STAFF REPORTS

NO. 1085
JANUARY 2024

# Self-Employment and Labor Market Risks

Richard Audoly

FEDERAL RESERVE BANK *of* NEW YORK

**Self-Employment and Labor Market Risks**
Richard Audoly
*Federal Reserve Bank of New York Staff Reports*, no. 1085
January 2024
https://doi.org/10.59576/sr.1085

**Abstract**

I study the labor market risks associated with being self-employed. I document that the self-employed are subject to larger earnings fluctuations than employees and that they frequently transition into unemployment. Given that the self-employed are not eligible to unemployment insurance, I analyze the provision of benefits targeted at these risks using a calibrated search model with (i) precautionary savings, (ii) work opportunities in paid and self-employment, and (iii) skill heterogeneity. This exercise suggests that extending the current U.S. unemployment insurance scheme to the self-employed comes with a clear increase in the transition rate from self-employment to unemployment and an unequal benefits-to-contributions ratio across skill groups. At the calibrated parameters, the self-employed in the middle of the skill distribution lose welfare.

JEL classification: J40, J64, J65
Keywords: self-employment, unemployment insurance, earnings dynamics

---

Audoly: Federal Reserve Bank of New York (email: audolyr@gmail.com). This paper is a revised version of the second chapter of the author's doctoral dissertation. He thanks his supervisors at University College London, Fabien Postel-Vinay and Jan Eeckhout, for their continual guidance. In addition, he thanks Rory McGee and Ludo Visschers, as well as his examiners, Rasmus Lentz and Wei Cui, for extremely useful feedback. He also thanks workshop and seminar participants at Aarhus University (Sandbjerg Manor), the EEA 2020 Online Conference, UCL, and NHH for their comments.

# 1 Introduction

Job loss ranks amongst the most significant risks that workers face over the course of their career. Many countries target sizable transfers to the unemployed in the form of unemployment insurance (UI), and a substantial literature aims at characterizing the optimal UI contract.[1] By contrast, while the self-employed account on average for fifteen percent of employment across OECD countries, there is little evidence on the labor market risks associated with self-employment.[2] Besides, because the majority of the self-employed are not eligible to UI in these countries, traditional social insurance programs are not well-suited to alleviate these risks.[3]

This paper studies the labor market risks associated with self-employment and the provision of benefits targeted at these risks. I first provide new empirical evidence on the earnings risks faced by the self-employed using US monthly survey data. The use of data at monthly frequency is important to accurately measure the drivers of these risks for the self-employed. It allows to separate earnings fluctuations within a given self-employment spell from transitions to unemployment or wage work. I show that (i) earnings are substantially more volatile during self-employment spells than during paid-employment spells, and (ii) there are frequent direct transitions from self-employment to unemployment.

My second contribution is to build a framework to assess the impact of extending UI benefits to the self-employed. I develop and calibrate a search model with precautionary savings that incorporates the patterns of labor market risks I document in the data. I use

[1]Following Jacobson et al. [1993], many studies have confirmed that job loss has long- term negative effects on workers' earnings. Examples of work on the optimal UI contract include Acemoglu and Shimer [1999], Chetty [2008], and Kolsrud et al. [2018].

[2]Throughout the paper, I use the terms "self-employed" and "self-employment" to designate all workers who get most of their labor income from a business. This definition is further clarified when I introduce the data in Section 2.

[3]Among these countries, a handful offer some form of public unemployment insurance for some narrow group of self-employed workers, such as artists and writers in Germany. These schemes are reviewed in detail in OECD [2018]. In the US, some owners of incorporated businesses (S- and C-corps) can become eligible under very specific conditions. Sole proprietors, partnership-owned business, Limited Liability Corporations, and independent contractors do not qualify. This restriction was abolished as part of the US government response to Covid-19 (CARES Act).

this framework to quantify the welfare changes and distortions in labor supply resulting from the introduction of a UI scheme for the self-employed.

A key pre-requisite to my calibration is to allow for substantial worker heterogeneity in the model. This heterogeneity is important because the data show that self-employed workers are over-represented in the tails of the earnings distribution, so the underlying skills are likely to differ widely across workers. To discipline this feature of the model, I follow Bonhomme et al. [2022] and use a k-means algorithm to partition workers based on their observed labor income and likelihood of becoming unemployed. These two measures are strongly correlated in the resulting clusters of workers. For instance, relative to the group of self-employed with high earnings, the low earners are close to four times as likely to make a transition from self-employment to unemployment. Building heterogeneity in such a way into the model allows to study how self-employed workers with different earnings potential respond to additional benefit entitlements.

The model is calibrated to replicate the empirical evidence on the exposure of the self-employed to labor market risks. Paid-employment and self-employment each come with specific labor income and unemployment shocks. In line with the data, the model captures the substantial flows between paid-employment, self-employment, and unemployment, thus allowing for moves between paid- and self-employment to represent a response to unemployment shocks or bad realizations of earnings. In addition, workers can also partially insure against labor market risks by borrowing and drawing down their savings, and by relying on household-level income (spousal income and welfare transfers). My calibration matches the large fraction of households with low wealth holdings in the data. Through the lens of the model, many self-employed households therefore have limited means to self-insure.

I use the calibrated model to study several alternative UI schemes targeted at the self-employed. These alternative schemes have different contribution regimes to highlight the redistributive dimension of UI in this setting. For instance, a standard UI scheme with a contribution ceiling implies that the burden of funding the policy falls in large part on workers with lower earnings, and the contributions to benefits ratio is less than one for workers in the middle of the skill distribution at the calibrated parameters. For this reason, I also report results for a scheme with a flat contribution rate and a scheme with no transfers across skill groups.

This exercise suggests that there is a clear labor supply response in all alternative UI scenarios. Unemployment increases for all earnings groups, in large part driven by an increase in the rate of self-employment to unemployment transitions. By contrast, the welfare effects of the counterfactual UI schemes depend on how the policy is funded. For instance, workers in the middle of the skill distribution lose welfare in the baseline UI scheme, which extends the actual policy for wage employees in the US to the self-employed. In the hypothetical scenario where the scheme would be actuarially fair for self-employed workers at the bottom of the skill distribution, their welfare decreases. Overall, none of the alternative UI schemes considered yield welfare gains for all earnings groups. These results point to the importance of the redistributive dimension of these UI schemes to properly account for their welfare effects.

More generally, the analysis developed in this paper is relevant for the design of policy in several dimensions. Like regular paid employees, first, I document that a fraction of the self-employed are not insulated from labor market risks. Second, the rise of alternative work arrangements with the emergence of online labor platforms (firms that match workers to customers without being bound to them by an employment contract) may further increase the number of self-employed with low earnings in the future.[4] My framework highlights some of the key trade-offs to providing additional benefits to this group of workers.

**Related literature** This work is related to the large literature studying the risks faced by wage employees in the labor market, most notably unemployment, and the associated optimal

[4]Collins et al. [2019] precisely measure the rise in platform work in tax data, which allow them to precisely pinpoint self-employment earnings derived from work for or intermediated by specific online platform firms. While these authors find a marked increase in this type of work arrangements, their analysis also shows that it primarily represents auxiliary income for individuals with traditional jobs.

provision of unemployment insurance [Chetty, 2008, Kolsrud et al., 2018]. My approach is closest to the series of papers that study UI benefits within the context of a fully specified structural model [Hansen and Imrohoroğlu, 1992, Acemoglu and Shimer, 1999, Lentz, 2009, Krusell et al., 2010]. I depart from these studies by focusing on the labor market risks specifically associated with being self-employed.

This work also relates to the extensive literature estimating income processes [e.g., Meghir and Pistaferri, 2004, Guvenen, 2009] and the degree of insurance households can achieve in response to the resulting income shocks [e.g., Blundell et al., 2008, Kaplan and Violante, 2010]. While most papers in this literature drop the self-employed, whose earnings are found to be substantially more volatile, I specifically center on this category of workers.[5] Using data at monthly frequency, I can further unpack the drivers of earnings fluctuations by separately studying transitions between employment states and earnings shocks within paid-employment and self-employment spells.

This paper also contributes to a growing body of work on self-employment. This literature primarily studies the decision to become self-employed [Hamilton, 2000, Hurst and Lusardi, 2004, Levine and Rubinstein, 2017, Humphries, 2017, Catherine, 2022, Jones and Pratap, 2020]. This line of research does not directly consider the insurance dimension of self-employment at the household level. Perhaps the closest paper to mine in that regard is Catherine [2022], who quantifies the value of paid-employment as a back up option for workers deciding whether to start a business. This last paper uses data on yearly earnings and does not consider the risk of becoming unemployed.

A branch of the self-employment literature also studies self-employment as a path out of unemployment by analyzing the outcomes of the ensuing businesses [Hombert et al., 2020, Camarero Garcia and Murmann, 2020]. Jackson [2022] specifically studies the role of gig economy work. While this channel is present in my model, I instead stress the risks to earnings conditional on workers being self-employed. I can then evaluate several policies

[5]In the income process section of their review paper on earnings and consumption, Meghir and Pistaferri [2011] write: "the focus is mainly on employed workers and self-employed workers are typically also dropped."

targeted at these risks within my framework.

**Outline** The next section presents the data. Section 3 introduces the model. Section 4 discusses the calibration procedure. Section 5 describes the policy experiments, and Section 6 concludes.

# 2 Data

This section provides empirical evidence on the earnings risks associated with being self-employed. An accurate description of these risks in the data requires: (i) frequent records of each individual's labor market history to capture transitions in and out of employment, (ii) information on household income and wealth, which represent key self-insurance channels, and (iii) a large enough sample since the self-employed represent a small fraction of total employment. The Survey of Income and Program Participation meets these requirements [Census Bureau, 2014, SIPP thereafter]. I pool together the four panels spanning 1996-2013. All statistics reported throughout the paper are obtained using longitudinal weights.[6] I stress two key dimensions of earnings risks in these data: shocks to labor income while employed (an intensive margin risk) and unemployment (an extensive margin risk). I start by showing how individuals with at least some experience of self-employment over the sample period differ from individuals always working as paid employees. In a second step, I document the drivers of earnings risks for the self-employed.

## 2.1 The self-employed in the SIPP

Workers are assigned to one of three labor market states $s$ in each month: unemployed ($s = U$), paid-employed ($s = P$), or self-employed ($s = S$). I use a standard definition

[6] One of the primary aims of the SIPP is to gather data on participation in welfare programs, and it therefore over-samples less affluent areas. In addition, the data are frequently missing (the attrition rate ranges from 26 percent to 35 percent across the panels spanning 1996-2013 in the raw data), and the weights are intended to adjust for sample non-response.

of unemployment. Workers are categorized as unemployed if they declare to be searching for work at any point during a non-employment spell. When in employment, workers are categorized as paid- or self-employed based on their primary source of earnings in a month.[7] For example, $s = S$ if the person declares that most of their earnings come from their business in a given month, so a person can therefore be paid- or self-employed in different months over the duration of the survey. I focus on earnings because hours are frequently missing in the SIPP.[8] Earnings also more directly relate to the resources workers can actually set aside in anticipation of labor market shocks. For an overwhelming majority of workers, this definition clearly singles out one labor form (paid- or self-employment) as an individual's main source of earnings in any given month.[9] Less than 3 percent of worker-month observations declare working both as employees and at their own business in the same month.

I restrict the sample to the working-age individuals with the largest earnings in each household over the duration of the survey. This restriction is imposed both to focus on the individuals with the strongest ties to the labor market and to account for household-level insurance channels since shared assets, earnings from other household members, and welfare programs represent alternative sources of insurance against earnings risks. I incorporate these insurance channels in the model introduced in Section 3.

Within the sample of main earners, I define the self-employed as workers who (i) get most of their earnings from their business in at least one month over the duration of the survey (labor market state $s = S$), and (ii) are always unincorporated when in self-employment. I focus on the unincorporated self-employed because some incorporated self-employed are already eligible to UI if they pay themselves a wage through their business, and I do not

[7] I check that this definition yields a sensible aggregate self-employment rate. Figure A.2b confirms that the self-employment rate derived from the SIPP is very close to the BLS series when taking all working age individuals into account.

[8] Starting with the 2004 panel, the survey allows respondents either to enter a number or to declare that "hours vary" when reporting how many hours were worked at each job and business in the current month. 25 to 30 percent of business owners and 10 to 15 percent of employees respond that "hours vary," which makes a definition based on hours impractical.

[9] In what follows, I use the expression "labor form" to designate whether the individual works as an employee ($s = P$) or at their own business ($s = S$).

| Business characteristics | Value |
|---|---|
| Reported business wealth (Median, $2009) | |
| Assets | 3,444 |
| Debt | 0 |
| Net business wealth | 2,268 |
| | |
| Share incorporated | 0.000 |
| Share with more than 25 employees | 0.015 |

**Table 1:** Business characteristics in the self-employed sample. Business wealth is obtained by multiplying the owner's share and summing the data for all businesses, if several businesses are reported. "Assets" is defined as the value of the business before any debts owed against it. "Debts" is defined as any debts owed against the business. "Net business wealth" is defined as assets minus debts. The incorporation and employment variables take value one if any of the reported businesses satisfies the condition.

observe eligibility in the data. Previous work has also established that the incorporated self-employed are distinct from the unincorporated self-employed along several dimensions, such as their cognitive and non-cognitive skills, their level of education, and the size of their business [Levine and Rubinstein, 2017]. In the remainder of the paper, I refer to the sample of unincorporated self-employed as "self-employed" for simplicity. Appendix A provides additional details on the data and on the definition of each labor market state.

Within the self-employed sample, the majority of businesses from which self-employment income is derived are limited operations. Table 1 reports some key business characteristics. The reported median business assets for workers in self-employment is slightly below $3,500, and only 1.5 percent of the self-employed declare that their business has ever had more than 25 employees since it was established.

## 2.2 Self-employed and paid employees

To better understand the self-employed's exposure to labor market risks, I compare them to the sample of main earners who are always paid employees over the duration of the survey. The self-employed sample represents about seven percent of the entire sample of

| | Paid employees sample | Self-employed sample |
|---|---|---|
| Age | 38.181 | 40.069 |
| Gender (woman=1) | 0.399 | 0.283 |
| Married (married=1) | 0.533 | 0.542 |
| Race (non-white=1) | 0.192 | 0.130 |
| HS Graduate | 0.907 | 0.863 |
| College Graduate | 0.471 | 0.422 |
| Post-graduate | 0.103 | 0.102 |
| Number of individuals | 100,312 | 8,623 |
| Average months in panel | 35.347 | 35.198 |

**Table 2:** Demographic characteristics by subsample. "Paid employees" is the subsample of main earners who are never self-employed (labor force status $s = S$) over the duration of the sample. "Self-employed" is the subsample of main earners who are (i) self-employed in at least one survey month (labor force status $s = S$ in at least one month) and (ii) always unincorporated when in self-employment.

main earners.

Table 2 gives some basic demographic statistics on these two groups of workers. The self-employed are more likely to be older men. They are also less likely to belong to a minority and to have a college degree.

Table 3 compares the paid employees and self-employed samples in terms of income and wealth. In Table 3 and throughout the paper, all monetary values are given in 2009 real dollars. To account for differences in monthly fluctuations in income across samples, all income measures are reported as twelve-month averages (including any potential zeros). A portion of the self-employed sample is characterized by lower earnings than paid-employees. For instance, the tenth percentile of total household income is \$1,103 (\$1,898), respectively for workers with (without) self-employment. In terms of liquid wealth, most self-employed have not accumulated more wealth than paid employees. The median net liquid wealth (net worth excluding business, pension, home, and vehicle equity, aggregated at the household level) is zero for both the self-employed and paid employees. Access to credit also appears

| | Sample | Percentile | | | | |
|---|---|---|---|---|---|---|
| | | p10 | p25 | p50 | p75 | p90 |
| Income ($2009) | | | | | | |
| (12-month average) | | | | | | |
| Main earner | Paid employees | 1,009 | 1,854 | 3,088 | 4,804 | 7,061 |
| | Self-employed | 312 | 994 | 2,131 | 3,953 | 7,255 |
| Other earners | Paid employees | 0 | 0 | 234 | 1,981 | 3,518 |
| | Self-employed | 0 | 0 | 0 | 1,422 | 2,994 |
| Welfare | Paid employees | 0 | 0 | 0 | 262 | 1,169 |
| | Self-employed | 0 | 0 | 0 | 264 | 1,169 |
| Household | Paid employees | 1,898 | 3,100 | 4,986 | 7,583 | 10,806 |
| | Self-employed | 1,103 | 2,117 | 3,929 | 6,892 | 11,171 |
| Wealth ($2009) | | | | | | |
| Net liquid wealth | Paid employees | -28,747 | -7,445 | 0 | 7,417 | 53,399 |
| | Self-employed | -23,799 | -5,307 | 0 | 9,021 | 71,421 |
| Unsecured debt | Paid employees | 0 | 0 | 1,621 | 9,166 | 24,470 |
| | Self-employed | 0 | 0 | 1,250 | 9,788 | 26,588 |

**Table 3:** Earnings and wealth: self-employed vs employee sample.

similar: the distribution of unsecured debt is very close in both samples. Taken together, these statistics suggest that, in terms of household finance, a large fraction of the self-employed is no more insulated from earnings risks than paid employees.

Figure 1 shows the distribution of yearly earnings growth year-on-year in each sample. Large changes in earnings are markedly more common in the self-employed sample. This pattern holds both for the earnings of the main earner (Figure 1a) and total household income (Figure 1b). Across these two measures, the self-employed appear to be exposed to more substantial earnings risks than paid employees.

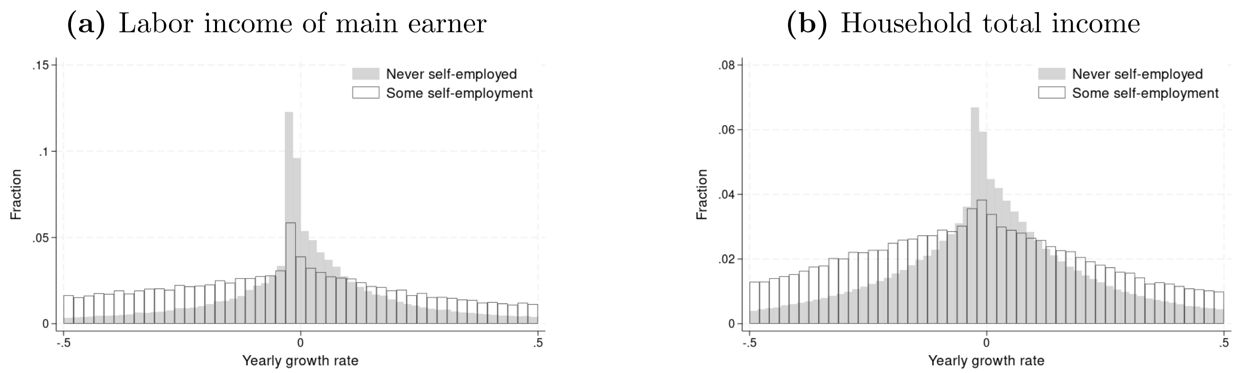



**Figure 1:** 12-month income growth: self-employed vs paid employee sample.

## 2.3 The self-employed's exposure to labor market risks

I now zoom in on the drivers of earnings changes in the self-employed sample. These empirical regularities represent some of the key moment targets used to calibrate the model introduced in Section 3.

Table 4 gives information on the monthly flows between paid-employment ($P$), self-employment ($S$), and unemployment ($U$). In the self-employed sample, workers are about equally likely to exit unemployment in paid-employment ($UP$ rate of 14 percent) or self-employment ($US$ rate of 7.5 percent). The implied aggregate transition rate from unemployment (21.2 percent) is of the same magnitude as in the paid employee samples (19.6 percent).[10]

In terms of transitions to unemployment, the self-employed do make direct transitions from self-employment to unemployment ($SU$ rate of 0.4 percent). While their chance to become unemployed after a spell as an employee is larger ($PU$ rate of 2 percent), workers in the self-employed sample are on average markedly more likely to experience a $PU$ transition than workers in the employee sample: the $PU$ rate in the employees sample is 0.7 percent. Using this figure as a benchmark, workers in self-employment then make a $SU$ transition at about half the rate of regular wage workers. Transitions to unemployment then represent an important driver of earnings risks in the self-employed sample.

[10] By definition, workers in the "paid employees" sample do not become self-employed.

| | Paid employees | Self-employed |
|---|---|---|
| Transition between emp. states $s$ (monthly rates, origin to destination) | | |
| $UP$ | 0.196 | 0.137 |
| $US$ | — | 0.075 |
| $PU$ | 0.007 | 0.020 |
| $SU$ | — | 0.004 |
| $SP$ | — | 0.011 |
| $PS$ | — | 0.021 |
| Share by labor force state | | |
| $U$ | 0.039 | 0.048 |
| $P$ | 0.961 | 0.339 |
| $S$ | — | 0.614 |

**Table 4:** Labor force states: transition rates and shares in each sample. By construction, workers in the paid-employees sample never become self-employed ($s = S$) over the duration over the survey.

There are also large direct flows (without an unemployment spell in between) between paid- and self-employment. In a typical month, there is a 2.1 percent chance that a paid-employed worker makes a direct transition to self-employment ($PS$) and a 1.1 percent chance that a self-employed worker makes a direct transition to paid-employment ($SP$).

Table 5 reports summary statistics on labor income and wealth for the self-employed sample. Within this sample, earnings show more dispersion in self-employment than in paid-employment. The tenth percentile of labor income is almost $300 lower in self-employment. Workers appear to start or to continue working at their own business with lower associated labor income than as an employee.

In addition, conditional on being continuously employed in the same labor form, earnings are substantially more volatile during a self-employment spell than during a paid-employment spell. Over a twelve-month period, for instance, there is a 25 percent chance that one's earnings drop by more than 30 percent in self-employment. The corresponding drop in

| | Percentile | | | | |
|---|---|---|---|---|---|
| | p10 | p25 | p50 | p75 | p90 |
| Labor income ($2009, monthly) | | | | | |
| Paid-employed ($P$) | 955 | 1,592 | 2,643 | 4,266 | 7,164 |
| Self-employed ($S$) | 630 | 1,263 | 2,524 | 4,894 | 8,974 |
| | | | | | |
| Labor income ($2009, 12-month average) | | | | | |
| Paid-employed ($P$) | 1,200 | 1,791 | 2,925 | 4,519 | 7,484 |
| Self-employed ($S$) | 728 | 1,511 | 2,759 | 4,996 | 8,498 |
| | | | | | |
| Labor income growth within emp. type (12-month growth) | | | | | |
| Paid-employed ($P$) | -0.352 | -0.081 | -0.010 | 0.211 | 0.745 |
| Self-employed ($S$) | -0.641 | -0.312 | -0.009 | 0.594 | 1.976 |
| | | | | | |
| Labor income growth within emp. type (12-month growth, 12-month average) | | | | | |
| Paid-employed ($P$) | -0.226 | -0.060 | 0.008 | 0.162 | 0.440 |
| Self-employed ($S$) | -0.493 | -0.222 | 0.030 | 0.441 | 1.272 |
| | | | | | |
| Net liquid wealth ($2009) | | | | | |
| Paid-employed ($P$) | -25,685 | -6,551 | 0 | 5,260 | 50,225 |
| Self-employed ($S$) | -22,854 | -4,635 | 12 | 12,769 | 88,092 |

**Table 5:** Earnings and wealth by labor force status in the self-employed sample.

earnings for workers in paid-employment is 8 percent. These findings are robust to taking averages of earnings within an employment type over several periods. Table 5 also reports the distribution of labor income and labor income changes using wages and business earnings averaged over twelve months. The larger dispersion of earnings in self-employment, both for income in levels and income growth, still holds in this case.

In terms of net liquid wealth holdings, finally, a substantial fraction of these workers do not have access to substantial liquid savings to self-insure against unemployment risk and fluctuations in income. In particular, among workers currently in self-employment, who are not eligible to UI benefits, more than fifty percent have negative liquid wealth.

**Summary** Three main points emerge from the data with regard to the self-employed's exposure to labor market risks. First, the self-employed are over-represented at the lower end of the earnings distribution. Second, there are some clear labor market risks associated with self-employment. When self-employed, workers become unemployed at a non-trivial rate, and they experience higher labor income volatility than in paid-employment. Third, many of the self-employed have limited wealth reserves that they can draw on in the face of these events. The next sections introduce a quantitative model that can account for these empirical regularities.

# 3 A search model with self-employment

I build a search model in which risk-averse workers can save and borrow to jointly capture (i) workers' transitions between paid-employment ($P$), self-employment ($S$), and unemployment ($U$), (ii) fluctuations in labor income, and (iii) a self-insurance motive. This framework relates to several studies that describe a frictional labor market where workers can move across several forms of employment, such as formal and informal employment [Meghir et al., 2015] or public sector and private sector employment [Bradley et al., 2017]. I depart from these prior studies by making agents risk-averse in an incomplete market framework.[11] Agents can self-insure against labor market risks by borrowing or drawing down their savings, as well as by relying on additional sources of income at the household level, such as spousal earnings and welfare transfers.

## 3.1 Environment

Time is discrete. The labor force is represented by a continuum of working age individuals with measure one. These individuals are the model counterpart to the self-employed defined in Section 2 (main earners and unincorporated when in self-employment). They are risk-averse and discount the future at rate $\beta \in (0, 1)$. Their per-period utility of consumption $c > 0$ is given by the utility function $u$.[12] Workers are allowed to borrow and save using a risk-free asset $a$, with rate of return $r$, subject to an exogenous borrowing limit $\underline{a} \leq 0$.

Workers can be in one of four labor market states $s$: paid-employment ($P$), self-employment ($S$), unemployed on UI benefits ($B$), or unemployed not eligible to benefits ($U$). They can search for work opportunities either as paid- or self-employed when unemployed. While employed, they can only search for opportunities in the alternative employment form.[13] Workers in paid-employment earn a wage $w$ in each period, while the self-employed earn business income $y$. In the baseline model, workers in unemployment are only eligible to UI benefits if they were previously employees. Self-employed workers terminating their business are not eligible.

Workers differ in labor market skills, which translate in permanent differences in earnings potential. I index this heterogeneity by $k \in \{1, \ldots, K\}$. A worker's skill level conditions their earnings level in the labor market. It first conditions the offer distributions from which they draw. These distributions are denoted by $F_k^P$ and $F_k^S$, respectively for workers of type $k$ drawing a wage $w$ or "self-employment income" $y$. There is no recall of past jobs or business

[11] See also Lise [2013] for a search model with savings in which workers climb a single job ladder.
[12] $u : \mathbb{R}_+^* \to \mathbb{R}$ is assumed to satisfy $u' > 0$, $u'' < 0$, and $\lim_{c\to 0} u'(c) = \infty$.
[13] $SS$ and $PP$ transitions are captured by earnings shocks.

opportunities. In addition, workers experience fluctuations in labor income while working at a job or at their business. The income processes governing these shocks, conditional on current labor income, also depend on $k$ and are given by $Q_k^P(.|w)$ and $Q_k^S(.|y)$.

A worker's skill level also conditions the destruction rate of employment opportunities. Jobs and businesses disappear, respectively, with exogenous probability $\delta_k^P$ and $\delta_k^S$. These destruction rates are indexed on ability $k$ to allow for differential exposure to unemployment risk by worker type. In addition, workers are always free to leave their current job or business, in which case they become unemployed without access to unemployment benefits (they go straight to labor market state $s = U$).

Unemployment insurance (UI) has two dimensions in the model: $b$ and $T$. $b$ is a benefit function, mapping workers' wage in their last job to some benefit level. $T \geq 1$ controls the potential duration of benefits payments, with the eligibility to benefits expiring at rate $T^{-1}$.[14] These assumptions restrict UI to belong to the "constant benefit, finite duration" class of UI policies, in line with the US system. Wages $w$ in paid-employment are assumed to be net of UI contributions, in line with the data used to calibrate the model, so I do not explicitly model the corresponding contribution rate.[15]

There are additional income sources that accrue to the household beyond the main earner's labor income and the (potentially negative) returns to wealth $ra$. There can be additional earners in the household. On top of unemployment insurance, there can also be other welfare programs to which the household is eligible. Total household income is then modeled as a tuple of functions $\left\{Y_k^s,\ s \in \{U, P, S\}\right\}$ of the main earner's labor income in the corresponding labor market state $s$ and worker type $k$. $Y_k^s$ is a reduced-form representation of the phasing-out of welfare programs as earnings increase and of added worker effects.[16]

[14]This simplification is introduced to avoid making the eligibility period to unemployment benefits a state variable in state $B$. Because agents are risk-averse in the model, this simplification underestimates the value of unemployment benefits relative to a model with a deterministic eligibility period.

[15]UI contributions are paid by employers in the US on the basis of their total taxable wage bill. As a result, the data on gross pay collected for employees in the SIPP already implicitly takes into account UI contributions.

[16]It also potentially capture a labor supply response of the main earner to a negative shock to their main source of labor income. As an example, Koustas [2018] documents that some wage workers smooth income

## 3.2 Timing

Each period $t$ unfolds as follows.

1. Earnings realization. Workers in paid- and self-employment get a new labor income draw, respectively from the conditional distributions $Q_k^P(.|w)$ and $Q_k^S(.|y)$.

2. Quits and separations. Conditional on their earnings realization in the current period, workers can decide to voluntary quit, in which case they become unemployed with no access to UI benefits ($s = U$). Otherwise, they are hit by an exogenous destruction shock with probability $\delta_k^s,\ s \in \{P, S\}$.

3. Search. Wage workers who are not separated get a draw from the distribution of self-employment opportunities $F_k^S$ at rate $\lambda^{PS}$. Similarly, self-employed workers who keep their business sample from the distribution of job offers $F_k^P$ at rate $\lambda^{SP}$.

   Workers who become unemployed do not search in the current period. Previously unemployed workers sample work opportunities in paid-employment and self-employment, respectively at rate $\lambda^{UP}$ and $\lambda^{US}$. These probabilities are assumed to be mutually exclusive, and, therefore, unemployed workers get at most one labor income draw in each period, either from $F_k^P$ or $F_k^S$. If they choose to pursue this job or business opportunity, they immediately switch to this new labor form and earn the associated wage or business income.

4. Consumption and savings. Household income accrues to all workers. UI benefits are paid out to the eligible fraction of unemployed workers. Agents then choose consumption $c$ and next period's net wealth $a'$.

shocks by taking up auxiliary jobs as rideshare drivers.

### 3.3 Worker's problem

**Notations** Let $R_k^s(a, y)$ be the present value of being in state $s$ with net wealth holdings $a$ and labor income $y$ for a worker of type $k$ at the start of the quits and separations stage. Let $V_k^s(a, y)$ stand for the worker's present value at the start of the consumption and savings stage. I denote the net gains from getting a job or business opportunity (a draw from $F_k^{s'}$) as

$$\phi_k^{ss'}(a, y) := \int \max\Big\{V_k^{s'}(a, \tilde{y}) - V_k^s(a, y), 0\Big\} dF_k^{s'}(\tilde{y}),$$

where $s$ is the worker's current state, and $s' \in \{P, S\}$, which by assumption is only defined for $s \neq s'$.

The notation $\phi_k^{ss'}(a, y)$ implicitly defines a reservation income given a worker's state variables. For instance, a worker with current assets $a$ and currently in unemployment (labor force state $s = U$) who gets a business idea takes on any business opportunity generating income at least greater than $\underline{y}_k^{US}(a)$: $V_k^S(a, \underline{y}_k^{US}(a)) = V_k^U(a)$. The net gains from getting such a business opportunity can then be written

$$\phi_k^{US}(a) = \int_{\underline{y}_k^{US}(a)}^{\overline{y}} V_k^S(a, \tilde{y}) - V_k^U(a) dF_k^S(\tilde{y}).$$

**Value functions: paid-employed** The value of holding a job with current wage $w$ and wealth $a$ at the beginning of the period is given by

$$R_k^P(a, w) = \max\left\{V_k^U(a),\ \delta_k^P V_k^B(a, w) + (1 - \delta_k^P)\Big[V_k^P(a, w) + \lambda^{PS}\phi_k^{PS}(a, w)\Big]\right\}. \quad (1)$$

The max operator corresponds to the worker's choice to remain in paid-employment given the current realization of wages. The worker's outside option is to become unemployed and start searching next period. Because only layoffs ($\delta_k^P$-shocks) entitle workers to unemployment benefits, voluntary quits result in unemployment without benefits ($V_k^U(a)$).

The value of being in paid-employment with wage $w$ and wealth $a$ at the consumption and savings stage writes

$$\begin{aligned} V_k^P(a,w) = \max_{c,\tilde{a}} & \left\{ u(c) + \beta \int R_k^P(\tilde{a},\tilde{w}) dQ_k^P(\tilde{w}|w) \right\} \\ \text{s.t.} \quad & c + \frac{\tilde{a}}{1+r} = Y_k^P(w) + a; \\ & \tilde{a} \geq \underline{a}. \end{aligned} \tag{2}$$

The budget constraint in the employee's consumption-saving problem (2) does not explicitly model UI insurance contribution because the wage $w$ is assumed to be net of UI contributions.

**Value functions: self-employed** The value of a business at the beginning of the period mirrors that of job holders in equation (1), with the difference that unemployment transitions are all to the no-benefits state ($U$):

$$R_k^S(a,y) = \max \left\{ V_k^U(a),\ \delta_k^S V_k^U(a) + (1-\delta_k^S) \Big[ V_k^S(a,y) + \lambda^{SP} \phi_k^{SP}(a,y) \Big] \right\}. \tag{3}$$

Similarly to the case of paid-employment ($P$), workers can decide to terminate their business given the current realization of income and become unemployed (first term in the max operator). Otherwise, they continue with their current business and are subject to exogenous destruction shocks with probability $\delta_k^S < 1$.

The self-employed's savings problem is given by

$$\begin{aligned} V_k^S(a,y) = \max_{c,\tilde{a}} & \left\{ u(c+\kappa) + \beta \int R_k^S(\tilde{a},\tilde{y}) dQ_k^S(\tilde{y}|y) \right\} \\ \text{s.t.} \quad & c + \frac{\tilde{a}}{1+r} = Y_k^S(y) + a; \\ & \tilde{a} \geq \underline{a}. \end{aligned} \tag{4}$$

The key difference with the savings problem of the paid-employed is that self-employment comes with non-pecuniary benefits, which are modelled as a parameter $\kappa \geq 0$ in units of consumption. [see, for example, Humphries, 2017, Catherine, 2022]. This additive formulation $u(c+\kappa)$ is commonly used in the self-employment and entrepreneurship literature.[17] The non-pecuniary preference parameter captures a preference for being self-employed, which could stem from the intrinsic value of being one's own boss, from unreported earnings, or from other tax advantages.[18] Empirically, this feature is motivated in the data by the non-trivial growth in earnings observed following a direct $SP$ transition, which is interpreted as the compensating differential of self-employment. Non-pecuniary benefits can potentially lead workers to turn down a job offering a better pay than what they get from their current business.

**Value functions: unemployed** The probabilities of getting a job opportunity ($\lambda_{UP}$) or business opportunity ($\lambda_{US}$) are assumed to be mutually exclusive. The values of searching for employment opportunities are then given by

$$R_k^B(a,w) = V_k^B(a,w) + \lambda^{UP}\phi_k^{BP}(a,w) + \lambda^{US}\phi_k^{BS}(a,w), \tag{5}$$

and

$$R_k^U(a) = V_k^U(a) + \lambda^{UP}\phi_k^{UP}(a) + \lambda^{US}\phi_k^{US}(a), \tag{6}$$

respectively for workers eligible (state $s = B$) and not eligible (state $s = U$) to unemployment benefits. In equation (5), $w$ denotes the individual's last wage before becoming unemployed, which is part of the state-space as it determines their UI benefit income.

Equations (5) and (6) imply an indirect comparison between work opportunities in paid-

[17]Most structural models of self-employment similarly allow for some degree of preference for business ownership [Humphries, 2017, Catherine, 2022, Jones and Pratap, 2020]. Humphries [2017] and Catherine [2022] use a similar additive formulation.

[18]Flexibility in one's schedule is an example of the non-pay value of not being an employee. For instance, Lim [2019] presents evidence that young mothers are more likely to be self-employed and that self-employment is associated with additional flexibility in their work location and schedule. Private consumption through one's business is one example of such a tax advantage.

and self-employment. Consider a currently unemployed worker with a potential job $w$ in hand (a draw from $F_k^P$ which occurs with chance $\lambda^{UP}$). This worker decides whether to wait for better options either working for others ($P$) or for themselves ($S$). This decision is captured by the functions $\phi_k^{BP}(a,w)$ (or $\phi_k^{UP}(a)$ if the worker is no longer eligible to UI benefits), which denote the net gains from a draw from $F_k^P$.

The consumption-savings problem for unemployed workers on benefits is given by

$$\begin{aligned} V_k^B(a,w) &= \max_{c,\tilde{a}} \Big\{ u(c) + \beta \big[ \delta^B R_k^U(\tilde{a}) + \left(1-\delta^B\right) R_k^B(\tilde{a},w) \big] \Big\} \\ \text{s.t.} \quad & c + \frac{\tilde{a}}{1+r} = Y_k^U + a + b(w); \\ & \tilde{a} \geq \underline{a}, \end{aligned} \tag{7}$$

where the term in squared brackets after the discount factor comes from benefits expiring at rate $\delta^B := T^{-1}$. Similarly to equation (5), $w$ is part of the state-space because benefits are indexed on the last wage received by the worker.

The consumption-savings problem for unemployed workers not eligible to benefits (state $s = U$) follows directly from adapting equation (7) and is given by

$$\begin{aligned} V_k^U(a) &= \max_{c,\tilde{a}} \Big\{ u(c) + \beta R_k^U(\tilde{a}) \Big\} \\ \text{s.t.} \quad & c + \frac{\tilde{a}}{1+r} = Y_k^U + a; \\ & \tilde{a} \geq \underline{a}. \end{aligned} \tag{8}$$

## 3.4 Stationary equilibrium

Taken together, the optimal choices of consumption and savings, as well as the reservation strategies implied by workers' decisions to quit and to follow through on alternative employment opportunities imply a stationary distribution over employment states $s \in \{U, B, S, P\}$, labor income $(y, w)$, and net wealth $(a)$ for each skill level $k$. These distributions are denoted $\Gamma_k^s$ in what follows. A formal statement of this definition is given in Appendix B.

### 3.5 Discussion

I conclude the exposition of the model by briefly discussing some of my modeling choices. First, I do not model investment and the associated accumulation of business capital. Hurst and Lusardi [2004] argue that liquidity constraints are unlikely to be binding for a majority of entrepreneurs in the US, which would imply either that rental markets for capital function well or that these entrepreneurs have limited capital requirements. Given my focus on the sample of unincorporated self-employed, whose median business assets are $3,444 (see Table 1), it seems plausible that liquidity constraints are limited for many self-employed. Besides, the arrival rate parameters $\lambda_{US}$ and $\lambda_{PS}$ represent frictions that prevent workers from immediately starting their own business, which can in part be interpreted as the time required to secure financing. For the subset of the self-employed operating on a larger scale, a more general model would consider how these workers allocate their portfolio between liquid assets, their business, and other illiquid assets, given the labor market risks associated with this employment form.

Second, I abstract from general equilibrium effects. Krusell et al. [2010] find that large welfare gains can be achieved by lowering unemployment benefits in their general equilibrium search model with precautionary savings, and that these welfare gains in part go through general equilibrium channels (higher capital stock, job creation, higher wages). These effects are not present in my framework. The primary goal of the model introduced here is to capture the labor market risks and precautionary savings behavior of the self-employed as defined in Section 2. Given the self-employed account for less than eight percent of the sample of main earners, the partial equilibrium counterfactuals considered below seem a reasonable first step to study the welfare effects of policies targeted at the labor market risks associated with self-employment.

Lastly, I model worker skills $k$ as uni-dimensional. As such, I implicitly assume that these skills are transferable across employment forms. By focusing on the self-employed sample in the calibration, I ensure that workers are skilled enough in self-employment that it becomes their main source of labor income at some point. Because many self-employed in the data are not observed working as employees, convincingly recovering a joint distribution of skills in both labor forms is not feasible. The life cycle income profiles reported in Humphries [2017] for various groups of self-employed suggest that experience as a wage worker is, to a large extent, transferable to self-employment, conditional on making the transition.

# 4 Calibration

My calibration strategy proceeds in two steps. Worker unobserved heterogeneity in skills is first introduced by grouping workers with similar labor market outcomes. In a second step, I target key moments from the data to calibrate the remaining parameters, conditional on the heterogeneity uncovered in the first step. The underlying data is always from the self-employed sample (main earners (i) with at least some experience of self-employment and (ii) always unincorporated when in self-employment).

## 4.1 Worker heterogeneity

To discipline worker heterogeneity in the model (the skill index $k$ in the notation introduced in Section 3), I rely on a clustering tool from machine learning to partition workers into distinct groups.[19] The logic behind this step is that, through the lens of the model, this unobserved skill heterogeneity translates into differences in the distribution of labor income and employment rate at the worker level. The evidence on earnings shown in Section 2 suggests that these differences are substantial in the self-employed sample (see Table 3). A key advantage of clustering over conditioning on standard observables, such as education categories, is to be agnostic about which covariate best capture this unobserved heterogeneity in the self-employed sample.

[19] Thereafter, I use the terms "class," "group," "type," and "cluster" interchangeably to designate the outcome of the clustering procedure.

**Clustering algorithm** I use a k-means algorithm to create groups of workers with similar skills in the self-employed sample [Bonhomme et al., 2019, 2022]. This procedure finds the best partition of the data implied by the objective function

$$\underset{\tilde{h},k_1,...,k_N}{\arg\min} \sum_{i=1}^{N} \left\| \hat{h}_i - \tilde{h}(k_i) \right\|^2, \tag{9}$$

where $i \in \{1, ..., N\}$ denotes individuals, $k_i \in \{1, ..., K\}$ indicates the group to which individual $i$ is assigned, with $1 < K \leq N$, and $h$ is a vector of variables, with $h_i$ and $h(k_i)$ denoting the vectors for individual $i$ and group $k_i$ respectively. Each element in $\tilde{h}(k_i)$ is computed by averaging over the members of the group. The solution to the minimization problem (9) then assigns a cluster to each $i$ such that the squared Euclidean distance between $i$'s vector of characteristics and the average of these characteristics in $i$'s group is minimized.[20]

The rationale behind this clustering step is conceptually the same as the strategy described in Bonhomme et al. [2019] in the context of a model with two-sided unobserved heterogeneity. These authors first cluster employers based on their empirical distribution of earnings amongst employed workers before using the resulting groups as inputs in a series of mixture models. The partition of firms they obtain is based on an outcome variable (firm-level earnings), and it therefore captures both observed and unobserved firm heterogeneity. I adopt a similar strategy to discipline unobserved worker heterogeneity in the model. Here, this partition translates into group-specific distributions of labor income in paid- and self-employment $\left\{Q_k^P, F_k^P, Q_k^S, F_k^S\right\}$ and unemployment shocks $\left\{\delta_k^P, \delta_k^S\right\}$.

**Implementation** In practice, one needs to choose the vector of variables on which the classification operates $\hat{h}_i$ and the number of clusters $K$. I include two key outcomes in the vector $\hat{h}_i$: a measure of labor income when employed and a measure of exposure to unemployment. The measure of labor income is the empirical cumulative distribution function

[20] Standard algorithms to efficiently solve this global minimization problem are readily available in standard packages. I use the Matlab implementation of the "K-means++" procedure [Arthur and Vassilvitskii, 2007].

(cdf)

$$\text{ECDF}_i(y_d) := \frac{1}{\sum_t \mathbb{1}\{\text{Employment}_{it} = 1\}} \sum_t \mathbb{1}\{y_{it} \leq y_d\}, \tag{10}$$

with $y_{it}$ denoting labor income for individual $i$ in month $t$ and $\text{Employment}_{it}$ the corresponding indicator for employment. This distribution is based on all earnings, irrespective of whether they come from paid- or self-employment at a point in time.[21] It is computed for each decile $y_d$ of the empirical distribution of labor income in the sample (the black markers in the right panel of Figure 2). The measure of exposure to unemployment is the fraction of time workers spend in non-employment over the duration of the survey. Letting $T_i$ be the number of months the individual is observed in the panel (typically slightly under three years in the 1996-2008 SIPP panels), it is given by

$$\text{Unemployment Exposure}_i := \frac{1}{T_i} \sum_t \mathbb{1}\{\text{Employment}_{it} = 0\}.$$

The vector of clustering variables is then given by

$$\hat{h}_i = \big(\text{Unemployment Exposure}_i, \text{ECDF}_i(y_{p10}), \ldots, \text{ECDF}_i(y_{p90})\big),$$

where $y_{p10}, \ldots, y_{p90}$ denote earnings decile for the self-employed sample as a whole.

For the number of clusters, I set $K = 4$ taking into account three considerations: how much of the sample variation (9) is explained at a given $K$, computational burden, and ease of exposition. Intuitively, more of the variation is explained with a larger number of groups $K$, but this also entails more group-specific parameters to calibrate. In Appendix C, I show that $K = 4$ captures a reasonable degree of heterogeneity of the variables in $\hat{h}_i$.

[21] I do not further distinguish between labor income from paid- versus self-employment since many workers are never observed in paid-employment in my sample.

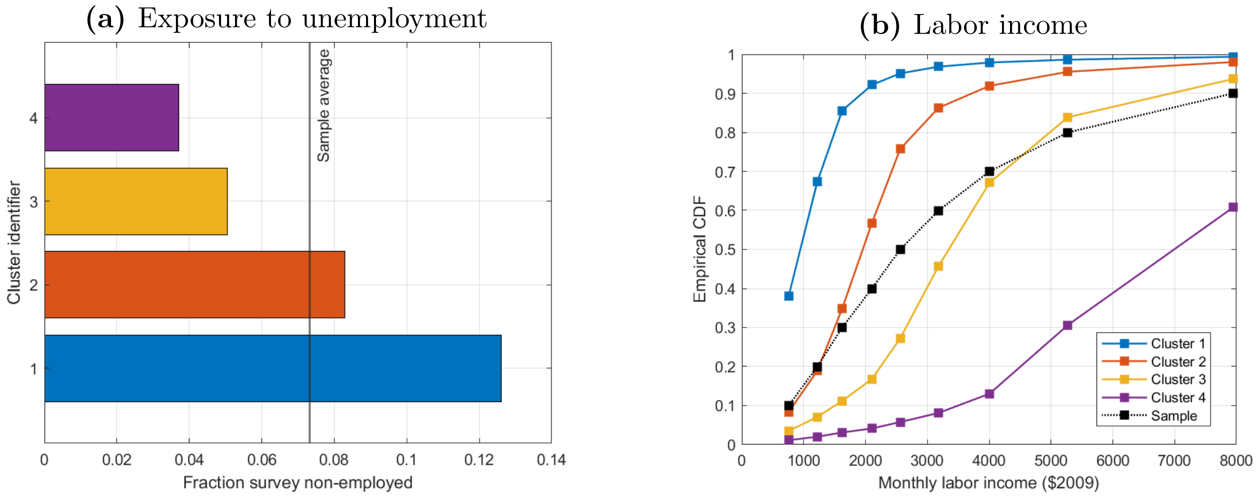



**Figure 2:** Worker clusters obtained from the k-means algorithm with $K = 4$. The figure shows the average of the clustering variables in each group (the elements of the vector $\tilde{h}(k)$ in Equation 9). The clusters are ordered based on the average value of the exposure to non-employment variable in each group. The black lines denote the value for the self-employed sample as a whole.

**Results** The resulting centers in each worker group $\tilde{h}(k_i)$ are shown in Figure 2. The figure shows a clear relationship between the two measures used to build the clusters: the group with the largest earnings is also least likely to be unemployed. This negative relationship between earnings when employed and unemployment risk holds for all worker groups.

I subsequently label the four clusters of workers based on the median earnings in each group. As shown in Table 6, the "Low", "Med-Low", "Med-High", and "High" groups correspond to clusters with a median monthly labor income of, respectively, $992, $1,922, $3,389, and $6,996.

Table 6 further shows how worker clusters differ along several observables not directly used in the clustering procedure. In terms of demographic composition, workers with lower earnings are more likely to be women and belong to a minority, and are less likely to be married and have a college education. (Recall that the sample is restricted to the main earner in each household.) I also find that the average age is very similar across worker clusters.[22] The resulting worker clusters also relate to the large variation in the scale of the

[22]Figure D.1 shows that the pseudo-cohort age profile of earnings is relatively flat within cluster, except for

corresponding businesses in terms of business wealth and number of employees. For example, median net business wealth is \$9,000 in the group of highest earners and close to zero in the group with the lowest earnings.

The worker types resulting from the algorithm therefore capture meaningful observed characteristics, as well as potentially unobserved ability traits determining the labor market outcomes of the self-employed. While Table 6 shows that some characteristics are more common in some groups, standard cuts of the data, such as education, would not capture as much of the relevant heterogeneity. For instance, while a post-graduate qualification is a good predictor of being in the "High" worker group, more than 30 percent of workers in this group do not hold even a college degree.[23] Because it is unclear which is the most relevant cut of the data, the clustering approach described here allows to capture a lot of heterogeneity while keeping the model parsimonious.

## 4.2 Model parametrization

**Utility and labor income distributions** Utility from consumption is assumed to be given by $u(c) = \ln(c)$. The income process governing the shocks within each labor form, $Q_k^s(.|y)$ for $s \in \{P, S\}$, is assumed to follow

$$\ln y_{it}^s = (1 - \rho^s)\mu_k^s + \rho^s \ln y_{it-1}^s + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^s), \tag{11}$$

and therefore the long-run mean implied by the process is allowed to be specific to each labor form $s$ and worker group $k$ through the parameter $\mu_k^s$. The distribution of labor income draws $F_k^s$ is set to the stationary distribution implied by the income process (11).[24]

the "High" group. I also report the same age profile by college graduation status in Figure D.2. The profiles in Figure D.2 also suggest that, while earnings are increasing with age, especially across college graduates, there is substantial permanent heterogeneity in earnings in the self-employed sample.

[23] Similarly, I also find that industries partially correlate with a worker's assigned cluster. See Appendix Figure D.3.

[24] In practice, the income process (11) is discretized, and its implied stationary distribution is derived numerically.

| | Worker earnings group | | | |
|---|---|---|---|---|
| | Low | Med-Low | Med-High | High |
| Median earnings ($2009) | 992 | 1,922 | 3,389 | 6,996 |
| Demographic characteristics | | | | |
| Age | 41 | 39 | 40 | 42 |
| Gender (woman=1) | 0.473 | 0.269 | 0.217 | 0.197 |
| Married (married=1) | 0.375 | 0.472 | 0.631 | 0.699 |
| Race (non-white=1) | 0.162 | 0.150 | 0.111 | 0.094 |
| College Graduate | 0.279 | 0.337 | 0.445 | 0.694 |
| Post-graduate | 0.045 | 0.043 | 0.081 | 0.304 |
| Business characteristics | | | | |
| Median net business wealth ($2009) | 18 | 1,136 | 5,343 | 9,084 |
| More than 25 employees | 0.008 | 0.008 | 0.014 | 0.032 |
| Number of workers ($N_k$) | 1,812 | 2,504 | 2,613 | 1,694 |

**Table 6:** Characteristics of worker clusters.

**UI benefits** The actual unemployment insurance system in the US is determined at the state level. I follow Chetty [2008] and Saporta-Eksten [2014] and approximate these policies as a fifty percent replacement rate for a maximum duration of six months. I also cap the maximum monthly benefit payments at \$2,000.[25] In the notation of the model, the UI system is then given by the replacement rate function $b(w) = \max\left\{2,000; 0.5 \cdot w\right\}$ and $T = 6$. The relevant model wage $w$ is assumed to be the last wage realization before the worker becomes unemployed. To reiterate, I do not model UI taxes because the wage $w$ is net of UI contributions in the data.

**Household income functions** I obtain a mapping between an individual's current labor income and their household income by estimating the following regressions

$$\ln Y_{it}^s = \alpha_k^s + \beta_k^s \cdot \ln y_{it}^s + \epsilon_{it}, \quad k = 1, \ldots, K, \quad s \in \{P, S, U\}, \tag{12}$$

where $Y_{it}^s$ is total household income, including other earners and welfare transfers, but net of UI payments, $\ln y_{it}^s$ is the main earner's labor income, and $\epsilon_{it}$ is an error term. The index $k$ refers to the worker's assigned cluster, and $s$ denotes their current state in the labor market. (The slope $\beta_k^s$ is omitted for $s = U$, since labor income is zero in this case by definition.) The estimated coefficients $\left(\widehat{\alpha}_k^s, \widehat{\beta}_k^s\right)$ give the corresponding household income function by cluster and labor force status.[26] This functional form assumption allows to capture some degree of household heterogeneity in total household income by worker cluster without introducing additional state variables in the model. For example, by running the regressions (12) separately by $k$, this specification allows for assortative mating, as spousal income is allowed to depend on the skill level of the household's main earner. This heterogeneity in income at the household level represents another difference in the scope for insurance.

[25] This further approximates the caps on UI benefit payments enacted by most states. See the "Significant Provisions Of State Unemployment Insurance Laws" tables from the Department of Labor for the corresponding weekly ceilings by state.

[26] These coefficients are reported in Table D.1.

## 4.3 Calibration procedure

I set the returns on net liquid wealth $r$ to zero. The lower bound on net liquid wealth is allowed to differ by worker group. I set $\underline{a}_k$ to the observed 10th percentile of the net liquid wealth distribution in each worker group $k$. In the data, this percentile decreases monotonically by group from around $-\$6,000$ for the group of Low Earners to $-\$12,000$ for the group of High Earners. The borrowing constraint is therefore tighter at the margin for households with lower earnings in the calibration.

The remaining parameters are calibrated by matching a set of moments from the data to their counterpart in the simulated model. Starting from the invariant distribution of workers in each group $\Gamma_k$, I simulate a panel of $N_k$ workers (the number of observations in each cluster $k$ in the data) for 48 months. I stack all cluster-specific panels and obtain the moments simulated from the model using the exact same definitions as those used to construct the moments derived from the data.

Though all parameters are jointly calibrated, I motivate my choice of moment targets by mapping each set of calibrated parameters to specific moments. The job and business arrival rates $\left(\lambda^{US}, \lambda^{UP}, \lambda^{PS}, \lambda^{SP}\right)$ are chosen to replicate the corresponding monthly transition rates. The job and business destruction rates $\left(\delta_k^P, \delta_k^S\right)$ directly relate to the $k$-specific transition rates into unemployment, respectively from paid- and self-employment. The parameters governing the long-run mean $\left(\mu_k^P, \mu_k^S\right)$ in the income process (11) are set to replicate the average of log-earnings by worker group and labor form. The discount factor $\beta$ is set to match the median net liquid wealth in the data. The dispersion parameters $\left(\sigma^P, \sigma^S\right)$ in the income process (11) are informed by the standard deviation of log-earnings in paid- and self-employment, respectively, while the persistence parameters $\left(\rho^P, \rho^S\right)$ are disciplined by the twelve-month autocorrelation of log-income for workers continuously employed. Finally, the non-pecuniary benefits of self-employment $\kappa$ are disciplined by the gap in median labor income growth following a direct $PS$ and $SP$ transition, where labor income growth is computed only for transitions with at least twelve months of continuous employment history on either side of the transition.

The classification of workers is implemented by relating group-specific parameters to group-specific targets. In line with my choice of clustering variables (unemployment risk and earnings), I make the job and business destruction parameters and long-run mean parameters in (11) group-specific.[27] I stress that, while there is a tight link between these parameters and the corresponding moments, they still need to be calibrated jointly due to the endogenous selection across $P$, $S$, and $U$, implied by the model. As an example, the simulated distribution of labor income in self-employment is shaped by workers' reservation income strategy with respect to paid-employment and unemployment given their current wealth level.

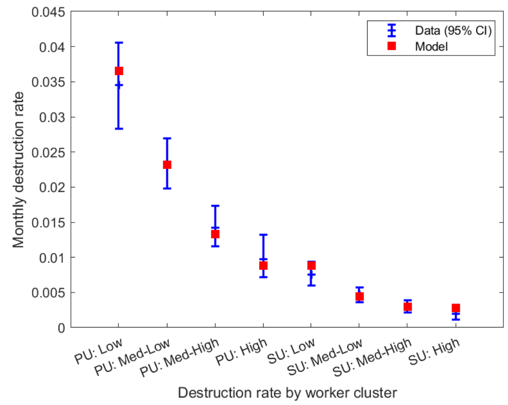
## 4.4 Results

I report the model fit to the targeted moments in Figure 3 and the calibrated parameters in Table 7. I start by describing the fit to the group-specific moments. Both for $PU$ and $SU$ transitions, there is a clear decreasing pattern where worker groups with larger labor earnings, defined by median earnings in their group, are much less likely to transition to unemployment (Figure 3a). For example, there is close to a one percent chance for the lowest earners to make a $SU$ transition in a given month, while the highest earners face a 0.2 percent chance of a similar event. Converserly, there is a clear increasing pattern in workers' average log-earnings both in paid- and self-employment (Figure 3b). Average log-earnings in self-employment are slightly lower than in paid-employment across worker groups, but this gap is small compared to the difference between groups. With group-specific parameters, the model very closely replicates these two dimensions of worker heterogeneity.

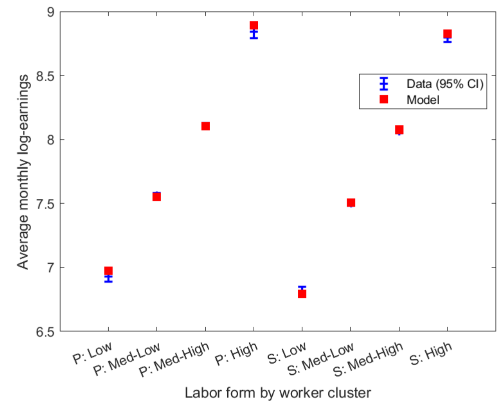
The model can reproduce the flows across unemployment, paid-employment, and self-employment observed in the data (Figure 3c). In terms of the arrival rates of job and business opportunities for employed workers ($\lambda^{SP}$ and $\lambda^{PS}$), the model sees paid-employment

[27] The $UP$ and $US$ transition rates, which also relate to unemployment, do not exhibit a clear pattern by worker group in the data, so the corresponding arrival rates are common to all workers.

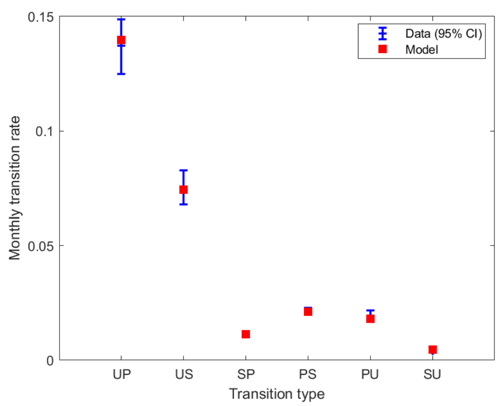
(a) Destruction rates by worker group



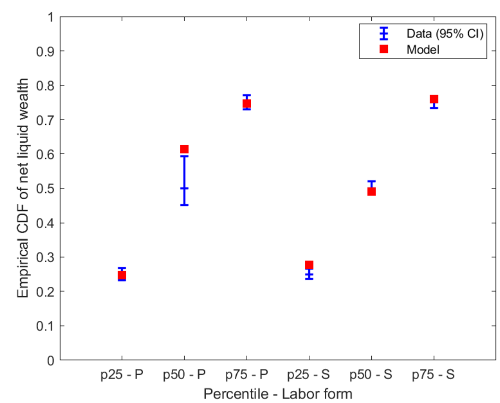

(b) Average log-earnings by worker group



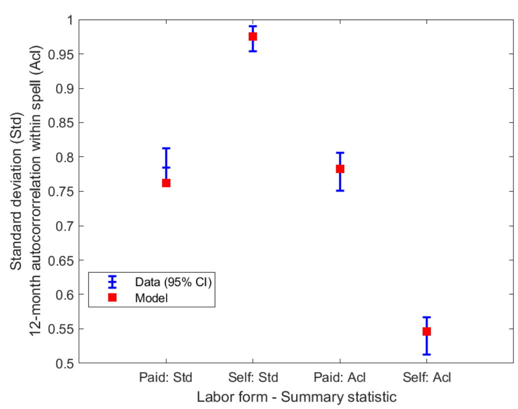

(c) Transition rates




(d) Distribution of net liquid wealth




(e) Log-earnings dispersion and persistence




(f) Log-earnings change after transition

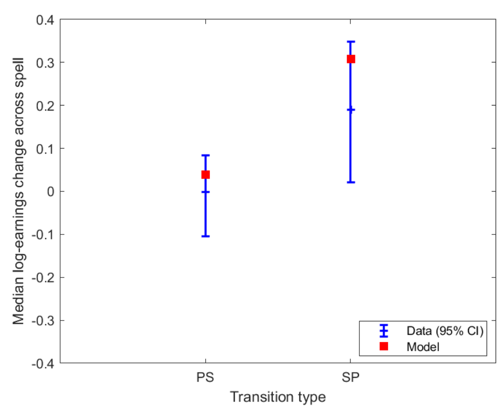



**Figure 3:** Model fit to targeted moments. 95 percent bootstrap confidence interval based on 500 replications. See main text for variable definitions.

| Parameter | Description | Value | Main target | Fit |
|---|---|---|---|---|
| **Group-specific parameters** | | | | |
| $\delta^P_{\text{Low}}$ | Exog. destruction rate from $P$ | 0.035 | $\widehat{\mathrm{E}}(PU_{it} \vert s_{it-1} = U, k = \text{Low})$ | Fig. 3a |
| $\delta^P_{\text{Med-Low}}$ | | 0.022 | $\widehat{\mathrm{E}}(PU_{it} \vert s_{it-1} = U, k = \text{Med-Low})$ | |
| $\delta^P_{\text{Med-High}}$ | | 0.014 | $\widehat{\mathrm{E}}(PU_{it} \vert s_{it-1} = U, k = \text{Med-High})$ | |
| $\delta^P_{\text{High}}$ | | 0.010 | $\widehat{\mathrm{E}}(PU_{it} \vert s_{it-1} = U, k = \text{High})$ | |
| $\delta^S_{\text{Low}}$ | Exog. destruction rate from $S$ | 0.008 | $\widehat{\mathrm{E}}(SU_{it} \vert s_{it-1} = U, k = \text{Low})$ | Fig. 3a |
| $\delta^S_{\text{Med-Low}}$ | | 0.002 | $\widehat{\mathrm{E}}(SU_{it} \vert s_{it-1} = U, k = \text{Med-Low})$ | |
| $\delta^S_{\text{Med-High}}$ | | 0.003 | $\widehat{\mathrm{E}}(SU_{it} \vert s_{it-1} = U, k = \text{Med-High})$ | |
| $\delta^S_{\text{High}}$ | | 0.002 | $\widehat{\mathrm{E}}(SU_{it} \vert s_{it-1} = U, k = \text{High})$ | |
| $\mu^P_{\text{Low}}$ | Mean of income shocks in $P$ | 6.894 | $\widehat{\mathrm{E}}(\ln w_{it} \vert s_{it} = P, k = \text{Low})$ | Fig. 3b |
| $\mu^P_{\text{Med-Low}}$ | | 7.491 | $\widehat{\mathrm{E}}(\ln w_{it} \vert s_{it} = P, k = \text{Med-Low})$ | |
| $\mu^P_{\text{Med-High}}$ | | 8.029 | $\widehat{\mathrm{E}}(\ln w_{it} \vert s_{it} = P, k = \text{Med-High})$ | |
| $\mu^P_{\text{High}}$ | | 8.836 | $\widehat{\mathrm{E}}(\ln w_{it} \vert s_{it} = P, k = \text{High})$ | |
| $\mu^S_{\text{Low}}$ | Mean of income shocks in $S$ | 6.735 | $\widehat{\mathrm{E}}(\ln y_{it} \vert s_{it} = S, k = \text{Low})$ | Fig. 3b |
| $\mu^S_{\text{Med-Low}}$ | | 7.363 | $\widehat{\mathrm{E}}(\ln y_{it} \vert s_{it} = S, k = \text{Med-Low})$ | |
| $\mu^S_{\text{Med-High}}$ | | 8.016 | $\widehat{\mathrm{E}}(\ln y_{it} \vert s_{it} = S, k = \text{Med-High})$ | |
| $\mu^S_{\text{High}}$ | | 8.708 | $\widehat{\mathrm{E}}(\ln y_{it} \vert s_{it} = S, k = \text{High})$ | |
| **Common parameters** | | | | |
| $\beta$ | Discount factor | 0.988 | Distribution of net liquid wealth | Fig. 3d |
| $\kappa$ | Non-pecuniary benefits in $S$ | 192.6 | Earnings growth after $SP_{it}$ or $PS_{it}$ | Fig. 3f |
| $\lambda_{UP}$ | Job offer arrival rate in $U$ | 0.141 | $\widehat{\mathrm{E}}(UP_{it} \vert s_{it-1} = U)$ | Fig. 3c |
| $\lambda_{US}$ | Business opp. arrival rate in $U$ | 0.076 | $\widehat{\mathrm{E}}(US_{it} \vert s_{it-1} = U)$ | Fig. 3c |
| $\lambda_{SP}$ | Job offer arrival rate in $S$ | 0.039 | $\widehat{\mathrm{E}}(SP_{it} \vert s_{it-1} = S)$ | Fig. 3c |
| $\lambda_{PS}$ | Business opp. arrival rate in $P$ | 0.034 | $\widehat{\mathrm{E}}(PS_{it} \vert s_{it-1} = P)$ | Fig. 3c |
| $\sigma^P$ | Dispersion of income shocks | 0.184 | $\widehat{\mathrm{std}}(\ln w_{it} \vert s_{it} = P)$ | Fig. 3e |
| $\rho^P$ | Persistence of income shocks | 0.896 | $\widehat{\mathrm{corr}}(\ln w_{it}, \ln w_{it-12} \vert s_{it}, \ldots, s_{it-12} = P)$ | Fig. 3e |
| $\sigma^S$ | Dispersion of income shocks | 0.370 | $\widehat{\mathrm{std}}(\ln y_{it} \vert s_{it} = S)$ | Fig. 3e |
| $\rho^S$ | Persistence of income shocks | 0.861 | $\widehat{\mathrm{corr}}(\ln y_{it}, \ln y_{it-12} \vert s_{it}, \ldots, s_{it-12} = S)$ | Fig. 3e |

**Table 7:** Calibrated model parameters. The "main target" column gives the key moment target for each parameter. The "fit" column lists the corresponding figure. See main text for details on parameter and moment definitions.

opportunities as relatively common for self-employed workers (a four percent chance in every month), though many of these opportunities are turned down (the actual transition rate is slightly above one percent). Business opportunities for employess are slightly less frequent (a 3.5 percent chance every month) and also more likely to be accepted (the actual transition rate is 2.1 percent). The arrival rates of work opportunities in unemployment ($\lambda_{UP}$ and $\lambda_{US}$) are very close to their respective transition rate in the data, and therefore workers turn down few of these opportunities at the calibrated parameters.

The model also replicates well the distribution of net liquid wealth, with the exception of the fraction of workers below the 50th percentile in paid-employment, which it tends to slightly over-estimate (Figure 3d). Despite assuming a low degree of risk aversion (log-utility) and despite the degree of insurance available to agents through additional welfare programs and the income of other household members in unemployment (captured by $Y_k^U$ in the model), the yearly discount rate required to match the fraction of workers with low wealth levels is high (approximately 15 percent). The gap between the discount rate and the interest rate is implicitly the cost of using savings as insurance, and the gap required to match median net liquid wealth holdings in the data is large. Lentz [2009] notes that workers tend to value unemployment benefits more when this gap is larger in his structural model estimated on Danish data. A similar argument therefore suggests that additional social insurance is potentially quite valuable for some groups of self-employed.

The model also reproduces the dispersion and persistence of log-earnings during a continuous paid- and self-employment spell (Figure 3e). Earnings in self-employment are markedly more volatile, as documented in Section 2. As a result, the calibrated dispersion of earnings shocks in self-employment ($\sigma^S$) is almost twice as large as the equivalent parameter for workers in paid-employment ($\sigma^P$).

Finally, the median growth rate of earnings following an employment-to-employment transition differs starkly between $PS$ and $SP$ transitions. It is close to zero for $PS$ transitions, but above 15 percent for $SP$ transitions. The model tends to overshoot these targets, but they remain well within the confidence interval. The non-pecuniary benefits of self-employment (the parameter $\kappa$) required to replicate this asymmetry is in the order of \$200 a month. Despite these benefits, the decision to move from paid-employment to self-employment (a $PS$ transition) is not necessarily associated with negative median earnings growth in the model. This is because this decision takes into account the difference in earnings risk between these two labor forms, which differ across labor forms through the exogenous destruction shocks $\delta_k^s$ and the earnings process (11). The optimal decision weighing these risks is summarized by the earnings threshold $\underline{y}_k^{PS}(a, w)$.

# 5 Policy analysis

What are the labor market and welfare effects of providing benefits targeted at the labor market risks associated with self-employment? This section assesses the impact of extending UI to the self-employed through the lens of the calibrated model. I emphasize the labor market and welfare implications of this policy for heterogeneous groups of self-employed workers.

## 5.1 UI for the self-employed

I use a similar formulation as the one in the baseline model. UI for the self-employed is given by $b(y) = \max\{0.5 \cdot y;\ 2,000\}$: benefits are paid at a 50 percent replacement rate of workers' last income capped at \$2,000. Benefits expire at rate $T^{-1}$ with $T = 6$, so the average benefit duration is six months. These benefits are financed by taxing the labor income of workers in self-employment according to the tax schedule $\tau^S$. Introducing some notation, $\tau^S$ must satisfy

$$\sum_k \omega_k \left[ \int \tau^S(y) d\Gamma_k^S(a, y) - \int b(y) d\Gamma_k^C(a, y) \right] = 0, \tag{13}$$

where state $s = C$ denotes previously self-employed workers currently on benefits (the counterpart to state $s = B$), $\Gamma_k^s$ denotes the measure of workers in labor force state $s$ and worker class $k$, and $\omega_k := N_k/N$ is the share of workers in skill group $k$.

Equation (13) imposes that the program is balanced overall, but it does not imply that it is actuarially fair within each group $k$. To better understand the role of redistribution across worker types, I also consider a scenario where UI benefits are financed separately within groups. In this case, the parameters of the policy must satisfy

$$\int \tau_k^S(y)d\Gamma_k^S(a,y) - \int b_k(y)d\Gamma_k^C(a,y) = 0, \qquad \forall k \tag{14}$$

where the tax schedule $\tau_k^s$ and benefit schedule $b_k$ are allowed to be specific to each group. This policy scenario should be seen as hypothetical in that it requires the tax authority to know workers' unobserved types.

In the case of UI after a self-employment spell, I assume that, contrary to paid-employment, the self-employed can choose to terminate their activity and still be eligible for unemployment benefits. The distinction between layoffs and voluntary quits, which conditions eligibility in many UI systems, is irrelevant for the self-employed. Formally, the value of self-employment at the beginning of the search stage (Equation 3 in the baseline model) is now given by

$$R_k^S(a,y) = \max\left\{V_k^C(a,y),\ \delta_k^S V_k^C(a,y) + (1-\delta_k^S)\Big[V_k^S(a,y) + \lambda^{SP}\phi_k^{SP}(a,y)\Big]\right\}, \tag{15}$$

where $V_k^C(a,y)$ is the value of becoming unemployed when benefits are paid out. There are then two ways for the self-employed to cash UI benefits: either following a $\delta_k^S$-shock (involuntary) or by "choosing" to become unemployed (voluntary).

This formulation abstracts from some of the challenges associated with designing a UI program specifically for the self-employed. Many of the self-employed are likely to run fairly informal businesses, even more so in the unincorporated self-employed sample, and such a program would require many self-employed to make their business more formal in legal terms. A more formal business entity would form the basis to both collect UI contributions (a role taken up by the employer for wage employees) and record business closures (the condition to claim benefits).[28] Keeping track more formally of workers' employment history in self-employment would further allow to introduce specific experience ratings and search requirements for these workers.

### 5.2 Contributions and benefits

I consider three alternative UI policy experiments. All three maintain the same baseline benefit level $b(y) = \max\{0.5 \cdot y;\ 2,000\}$ as workers in paid-employment but vary the contribution schedule. The first experiment (UI-A) aligns the contribution schedule to the actual US system for wage employees. Specifically, the tax schedule to finance UI is set to $\tau^S(y) = \min\{\tau \cdot y;\ 7,000/12\}$. In practice, all states cap the annual earnings subject to UI contributions.[29] Because UI contributions are infra-marginal for all but the lowest earners in such a system, I also consider a second, more progressive, schedule (UI-P) where UI contributions are paid out on all self-employment income at a constant rate $\tau^S(y) = \tau \cdot y$, without a cap. The benefit entitlements are kept identical to the first UI scenario. Finally, I also study the hypothetical policy scenario where the budget is balanced within group (UI-W). In this last case, UI contributions are paid out on all self-employment income at a constant rate $\tau_k^S(y) = \tau_k \cdot y$ specific to each worker group $k$.

These alternative UI policies are implemented by finding the tax rates satisfying the budget balance conditions in equations (13) and (14). In addition, I allow these tax rates to have an effect along the intensive margin of labor supply. I simply input an elasticity of labor income to the marginal tax rate of 0.5.[30] Table 8 summarizes the contributions schedules and corresponding parameters. The contribution rate required to satisfy the budget balance

[28] Requiring businesses operating on a limited scale to file accounting information seems potentially less demanding given transactions are increasingly digitalized.

[29] \$7,000 corresponds to the states with the lowest threshold. Detailed tables on UI contribution rates by state can be found at `https://oui.doleta.gov/unemploy/avg_employ.asp`.

[30] This value is an upper bound on the steady-state intensive margin labor supply elasticity in Chetty et al. [2011], who report 0.33 as their central estimate.

| UI Policy | Contribution schedule | Contribution rate ($\tau$) | Transfer size (rel. to UI-A) |
|---|---|---|---|
| Align (UI-A) | $\tau \cdot \min\{y^S;\ 7,000/12\}$ | 0.020 | 1.000 |
| Progressive (UI-P) | $\tau \cdot y^S$ | 0.003 | 0.997 |
| Within (UI-W) | $\tau_k \cdot y^S,\ \forall k$ | [0.012,0.003,0.003,0.001] | 1.028 |

**Table 8:** Summary of policy scenarios. All policies are for the same baseline benefit level $b(y) = \max\{0.5 \cdot y;\ 2,000\}$ as workers in paid-employment. The "Contribution rate" column gives the equilibrium contribution rate in each scenario. There are four contribution rates $[\tau_{\text{Low}}, \tau_{\text{Med-Low}}, \tau_{\text{Med-High}}, \tau_{\text{High}}]$ for the UI-W policy because the budget is balanced separately in each group. The "Budget" column reports the total transfers relative to the UI-A policy.

condition (13) in the first policy scenario (UI-A) is 2 percent, much larger than the 0.3 percent contribution rate required in the UI-P scenario, in which contributions are not capped. The contribution rates in the UI-W policy range from 1.2 percent (Low earnings group) to 0.1 percent (High earnings group).

The contribution schedules reported in Table 8 imply some degree of redistribution across worker types, as some groups contribute more than they receive overall. In Figure 4, I report the benefits to contributions ratio

$$\int b_k(y) d\Gamma_k^C(a,y) \cdot \left( \int \tau_k^S(y) d\Gamma_k^S(a,y) \right)^{-1} \tag{16}$$

for each worker group $k$ and in each policy scenario. In the UI-A policy, most of the tax burden falls on the "Med-Low" earnings group. The UI-P policy makes UI for the self-employed more redistributive, with the "High" earnings group contributing in excess and the "Low" group benefiting. (By definition, the within scenario (UI-W) has a ratio of one in each group.)

## 5.3 Unemployment response

The unemployment response to the introduction of unemployment benefits goes through two channels in the model. First, though the arrival rates of employment opportunities are

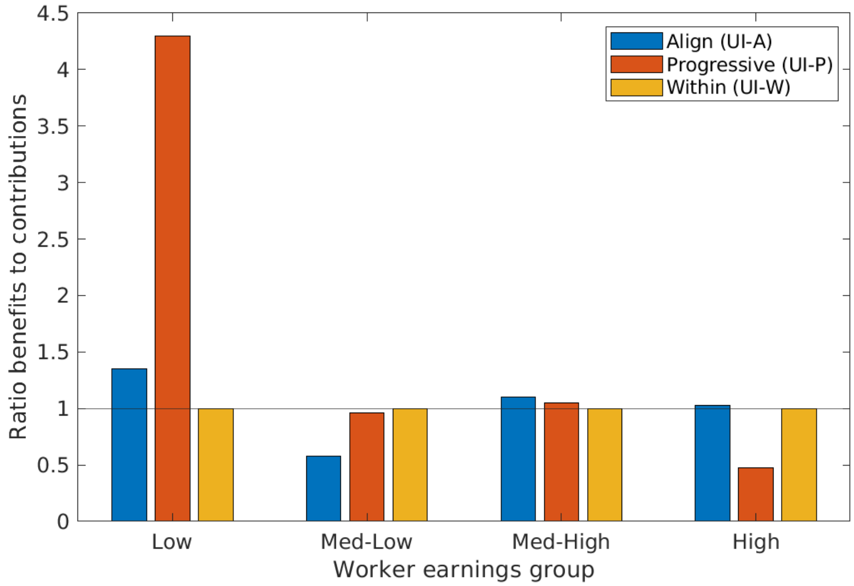



**Figure 4:** Benefits to contributions ratio by worker earnings group in each policy scenario. The benefits to contribution ratio is defined in Equation (16). The horizontal black line denotes a ratio of one.

exogenous, the reservation income at which workers decide to stop searching responds to policy changes. Second, workers in part choose to become unemployed after bad realizations of earnings, as described by equations 1 and 3. While the first channel has been a key focus of the UI literature (see Schmieder and Von Wachter [2016] for a review), there is also empirical evidence on how eligibility to UI affects the probability of becoming unemployed for wage employees (see Khoury [2023] for instance). This second channel is highly relevant for the self-employed in the context of the model, since they choose whether to terminate their activity and claim benefits (as captured by equation 15).

Table 9 reports the response to each parametrization of the UI policy of (i) the flows to and from unemployment, and (ii) the share of workers in each labor form. For concision, I contrast the response of the "Low" and "Med-Low" groups to the overall response pooling all worker types together.

Focusing first on the aggregate response, the right columns in Table 9 show that the unemployment rate increases by 3 to 5 percent across policies, while the self-employment rate

| | Worker earnings group | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Low | | | Med-Low | | | All workers | | |
| | UI-A | UI-P | UI-W | UI-A | UI-P | UI-W | UI-A | UI-P | UI-W |
| Transition rates (% change relative to baseline) | | | | | | | | | |
| $UP$ | -2.11 | -0.92 | -2.23 | 2.80 | 4.20 | 4.28 | 0.62 | 1.55 | 1.10 |
| $US$ | -1.88 | -1.80 | -2.12 | -2.97 | -2.66 | -3.13 | -1.97 | -1.76 | -2.16 |
| $PU$ | -1.16 | -2.60 | -1.99 | 0.49 | 0.48 | 0.25 | 0.14 | -0.77 | -0.24 |
| $SU$ | 1.37 | 1.99 | 1.98 | 44.38 | 31.37 | 31.35 | 13.57 | 10.63 | 10.35 |
| Share of labor force (% change relative to baseline) | | | | | | | | | |
| $U$ | 2.55 | 0.44 | 2.50 | 9.56 | 5.65 | 5.57 | 5.16 | 2.82 | 3.56 |
| $P$ | 0.94 | -2.01 | 1.13 | 2.53 | 1.44 | 1.44 | 1.04 | 0.40 | 0.63 |
| $S$ | -0.51 | 0.55 | -0.56 | -2.72 | -1.57 | -1.56 | -0.88 | -0.40 | -0.56 |

**Table 9:** Unemployment response to alternative UI policies. Each entry shows the percentage change in either the transition rate to and from unemployment or the share of workers in each labor force state.

decreases. The table shows that this is primarily driven by an increase in the rate at which the self-employed become unemployed ($SU$), which goes up by 10 to 13 percent. There is also a slight decrease in the rate at which unemployed workers become self-employed ($US$), which drops by approximately 2 percent. These results are intuitive: self-employment is less valuable because of UI contributions and unemployment is more valuable with additional insurance. Second, the heterogeneity results suggest that the strength of these two mechanisms differ markedly by worker earnings group (left columns). While the results are qualitatively similar between the "Low" and "Med-Low" groups, the response of low earners is relatively muted, notably in terms of the transition rate from self-employment to unemployment (less than a 2 percent increase). By contrast, the $SU$ rate in the "Med-Low" earnings group goes up by more than 30 percent. Third, I find that the unemployment response within a group is broadly similar across policy scenarios despite the differences in the ratio of benefits to contributions documented in Figure 16. This suggests that permanent worker heterogeneity is a more important driver of workers' decisions along the unemployment entry and exit margin than the exact parametrization of the benefit schedule.

## 5.4 Welfare

My main measure of welfare changes is the compensating cash grant, $\mathrm{CG}_k^s(a, y)$, which I define as

$$V_k^s(a + \mathrm{CG}_k^s(a, y), y) = \tilde{V}_k^s(a, y).$$

In this last expression, $\mathrm{CG}_k^s(a, y)$ denotes the cash transfer making the agent indifferent between the baseline economy (value function $V_k^s$) and the economy under one of the alternative policy scenarios (value function $\tilde{V}_k^s$). I report the average of these measures in each worker group.

Figure 5 reports the welfare effects of each counterfactual policy, broken down by worker group. The measure of welfare changes, $\mathrm{CG}_k$, is scaled by median earnings to make it comparable across clusters. Detailed results broken down by worker group $k$ and labor market state $s$ are reported in Appendix Table D.2.[31]

The results suggest that introducing a UI policy close to the current UI scheme for the paid-employed has a small, but mostly positive effect on welfare. Across policies and groups, workers do not value the compensating cash grant at more than 20 percent of median earnings. Despite the relatively high discount rate, which makes precautionary savings costly for these households, the low probability of $SU$ transitions (0.2 to 1 percent chance, see Figure 3a) and low degree of risk aversion (log utility) appear to limit the demand for UI in the various policy scenarios considered.

Figure 5 also shows that the degree of redistribution implied by each policy is important to account for welfare gains and losses. Focusing first on the group with the lowest earnings ("Low"), their welfare gains are almost four times as large in the progressive policy (UI-P)

[31] I also report results for the equivalent variation in consumption used in Krusell et al. [2010] with very similar results.

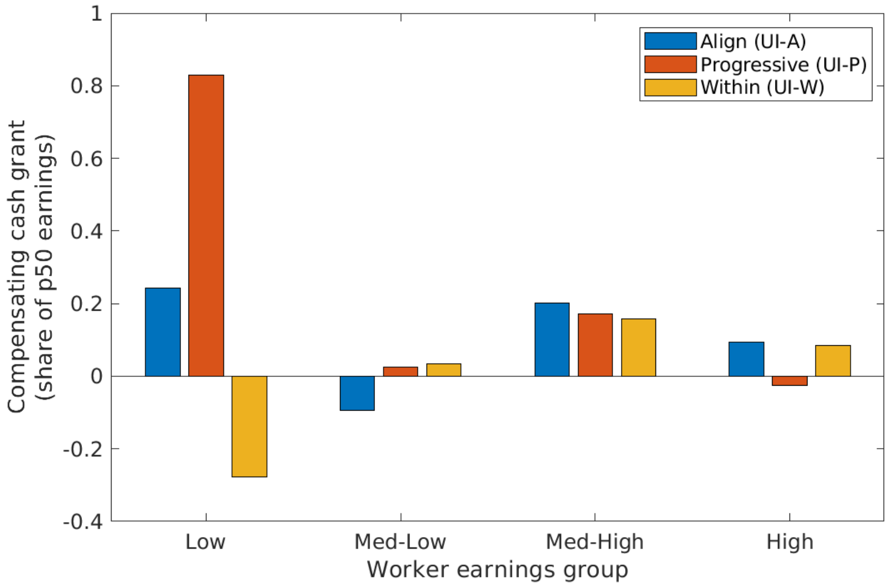



**Figure 5:** Change in welfare in each policy scenario.

than in the standard UI scheme with a cap on contributions (UI-A). In the hypothetical scenario where the policy is actuarially fair within group (UI-W), the compensating differential is negative in the "Low" group. The "Med-Low" type tends to experience very limited welfare gains across policy scenarios, and they lose welfare under the UI-A policy. In this last scenario, this group of workers contributes more than it receives, as shown in Figure 4. The "Med-High" type of workers is the only one with positive welfare gains across all three contribution schedules. The "High" group, finally, loses welfare in the progressive policy scenario (UI-P), since the tax burden is highest for them when there is no cap on contributions. Overall, none of the three policy scenarios considered yield welfare gains across all worker groups.

## 5.5 Summary and discussion

In the model, extending UI benefits to the self-employed involves some clear policy tradeoffs. In terms of unemployment outcomes, I find that transitions from self-employment to unemployment are an important channel for the unemployment response to UI benefits for the self-employed. Given the distinction between layoffs and quits does not readily apply to self-employment, this channel should be central to the design of any actual policy extending eligibility to the self-employed. In terms of welfare effects, my results suggest that there is a clear redistribution dimension to UI benefits for the self-employed. Depending on the details of the contribution schedule, welfare gains are not necessarily increasing in the earnings groups elicited from the data.

# 6 Conclusion

In this paper, I study the labor market risks associated with self-employment. My first set of results is empirical. I show that the large earnings fluctuations experienced by the self-employed are driven both by labor income fluctuations during self-employment spells and transitions to unemployment. In addition, many self-employed do not earn more or have more liquid wealth than employees, thus potentially limiting their resources to self-insure. These findings are for a sample of individuals who are the main earners within their household.

I then calibrate a search model with precautionary savings that can replicate these empirical regularities. I use the model to assess how different groups of self-employed value UI benefits similar to the UI system in place for wage employees. My second set of results relates to these policy experiments. I find (i) a strong unemployment response in large part coming from transitions from self-employment to unemployment, and (ii) strong heterogeneity in the welfare gains achieved by different types of workers. More generally, this framework can inform the current policy debate on the rise of "gig economy" work and the degree to which this category of workers can be reached by conventional social insurance programs.

The model in this paper offers a framework to study the labor market risks faced by the unincorporated self-employed. In doing so, it abstracts from explicitly modeling some dimensions of running a business that are potentially relevant for a subset of the self-employed.

An important direction for future work would be to understand how business owners allocate their portfolio between liquid assets, illiquid assets, and their business. Given the inherently risky nature of owning a business, such an extension would provide additional insights into the self-insurance dimension of entering and exiting self-employment.

# References

Daron Acemoglu and Robert Shimer. Efficient Unemployment Insurance. *Journal of Political Economy*, 107(5):893–928, October 1999. ISSN 0022-3808. doi: 10.1086/250084.

David Arthur and Sergei Vassilvitskii. K-means++ the advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035, 2007.

Richard Blundell, Luigi Pistaferri, and Ian Preston. Consumption inequality and partial insurance. *American Economic Review*, 98(5):1887–1921, 2008.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. A Distributional Framework for Matched Employer Employee Data. *Econometrica*, 87(3):699–739, 2019. ISSN 1468-0262. doi: 10.3982/ECTA15722.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. Discretizing unobserved heterogeneity. *Econometrica*, 90(2):625–643, 2022.

Jake Bradley, Fabien Postel-Vinay, and Hélene Turon. Public sector wage policy and labor market equilibrium: A structural model. *Journal of the European Economic Association*, page jvw026, 2017.

Sebastian Camarero Garcia and Martin Murmann. Unemployment benefit duration and startup success. *ZEW-Centre for European Economic Research Discussion Paper*, (20-033), 2020.

Sylvain Catherine. Keeping options open: What motivates entrepreneurs? *Journal of Financial Economics*, 144(1):1–21, 2022.

Census Bureau. Survey of Income and Program Participation. `https://www.census.gov/programs-surveys/sipp.html`, 2014. 1996, 2001, 2004, and 2008 panels.

Raj Chetty. Moral Hazard versus Liquidity and Optimal Unemployment Insurance. *Journal of Political Economy*, 116(2):173–234, April 2008. ISSN 0022-3808. doi: 10.1086/588585.

Raj Chetty, Adam Guren, Day Manoli, and Andrea Weber. Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review*, 101(3):471–75, 2011.

Brett Collins, Andrew Garin, Emilie Jackson, Dmitri Koustas, and Mark Payne. Is gig work replacing traditional employment? evidence from two decades of tax returns. *Unpublished paper, IRS SOI Joint Statistical Research Program*, 2019.

Fatih Guvenen. An empirical investigation of labor income processes. *Review of Economic dynamics*, 12(1):58–79, 2009.

Barton H. Hamilton. Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political Economy*, 108(3):604, June 2000. ISSN 00223808.

Gary D Hansen and Ayşe Imrohoroğlu. The role of unemployment insurance in an economy with liquidity constraints and moral hazard. *Journal of political economy*, 100(1):118–142, 1992.

Trevor Hastie, Robert Tibshirani, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

Johan Hombert, Antoinette Schoar, David Sraer, and David Thesmar. Can unemployment insurance spur entrepreneurial activity? evidence from france. *The Journal of Finance*, 75(3):1247–1285, 2020.

John Eric Humphries. The causes and consequences of self-employment over the life cycle. *Yale University*, 2017.

Erik Hurst and Annamaria Lusardi. Liquidity Constraints, Household Wealth, and Entrepreneurship. *Journal of Political Economy*, 112(2):319–347, April 2004. ISSN 0022-3808. doi: 10.1086/381478.

Emilie Jackson. Availability of the gig economy and long run labor supply effects for the unemployed. In *2021 APPAM Fall Research Conference*. APPAM, 2022.

Louis S Jacobson, Robert J LaLonde, and Daniel G Sullivan. Earnings losses of displaced workers. *The American economic review*, pages 685–709, 1993.

John Bailey Jones and Sangeeta Pratap. An estimated structural model of entrepreneurial behavior. *American Economic Review*, 110(9):2859–98, 2020.

Greg Kaplan and Giovanni L Violante. How much consumption insurance beyond self-insurance? *American Economic Journal: Macroeconomics*, 2(4):53–87, 2010.

Laura Khoury. Unemployment benefits and redundancies: Incidence and timing effects. *Journal of Public Economics*, 226:104984, 2023.

Jonas Kolsrud, Camille Landais, Peter Nilsson, and Johannes Spinnewijn. The optimal timing of unemployment benefits: Theory and evidence from sweden. *American Economic Review*, 108(4-5):985–1033, 2018.

Dmitri Koustas. Consumption insurance and multiple jobs: Evidence from rideshare drivers. *Unpublished working paper*, 1(2), 2018.

Per Krusell, Toshihiko Mukoyama, and Ayşegül Şahin. Labour-Market Matching with Precautionary Savings and Aggregate Fluctuations. *Review of Economic Studies*, 77(4):1477–1507, October 2010. ISSN 00346527. doi: 10.1111/j.1467-937X.2010.00700.x.

Rasmus Lentz. Optimal unemployment insurance in an estimated job search model with savings. *Review of Economic Dynamics*, 12(1):37–57, 2009.

Ross Levine and Yona Rubinstein. Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More? *The Quarterly Journal of Economics*, 132(2):963–1018, May 2017. ISSN 0033-5533. doi: 10.1093/qje/qjw044.

Katherine Lim. Do american mothers use self-employment as a flexible work alternative? *Review of Economics of the Household*, 17(3):805–842, 2019.

Jeremy Lise. On-the-Job Search and Precautionary Savings. *The Review of Economic Studies*, 80(3):1086–1113, January 2013. ISSN 0034-6527, 1467-937X. doi: 10.1093/restud/rds042.

Costas Meghir and Luigi Pistaferri. Income variance dynamics and heterogeneity. *Econometrica*, 72(1):1–32, 2004.

Costas Meghir and Luigi Pistaferri. Earnings, consumption and life cycle choices. In *Handbook of labor economics*, volume 4, pages 773–854. Elsevier, 2011.

Costas Meghir, Renata Narita, and Jean-Marc Robin. Wages and Informality in Developing Countries†. *American Economic Review*, 105(4):1509–1546, April 2015. ISSN 00028282. doi: 10.1257/aer.20121110.

OECD. *The Future of Social Protection*. 2018.

Itay Saporta-Eksten. Job loss, consumption and unemployment insurance. *Unpublished manuscript*, pages 1–50, 2014.

Johannes F Schmieder and Till Von Wachter. The effects of unemployment insurance benefits: New evidence and interpretation. *Annual Review of Economics*, 8:547–581, 2016.

# Appendix

# A Data

The main data source is the Survey of Income and Program Participation (SIPP), a survey with detailed information on households' use of welfare programs [Census Bureau, 2014]. I use data from the 1996, 2001, 2004, and 2008 panels. The attrition rate in the raw data ranges from 26 percent (2008 panel) to 35 percent (2001 panel). The share of individuals with no missing month in their record ranges from 65 percent to 84 percent across panels. All summary statistics reported in the main text are computed using longitudinal weights, which should in part correct the bias from non-random attrition. This Appendix expands on the definition of paid-employment and self-employment, and on the construction of the "main earners" and "self-employed" samples.

## A.1 Labor force status

The SIPP contains two key pieces of information to classify individuals as unemployed, paid-, or self-employed: a week-by-week account of their employment state (employed, on layoff, unemployed, non-participating) and information on up to two jobs and two businesses (such as job/business identifier, wages, profits, incorporation status).

The SIPP questionnaire distinguishes between three types of non-employment week-by-week: on layoff, no job looking for work, no job not looking for work. I define as unemployment spells all non-employment spells of at most fifty weeks if the individual declares to be looking for a job at some point during the spell. This requirement is standard in the unemployment duration literature [see for instance Chetty, 2008, Appendix B]. Despite the SIPP questionnaire only asking whether the individual is actively looking for work, this definition is potentially conservative in that some self-employed may associate the question with the receipt of unemployment benefits.

Employed individuals are classified as paid- or self-employed on the basis of the jobs or businesses for which they report most of their earnings. I start by cleaning the job and business identifiers provided in the survey to make them consistent with the start and end dates reported for each job or business. For a large majority of workers, only one type of employment is reported in a week: wage work or self-employment. Conditional on workers being employed, around 2.5 percent of worker-week observations are employed both as employees and at their own business. Figure A.1 shows the share of earnings in the individual's assigned labor form over these overlapping spells. Among the individuals working in both paid- and self-employment at the same time, the figure shows that there is a main labor form for a large number of these spells: more than 50 percent of these spells have at least 80 percent of their earnings coming from their assigned labor form. In the case where workers report both working as an employee and having a business, they are assigned the labor form status in which they report the largest earnings over the duration of the overlapping spell. If workers report more than one job or business (there is information on up to two jobs and/or businesses in the SIPP data), I simply add the corresponding wages/profits to get a measure of earnings in paid- or self-employment. I proceed similarly to define the business characteristics reported in the main text. For categorical variables (incorporation status and number of employees), I use the maximum across the two businesses.

Finally, I follow the convention in the Current Population Survey and build a monthly panel of employment state for each individual based on the second week (the first full week) in each month. Keeping only working-age individuals, I check the validity of my definitions by plotting the implied unemployment rate (Figure A.2a) and self-employment rate (Figure A.2b) against data from the Current Population Survey (BLS).[32] The data from the SIPP are aggregated using cross-sectional weights.[33] The figures show that my definitions yield sensible unemployment and self-employment rates.

[32] The corresponding series from the Bureau of Labor Statistics are LNU02048984 ("Incorporated Self-Employment"), LNU02027714 ("Non-incorporated Self-Employment"), LNU03000000 ("Unemployment"), and LNU02000000 ("Employment").

[33] The rest of the paper uses longitudinal weights.

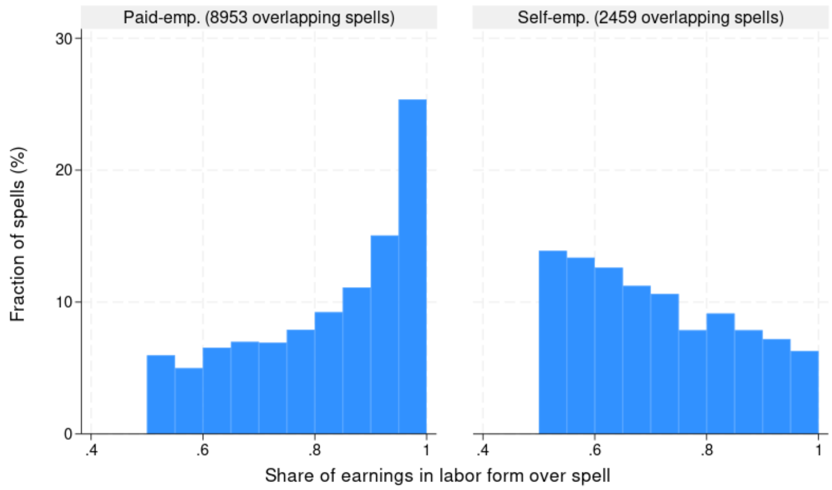



**Figure A.1:** Share of labor earnings over overlapping employment spell conditional on assigned labor form. An overlapping employment spell is defined as consecutive weeks during which an individual report working both as an employee and at their own business. By construction, the proportion of earnings in the individual's assigned labor form is greater than 50 percent.

## A.2 Main earners sample

I retain individuals aged at least 25 or at most 65 across the survey. In each household, I only keep the main earner, defined as the individuals with the largest labor earnings across the survey, irrespective of their gender. I also exclude individuals who are mostly out of the labor force over the survey period. The aim of these restrictions is to reduce the sample to the individuals most strongly attached to the labor market. Taken together, these restrictions yield a dataset of 117,051 individuals, who are followed on average for 36 months. In this sample, around 3 percent of worker-month observations declare working at the same time as an employee and at their own business in a given month.

## A.3 Self-employed sample

The key analysis and estimation sample is made of individuals who are part of the main earner sample, and who (i) spend at least one month in self-employment over the duration of the panel and (ii) never declare that their business is unincorporated over the duration of

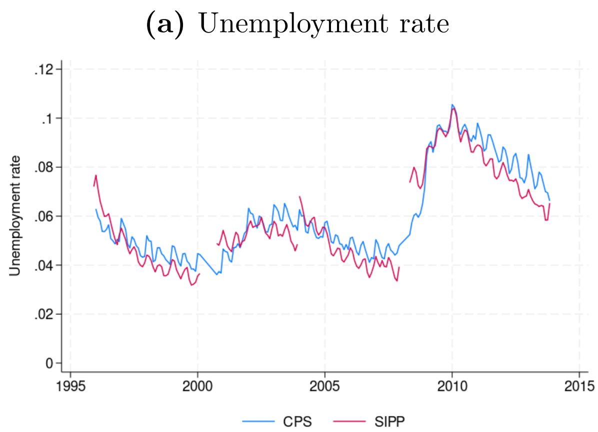




**Figure A.2:** Benchmark SIPP vs Bureau of Labor Statistics Series (BLS). The self-employment and unemployment rate in the SIPP are based on the definitions of labor force status given in Section 2 and Appendix A. The sample used to build the series include all working age individuals, not just the main earner in each household. Observations are grossed using cross-sectional weights.

the survey. In this sample, around 8 percent of worker-month observations declare working at the same time as an employee and at their own business in a given month. This additional restriction gives a sample of 8,623 workers who are followed on average for 35 months. All summary statistics targeted in the calibration procedure are drawn from the self-employed sample.

# B Equilibrium definition

I introduce some additional notation for workers' reservation strategies to formally describe the equilibrium. Let $d_k^s$ be an indicator denoting workers' decision to quit their job or shut down their business given the realization of income. For instance, for a self-employed worker deciding to terminate their activity,

$$d_k^S(a,y) := \mathbb{1}\Big\{V_k^U(a) > \delta_k^S V_k^U(a) + (1-\delta_k^S)\Big[V_k^S(a,y) + \lambda^{SP}\phi_k^{SP}(a,y)\Big]\Big\}.$$

Let $\rho_k^{ss'}(a, y^s)$ be the reservation income draw from $F_k^{s'}$ making workers indifferent between their current labor force status s and employment in $s'$. This reservation income is implicitly given by $V_k^s\big(a, y^s\big) = V_k^{s'}\big(a, \rho_k^{ss'}(a, y^s)\big)$.

For each $k \in \{1, ..., K\}$ and $s \in \{U, B, P, S\}$, a stationary equilibrium is a set of value functions $V_k^s$ and $R_k^s$, decision rules $d_k^S$, $\rho_k^{ss'}$, $\hat{a}_k^s$ and $\hat{c}_k^s$, and a distribution $\Gamma_k$ across labor force states, wealth and income, such that

1. The $R_k^s$ functions are defined by equations (1), (3), (5), and (6);
2. The choice to terminate employment, $d_k^S(a, y)$, and the reservation income functions $\rho_k^{ss'}(a, y^s)$ solve (1), (3), (5), and (6);
3. The $V_k^s$ functions are defined by equations (2), (4), (7), (8);
4. The asset and consumption choice functions, $\hat{a}_k^s$ and $\hat{c}_k^s$, solve equations (2), (4), (7), (8);
5. Finally define $\boldsymbol{Q}_k$ the operator mapping the current distribution of workers to that in the next period. This operator arises from workers' optimal consumption and savings decisions, as well as their reservation strategies. The associated stationary distribution of workers $\Gamma_k$ solves $\Gamma_k = \boldsymbol{Q}_k \circ \Gamma_k$.

# C Choice of number of clusters

The number of clusters $K$ is typically chosen by examining the within sum of squares as it evolves as a function of $K$ [see Hastie et al., 2009, Chapter 14.3.11]. There is no explicit objective function, but the intuition is to look for the point where the reduction in the within sum of squares become more marginal with an additional cluster.

Figure C.1 reports several statistics on the evolution of the within sum of squares as a function of $K$: the within sum of squares (Figure C.1a), its log (Figure C.1b), the coefficient of determination (Figure C.1c), and the proportional reduction of error (Figure C.1d). There is no clear break across these different measures when applied to the self-employed sample as described in Section 4.1. But these statistics suggest that the number $K = 4$ used in the model captures a reasonable amount of heterogeneity in the clustering variables, while keeping the model relatively parsimonious, since increasing $K$ requires to estimate another set of cluster-specific parameters. For example, increasing the number of clusters from 4 to 10 results in an increase in the coefficient of determination from 75 percent to 85 percent (Figure C.1c).

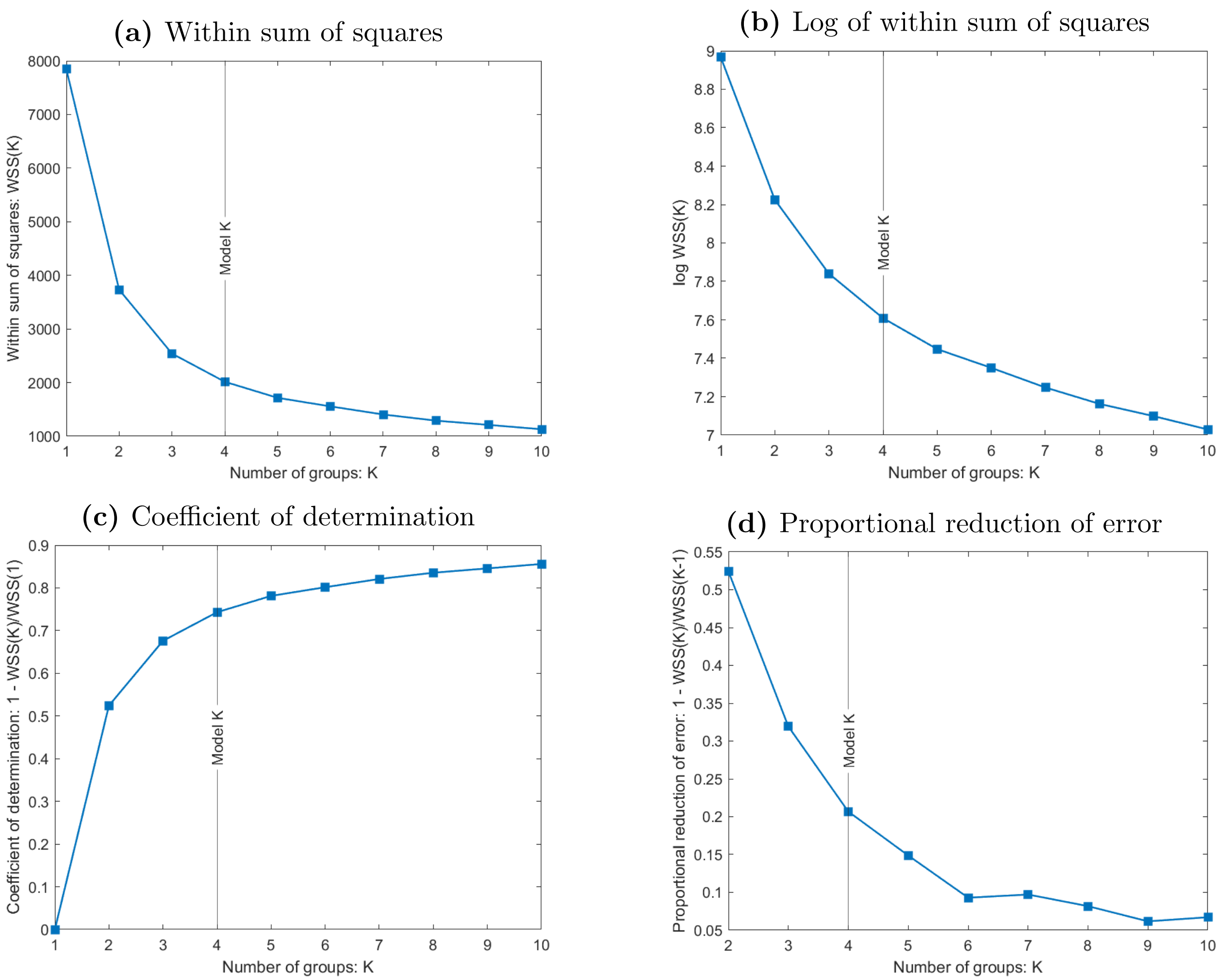



**Figure C.1:** Change in k-means fit with the number of clusters $K$. The vertical line denotes the number of clusters used in the model.

# D Additional figures and tables

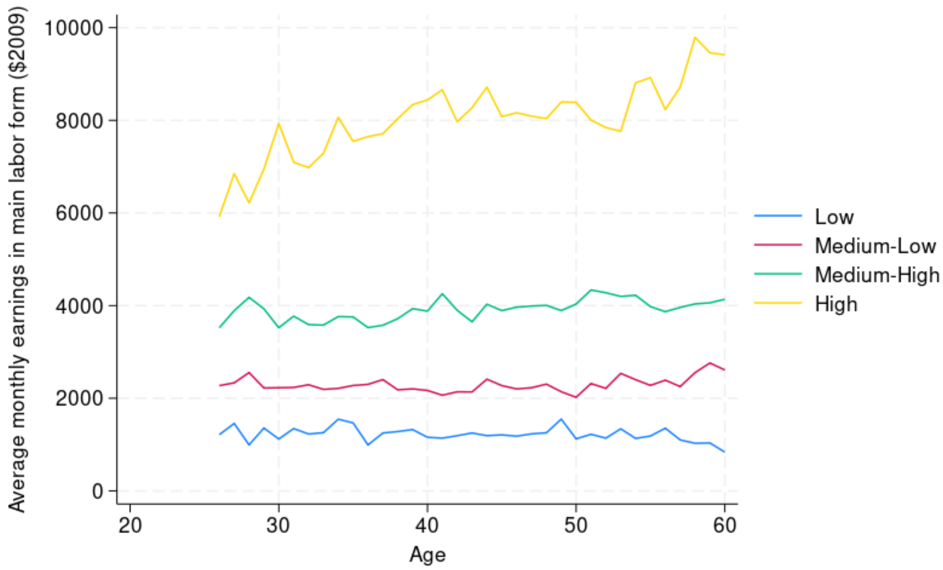



**Figure D.1:** Pseudo-cohort age profile of earnings by cluster.

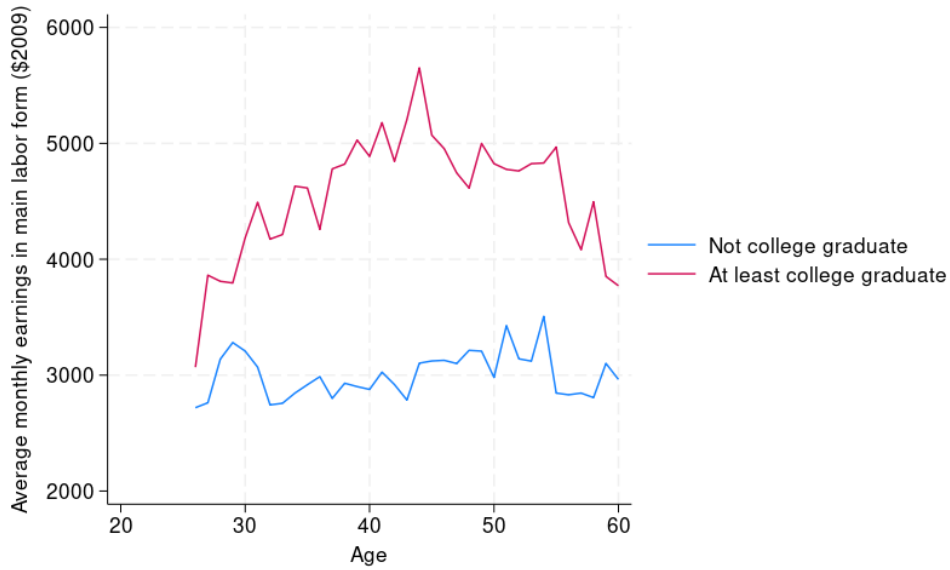



**Figure D.2:** Pseudo-cohort age profile of earnings by college graduation status.

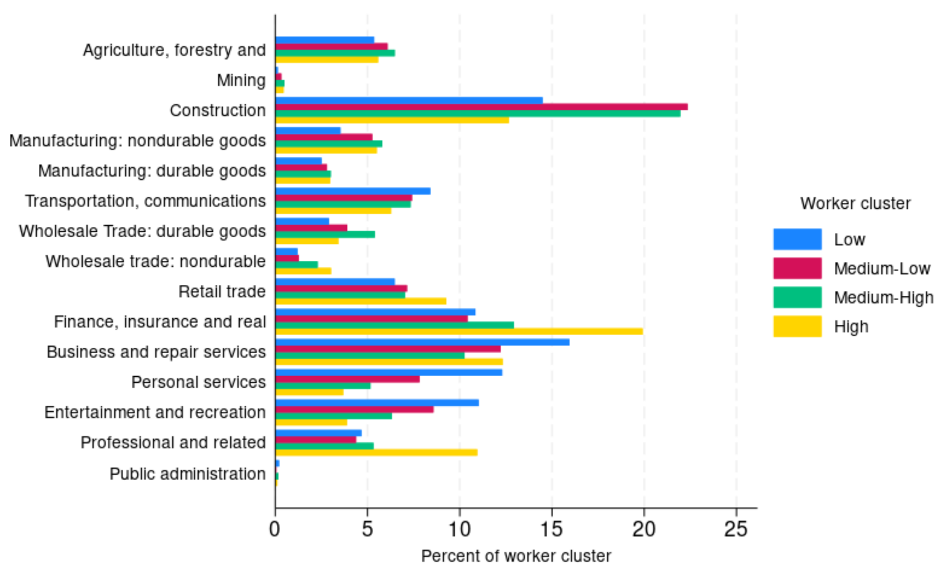



**Figure D.3:** Industry composition by worker class.

| | $\ln Y_{it}^s = \alpha_k^s + \beta_k^s \cdot \ln y_{it}^s + \epsilon_{it}$ | | | |
|---|---|---|---|---|
| | $\alpha_k^s$ | $\beta_k^s$ | $R^2$-adj | Obs. |
| Unemployment $(s = U)$ | | | | |
| Low | 6.560 | — | — | 1,860 |
| Med-Low | 6.667 | — | — | 1,685 |
| Med-High | 6.473 | — | — | 1,187 |
| High | 7.114 | — | — | 657 |
| Paid-employment $(s = P)$ | | | | |
| Low | 3.711 | 0.557 | 0.174 | 8,259 |
| Med-Low | 3.716 | 0.580 | 0.202 | 14,813 |
| Med-High | 3.229 | 0.655 | 0.323 | 15,261 |
| High | 2.660 | 0.735 | 0.535 | 8,924 |
| Self-employment $(s = S)$ | | | | |
| Low | 3.268 | 0.631 | 0.273 | 14,868 |
| Med-Low | 2.827 | 0.690 | 0.449 | 21,268 |
| Med-High | 2.864 | 0.701 | 0.527 | 25,639 |
| High | 2.945 | 0.706 | 0.631 | 15,324 |

**Table D.1:** Household income functions.

| | Worker earnings group | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Low | | | Medium-Low | | | Medium-High | | | High | | |
| | UI-A | UI-P | UI-W | UI-A | UI-P | UI-W | UI-A | UI-P | UI-W | UI-A | UI-P | UI-W |
| Equivalent variation in $c$ – $\mathrm{EV}_k^s$ (percent) | | | | | | | | | | | | |
| All $s$ | 0.13 | 0.43 | -0.15 | -0.06 | 0.02 | 0.03 | 0.15 | 0.13 | 0.12 | 0.07 | -0.02 | 0.07 |
| $P$ | 0.09 | 0.33 | -0.14 | -0.06 | 0.01 | 0.01 | 0.11 | 0.09 | 0.08 | 0.05 | -0.02 | 0.05 |
| $S$ | 0.11 | 0.43 | -0.19 | -0.08 | 0.02 | 0.02 | 0.14 | 0.13 | 0.11 | 0.08 | -0.03 | 0.07 |
| $B$ | 0.10 | 0.36 | -0.14 | -0.06 | 0.01 | 0.02 | 0.12 | 0.11 | 0.10 | 0.06 | -0.02 | 0.06 |
| $U$ | 0.09 | 0.35 | -0.14 | -0.06 | 0.01 | 0.02 | 0.12 | 0.10 | 0.09 | 0.06 | -0.02 | 0.05 |
| $C$ | 1.19 | 1.47 | 0.95 | 0.64 | 0.72 | 0.73 | 1.96 | 1.95 | 1.94 | 1.18 | 1.11 | 1.18 |
| Compensating grant in $a$ – $\mathrm{CG}_k^s$ (ratio to median earnings) | | | | | | | | | | | | |
| All $s$ | 0.24 | 0.83 | -0.28 | -0.09 | 0.03 | 0.04 | 0.20 | 0.17 | 0.16 | 0.09 | -0.03 | 0.09 |
| $P$ | 0.17 | 0.61 | -0.22 | -0.08 | 0.02 | 0.02 | 0.15 | 0.13 | 0.12 | 0.07 | -0.02 | 0.07 |
| $S$ | 0.23 | 0.87 | -0.35 | -0.12 | 0.02 | 0.03 | 0.21 | 0.18 | 0.16 | 0.11 | -0.04 | 0.10 |
| $B$ | 0.14 | 0.52 | -0.18 | -0.06 | 0.02 | 0.02 | 0.10 | 0.09 | 0.08 | 0.05 | -0.01 | 0.05 |
| $U$ | 0.16 | 0.59 | -0.23 | -0.07 | 0.02 | 0.02 | 0.13 | 0.11 | 0.10 | 0.07 | -0.02 | 0.07 |
| $C$ | 1.71 | 2.10 | 1.37 | 0.57 | 0.65 | 0.66 | 1.45 | 1.44 | 1.43 | 0.72 | 0.66 | 0.72 |

**Table D.2:** Welfare changes by worker class in response to each alternative policy. The equivalent variation in consumption, $\mathrm{EV}_k^s(a,y)$, is defined as $\mathbb{E}_t \sum_{\tilde{t}=0}^{\infty} \beta^{\tilde{t}} u((1+\mathrm{EV}_k^{s_t}(a_t,y_t))c_{t+\tilde{t}}) = \mathbb{E}_t \sum_{\tilde{t}=0}^{\infty} \beta^{\tilde{t}} u(\hat{c}_{t+\tilde{t}})$ where $\hat{c}_t$ denotes consumption choices with one of the policies and $c_t$ is consumption in the baseline economy. The compensating cash grant, $\mathrm{CG}_k^s(a,y)$, is defined as $V_k^s(a+\mathrm{CG}_k^s(a,y),y) = \hat{V}_k^s(a,y)$, where $V_k^s$ and $\hat{V}_k^s$ denote the worker's value function, respectively, in the baseline economy and in the economy with one of the policies.