Self-employment and Unemployment Risk

Richard Audoly
UCL and IFS*

September 2018

Preliminary and incomplete

Abstract

Public unemployment insurance (UI) schemes represent an important feature of the social safety net in most advanced economies. However, the self-employed are generally excluded from these programs. This paper calibrates a job search model to US data to evaluate the potential welfare gains from extending UI benefits to the group of self-employed workers. The model features workers in both paid- and self-employment who face the risk of becoming unemployed; it also allows them to privately save to self-insure. Preliminary results suggest that extending UI benefits to the self-employed yields modest, but positive, welfare gains.

1 Introduction

A substantial literature aims at characterizing the optimal unemployment insurance contract (e.g., Chetty, 2006; Acemoglu and Shimer, 1999; Shimer and Werning, 2008). By comparison, the unemployment risks faced by the self-employed have received scant attention, since, in practice, the group is barred access to public unemployment insurance (UI) schemes in most OECD economies.[1] But is it clear that the self-employed face different unemployment risks from traditional wage workers? Can they self-insure against them? And if not, could opening unemployment insurance to the group potentially improve welfare? Guided by evidence from US data, this paper describes a job search model with workers in both paid- and self-employment, who all face the risk of becoming unemployed. I then leverage this framework to assess the welfare effects of opening unemployment benefits to the self-employed.

*University College London and Institute for Fiscal Studies, richard.audoly@ifs.org.uk. Thanks to my advisors at UCL, Fabien Postel-Vinay and Jan Eeckhout, for their support and comments. For extremely valuable feedback, I also wish to thank Ludo Visschers, Cormac O'Dea, and Rasmus Lentz.

[1] Among these countries, a handful offer some form of public unemployment insurance for some subgroups of independent workers, such as artists and writers in Germany. These schemes are reviewed in details in OECD (2018). In terms of private unemployment insurance, anecdotal evidence suggests that when a private UI market does exist, the self-employed cannot buy these policies.

In the model, risk-averse workers search both for traditional wage jobs and business opportunities. They are allowed to privately save and borrow, and they can use these savings to smooth their consumption in the event of a shock. The trade-off highlighted in the UI literature between the insurance value of unemployment benefits and how workers adjust their search behavior in response is still at play here, but the distinction between unvoluntary layoffs and voluntary quits, central to UI systems in most advanced economies, does not readily translate to the self-employed. Since it is difficult to argue that business shutdown is due to circumstances strictly beyond the owner's control, this paper studies UI policies in which the planner cannot distinguish between voluntary and unvoluntary unemployment for the self-employed. I use the calibrated model to compute the insurance value of extending to the self-employed a mandatory public UI scheme with a 50 percent replacement rate. My preliminary results suggest that this system yields positive gains for the self-employed who become unemployed, of the order of a one-time $9,000 payment.

The model also allows for substantial worker heterogeneity to capture the large differences in ability underlying earnings data. To discipline this heterogeneity, I follow Manresa et al. (2017) and Bonhomme et al. (2019) in using a clustering algorithm to discretize the earnings potential of workers in a first, pre-estimation, step. I use a k-means algorithm to partition workers in ability groups based on their observed labor earnings. Besides improving the fit of the model to the data, such a partition allows to decompose the response of different groups of workers to extending UI benefits to the self-employed. As an example, my preliminary results point to the insurance value of the policy being largest for high earners.

Another contribution of this paper is to offer empirical evidence on the unemployment risks faced by the self-employed. Self-employment represents a sizable share of total employment in advanced economies. In the United States, where my data are drawn from, more than one employed worker in ten was self-employed between 2008 and 2014. I further document in the Survey of Income and Program Participation (SIPP) the existence of substantial transitions between labor forms, as well as in and out of unemployment. For instance, the chance for self-employed (paid-employed) workers to find themselves unemployed the next month is .6 (1.2) percent on average. The formerly self-employed do not exit unemployment significantly faster than the previously paid-employed. Conditional on becoming unemployed, the group of self-employed workers also have less liquid assets to self-insure. To the best of my knowledge, these empirical patterns have not been documented elsewhere.

From a policy perspective, lastly, this paper provides a framework to evaluate the welfare effects of extending a central feature of the social safety net to the self-employed. This is relevant for policy in at least two dimensions. First, as noted above, almost all OECD countries exclude the self-employed from UI public schemes. While there are clear moral hazard concerns specific to this labor form, they need to be quantified and balanced with the potential insurance value of UI to these workers. Second, the rise of alternative work arrangements with the emergence of labor platform companies – firms that match workers to customers without being bound to them by a traditional labor contract – also brings about a series of questions regarding the exact status of these workers and their treatment by the welfare state. It is difficult at this stage to evaluate the prevalence of these arrangements in the data, not to mention forecasting their growth over the next decade. But my framework does make progress in quantifying the trade-offs to extending UI rights to labor market participants beyond traditional wage workers.

Related literature. Recent work on the welfare effects of UI benefits falls into two main categories. Following Chetty (2006), numerous studies have characterized optimal UI benefits in terms of sufficient statistics, a reduced number of elasticities, which can be computed independently from a model's primitives (see Kolsrud et al., 2015, for a recent example). Using this approach, Chetty (2008) further makes the point that the response to changes in UI benefits also goes through a liquidity channel on top of the moral hazard channel generally put forward in the literature. My paper directly builds on this finding and explicitly models a self-insurance motive.

The second strand of this literature studies the welfare effects of UI by specifying a full structural model. Acemoglu and Shimer (1999) show that workers search for better-paying jobs with higher UI benefits in a model with directed search. Hansen and Imrohoroglu (1992) find that, in the presence of moral hazard, the optimal level of UI benefits is close to zero in an economy with liquidity accumulation and risk-averse agents. Krusell et al. (2010) integrate consumption-saving choices into a search framework and show how UI benefits stifles job creation, as it increases workers' outside option.

This paper takes a structural approach to analyze the liquidity-moral hazard trade-off highlighted by Chetty (2008) for the self-employed. Because most advanced economies limit UI public schemes to wage workers, exogenous policy variation affecting the group is not readily available. Instead, this study takes a stand on the primitives underlying the sufficient statistics approach, which allows to directly derive the agents' responses to a change in policy. While misspecification remains a concern, the calibrated model replicates a large number of moments that directly relate to workers' choice of labor form.

This study also contributes to the literature on the determinant of self-employment. Hamilton (2000) shows that opening a business is not a guarantee of higher income. Humphries (2018) and Catherine (2017) develop life cycle models of the decision to move across labor forms. They focus on the impact of policies to promote entrepreneurial entry, emphasizing the importance of heterogeneous skill accumulation (Humphries, 2018) and the option value of paid-employment (Catherine, 2017) in properly accounting for workers' choice of labor form over their life cycle. My paper is complementary to these studies in the sense that it centers on a specific aspect of the self-employed's working life: unemployment spells. It stresses policies aimed at providing insurance during these episodes.

Finally, my framework combines two aspects of the recent search literature. First, a series of models have relaxed the assumption that workers are risk-neutral (for example, Shi, 2009; Lamadon, Lamadon); a subset of these models further allows workers to privately save (Chaumont and Shi, 2017; Lise, 2013). Second, several papers develop models of a segmented labor market, such as an informal and formal sector (Meghir et al., 2015), public versus private employment (Bradley et al., 2017), or paid- and self-employment (Visschers et al., 2014; Bradley, 2016). The model described in this paper features both a two-ladder structure, to capture workers' outside options as they move across labor market states, and risk-aversion, to let them self-insure against labor market shocks.

Outline. The next section documents to which extent the self-employed are exposed to unemployment risks in the Survey of Income and Program Participation (SIPP). Section 3 describes the model. Section 4 details the procedure to discretize worker heterogeneity. Section 5 shows the model fit, and Section 6 analyzes the effects of extending UI benefits to the self-employed.

2 Self-employment and unemployment in the data

This section provides empirical evidence on the exposure of the self-employed to labor market risks in the Survey of Income and Program Participation (SIPP).[2] Because the term "self-employed" covers a variety of situations in the labor market, I start with a brief description of the group in the SIPP. I then proceed with a comparison of the paid- and self-employed in terms of their exposure to unemployment risk.

2.1 The self-employed in the SIPP data

The sample is restricted to the individuals with the largest labor earnings in each household over the duration of the survey. These workers are assigned to one of three labor market states depending on their primary source of labor income over the duration of each job or

[2]The data can be obtained from the Census Bureau's website: `https://www.census.gov/programs-surveys/sipp/`

	paid-employed	self-employed
age	43.05	46.53
woman	0.437	0.266
married	0.568	0.627
not white	0.198	0.149
college graduate	0.569	0.574
post-graduate	0.130	0.161

Table 1: Summary statistics by labor form for the selected sample.

business: paid-employed (P), self-employed (S), or unemployed (U).[3]

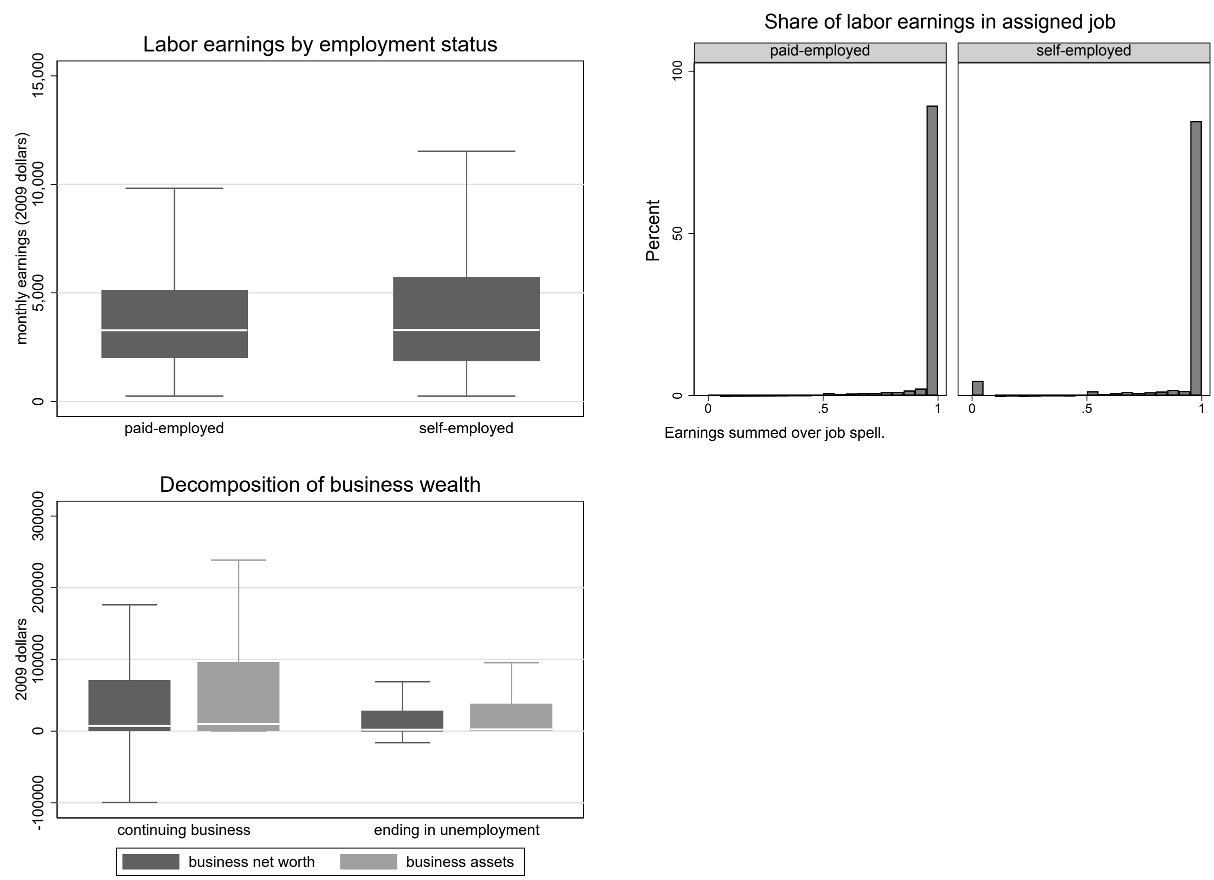
Table 1 provides some basic demographic information on the two groups. It shows that the self-employed are slightly more likely to be older men, married, and not to belong to a minority. Figure 1 further compares these groups in terms of labor income. Though there is more heterogeneity among the self-employed, in particular with more mass at the top, there is no marked difference between them (the median income is about the same for both labor form). Figure 1 also gives the share of total labor earnings the assigned job or business accounts for. It shows that despite some evidence of these workers having an auxiliary source of labor income, most of them have a clear main activity irrespective of their choice of labor form.

The information available in the SIPP is also consistent with the view that entrepreneurs – defined as business owners aiming to grow by introducing new products or processes – account for a small fraction of the self-employed.[4] The distribution of business wealth, displayed in the bottom left panel of Figure 1, suggests that most business owners run mildly capital-intensive operations (the median reported business equity is about 7,000 dollars for continuing businesses, 2,000 dollars for business owners ending up in unemployment). Taken together, only 55 percent of all businesses are incorporated and only 5 percent ever had more than 25 employees at any point since their creation.

[3] Appendix A provides a complete description of the SIPP and the procedure to ascribe workers to each labor market state.

[4] Hurst and Pugsley (2011) find that most new business owners do not intend to grow or bring new products to the market.

Figure 1: Characteristics of the self-employed in the SIPP



Notes: Top left panel: Distribution of monthly labor earnings in primary activity (2009 dollars). Top right panel: Share of total labor earnings as a total of the person's labor income over the spell, a measure of moonlighting. Bottom left panel: Distribution of business wealth (gross and net), broken down by whether the spell ended up in unemployment. Last recorded value before the end of the spell/survey.

Orig./Dest.	P	S	U
P	0.014	0.001	0.012
S	0.010	0.005	0.006
U	0.150	0.010	

Table 2: Monthly transition rates in the SIPP (2008-2013). Transition matrix between the three labor states identified in the SIPP data. "PP" and "SS" denote job-to-job and business-to-business transitions respectively. Rows do not sum to one by construction: the

Table 2 displays monthly transition rates across labor states in the SIPP. The data come with job and business identifiers, so job-to-job (PP) and business-to-business (SS) transitions can be computed. The table shows that paid- and self-employment, as defined by a person's main source of earnings, are connected labor forms. There is, for example, a one percent chance that a worker in self-employment will end up in paid-employment the next month. Any policy that affects the value of self-employment should therefore take into account the potential for workers to change labor form.

2.2 Unemployment risk

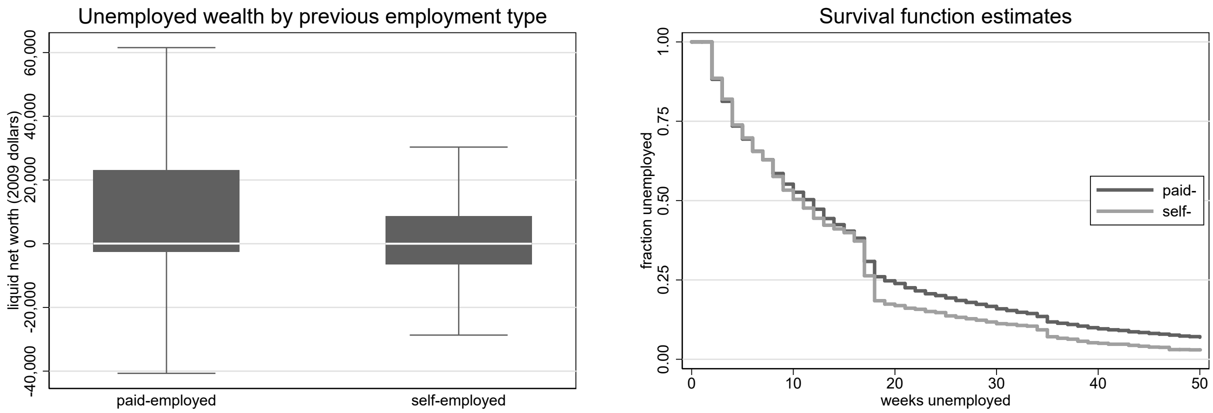
Table 2 also shows that workers in both paid- and self-employment are exposed to unemployment risk. 1.2 percent of workers in paid-employment end up in unemployment each month on average; the figure is .6 percent for workers in self-employment. Here I briefly describe how they differ in terms of their ability to self-insure.

The level of liquid wealth workers can draw on to smooth consumption over an unemployment spell is a first key dimension of self-insurance to unemployment risks. I follow Chetty (2008) in defining a class of liquid assets that individuals can easily access to smooth consumption in the event of job loss or business termination. Net liquid wealth is defined as net worth minus home, vehicle, and business equity at the household level. I use the last reported measure of wealth at or before the start of the spell. Figure 2, left panel, displays the interquartile range for unemployed workers' net liquid assets, broken down by their previous labor form. It shows that workers previously in self-employment who find themselves in unemployment have markedly lower liquid assets.

Another way for workers to self-insure is to adjust their search strategy, hence reducing the length of their unemployment spell. Figure 2, right panel, plots the survival curve in unemployment for each type of worker.[5] Workers previously in self-employment do not appear

[5] I restrict the sample to workers with at least thirteen weeks of employment history. I also truncate the spells after fifty weeks.

Figure 2: Exposure to labor market risks by previous employment type.



Notes: Left panel: Liquid net worth by previous employment status (paid- or self-employed). "Liquid" excludes business, housing, and vehicle equity. Right panel: Kaplan-Meier estimates of survival probability in unemployment by previous employment status. Adjusted for seam effect.

to exit unemployment markedly faster.[6] I confirm this finding in a series of proportional hazard models. Table 3 tests the association between a worker's exit rate from unemployment, her previous employment status and a battery of additional controls. It shows in particular that while the self-employed appear to exit unemployment at a slightly faster rate (first two columns), this pattern is not robust to the inclusion of industry controls.

To sum up, the empirical evidence outlined here points to the self-employed – defined as workers deriving their primary income from a business – not being exceedingly different from traditional wage workers. First, they have a main source of earnings that makes up for most of their labor income. Second, most businesses are relatively small and do not drive large differences in household wealth. Third, there exists substantial transitions across labor market states, including business owners becoming unemployed. Fourth, when in unemployment workers previously in paid- or self-employment do not self-insure in markedly different ways.

[6]Though the two curves appear to diverge after 20 weeks, a Wilcoxon test of equality gives $p = .13$, implying no rejection even at the 10 percent level.

	(1)	(2)	(3)	(4)
self-employed	0.117**	0.126**	0.113**	0.111**
	(0.0359)	(0.0393)	(0.0420)	(0.0426)
non-white	-0.130***	-0.131***	-0.135***	-0.132***
	(0.0271)	(0.0309)	(0.0307)	(0.0309)
woman	-0.0661***	-0.0278	-0.0769***	-0.0427**
	(0.0181)	(0.0173)	(0.0167)	(0.0161)
married	0.138***	0.126***	0.131***	0.126***
	(0.0250)	(0.0261)	(0.0253)	(0.0253)
age controls	Yes	Yes	Yes	Yes
education controls	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
state	Yes	Yes	Yes	Yes
occupation	No	Yes	No	Yes
industry	No	No	Yes	Yes
Observations	13113	11774	11774	11774

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Proportional Hazard models for unemployment duration. All labor force variables (labor form, industry, occupation) refer to the worker's previous employment spell. Occupation and industry are missing for some spells.

3 A search model with self-employment and savings

Based on the empirical evidence outlined in Section 2, I construct a random search model with asset accumulation to jointly capture i) workers' transitions between paid-employment, self-employment, and unemployment; ii) a self-insurance motive. The model builds on Lise (2013) to incorporate a self-employment ladder on top of the traditional wage ladder.[7]

3.1 Model environment

Time is discrete. The labor force is represented by a continuum of working age individuals with measure one. Workers are risk-averse, discount the future at rate $\beta < 1$, and are allowed to borrow and save using a risk-free asset, $a \geq \underline{a}$, with exogenous rate of return r. Their per-period utility of consumption $c > 0$ is given by utility function u.[8]

Workers can be in one of four labor market states: paid-employment (P), self-employment (S), unemployed on UI benefits (B), or unemployed not eligible to benefits (U). Workers can search in either sectors both when they are employed and unemployed. In the baseline model, workers in unemployment are eligible to UI benefits only if they were previously employed by a firm. Self-employed workers terminating their business are not eligible to such transfers.

Search is random. Workers differ in terms of their earnings potential in the labor market. I index worker heterogeneity by $k = 1, ..., K$. This heterogeneity conditions the distribution of labor income from which workers draw when searching. There is a chance that workers will receive an offer in each period. In this event, they draw from the corresponding labor income distributions, F_k^P or F_k^S, respectively for worker of type k drawing a wage or a "self-employment income". This last distribution is the self-employment counterpart to the wage offer distribution in the standard McCall model (McCall, 1970). There is no recall of past offers. Workers are always free to leave their current job or business. In addition, jobs and businesses are exogenously destroyed with probability δ_P and δ_S.

3.2 Timing

Each period can be decomposed into two main stages:

1. Separation and search stage. Workers in employment, both in paid- and self-employment, can choose to quit their job or shut down their business. If not, they are hit by an exogenous separation shock with respective probability $\delta_s, s \in \{P, S\}$. Workers not sepa-

[7] See also Visschers et al. (2014) for a search model with transitions between paid- and self-employment in a directed search framework. Other search models with two separate ladders include Meghir et al. (2015) (formal vs informal employment) and Bradley et al. (2017) (private vs public employment).

[8] $u : \mathbb{R}_+^* \to \mathbb{R}$ is assumed to satisfy $u' > 0$, $u'' < 0$, and $\lim_{c \to 0} u'(c) = \infty$.

rated, as well as those previously in unemployment sample job offer and self-employment ideas with probabilities λ_{sP} and λ_{sS} respectively, where $s \in \{U, S, P\}$. These probabilities are assumed to be mutually exclusive, so that, conditional on searching, workers get at most one labor income draw in each period, either from F_k^P or F_k^S.

2. Consumption and savings stage. Labor income accrues to employed workers. UI benefits get paid to the eligible fraction of unemployed workers. Agents then choose consumption c and next period's assets a'.

3.3 Worker's problem

Notations Let $R_k^s(a, y)$ be the present value of being in state s with net liquid wealth holdings a and labor income y for a worker of type k at the start of the search stage. Similarly, 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 value of getting a draw from paid- or self-employment as

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

where s is the person's current state and $s' \in P, S$. Finally let $\rho_k^{ss'}$ denote the reservation income functions, which condition workers' change of jobs or businesses. These are implicitly defined as $V_k^{s'}(a, \rho_k^{ss'}(a, y)) = V_k^s(a, y)$, the income draw that makes them indifferent between their current labor state s and a draw in labor form s'. (For a draw within the same labor form, the reservation income is simply the person's current labor income irrespective of her asset holdings.)

Paid- and Self-employment The problem faced by workers in paid- and self- employment are similar except for the probability to see the job or business discontinued and the chance to sample from the respective distributions. For example the value of being self-employed at the beginning of the search stage writes

$$R_k^S(a, y) = \max\left\{V^U(a),\ \delta_S V^U(a) + (1 - \delta_S)\Big[V_k^S(a, y) + \lambda_{SP}\mu_k^{SP}(a, y) + \lambda_{SS}\mu_k^{SS}(a, y)\ \Big]\right\}, \quad (1)$$

with the first term in the max operator denoting the option to shut down one's business and become unemployed and the second giving the value of searching. Again, in the baseline model, the self-employed cannot collect UI benefits when becoming unemployed. This is in contrast with workers in paid-employment, who solve

$$R_k^P(a,w) = \max\left\{V^U(a),\ \delta_P V^B(a,w) + (1-\delta_P)\Big[V_k^P(a,w) + \lambda_{PP}\mu_k^{PP}(a,w) + \lambda_{PS}\mu_k^{PS}(a,w)\ \Big]\right\}. \quad (2)$$

Just like in the actual UI system in the US, voluntary quits (the first term in the max operator) do not give rights to unemployment compensation. Displaced workers (an event occurring with probability δ_P in the model), on the other hand, are entitled to benefit payments in proportion to their last wage.

At the consumption and savings stage, an employed worker chooses consumption, c, and next period's assets, $\tilde{a}$, subject to her budget constraint, which yields

$$\begin{aligned} & V_k^s(a,y) = \max_{c,\tilde{a}}\left\{u(c) + \beta R_k^s(\tilde{a},y)\right\} \\ \text{s.t} \quad & c + \frac{\tilde{a}}{1+r} = y + a + y_k^{HH} \\ & \tilde{a} \geq \underline{a}. \end{aligned} \quad (3)$$

The y_k^{HH} term in the budget constraint denotes additional sources of income, which originate either from i) other earners in the household, or ii) welfare payments other than unemployment benefits.

Unemployment To avoid keeping track of unemployment duration when workers become eligible to benefits, I assume that UI payments come as a one-time transfer upon separation.[9] When hit by a δ_P-shock, workers get UI payments $UI(w)$. Their present value at the consumption stage is then given by

$$\begin{aligned} & V_k^B(a,w) = \max_{c,\tilde{a}}\left\{u(c) + \beta R_k^U(\tilde{a})\right\} \\ \text{s.t} \quad & c + \frac{\tilde{a}}{1+r} = UI^P(w) + a + y_k^{HH} \\ & \tilde{a} \geq \underline{a}. \end{aligned} \quad (4)$$

[9] This could be extended to spread out benefits over T periods by defining values $V^{B_1}, ..., V^{B_T}$ for workers eligible to benefits.

In the search stage, the value of being unemployed then simply writes

$$R_k^U(a) = \lambda_{UP}\mu_k^{UP}(a) + \lambda_{US}\mu_k^{US}(a), \tag{5}$$

where when drawing a potential labor income, workers' outside option is $V_k^U(a)$, the counterpart to (4) in the absence of unemployment insurance payments.

3.4 Stationary Equilibrium

Taken together, optimal savings decisions and the reservation incomes across the different states imply a stationary distribution of workers over $\{U, B, S, P\}$, labor income (y, w), and net liquid wealth (a). I denote these distributions Γ_k in what follows, where k indexes worker types. Based on Γ_k a number of statistics with a direct counterpart in the data can be computed, such as transition rates across labor market states and wealth/income distributions.

Formally, for $k \in \{1, ..., K\}$ and $\forall s \in \{U, B, P, S\}$, a stationary equilibrium is a set of value functions V_k^s and R_k^s, policy functions $\hat{a}_k^s$, $\hat{c}_k^s$, reservation labor income functions ρ_k^{sP} and ρ_k^{sS} and a distribution Γ_k^*, such that

1. V_k^s and R_k^s are defined by equations (1)-(5);

2. The asset and consumption choices, $\hat{a}_k^s$ and $\hat{c}_k^s$, solve equations (3) and (4). The reservation wage functions $\rho_k^{ss'}$ are defined by the solutions to $\mu_k^{ss'}$ in equations (1), (2), and (5);

3. Finally define $Q_k : \Gamma_k \mapsto \Gamma'_k$, where Γ'_k is the distribution of workers in the next period, the operator mapping the current distribution of workers to the future one. This operator arises from workers' optimal choices and the transition parameters. The associated stationary distribution of workers is such that $\Gamma_k^* = Q_k \circ \Gamma_k^*$.

4 Unobserved heterogeneity in worker ability

To discipline worker heterogeneity in earnings potential, I rely on tools from machine learning to classify workers in a first step. The model is then solved in a second step, conditional on this discretization. This section starts by describing the classification procedure used, before turning to the details of its implementation in the SIPP data.

4.1 Using k-means to classify workers

Following Manresa et al. (2017) and Bonhomme et al. (2019), I use a k-means algorithm to classify workers in ability groups, or labor earnings potential, in the SIPP data. This procedure is a standard clustering method that finds the best partition of the data according to the following objective function

$$\underset{\tilde{h},k_1,...,k_N}{\arg\min} \sum_{i=1}^{N} \left\| \hat{h}_i - \tilde{h}(k_i) \right\|^2, \tag{6}$$

where, using their notation, N is sample size, $k_i \in \{1, ..., K\}$ are partitions of $\{1, ..., N\}$ with $1 < K \leq N$, $\hat{h}_i$ is a vector of features used for classification, and $\tilde{h}(k_i)$ is the corresponding vector of features for group k to which i is assigned. Each element in $\tilde{h}(k_i)$ is computed by averaging over the members of the group. The solution to Equation (6) 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.[10]

The logic behind this classification step is closest to that described in Bonhomme et al. (2019). These authors first cluster firms based on their empirical distribution of earnings amongst employed workers, before using the resulting classes in a series of mixture models. The estimated partition then captures both observed and unobserved firm heterogeneity. In my framework, I rely on a similar partition of the data to discipline workers' earning potential, which in the model translates into class-specific distributions of employment opportunities in paid-employment F_k^P and self-employment F_k^S, where k indexes worker type. This set of distributions can be seen as the equivalent to the firm-class fixed effects in Bonhomme et al. (2019); they similarly capture heterogeneity in labor earning potentials that can be correlated with both observed and unobserved characteristics.

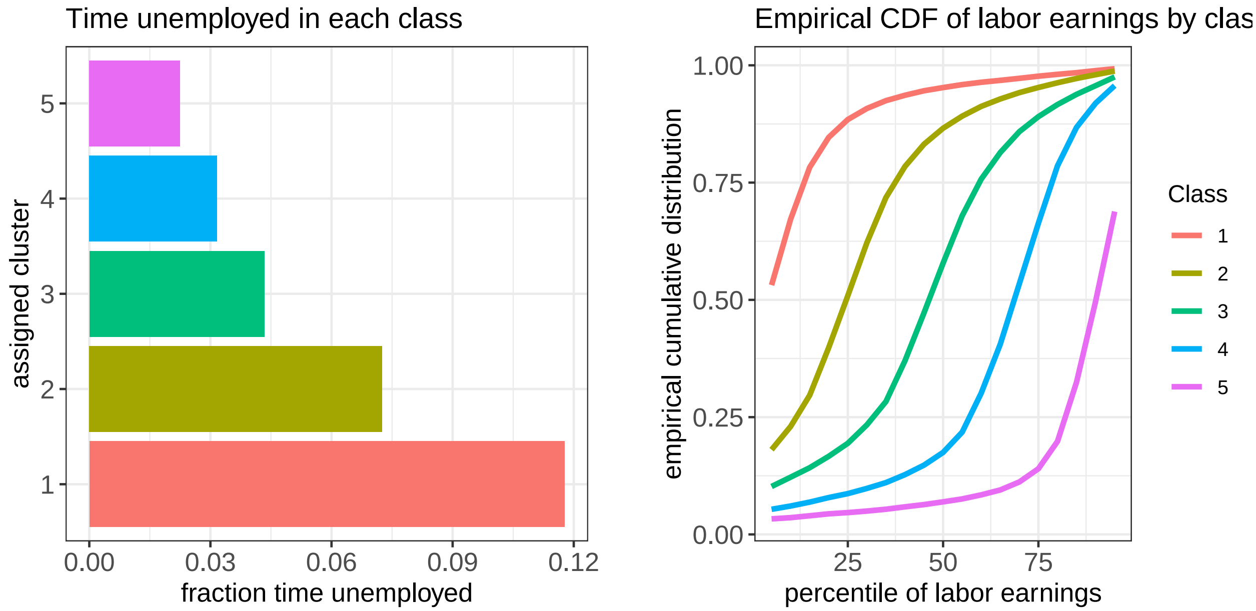
4.2 Ability groups in the SIPP data

To implement this classification step in practice, one needs to choose the number of classes (K) and the vector of features $\hat{h}_i$ on which the classification operates. I tentatively set $K = 5$, which offers a good trade-off between capturing a reasonable degree of heterogeneity in worker type and computation time.[11] The vector of features $\hat{h}_i$ includes first a measure of

[10]Standard algorithms to efficiently solve this global minimization problem are available in standard packages. In practice, I use the implementation of the "Hartigan-Wong" algorithm in the base R function "kmeans".

[11]Figure 7 shows how the total within sum of square changes with the number of clusters. Improvements in fit appear to flatten out after about five-six clusters.

Figure 3: Worker labor earnings classification in the SIPP data



Notes: The figure depicts the estimated clusters ("centers") obtained when classifying workers in five groups. The left-panel shows a measure of labor market attachment as the fraction of time (in months) the person spends unemployed over the survey period. The right panel depicts the empirical CDF of labor earnings. Classes were ordered from weakest to strongest labor market attachment.

labor market attachment

$$\widehat{LMA}_i = \frac{1}{T_i} \sum_t \mathbb{1}\{U_{it} = 1\}$$

where T_i is the number of months the individual spends in the panel (typically about five years in the SIPP 2008 panel) and $U_{it} = 1$ if she is unemployed in period t. Its second key component is the empirical CDF of labor market earnings, irrespective of whether this income comes from paid- or self-employment,

$$\widehat{ECDF}_i(y_p) = \frac{1}{T_i - \sum_t \mathbb{1}\{U_{it} = 1\}} \sum_t \mathbb{1}\{y_{it} \leq y_p\}$$

with y_{it} denoting labor income for i in period t. I compute this empirical CDF for twenty quantiles of the empirical distribution of labor earnings in the sample.

These variables were chosen as they directly relate to the distribution of earning draws, F_k^P and F_k^S, in the model. The resulting estimated clusters for labor market attachment (left) and the empirical CDF of labor market earnings (right) are shown in Figure 3. It suggests that a clear partition exists, where low earners also have a weaker labor market attachment.

5 Calibration

5.1 Functional forms

I first assume that utility is given by a standard CRRA specification $u(c) = (1-\sigma)^{-1}c^{1-\sigma}$. I set the risk-aversion parameter σ to two throughout, a standard value in the literature (Lise, 2013; Saporta-Eksten, 2014).

The labor income initial draws, which capture unobserved worker heterogeneity, are assumed to follow a truncated Pareto distribution. This assumption gives the following set of parameters to be calibrated $\{\underline{y}_k^s,\ \overline{y}_k^s, \alpha_k^s\}$ for each worker type $k \in \{1, ..., K\}$ and labor form $s \in \{P, S\}$. The first two elements denote the bounds of the support of the distribution, while α_k^s is its shape parameter. I set the support of the distribution to lie between the 2nd and 98th percentile of the empirical income distribution in the SIPP, for each class $k \in \{1, ..., K\}$ and employment form $s \in P, S$. The shape parameters are internally calibrated as described below.

5.2 Baseline welfare system

The baseline welfare system in this economy is made of two key components. As noted above, unemployment benefits $UI(w^{\text{last}})$ are modeled as a one-time payment to avoid keeping track of unemployment duration in the worker's problem. Following Saporta-Eksten (2014), I further approximate the unemployment insurance system currently in place in the United States as

$$UI^P(w^{\text{last}}) = .5\ *\ \text{present value of last } w \text{ over 6 months}$$

since most workers are eligible to a 50% replacement rate for six months when being laid-off.[12]

Besides, the auxiliary income in the household's budget constraint also captures some additional welfare payments to which the household may be eligible. It is parameterized directly from the data, using total household income including welfare payments. I set auxiliary income to the median of that measure of household income within each worker class, when the main earner is in unemployment, subtracting any reported unemployment benefits.

[12]This formula is a very coarse approximation of the actual UI system, which varies by State, has some employment duration requirements, and some ceiling for high earners.

Table 4: Externally calibrated parameters

Parameter	Description	Value	Target/Source
σ	CRRA utility parameter	2	(Lise, 2013; Saporta-Eksten, 2014)
r	risk-free rate	$(1+.045)^{1/12}$	4.5% annual return
δ_p	separation rate (paid-)	0.012	SIPP
δ_s	termination rate (self-)	0.0058	SIPP

5.3 Externally calibrated parameters

Several parameters are either calibrated externally based on commonly accepted values in the literature or taken directly from the data. They are listed in Table 4, which also reports the corresponding targets.

5.4 Internally calibrated parameters

Table 5 displays the seventeen remaining parameters that are calibrated internally by minimizing the distance between a set of model simulated moments and their empirical counterpart.[13] These features of the data are chosen to capture the self-employed's exposure to labor market risk (transitions in and out of unemployment) and their ability to self-insure (net liquid assets and income distributions).

Table 6 and Figure 4 report the model fit to the targeted moments. Table 6 displays the fit to workers' transition rates across the three labor forms. Figure 4 shows the net liquid asset (left column) and labor income (right column) distributions in the model and the data.

[13]Given a parameter vector θ, the objective function is

$$(\mathbf{m}_{\text{data}} - \mathbf{m}_{\text{model}}(\theta))^{\intercal}\mathbf{\Lambda}(\mathbf{m}_{\text{data}} - \mathbf{m}_{\text{model}}(\theta))$$

with $\mathbf{m}_{\text{data}}$ the vector of data moments and $\mathbf{m}_{\text{model}}(\theta)$ the corresponding model generated vector of moments. Each moment is rescaled by the inverse of the square of its empirical value: $\mathbf{\Lambda} = \text{diag}(1/\mathbf{m}^2_{\text{data}})$.

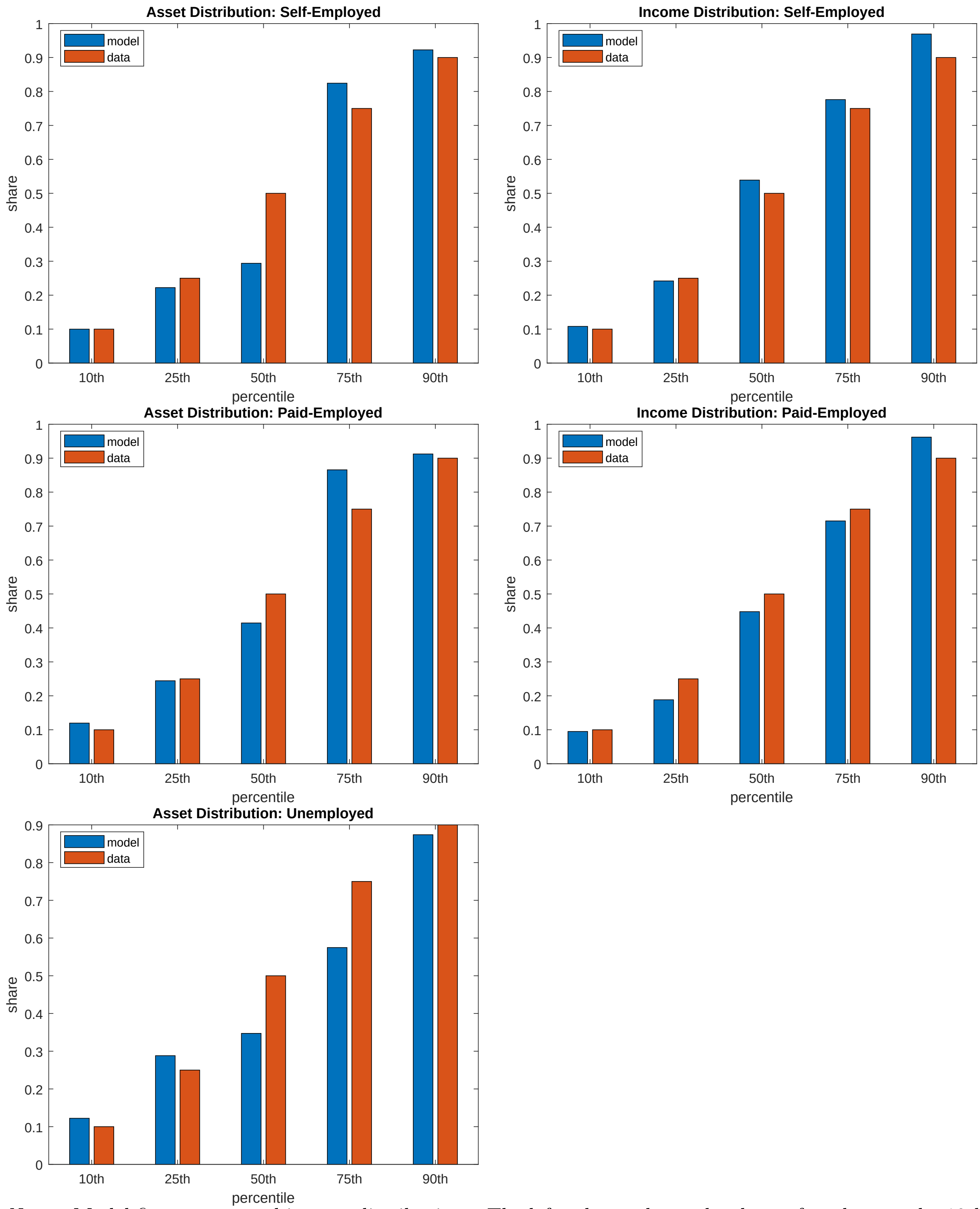
Table 5: Internally calibrated parameters

Parameter	Value	Description
λ_{PP}	0.093	
λ_{SS}	0.200	$\lambda_{ss'}$ is chance to sample from $F_k^{s'}$, when in state s
λ_{SP}	0.060	s: origin state
λ_{PS}	0.066	$F_k^{s'}$: labor income draw for worker of type k in state s'
λ_{UP}	0.151	
λ_{US}	0.072	
β	0.995	Monthly discount factor
α_1^P	1.001	
α_1^S	2.655	
α_2^P	2.798	
α_2^S	4.997	Income draw is parametrized as truncated Pareto($\underline{y}_k^s$,$\overline{y}_k^s$,α_k^s)
α_3^P	2.304	$\underline{y}_k^s$: p02 of income distribution for type k in state s
α_3^S	2.333	$\overline{y}_k^s$: p98 of income distribution for type k in state s
α_4^P	1.431	α_k^s: shape parameter for type k in state s
α_4^S	2.264	
α_5^P	1.071	
α_5^S	2.912	

Table 6: Model fit to monthly transitions

	Model	Data
UP rate	0.1409	0.1486
US rate	0.0097	0.0097
SP rate	0.0095	0.0096
SS rate	0.0054	0.0054
PS rate	0.0009	0.0009
PP rate	0.0146	0.0139
PU rate	0.0112	0.0116
SU rate	0.0058	0.0058

Figure 4: Fit to assets and income distributions



Notes: Model fit to assets and income distributions. The left column shows the share of workers at the 10th, 25th, 50th, 75th and 90th percentiles of the net liquid assets distribution in the model and the data. (The 10th percentile is .1 in the data by construction.) The right column displays similar shares for the labor income distributions of these workers in paid- and self-employment.

6 Unemployment insurance for the self-employed

What are the welfare effects of extending benefit entitlements to the self-employed? This section describes the impact of one set of policies: mandating unemployment insurance contributions in exchange for a one-time payment in the event the person becomes unemployed. I use the calibrated model to decompose welfare changes for each group of workers.

6.1 Baseline policy

I analyze the impact of introducing a UI system similar to the one in place for workers in paid-employment in the model. It is made of a replacement rate, b, such that workers get

$$UI^S(y^{\text{last}}) = b \ * \ \text{present value of last } y \text{ over 6 months}$$

upon becoming unemployed and a mandatory rate of contribution, τ, applied to labor earnings for the self-employed.

τ is set to balance the policy within the group of self-employed

$$\tau \int y d\Gamma^S_{\tau,b}(a,y) = (1-\tau) \int UI^S(y) d\Gamma^C_{\tau,b}(a,y) \tag{7}$$

where $\Gamma^s_{\tau,b}$ denotes the stationary distribution of workers over wealth a and income y across types and in state s under self-employment UI policy (b, τ). Labor market $s = C$ stands for the group of self-employed who becomes unemployed.

In this baseline case, I assume that there is full moral hazard pass-through in the sense that the self-employed can choose to terminate their activity. Going back to the value of self-employment at the beginning of the search stage (2), this expression now reads

$$R^S_k(a,y) = \\ \max\left\{V^C_k(a,y),\ \delta_S V^C(a,y) + (1-\delta_S)\Big[V^S_k(a,y) + \lambda_{SP}\mu^{SP}_k(a,y) + \lambda_{SS}\mu^{SS}_k(a,y)\ \Big]\right\},$$

where V^C_k is the value of becoming unemployed when benefits are paid out. There are then two ways for the self-employed to cash these benefits: either by being hit by a δ_S-shock (involuntary) or by "choosing" unemployment in the last expression (voluntary).

Worker class (\$$y_k^{HH}$)	E(contributions)	E(benefits)	Ratio cont. to ben.
1208	3.7	6.2	1.69
1628	8.1	8.5	1.05
2068	0.4	5.3	12.45
2588	21.4	16.1	0.76
2879	13.4	10.9	0.81

Table 7: Expected contributions and benefits to the self-employment UI scheme. Worker class is indicated by y_k^{HH}, the median household income when the worker is in unemployment. E(contributions) and E(benefits) show the expected ex-ante monthly contributions and benefits, not conditioning on the person's labor market status. Amounts in 2009 dollars.

6.2 Equal treatment case: $b = .5$

I first center on the case where UI payments are brought in line with the paid-employed. The replacement rate is set to $b = .5$ and the UI contribution rate τ satisfies the budget balance condition (7).

Table 7 displays ex-ante expected contributions and payments to the self-employment UI scheme by worker type. It shows that the policy is clearly redistributive: workers in higher income groups are net contributors to the system.

To pin down the insurance value of introducing a UI scheme for the self-employed, I compute two statistics. Following Krusell et al. (2010), I first define the compensating variation in consumption as the constant $\Delta_{b,\tau}^{\text{comp}}$ solving

$$E_t \sum_{\tilde{t}=0}^{\infty} \beta^{\tilde{t}} u((1 + \Delta_{b,\tau}^{\text{comp}}) c_{t+\tilde{t}}) = E_t \sum_{\tilde{t}=0}^{\infty} \beta^{\tilde{t}} u(\hat{c}_{t+\tilde{t}}),$$

where $\hat{c}$ denotes the consumption path in the economy with the self-employment UI scheme. I also report $\Delta_{b,\tau}^{\text{transfer}}$ as the cash transfer that makes the worker indifferent between these two economies: $V(a + \Delta_{b,\tau}^{\text{transfer}}) = \hat{V}(a, y)$, where $\hat{V}$ is the present value of consumption in the economy with $b = .5$.

Table **??** displays these statistics by worker type, taking an unconditional average (left columns) and conditioning on the person's current labor market state. It suggests that the policy has a positive, if modest, insurance value for all groups of workers. Focusing on the self-employed who become unemployed (columns $s = C$), the liquidity value of being in the economy with the UI scheme averages to about a one-time \$9,000 cash grant (or about a .9% increase in consumption).

Worker class ($\$y_k^{HH}$)	replacement (b_S^*)	contribution (τ_S^*)
1208	0.90	0.106
1628	0.84	0.061
2068	0.44	0.043
2588	0.67	0.031
2879	0.77	0.030
all	0.80	0.052
equal treatment	0.50	0.013

Table 8: Optimal policies by worker type. The replacement rate (b^*) and social contribution rate (τ^*) are the solution to (8) for each class of worker taken in isolation. Worker class is indicated by y_k^{HH}, the median household income when the worker is in unemployment. "all" is the solution to (8) for the whole economy. "equal treatment" is the case where b is aligned with the paid-employed.

6.3 Optimal policies

One can also derive the socially desirable policy (b^*, τ^*) given some social welfare function.[14] In the model's notations, the utilitarian welfare function for workers of type k writes

$$\Omega_k(b, \tau) := \sum_s \int V_k^s(a, y) d\Gamma_k^s(a, y),$$

the expected discounted utility of consumption for type k. Letting ω_k be the share of each worker type, aggregate social welfare is then given by $\Omega(b, \tau) := \sum_k \omega_k \Omega_k(b, \tau)$. The utilitarian planner's objective is then to

$$\max_{b,\tau} \Omega(b, \tau) \quad \text{s.t.} \quad \tau \int y d\Gamma_{\tau,b}^S(a, y) = (1 - \tau) \int UI^S(y) d\Gamma_{\tau,b}^C(a, y). \tag{8}$$

Table 8 displays optimal policies, defined as the solution to (8), for all workers and for each class k taken separately. The goal of this last experiment is to determine what the planner would do were she able to mandate a different policy for each worker group. These results suggest that there are some substantial trade-offs between the level of benefits offered and the mandatory level of contributions to support the system. In particular, a lower replacement rate is optimal for individuals with lower earnings potential (first row), as social contributions would become prohibitively high for larger bs.

[14]Optimal within the set of policies studied in this paper.

References

Acemoglu, D. and R. Shimer (1999, October). Efficient Unemployment Insurance. *Journal of Political Economy 107*(5), 893–928.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica 87*(3), 699–739.

Bradley, J. (2016, June). Self-employment in an equilibrium model of the labor market. *IZA Journal of Labor Economics 5*(1), 6.

Bradley, J., F. Postel-Vinay, and H. Turon (2017). Public sector wage policy and labor market equilibrium: A structural model. *Journal of the European Economic Association*, jvw026.

Catherine, S. (2017). Keeping Options Open: What Motivates Entrepreneurs? pp. 57.

Chaumont, G. R. and S. Shi (2017, January). Wealth Accumulation, on the Job Search and Inequality. SSRN Scholarly Paper ID 2893777, Social Science Research Network, Rochester, NY.

Chetty, R. (2006, November). A general formula for the optimal level of social insurance. *Journal of Public Economics 90*(10–11), 1879–1901.

Chetty, R. (2008, April). Moral Hazard versus Liquidity and Optimal Unemployment Insurance. *Journal of Political Economy 116*(2), 173–234.

Fujita, S. and G. Moscarini (2017, December). Recall and Unemployment. *American Economic Review 107*(12), 3875–3916.

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

Hansen, G. D. and A. Imrohoroglu (1992, February). The role of unemployment insurance in an economy with liquidity constraints and moral hazard. *Journal of Political Economy 100*(1), 118.

Humphries, J. E. (2018). The Causes and Consequences of Self-Employment over the Life Cycle. pp. 83.

Hurst, E. and B. W. Pugsley (2011, May). What Do Small Businesses Do? Working Paper 17041, National Bureau of Economic Research.

Kolsrud, J., C. Landais, P. Nilsson, and J. Spinnewijn (2015). The Optimal Timing of Unemployment Benefits: Theory and Evidence from Sweden. STICERD - Public Economics Programme Discussion Paper 25, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Krusell, P., T. Mukoyama, and A. Şahin (2010, October). Labour-Market Matching with Precautionary Savings and Aggregate Fluctuations. *Review of Economic Studies 77*(4), 1477–1507.

Lamadon, T. Productivity Shocks, Long-Term Contracts and Earnings Dynamics. pp. 86.

Lise, J. (2013, January). On-the-Job Search and Precautionary Savings. *The Review of Economic Studies 80*(3), 1086–1113.

Manresa, E., T. Lamadon, and S. Bonhomme (2017, March). Discretizing unobserved heterogeneity. Technical report, The IFS.

McCall, J. J. (1970). Economics of Information and Job Search. *The Quarterly Journal of Economics 84*(1), 113–126.

Meghir, C., R. Narita, and J.-M. Robin (2015, April). Wages and Informality in Developing Countries†. *American Economic Review 105*(4), 1509–1546.

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

Saporta-Eksten, I. (2014). Job loss, consumption and unemployment insurance. *Unpublished manuscript*, 1–50.

Shi, S. (2009, March). Directed Search for Equilibrium Wage–Tenure Contracts. *Econometrica 77*(2), 561–584.

Shimer, R. and I. Werning (2008, December). Liquidity and Insurance for the Unemployed. *American Economic Review 98*(5), 1922–1942.

Visschers, L., A. Millan, and M. Kredler (2014). Great opportunities or poor alternatives: Self-employment, unemployment and paid employment over the business cycle. Technical Report 597, Society for Economic Dynamics.

A Data

The main data source is the Survey of Income and Program Participation (SIPP), a survey centred on how Americans use welfare programs.[15] This Appendix describes the sample selection and expands on the definition of paid-employment and self-employment.

[15] The data and corresponding documentation can be accessed from the Census Bureau's website: `https://www.census.gov/programs-surveys/sipp/`

Sample selection. All data in this paper come from the 2008 panel. This choice is guided by the greater consistency of job and business identifiers and the more comprehensive definition of business income in this specific panel.[16] I further restrict the analysis to individuals aged at least 25 or at most 65 across the survey. In each household, I only keep the main earner, defined as the individual with the largest labor earnings across the survey within each household and exclude workers who make transitions out of the labor force (see below for definition). The aim of these restrictions is to reduce the sample to the individuals most likely to be strongly attached to the labor market. Taken together, these conditions yield a dataset of 24,451 individuals which are followed on average for 51 months.

Definition of labor force status in the SIPP. 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 status (employed, on layoff, unemployed, non-participating) and information on up to two jobs and two businesses (such as job/business identifier, wages, profits).

Employed individuals are classified as paid- or self-employed on the basis of the job or business for which they report the largest average earnings. This is done on the basis of the job and business identifiers provided in the survey, which are cleaned in a first step to be consistent with the start and end date reported for each job or business.[17] However, these variables are generally only consistent for workers continuously in employment, so this assignment procedure identifies a worker's main activity within an employment spell (see Fujita and Moscarini, 2017, Section II for details).

The SIPP reports three types of non-employment week-by-week: on layoff, no job looking for work, no job not looking for work. To address potential differences in reporting between previously paid- and self-employed workers, I define as unemployment spells any non-employment spell of at most fifty weeks if the worker looks for a job at least 50 percent of the time or the spell ends up in employment over the duration of the survey.

Finally, I follow the convention in the Current Population Survey and build a monthly panel of employment status for each individual based on the second week (the first full week) in each month.

B Additional tables and figures

[16] See the following link for details: `https://www.census.gov/programs-surveys/sipp/tech-documentation/user-notes/core_notes_2008-General-User-Note.html`.

[17] I am grateful to Fabien Postel-Vinay for sharing his code to clean these identifiers.

Figure 5: Benchmark of aggregates derived from the SIPP with the corresponding Bureau of Labor Statistics series. The unemployment and self-employment rates in the SIPP are based on the definitions of labor force status given in the main text. Observations are weighted using the provided cross-sectional weights.

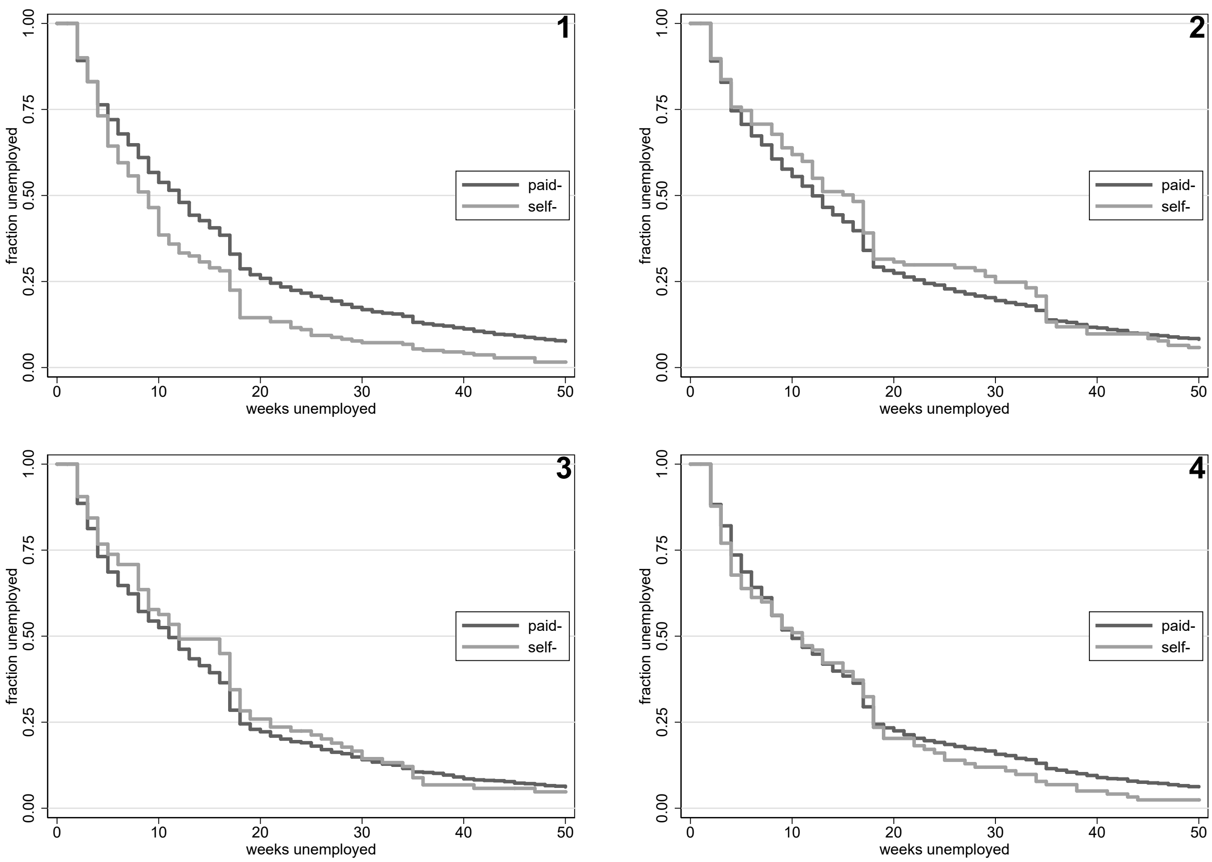


Figure 6: Survival function by net liquid wealth quartile. Kaplan-Meier estimates by net liquid wealth quartile and previous employment status (paid- and self-). The digit in the top-right corner corresponds to the individual's household's net liquid wealth at or before the start of the unemployment spell.

	(1)	(2)	(3)	(4)
self- x Q1 wealth	0.230*	0.187	0.204	0.193
	(0.107)	(0.108)	(0.110)	(0.107)
self- x Q2 wealth	0.0855	0.0424	0.0517	0.0398
	(0.0928)	(0.103)	(0.100)	(0.103)
self- x Q3 wealth	0.0337	0.0308	0.0304	0.0308
	(0.121)	(0.123)	(0.126)	(0.124)
self- x Q4 wealth	0.122	0.124	0.127	0.112
	(0.0928)	(0.0956)	(0.0923)	(0.0951)
non-white	-0.134***	-0.126***	-0.128***	-0.123***
	(0.0324)	(0.0355)	(0.0346)	(0.0348)
woman	-0.0550*	-0.0705*	-0.0110	-0.0271
	(0.0254)	(0.0280)	(0.0259)	(0.0286)
married	0.137***	0.128***	0.114**	0.115**
	(0.0373)	(0.0374)	(0.0374)	(0.0373)
age controls	Yes	Yes	Yes	Yes
education controls	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes
state	Yes	Yes	Yes	Yes
occupation	No	No	Yes	Yes
industry	No	Yes	No	Yes
Observations	7393	7090	7090	7090

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Proportional Hazard models for unemployment duration. All labor force variables (labor form, industry, occupation) refer to the worker's previous employment spell. Occupation and industry are missing for some spells.

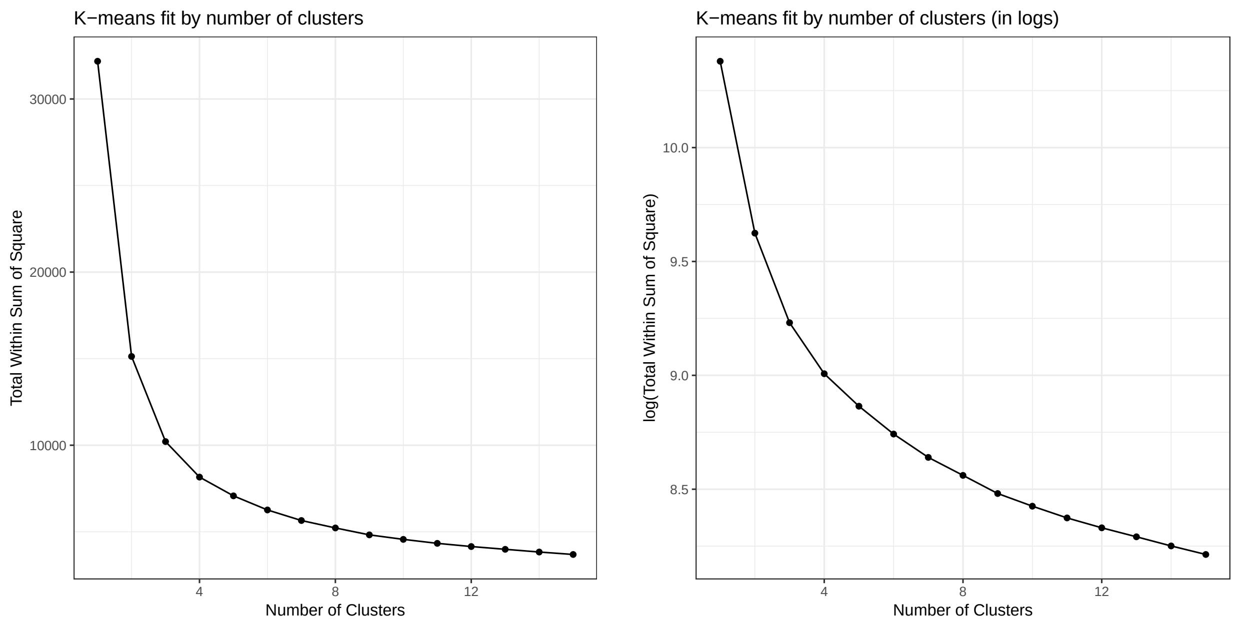


Figure 7: Change in kmeans fit with number of clusters. The left panel shows a measure of fit for k-means, the total within sum of squares, as the number of clusters increases. The right panel shows the same figure with the total sum of squares in logs.