allard and Polivka, 2018; Jackson et al., 2017; Katz and Krueger, 2019).

Measuring Self-Employment in Administrative Data

Administrative data are derived from tax reporting and can be used to identify wage and salaried employment separately from self-employment based on the types of income reported in tax filings. However, unlike administrative records of wage and salary income which come from third-party reporting by employers on Form W-2s, administrative records of self-employment income rely on both third-party reporting by firms on 1099-K and 1099-NEC forms[1] as well as taxpayer reporting of other self-employment income.

While 1099 forms do provide valuable information on some self-employment income, these forms suffer from incomplete coverage. Self-employed workers are only

[1] The Form 1099-NEC replaced the Form 1099-MISC in the 2020 tax year.

issued a 1099-NEC form by firms from which they received compensation of $600 or more. While the income threshold is relatively low, self-employed workers primarily receiving payment from non-firms or through payment processors, such as sellers on eBay or drivers on ride-share apps like Uber or Lyft,[2] are not captured. Additional coverage is provided through the 1099-K form, which requires payment processors to report a payment recipient's transactions if they exceed $20,000 and the number of transactions exceeds 200.

In the absence of third-party reporting, taxpayers report self-employment income through their tax filings on Schedules C and SE. Absent noncompliance, taxpayer-reported self-employment income should provide comprehensive and accurate estimates of self-employment. However, taxpayers face incentives to strategically report taxable income, and lacking third-party information reporting, are more likely to strategically report (Slemrod et al., 2017), leading to concerns regarding the accuracy of administrative data (Mortenson and Whitten 2020; Saez 2002) as the IRS estimates only a 44% tax reporting compliance rate among self-employed workers (Slemrod, 2016), consistent with the results of Abramowitz (2023) finding self-employment income underreported in administrative data as compared to the Health and Retirement Study. Furthermore, administrative data can be sensitive to reporting incentives as Garin et al. (2022) found that recent increases in taxpayer-reported self-employment in tax data are largely explained by changes in reporting behavior in response to reporting incentives rather than actual changes in self-employment activity.

[2] These workers may receive a 1099-NEC if, for example, they participated in a bonus or referral program. However, the bulk of their earnings would not be covered by the 1099-NEC.

Measuring Self-Employment in Surveys

Since individuals do not face the same economic incentives to misreport on surveys as on their tax filings, surveys provide a valuable complement to administrative data, especially in the context of measuring self-employment. Surveys can also capture greater breadth and detail of a variety of measures often absent in administrative data such as physical and mental health, education, and time use.

However, surveys were designed to capture traditional wage and salaried employment, and accordingly, an ample literature documents myriad issues as survey questions may not capture the way individuals perceive their work, especially in less traditional work arrangements (Allard and Polivka, 2018; Bureau of Labor Statistics, 2018; Abraham and Amaya, 2019; Bracha and Burke, 2021; Abraham et al., 2023), and can produce divergent estimates based on sampling modes, methods, and timing (Katz and Krueger, 2019).

To overcome these issues, some recent work has conducted independent surveys to ask respondents questions specific to the topic of interest. For example, Abraham at al. (2021b) and Abraham et al. (2023) conducted a Gallup telephone survey module to identify independent contracting. Conducting such surveys in combination with focus groups allows refining of questions to address common misalignments between the question intent and the respondent's response. However, response rates may be low and new surveys cannot provide historical data to understand changes in outcomes of interest over time.

An alternative approach is to use collected measures in existing surveys as proxies for identifying heterogeneity in self-employment. For example, Levine and Rubinstein (2017) examined differences in the characteristics of individuals engaged in incorporated self-employment and unincorporated self-employment, aiming to distinguish between entrepreneurs and other business owners. Similarly, Boeri et al. (2020) considered differences between the self-employed with employees and the solo self-employed to distinguish between more and less formal self-employment. Likewise, Moulton and Scott (2016) used broad occupation codes, number of employees, and the presence of household business assets to identify more and less desirable categories of self-employment.

A drawback to these approaches is the extent to which such proxies reflect their intended measures. For example, Light and Munk (2018) use data from the 1979 NLSY to show that the majority of reported self-employment does not reflect business ownership: they find that 68 percent of self-employment is not identified as business ownership and 30 percent of incorporated self-employment is associated with neither business ownership nor reported business income.

Novel Data on Self-Employment

While the aforementioned survey-based approaches provide valuable insights into alternative work arrangements and entrepreneurship, they underscore the potential benefit of having better data on these arrangements in large-scale and long-running surveys. The present study adds to the literature by exploring heterogeneity in self-employment work arrangements based on respondents' descriptions of their work responsibilities and employer names. By leveraging narrative survey information to capture the breadth of self-employment work arrangements, we can identify their prevalence and trends and understand their links with individual wellbeing. This work contributes to a more thorough understanding of the determinants and outcomes associated with different work arrangements. By using PSID data, this work benefits from a high survey response rate as well as the breadth of information collected in the PSID both contemporaneously and longitudinally.

Data and Methods

Data

This analysis uses the 2003-2019 PSID. The PSID is a longitudinal dataset that began in 1968 with a sample of approximately 5,000 U.S. households; it was updated annually through 1997 and bi-annually thereafter. As of 2017, it had grown to include over 10,000 families and 24,000 individuals. The PSID asks questions on a breadth of topics including employment, income, and physical and mental health. While the PSID collects some information on all household members, most measures are collected only for the reference person ("Head") and their spouse/long-term cohabitor. Relevant to our analyses, the PSID asks respondents to describe all of the work for money that the reference person and spouse have done since January 1 of the prior wave year. Respondents are subsequently asked whether the reference person and spouse are self-employed or employed by someone else on up to four jobs that they reported holding since the prior survey wave.[3]

In addition to publicly-available PSID data, the analysis leverages internal data collected in the 2003-2019 PSID on narrative descriptions of industry and occupation and employer names to classify work arrangements into a useful framework. The narratives include answers to the following open-ended questions: "What kind of business or industry is that [job] in?" and "In your work for [your employer] what is your occupation?" and tend to be 3-4 sentences long. The PSID collects this information for all of the respondent's and the spouse's jobs held since January 1 of the prior survey

[3] Respondents are generally the reference person or the spouse. In a small number of cases, when the reference person or the spouse is unavailable, another family unit member will complete the interview.

year through the time of the survey.[4] Among employed respondents and spouses, 99.9% provided current job narratives to the open-ended industry and type of work questions. Some respondents who self-report as not employed also reported current job narratives.[5]

While the narratives cover the two years prior to the time of the interview, we limit our primary analysis to narratives for main jobs held at the time of the interview, with some supplemental analysis of secondary jobs held at the time of interview. We focus on jobs held at the time of interview to frame our analysis at a given point in time. To distinguish currently-held main from secondary jobs, we rely on internal PSID coding of the job narratives as "current main jobs" as well as publicly-available information on the timing of job spells. By construction, individuals can hold multiple secondary jobs and these job types can overlap with the main job type. For 3.9% of job narratives, we cannot distinguish whether the job is currently or previously held, and we exclude these from our main analysis.[6]

In Figure 1, we report our sample criteria and the effect on overall sample size. We restrict our base sample to respondent-waves linked to a current job narrative or self-reporting as employed between 2003-2019, among respondents age 16 or older who are classified at least once as a reference person or spouse. This leaves us with 85,968 respondent-waves linked to 111,498 narratives. We then drop any respondents self-reporting as employed but lacking a current job narrative, reducing our sample to

---

[4] For example, in survey year 2019, respondents are asked about their jobs in the 2017 and 2018 calendar years as well in the survey year through the time of the interview.

[5] The PSID asks multiple times for employment status. If the respondent ever mentions being employed in the first three instances, the PSID codes these individuals as employed. If the respondent never answers that they are employed, the PSID asks a more general question of whether the respondent is "doing any work for money now?" An affirmative answer to this question leads to the creation of a current job narrative, despite the respondent being classified as not employed in the public PSID data.

[6] While narratives are available beginning in 1997 and the classification approach has been applied to the 1997-2019 data, we do not include data from the 1997–2001 survey waves in this paper's analyses due to changes in how the PSID coded main jobs that prevent us from consistently identifying main jobs separately from secondary jobs.

81,210 respondent-waves linked to 111,420 narratives. We then drop 7 respondents who provide some current job narrative, but no main job narrative. This leaves us with our main analysis sample of 81,203 respondent-waves linked to 111,403 narratives.[7]

Classification of Work Arrangements

The project makes use of Abramowitz et al.'s (2023) classification of work arrangements for the 2003-2019 waves of the PSID, whereby employer names and narrative responses to the open-ended industry and type of work questions were each coded as one of five work arrangements (platform-mediated gig work, informal self-employment, formal self-employment, business owners, wage and salaried employees) and a small number assigned no label due to insufficient information. "Platform-mediated gig work" includes work for app- or Internet-based platforms (e.g., Doordash, Uber, Lyft). "Informal self-employment" includes work respondents report as self-employment for non-business entities (e.g., cleaning, handyman) as well as itinerant forms of work (e.g., freelancer, babysitting, day laborer). "Formal self-employment" includes self-employment typically worked for another business entity, such as self-employment under an "umbrella" company (e.g., real estate agents, financial planners at an advisory company), consultants, independent contractors, or subcontractors. "Business ownership" includes reports of owning or running a business or family farm, being a partner in a firm or business, and being self-employed and managing their own or a family member's business or supervising employees. Finally, "wage and salaried employment" includes employees and employed supervisors including short-term employment and work at a temp agency.

[7] To estimate transitions across our work arrangements and nonemployment, we use a separate transitions sample that retains nonemployed respondent-waves. Our transitions sample contains 115,522 respondent-waves linked to 122,904 narratives.

The classification first distinguishes between wage and salaried work arrangements and self-employment work arrangements. The approach incorporates information from self-reports of employment status and self-employment status on a given job, but reclassifies work arrangements to align with the narrative information when it conflicts with the employment status or self-employment status report.[8] Among the self-employed, the classification further distinguishes business ownership (requiring investment and managerial responsibilities), working independently but typically for a business entity (providing greater structure to the employment relationship), and working independently but typically for an individual or on an electronic platform, or having itinerant work (offering less structure to the employment relationship). For most analyses, we aggregate platform-mediated gig work into the informal self-employment category to make inferences based on sufficient sample size. While platform-mediated gig workers are considered independent contractors for tax purposes, we aggregate platform-mediated gig work into the informal self-employment category because we observe that the characteristics of platform-mediated gig workers are most similar to workers engaged in informal self-employment.

We used machine learning to automate the classification. Two reviewers classified the same subset of 30% the data according to the described schema, with disagreements adjudicated by a third reviewer, to be used to train a BERT-based machine learning model to classify of the remainder of the data. Reviewers also classified records for which the model did not confidently predict a classification, following the same procedure as for producing the training data.

[8] For example, an Uber driver might mistakenly classify herself as an employee. Alternatively, an employee at a temp agency might mistakenly classify herself as self-employed.

Analysis

Using this classification, we examine (1) how the prevalence and nature of different self-employment work arrangements has changed over time, (2) how individuals transition across different types of self-employment work arrangements, and (3) the characteristics of individuals working in different types of self-employment work arrangements. We largely restrict our analysis to those age 16 or older with at least one current job narrative. While our sample largely overlaps with the sample of employed respondents as reported in the public PSID, our sample also includes respondents who self-report as not employed yet report a current job narrative. When estimating transitions, we expand our sample to those who reported a current job narrative in any wave, allowing us to observe extensive margin transitions. We deflate all measures of dollar amounts to 2019 dollars using the CPI-U. Finally, we weight all analyses using the PSID's cross-sectional individual weights.[9]

Results

Trends in Work Arrangements

We first examine how the shares of workers in self-employment have changed over time and compare our estimates to those from other sources. In Panel A of Figure 2, we present the share of self-employed workers among the employed, as reported in the PSID and CPS-ASEC. We show three different estimates of the self-employment share among the employed using the PSID to highlight how our classification approach allows us to capture a broader set of workers. First, in the dashed black line, we present the share of self-employed workers among the employed using self-reported self-

[9] For a detailed description of PSID cross-sectional individual weights, see Chang et al. (2019).

employment status from the public PSID.[10] In the dashed-and-dotted black line, we again show the share of self-employed workers among the employed using self-reported self-employment status from the public PSID, but include respondents with any current job narrative as employed. With this adjustment, we see a larger share of self-employed workers and a similar downward trend over our time period. The larger share is due to a disproportionate share of respondents with self-employed jobs self-reporting as non-employed.[11] In the solid black line, we show our preferred specification where we assign workers' self-employment status exclusively using our classification and assign respondents with any current job narrative as employed. This specification also shows a larger self-employment share, while maintaining a less pronounced downward trend from 14.4% in 2003 to 14.0% in 2019.

A key takeaway from Panel A of Figure 2 is that our measures of self-employment broadly capture similar trends as those measured by worker self-reports. Focusing on the solid and dashed-and-dotted black lines, these lines share a common sample—those who have a current job narrative—differing only in how they define self-employment status. We can see that up until roughly 2013, both our classification-based and the PSID's self-reported self-employment status capture similar aggregate trends with a modest divergence since. Overall, among those with a current job narrative, we find that our classification-based measure and the PSID's self-reports of primary self-employment match in 98.3% of cases. This provides supporting evidence that our approach is capturing a meaningful and common-sense notion of self-

[10] As discussed in the data section, the public PSID's definitions of employment and self-employment status rely solely on self-reports. In particular, respondents can be classified as nonemployed even in the presence of a current job narrative.
[11] For example, someone may self-report as "disabled" when asked the employment question in the PSID. This would lead them to be classified as non-employed, even if they have a current job narrative. Among those who self-report as nonemployed but provide a current main job narrative, more than half (55.8%) report being self-employed on their main job.

employment broadly, while allowing for us to distinguish across work arrangements and to identify employment not captured by worker self-reported employment status.

Comparing across data sources, Panel A of Figure 2 shows that relative to the share in either the CPS-ASEC, the public PSID data (and, in fact, all three of our estimates from the PSID) show a larger share of self-employment. The higher rates of self-employment in the PSID align with related work identifying undercounting of self-employment in the CPS-ASEC (Abraham et al. 2021a). Similar to the CPS-ASEC, the PSID data exhibit a downward trend, from 14.4% in 2003 to 12.5% in 2019. In Panel B of Figure 2, we benchmark our estimates of the share of workers reporting any current self-employment, using our preferred specification as defined above and plotted in the solid black line, to the share of the "tax workforce" that filed a Form 1099 or Schedule C or SE in a given tax year using data from Garin et al. (2023), plotted in the solid grey line.[12] While both sets of estimates include any self-employment, they differ in their reference periods, as our estimates include any current self-employment at the time of the survey, whereas the tax data estimates include any self-employment at any point in the tax year. As a result, our estimates mechanically represent a lower bound relative to the tax data. Indeed, we see a higher share of workers who report some self-employment in the tax data relative to our estimates. Moreover, the administrative data show a general upward trend, in contrast to the downward trend found in survey data. Recent work, however, suggests that rising self-employment in administrative tax data may largely reflect changes in strategic reporting rather than actual changes in labor market behaviour (Garin et al. 2022). Taken together, Panels A and B of Figure 2 show

[12] As defined by Collins et al. (2019), the "tax workforce" represents all individuals that report wage earnings (W-2), self-employment earnings (Schedule SE or C), or non-employee compensation (Form 1099) in their tax filings.

our estimates of self-employment fall within the range of plausible estimates found in other data sources.

The estimates in Panels A and B of Figure 2 show overall trends in self-employment, but they mask potential heterogeneity in trends across different self-employment work arrangements. To understand such trends, Figure 3 presents trends by work arrangement for main jobs in Panel A and for secondary jobs in Panel B. Figure 3 shows that informal self-employment increasingly has become the most common form of self-employment over our time period for both primary and secondary employment. For main jobs, in Panel A, we see a rise in informal self-employment, from 4.9% in 2003 to 5.9% in 2019, with much of this rise occurring after 2009.This is partly driven by trends in platform gig work—excluding platform gig workers, we find the share of informal self-employment to be 5.4% in 2019. However, roughly half of this rise remains unexplained by rising platform gig work. In contrast, we see a decline in formal self-employment from 5.0% in 2003 to 3.4% in 2019, with much of this decline occurring after 2009. For business ownership, we see an increase following the Great Recession that has subsequently returned to pre-recession levels. For secondary jobs, in Panel B, while the share of workers holding secondary jobs in formal self-employment or business ownership has been relatively constant, we see a rise in informal self-employment. Again, this rise is partly but not entirely explained by an increase in platform gig work.

To understand how trends in wellbeing vary across work arrangements, Figure 4 shows labor market and health outcomes by work arrangement and survey wave. Panels A-C show that labor earnings, weekly hours, and hourly wages are largely stable with a slight upward trend for wage and salaried employees; on the other hand, across all types of self-employment, we observe a downward trend along those same measures. While trends are similar across all self-employment categories, we see persistent differences in levels: informally self-employed workers on average had the lowest total earnings and wages and worked the fewest hours. Differences in self-reported health are stark as well. While all workers experienced a downward trend in self-reported health, informally self-employed workers are lowest in levels and exhibit the steepest downward trend.

How Workers Transitions across Work Arrangements

To understand the job transitions driving these trends, we next examine how workers transition across work arrangements from one survey wave to the next. Table 1 presents transition matrices across our four work arrangements and non-employment. To observe changes in transitions over time, we report separate wave-to-wave transitions for 2003-2009 and 2011-2019. The work arrangement in the prior survey wave is represented vertically while the survey wave work arrangement is represented horizontally. Each cell shows the percentage of respondents having that combination of prior and current work arrangements.

Table 1 shows that across all work arrangements, respondents are most likely to persist in the work status that they had in the prior survey wave. However, relative to non-employment and wage and salaried employment, self-employment is associated with greater diversity in transitions. Focusing first on 2003-2009 in Panel A, we see that 74.0% of the nonemployed and 85.3% of wage and salaried workers in the previous wave stay in their respective roles. On the other hand, we see that self-employed workers are not nearly as likely to remain in their roles across survey waves with only 36.9% of informally self-employed workers and 37.4% of formally self-employed workers staying in their roles across waves. Business owners are also less likely to persist in their roles than wage and salaried employees, but more likely to remain in their roles than any other type of self-employment, with 49.8% of business owners remaining in their roles across survey waves. In addition, we see that relative to all other work arrangements, the informally self-employed are more likely to enter non-employment: 20.2% of those informally self-employed in the previous wave became nonemployed in the current survey wave, nearly double the probability of any other work arrangement.

Examining transitions from 2011-2019 in Panel B, we see largely similar patterns to those that we document from 2003-2009. One notable change is a rise in persistence within work arrangements. For informally self-employed workers, we see that the probability of staying informally self-employed across waves increases from 36.9% in 2003-2009 to 47.1% in 2011-2019. For business owners, we see a larger rise with the probability of remaining a business owner increasing from 49.8% to 70.2%.

Characteristics of Workers by Work Arrangement

To better understand the characteristics of workers in different work arrangements, Table 2 presents demographic and economic characteristics and measures of wellbeing across work arrangements on main jobs. We break out platform gig workers to assess their comparability to informally self-employed workers. Examining demographic characteristics in Table 2, we can see that the informally self-employed are less educated than workers in all other work arrangements, though these differences are slight relative to wage and salaried employees and platform gig workers. The informally self-employed tend to be older than wage and salaried employees and slightly younger than or similar in age to other self-employed workers. We also see that the informally self-employed are more racially diverse than all other work arrangements, with the exception of platform gig workers. Table 2 further shows informal self-employment is associated with having lower labor earnings, fewer weekly hours worked, and lower wages relative to all other types of work other than platform gig work. In fact, platform gig workers and informally self-employed workers report similar labor market outcomes. Despite similar labor market outcomes, platform gig workers are more likely to report not owning a business relative to informally self-employed workers, instead reporting levels similar to wage and salaried employees. On the other hand, formally self-employed workers and business owners are more likely to report owning a business than the informally self-employed.

Finally in Table 2, we examine the extent to which different roles are associated with differential wellbeing. The informally self-employed are the least likely to report: (1) being in good health, (2) the absence of psychological distress, and (3) being very satisfied with their lives relative to all other work arrangements other than platform gig workers. As with labor market outcomes, we find that differences in self-reported well-being between platform gig workers and informally self-employed workers are statistically insignificant.

In Table 3, we examine how the characteristics of workers vary by secondary work arrangements.[13] The patterns we find in Table 2 largely hold. However, we find that the large gaps in economic outcomes—total labor earnings, weekly hours, and hourly wages—that we observed in Table 2 between informally self-employed workers and wage and salaried employees shrinks. In fact, those that hold secondary informal self-employment work largely similar hours to those who hold secondary wage and salaried employment.[14]

These results suggest salient differences in the composition of self-employment

---

[13] Whereas workers can only retain a single main job category by definition, workers can hold multiple secondary jobs. Thus, the same worker can be classified as a secondary job holder in multiple work arrangements.

[14] We note that this is also largely true for those who hold secondary platform gig work.

across the income distribution. To directly examine this, in Figure 5, we report self-employment shares for those that hold primary informal self-employment, primary non-informal self-employment, and only secondary self-employment by income quintiles. We find a U-shaped distribution with the largest share of workers engaged in self-employment at the bottom (28.9%) and top (17.8%) income quintiles. As we move along the income distribution, the composition of self-employment changes. In the bottom quintile, 15.3% of workers are engaged in informal self-employment as their main job, in contrast with 1.6% at the top income quintile. On the other hand, workers in the top income quintile are slightly more likely to be engaged in formal self-employment or business ownership (12.3%) than workers in the bottom quintile (11.3%).

Our classification also suggests important heterogeneity in the composition of self-employment by gender and age. In Panel B of Figure 5, we report the composition of self-employment across the age distribution separately for men and women. We see that the share of workers engaged in self-employment increases with age for both men and women. While men persistently have higher rates of overall self-employment across the age distribution than women, this is largely driven by the much larger rates of non-informal self-employment among men. Relative to men of comparable ages, women are consistently more likely to be informally self-employment than men. On average, 5.9% of women aged 31 – 64, and 11.9% of women aged 65 and older report being informally self-employed; in contrast, 4.3% and 11.0% of men of corresponding age groups report being informally self-employed. Overall, these findings show that the composition of self-employment is not constant across important subgroups. Accordingly, it is important to account for such heterogeneity when conducting causal analyses of the effects of self-employment.

Discussion

This paper used novel data to examine the breadth of self-employment work arrangements to understand: (1) how the prevalence and nature of different self-employment work arrangements has changed over 2003-2019, (2) how individuals transition across different types of self-employment work arrangements, and (3) the characteristics of individuals working in different types of self-employment work arrangements.

We demonstrate that our approach captures more self-employment, finding higher levels of self-employment than either the CPS-ASEC or the public PSID data, while matching the general downward trend in self-employment seen in both data sources. We further find evidence that divergent trends exist within self-employment: formal self-employment shares have fallen, whereas informal self-employment shares have risen. Business ownership experienced large fluctuations, rising following the Great Recession and subsequently returning to pre-Great Recession levels.

We also document that our classification captures meaningfully different types of workers. Relative to other self-employed workers, business owners are far more likely to remain business owners across survey waves; on the other hand, informal self-employed workers are nearly twice as likely to enter non-employment relative to other work arrangements. These patterns are stable over time with a rise in the probability of staying in one's work arrangements across survey waves. Next, we find that on a wide set of demographic, labor market, and well-being measures the informally self-employed are relatively similar to platform gig workers, but diverge sharply from the formally self-employed and business owners. Taken together, our results support the notion that the informally self-employed face less rewarding work prospects and decreased wellbeing than workers engaged in other types of self-employment.

Finally, we show that the composition of self-employment varies across meaningful subgroups. In particular, we see that women and low-income workers are far more likely to engage in informal self-employment than men or higher-income workers. These differences across subgroups are substantial in relative terms. Our findings suggest the importance of accounting for such heterogeneity in empirical analyses.

Taken together, our findings suggest salient differences in trends, transitions, and work and individual characteristics of the self-employed that would otherwise be masked in administrative data and other survey sources. Prior work has shown that administrative data miss substantial amounts of self-employment activity at both the intensive and extensive margins and surveys generally do not probe to the necessary extent to identify work arrangements of interest. Using novel data, we are able to identify these self-employment work arrangements and find that they do reflect substantially different work characteristics and are engaged in by individuals with different characteristics. It is important to identify these differences to understand how the nature of work is changing over time and to inform future policy making to improve the wellbeing of workers.

It is important to note that while the paper's approach is valuable, it does have limitations. The results are limited in that the classification can only be used to the extent the respondents provided sufficiently detailed narratives and there is some degree of subjectivity and error in reviewer coding of work arrangements. However, we have mitigated the latter by having every job record reviewed by at least two reviewers according to a standardized classification schema. Another limitation of this analysis is that we only examine current jobs and focus on current main jobs. Future work could examine all jobs held to develop a more nuanced understanding of how individuals hold and transition across multiple jobs over time.

The results of this study provide greater insight into the nature of self-employment work arrangements and permit future work more thoroughly considering the causes and implications of differences in these work arrangements. This work lays the groundwork for future research examining individuals' work trajectories leading to these roles, movement between different work arrangements, and how these are associated with different levels of economic, physical, and psychological wellbeing over the life course.

References

Abraham, Katharine G., and Ashley Amaya. 2019. ‘Probing for Informal Work Activity.’ Journal of Official Statistics 35(3):487–508.

Abraham, Katharine G., John Haltiwanger, Claire Hou, Kristin Sandusky, and James R. Spletzer. 2021a. ‘Reconciling Survey and Administrative Measures of Self-Employment.’ Journal of Labor Economics 39(4):825-860.

Abraham, Katharine, John Haltiwanger, Kristin Sandusky, and James Spletzer. 2018. ‘Measuring the Gig Economy: Current Knowledge and Open Issues.’ Working Paper No. 24950. Cambridge, MA: National Bureau of Economic Research.

Abraham, Katharine G., Brad Hershbein, and Susan N. Houseman. 2021b. ‘Contract Work at Older Ages.’ Journal of Pension Economics and Finance, 20(3):426–447.

Abraham, Katharine G., Brad Hershbein, Susan N. Houseman, and Beth Truesdale. 2023. ‘The Independent Contractor Workforce: New Evidence on its Size and Composition and Ways to Improve its Measurement in Household Surveys.’ Working Paper No. 30997. Cambridge, MA: National Bureau of Economic Research.

Abramowitz, Joelle, Jenna Kim, and Jinseok Kim. 2023. “C-NARS/P: An Open-Source Tool for Classification of Narratives in Survey Data Using Pre-Trained Language Models.” Technical Report.

Allard, Mary D., and Anne E. Polivka. 2018. ‘Measuring Labor Market Activity Today: Are the Words Work and Job Too Limiting for Surveys?’ Monthly Labor Review. U.S. Bureau of Labor Statistics.

Boeri, Tito, Giulia Giupponi, Alan B. Krueger, and Stephen Machin. 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.

Chang, Wen, Raphael Nishimura, Steven G. Heeringa, Kate McGonagle, and David Johnson. 2019. 'PSID Cross-Sectional Individual Weights, 1997–2017.' Ann Arbor, MI: University of Michigan Institute for Social Research.

Garin, Andrew, Emilie Jackson, and Dmitri Koustas. 2022. 'New Gig Work or Changes in Reporting? Understanding Self-Employment Trends in Tax Data.' Working Paper No. 2022-67. University of Chicago Becker Friedman Institute for Economics.

Garin, A., Jackson, E., Koustas, D.K. and Miller, A., 2023. 'The Evolution of Platform Gig Work, 2012-2021.' Working paper No. 31273. Cambridge, MA: National Bureau of Economic Research.

Jackson, Emilie, Adam Looney, and Shanthi P. Ramnath. 2017. 'The Rise of Alternative Work Arrangements: Evidence and Implications for Tax Filing and Benefit Coverage.' Working Paper No. 114. Internal Revenue Service Office of Tax Analysis.

Katz, Lawrence F., and Alan B. Krueger. 2019. 'Understanding Trends in Alternative Work Arrangements in the United States.' Working Paper No. 25425. Cambridge, MA: National Bureau of Economic Research.

Mortenson, Jacob A., and Andrew Whitten. 2020. 'Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule.' American Economic Journal: Economic Policy 12(3):402-32.

Levine, Ross, and Yona Rubinstein. 2017. 'Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?' The Quarterly Journal of Economics 132(2):963–1018.

Light, Audrey, and Robert Munk. 2018. 'Business Ownership versus Self-Employment.' Industrial Relations 57(3):435-468.

Moulton, Jeremy G., and John C. Scott. 2016. 'Opportunity or Necessity? Disaggregating Self-Employment and Entry at Older Ages.' Social Forces 94(4):1539–1566.

National Academies of Sciences, Engineering, and Medicine. 2020. 'Measuring Alternative Work Arrangements for Research and Policy.' Washington, DC: The National Academies Press.

Saez, Emmanuel. 2002. Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses. The Quarterly Journal of Economics 117(3):1039-1073.

Slemrod, Joel, Brett Collins, Jeffrey L. Hoopes, Daniel Reck, and Michael Sebastiani 2017. 'Does Credit-Card Information Reporting Improve Small-Business Tax Compliance?' Journal of Public Economics 149:1-19.

Slemrod, Joel. 2016. 'Caveats to the Research Use of Tax-Return Administrative Data.' National Tax Journal 69(4):1003-1020.

Tables

Table 1: Wave-to-Wave Transitions in Work Status

Panel A: Wave-to-Wave Transitions in Employment Status (2003-2009)

		Current Wave				
		Not Working	*W&S Employment*	*Informal SE*	*Formal SE*	*Business Ownership*
Prior Wave	*Not Working*	74.0%	21.2%	2.9%	1.1%	0.8%
	W&S Employment	10.5%	85.3%	1.6%	1.6%	1.0%
	Informal SE	20.2%	22.6%	36.9%	11.6%	8.7%
	Formal SE	7.9%	26.2%	11.9%	37.4%	16.6%
	Business Ownership	7.1%	16.8%	9.4%	17.0%	49.8%

Panel B: Wave-to-Wave Transitions in Employment Status (2011-2019)

		Current Wave				
		Not Working	*W&S Employment*	*Informal SE*	*Formal SE*	*Business Ownership*
Prior Wave	*Not Working*	82.2%	13.3%	2.7%	1.0%	0.8%
	W&S Employment	10.5%	86.5%	1.0%	1.0%	0.9%
	Informal SE	22.3%	17.2%	47.1%	8.1%	5.3%
	Formal SE	10.7%	24.3%	9.3%	43.4%	12.3%
	Business Ownership	8.5%	11.4%	4.0%	5.8%	70.2%

[a] *Source*: Internal PSID narrative data on industry and occupation and employer names (2003-2019) classified into work arrangement types. Additional information on work status comes from the public PSID (2003-2019) merged to the narrative data classified into work arrangement types.
[b] We report the probability of a respondent transitioning to a given current main job type conditional on their main job type in the prior survey wave. Estimates use cross-sectional PSID weights.
[c] Abbreviations: W&S, wage and salaried; SE, self-employment.

Table 2: Characteristics by Type of Work Arrangement on Main Job

	Platform Gig Work		Informal Self-Employment		Formal Self-Employment		Business Ownership		Wage and Salaried Employment	
	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error
Age	52.9*	2.47	48.6	0.57	50.0	0.66	50.8***	0.57	44.4***	0.12
Years of Education	13.8**	0.27	13.1	0.14	14.4***	0.10	14.3***	0.13	13.9***	0.06
% Male	65.6***	6.02	46.8	1.81	65.3***	2.16	74.8***	1.55	50.5**	0.46
% White, Non-Hispanic	38.6***	6.09	67.3	2.54	82.3***	2.43	87.6***	1.89	74.3**	1.74
% Black, Non-Hispanic	21.5**	4.72	10.5	1.22	5.9***	0.98	3.2***	0.66	10.5	1.21
% Hispanic	-	-	18.3	1.91	7.4***	1.56	5.5***	1.15	10.9***	0.94
Labor Income (000's) - Prior Year	25.7	3.89	27.7	1.20	75.4***	4.76	92.3***	5.90	60.2***	0.89
Weekly Hours - Prior Year	33.3	1.65	32.8	0.51	38.8***	0.60	44.8***	0.77	41.7***	0.10
Hourly Wages - Prior Year	16.5	3.73	19.6	0.67	38***	1.52	37.3***	1.65	28.8***	0.38
% Don't Own a Business - Prior Year	90.9***	1.38	63.4	1.67	43.6***	1.79	17.1***	1.48	90.4***	0.43
% Reporting Good Health	83.6	2.44	82.8	1.16	91.3***	0.84	91.8***	1.01	90.5***	0.32
% Not Psychological Distress	96	0.66	94.9	0.70	98.4***	0.41	98.1***	0.47	97.9***	0.17
% Very Satisfied	50	8.05	59.9	2.42	68.4**	2.53	78.2***	2.40	70.3***	0.61
Sample	68		4,106		2,923		3,310		70,796	

[a] *Source*: Internal PSID narrative data on industry and occupation and employer names (2003-2019) classified into work arrangement types. Demographics come from the public PSID (2003-2019) merged to the narrative data classified into work arrangement types.
[b] *** p<0.01, ** p< 0.05, * p < 0.10 for t-test of difference in means compared to workers classified as having informal self-employment for their current main job.
[c] We report demographics by current main job type. Reported observations represent true counts of observations in our data. % Hispanic is censored due to sample size falling below disclosure requirements. Estimates use cross-sectional PSID weights.

Table 1: Characteristics by Type of Secondary Work Arrangement

	Platform Gig Work		Informal Self-Employment		Formal Self-Employment		Business Ownership		Wage and Salary Employment	
	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error
Age	41.4**	1.48	45.0	0.54	47.3**	0.95	47.5**	1.07	43.9	0.42
Years of Education	14.4	0.26	14.3	0.12	15.1***	0.13	14.6	0.19	14.4	0.08
% Male	72.9*	7.38	58.4	2.53	66.4**	2.78	76.1***	2.74	48.6***	1.87
% White, Non-Hispanic	61.9**	8.30	82.5	1.74	85.4	2.43	86.4	2.56	75.9**	2.31
% Black, Non-Hispanic	23.8	9.97	7.7	1.21	7.7	1.82	4.6*	1.16	12**	1.81
% Hispanic	-	-	7.0	1.05	4.7	1.07	3.7**	1.12	8	1.10
Labor Income (000's) - Prior Year	52.2	5.92	51.8	1.76	87.6***	4.33	91.5***	5.87	61.3***	2.55
Hours - Prior Year	45.2	1.45	44.4	0.59	48.1***	1.04	52.8***	1.48	45.8*	0.46
Wages - Prior Year	23.9	2.47	23.4	0.57	35.5***	1.62	34.4***	2.05	26.1***	0.68
% Don't Own a Business - Prior Year	88.1***	6.60	58.8	2.23	46.1***	2.77	16.2***	2.06	78.8***	1.88
% Reporting Good Health	95.1**	0.95	91.7	0.92	95.1**	0.97	94.6	1.96	92.1	0.66
% Not Psychological Distress	99.4***	0.27	96.8	0.87	98.8*	0.62	99.5***	0.42	97.7	0.38
% Very Satisfied with Life	32.1***	8.48	67.0	2.59	73.4	4.01	78.6**	3.80	70.1	1.75
Sample	79		1,863		851		422		4,599	

[a] *Source*: Internal PSID narrative data on industry and occupation and employer names (2003-2019) classified into work arrangement types. Demographics come from the public PSID (2003-2019) merged to the narrative data classified into work arrangement types.
[b] *** p<0.01, ** p< 0.05, * p < 0.10 for t-test of difference in means compared to workers classified as having informal self-employment for their secondary job.
[c] We report demographics by current secondary job type. Since an individual can hold multiple secondary jobs, there is some overlap across columns. Reported observations represent true counts of observations in our data. % Hispanic is censored due to sample size falling below disclosure requirements. Estimates use cross-sectional PSID weights.

Figures

Figure 1. Sample Construction

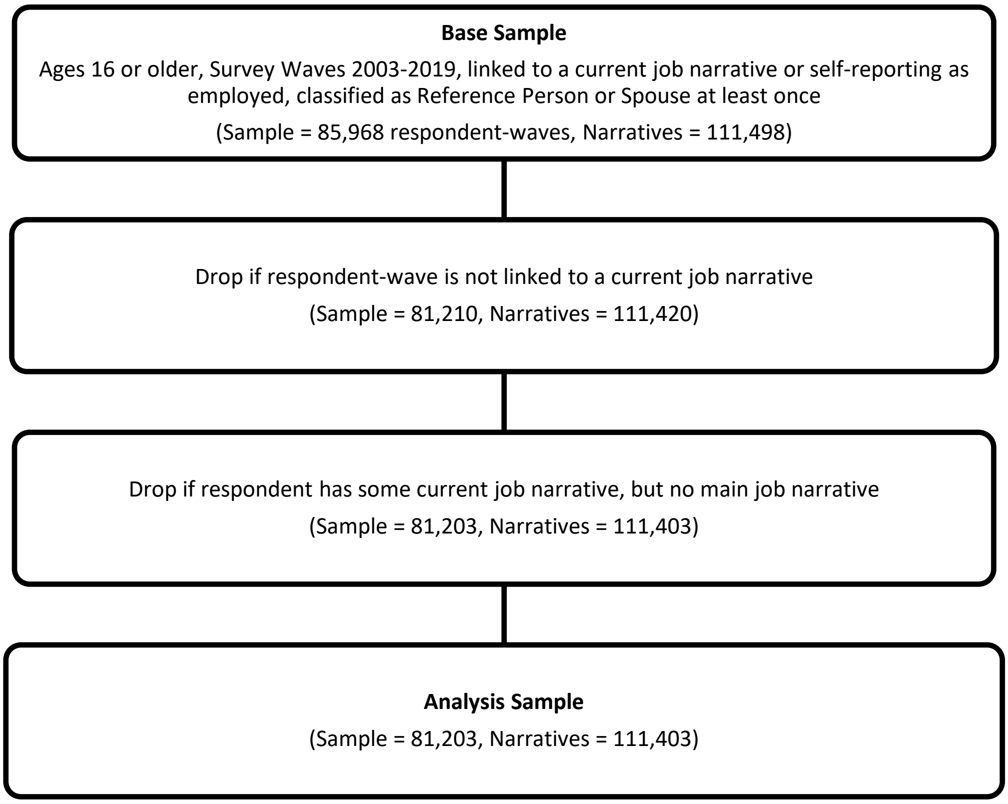
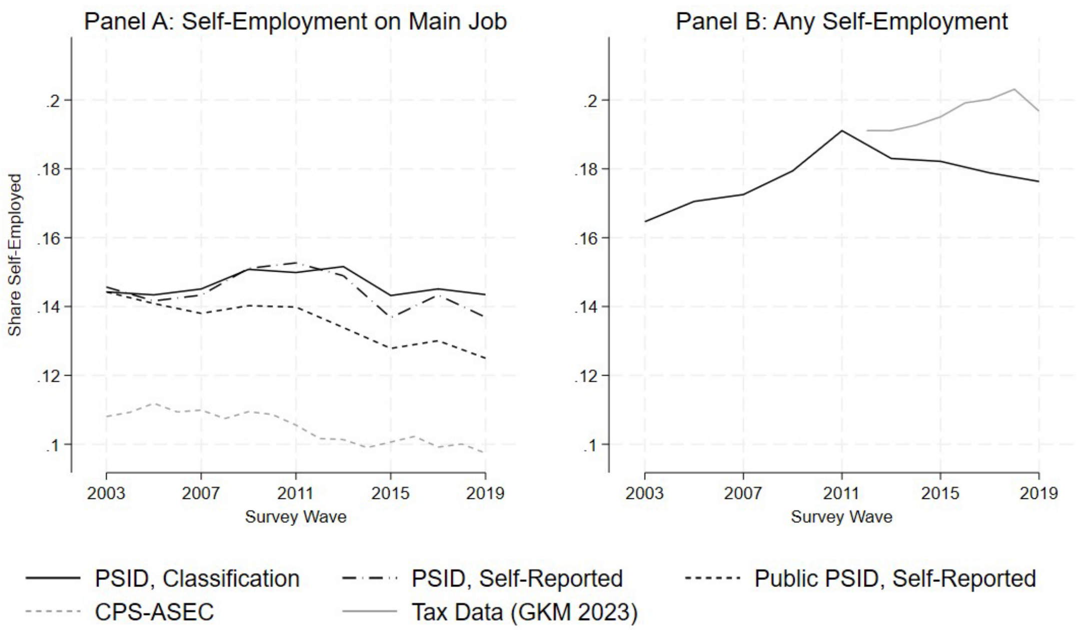
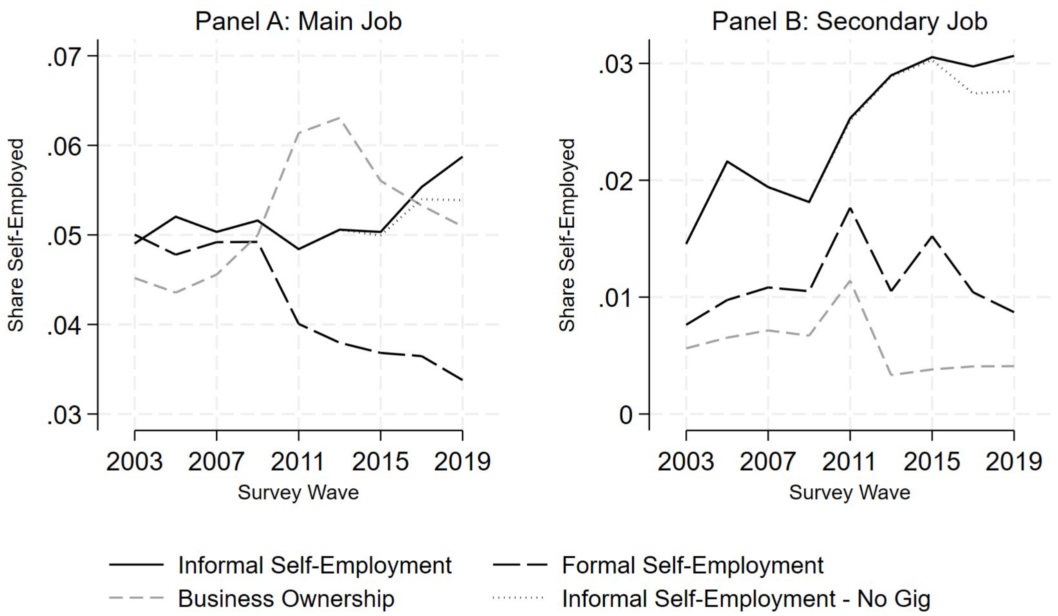


Figure 2: Share of Workers who are Self-Employed on Current Job by Survey Wave



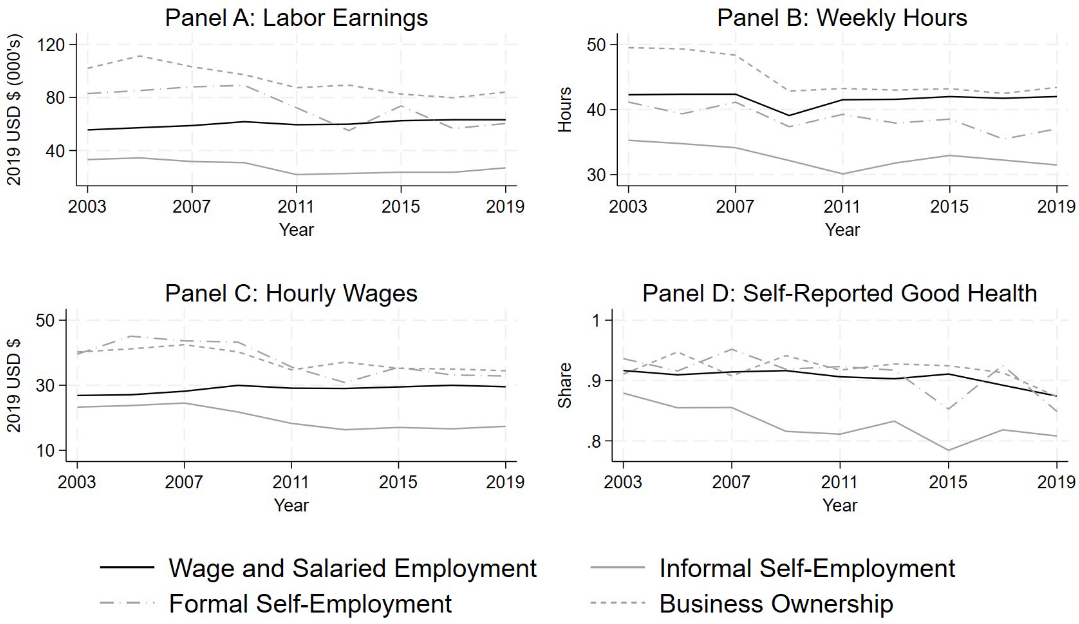
[a] *Source*: Internal PSID narrative data on industry and occupation and employer names (2003-2019) classified into work arrangement types and public PSID data (2003-2019).
[b] We report self-employment shares among employed workers. PSID estimates are weighted using cross-sectional weights. CPS-ASEC estimates are weighted using ASEC weights. Tax data estimates come from Garin et al. (2023). The dashed black line reports estimates using self-reported self-employment status from the public PSID. The dashed-and-dotted black line reports estimates using self-reported self-employment status from the public PSID, including respondents with any current job narrative as employed. The solid black line, our preferred specification, assigns workers' self-employment status exclusively using our work arrangements classification and includes respondents with any current job narrative as employed. Estimates use cross-sectional PSID weights.

Figure 3: Share of Workers who are Self-Employed by Work Arrangement and Survey Wave



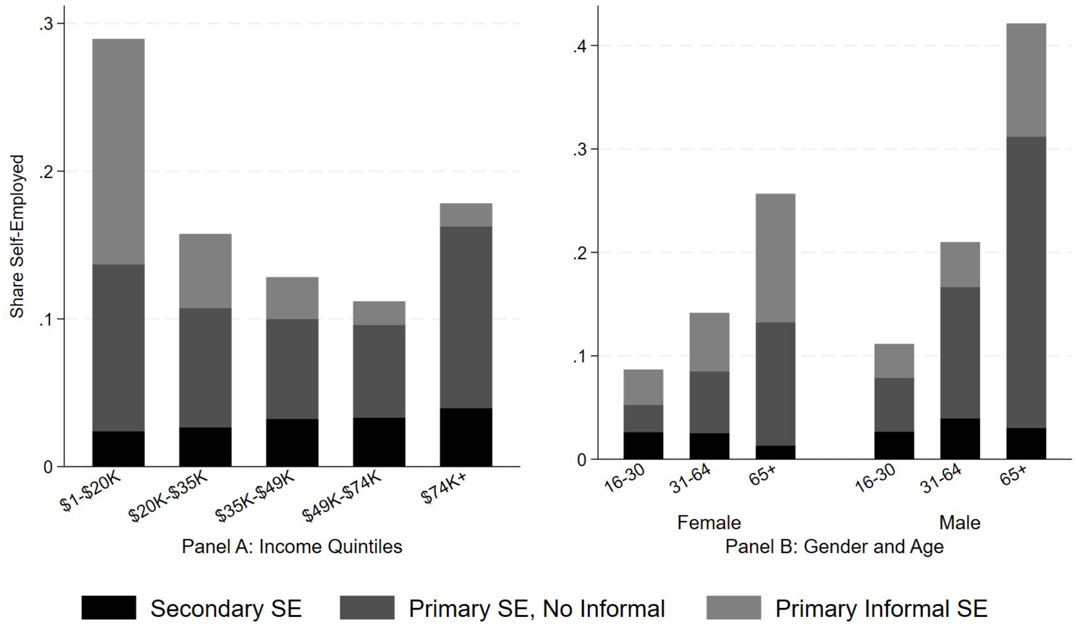
[a] *Source*: Internal PSID narrative data on industry and occupation and employer names (2003-2019) classified into work arrangement types and public PSID data (2003-2019).
[b] We report employment shares by current main job and secondary job type. We derive main and secondary job designations from the restricted PSID narrative data and public PSID data. Since workers can hold multiple secondary jobs, overlap across secondary job categories occurs. Our sample is restricted to respondent-waves in which a job narrative was given for a current main job. Estimates use cross-sectional PSID weights.

Figure 4: Trends in Well-being by Work Arrangement and Survey Wave



[a] *Source*: Internal PSID narrative data on industry and occupation and employer names (2003-2019) classified into work arrangement types and public PSID data (2003-2019).
[b] We report total labor earnings, weekly hours worked, hourly wages, and self-reported health by current main job type. We derive main job designations from the restricted PSID narrative data and public PSID data. Estimates use cross-sectional PSID weights.

Figure 5: Self-Employment Shares by Subgroups



[a] *Source*: Internal PSID narrative data on industry and occupation and employer names (2003-2019) classified into work arrangement types. Demographics come from the public PSID (2003-2019) merged to the narrative data classified into work arrangement types.

[b] We plot the share of self-employed workers among the employed by income quintiles and by gender and age. For this figure, workers classified as having secondary self-employment must report no primary self-employment. Earnings are rounded to the nearest $100 to maintain confidentiality. Estimates use cross-sectional PSID weights.

[c] Abbreviations: SE, self-employment.