UNIVERSITY COLLEGE LONDON

ESSAYS IN MACROECONOMICS AND LABOR ECONOMICS

Richard Audoly

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in the

Department of Economics

June 2020

Declaration of Authorship

I, Richard Audoly, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Richard Audoly
June 26, 2020

Abstract

This doctoral thesis is made up of three chapters at the intersection between macroeconomics and labor economics, all dealing with topics related to search frictions in the labor market.

In the first chapter, I develop a tractable model of firm and worker reallocation over the business cycle that emphasizes the interplay between firms with heterogeneous productivities and on-the-job search. I use this framework to study the role of search frictions in determining aggregate labor productivity following a large economic contraction. In the model, search frictions slow down worker reallocation after a recession, as employed workers face increased competition from a larger pool of unemployed workers. This crowding-out effect holds back the transition of employed workers from less to more productive firms, thus lowering aggregate productivity. Quantitatively, the model implies that worker reallocation has sizable and persistent negative effects on aggregate labor productivity. I provide evidence for this channel from data on the universe of British firms which show that the allocation of workers to firms has downgraded in the aftermath of the Great Recession.

In the second chapter, I study the unemployment risks faced by self-employed workers. Though public unemployment insurance (UI) schemes represent an important feature of the social safety net in most advanced economies, the self-employed are generally excluded from these programs. This chapter shows that, similarly to employees on a wage contract, the self-employed do go through unemployment spells in US data. It then calibrates a job search model to evaluate the potential welfare gains from extending UI benefits to this group of workers. The model features workers moving between paid- and self-employment who face the risk of becoming unemployed. Agents can also privately save and borrow to self-insure. My results suggest that extending UI benefits to the self-employed yields modest welfare gains.

In the third chapter, I use longitudinal data on patents to quantify sorting in knowledge production. The dimension of sorting I study is that

arising between inventors and their “firm” (private corporations, universities, public research institutes). My analysis points to the existence of clear, positive inventor-firm sorting. This mechanism accounts on average for five percent of the total variance of inventor output in the US between 1975 and 2010. This framework further suggests that the geographical sorting of inventors and firms is a key channel to explain regional disparities in inventor output.

Impact Statement

A common thread in this thesis is to understand how impediments to individual choices in the labor market explain aggregate distributions. These impediments, known as “search frictions” in the literature, capture the information problems that prevent workers from moving to better paying employers, self-employed workers to find a regular wage contract when their business is failing, and inventors to locate in the best research environments.

As shown in this work, the “search” paradigm is a particularly versatile framework, which can be used to grasp a variety of economic phenomena. It is therefore both at the center of a very active strand of academic research and an insightful tool to inform policymakers. I give some examples below of how my work in this thesis could have an impact in academia and in policymaking.

In academic terms, first, this thesis contributes to a line of research in labor economics and macroeconomics that aims at modeling the interaction between firms and workers to understand aggregate labor market outcomes. My first chapter is concerned with introducing a notion of a firm, a group of several workers, affected by both firm-level and macroeconomic events. I apply this framework to the Great Recession in the UK to analyze how the drop in worker transitions across firms feeds into the slow recovery in aggregate labor productivity observed following the last recession. Several other recent working papers have proposed models with similar features, but different underlying assumptions, in the last year (Elsby and Gottfries, 2019; Bilal et al., 2019). I therefore see this chapter as making a potentially useful methodological contribution on which other researchers can build in the future.

In terms of policy applications, next, my second chapter provides several insights that can guide policymakers in designing welfare policies for the self-employed, the group of workers not bound to an employer by a traditional wage contract. There is currently a lot of policy interest in the subgroup of self-employed workers matched to their customers through online platforms, who are sometimes collectively referred to as “gig economy" workers. I first show in the data that, even before the advent of these platforms, there is a subgroup of self-employed with low earnings, low liquid wealth, and who oscillate between wage work, self-employment, and unemployment. There is therefore a potential policy case to provide insurance to this group, who is not traditionally covered by government-managed unemployment insurance schemes. In addition, any such policy should take into account the diversity of labor market situations encompassed by self-employment, from small business owners to partners in law firms, to understand who would be the primary beneficiaries. This chapter of my thesis provides a quantitative framework to quantify who gains and who loses as a result of this type of policies.

Acknowledgments

I am deeply indebted to Fabien Postel-Vinay and Jan Eeckhout for their invaluable advice and guidance over the course of my graduate studies. I also wish to thank Vincent Sterk for being so generous with his time, as well as Giuseppe Moscarini for making a research visit at Yale University possible. Any errors are, of course, my own.

I am extremely grateful to everyone in the research community at the Department of Economics, University College London, and the Institute for Fiscal Studies. Evolving day-to-day in such a fantastic research environment has clearly made this work better.

I finally gratefully acknowledge funding from the Institute for Fiscal Studies through an ESRC Graduate Scholarship and from the Yale-UCL Student Exchange Programme.

Contents

List of Figures

List of Tables

Introduction

The underlying concept throughout this thesis is the existence of search-and-matching frictions in the labor market. These frictions prevent workers from immediately locating potentially better alternatives in the labor market, such as better employers (Chapter 1), business opportunities (Chapter 2), or research environments (Chapter 3). These three distinct applications show the versatility of this conceptual framework as a tool to analyze different facets of the labor market.

The opening chapter “Firm Dynamics and Random Search over the Business Cycle” proposes a tractable model of firm and worker reallocation over the business cycle. The three key features of this framework are: (i) firm dynamics to study the dynamic selection of firms, (ii) random search with on-the-job search to think through the reallocation of workers in-and-out of unemployment and between firms, (iii) aggregate shocks to analyze the response of this economy over the business cycle. I provide several analytical results that render the solution of the model tractable, though it is still necessary to use numerical methods to reduce the dimension of the problem in solving the model with aggregate shocks. I present an application to the Great Recession in the UK in which I use the calibrated model to analyze the persistent drop in labor productivity that followed the trough of the last economic downturn. The model allows to isolate two opposite forces that shape the evolution of aggregate labor productivity following a negative economy-wide shock. There is first a cleansing force, which makes the selection of businesses more stringent and improves aggregate productivity in the economy. Conversely, recessions also slow down the reallocation of workers from less to more productive firms in the model, as workers at surviving low productivity firms compete with a larger pool of unemployed workers for better jobs, a “sullying” effect. The model allows to isolate these two forces from the direct impact of aggregate shocks. In my application to the UK Great Recession, I show that the second “sullying” force dominates in the medium term, as search frictions entail a slow process of reallocation of workers to firms.

The second chapter "Self-employment and Unemployment Risks" studies the labor market risks faced by the self-employed. The motivation for this question comes from the fact that, in most OECD economies, the self-employed are excluded from public unemployment insurance schemes (OECD, 2018a). At the same time, the very large theoretical and empirical literature studying the optimal trade-off between insurance and moral hazard in providing benefits to unemployed workers has ignored the self-employed. This chapter first provides empirical evidence from longitudinal US data on the unemployment risks faced by the self-employed. I show that, using commonly accepted definitions of unemployment, the self-employed do go through unemployment spells, if less frequently than traditional wage workers. It also emerges that a substantial fraction of the self-employed have very little liquid wealth to self-insure. In the second part of the chapter, I set up a model to quantify the potential welfare gains of including the self-employed into a standard public unemployment insurance scheme: mandatory contributions while employed and payments contingent on previous earnings in the event of unemployment. The model features workers moving across three different labor force states (wage work, self-employment and unemployment), who can also accumulate wealth privately to self-insure. My calibration suggests that extending UI benefits to the self-employed yields modest positive welfare gains, which are concentrated amongst the lowest earners.

The third chapter "Sorting in Knowledge Production: Evidence from Patent Data" analyzes one specific dimension of sorting in knowledge production: to which extent the most productive inventors sort into the best research institutions? I use longitudinal patent data from the United States Patent Office to answer this question. These data make available inventor and "firm" (private corporations, research universities, public research institutes) identifiers, thus allowing to fit statistical models with two-sided unobserved heterogeneity. The notion of match-level output I consider is patent quality, which I define as the number of citations a patent receives. I can then decompose the variation in inventor output into an inventor component, a firm component, and a sorting term, mirroring the decomposition traditionally performed on matched employer-employee data. My results suggest that there is positive inventor-firm sorting. Across the time intervals I consider in my sample, which go from 1975 to 2009, sorting accounts for about five percent of the overall variance of inventor output. The correlation coefficient between inventor and firm fixed effects ranges from .12 to .25 across periods. I do not, however, find evidence of a clear trend in this sorting pattern over the period I study.

Chapter 1

Firm Dynamics and Random Search over the Business Cycle

1.1 Introduction

Aggregate labor productivity can be thought of as stemming from two components at the micro level: a distribution of firms with heterogeneous productivity levels and a distribution of workers across these firms. In this paper, I propose a tractable model in which both of these components evolve endogenously over the business cycle. This framework is motivated by a series of empirical regularities on firm and worker reallocation during recessions. On the firm side, the number of active businesses substantially drops during an economic contraction.[1] On the worker side, recessions both markedly increase flows into non-employment and decrease the pace at which unemployed workers find jobs. In addition, the rate at which employed workers make direct job-to-job transitions between employers also slows down sharply. But whether these changes in firm and worker flows reallocate workers to more productive firms is an open question.[2]

This paper argues that the slower pace of job-to-job transitions observed during recessions acts as a dampening channel for labor productivity. Conceptually, the economic mechanism relating these transitions to productivity is the existence of a productivity ladder in equilibrium: workers move from less to more productive firms when making a direct transition between employers since these firms offer more attractive jobs. Recessions

[1]The number of active firms shrank by, respectively, five (ten) percent in the US (UK) between 2008 and 2011.

[2]Classic models of the "cleansing" effect of recessions include Caballero and Hammour (1994) and Mortensen and Pissarides (1994). Barlevy (2002) and Ouyang (2009) are examples of papers putting forward a "sullying"/"scarring" channel. See Foster et al. (2016) for additional references.

slow down the transitions of employed workers up the productivity ladder, as these workers face increased competition for jobs from a larger pool of unemployed workers. I use the model to assess the quantitative significance of this mechanism in holding back labor productivity during the Great Recession in the UK. The model allows to quantify the magnitude of this worker reallocation channel (a negative effect) relative to the impact of firm selection (a positive effect) in shaping labor productivity after a recession.

To gage the importance of worker reallocation for aggregate productivity, I start by building a labor productivity index from the ground up, aggregating from British firm-level administrative data over the period 2000-2016. This data set makes output and employment available for all active firms in the economy, thus allowing to define labor productivity at the firm level. Besides, it also covers about a decade before and after the Great Recession – officially starting in 2008Q2 in the UK – the largest post-war economic contraction in Britain. I can then study separately the component of aggregate labor productivity coming from the productivity of individual firms and that arising from the allocation of labor to those firms before and after the Great Recession. I find that the allocation of labor to firms is significantly downgraded following the last recession. Firm-level regressions confirm that the positive relationship between the labor productivity of firms and their employment grow in the next period is weaker post-recession. I see these facts as evidence that less productive firms represent a dampening channel for labor productivity during the UK Great Recession and interpret them through the lens of the calibrated model.

A key contribution of this paper is to develop a tractable model of firm and worker reallocation to study these empirical patterns. My framework combines the three following features: aggregate shocks, search frictions, and firm dynamics. Aggregate shocks are a pre-requisite to studying the evolution of labor productivity over the business cycle. Search frictions in the labor market constrain the transition of workers out of unemployment. In the spirit of the random search framework with on-the-job search proposed by Burdett and Mortensen (1998), I also allow workers to search while employed. While it complicates the solution of the model, this addition is central since (i) about half of gross job creation and destruction flows originate in direct employer-to-employer transitions in the data, so these transitions matter quantitatively for worker reallocation,[3] (ii) Barlevy (2002) points out that allowing for on-the-job search can potentially

[3]See Haltiwanger et al. (2018) for a detailed analysis of these worker flows.

drag productivity down, as it gives unemployed workers the option to take bad jobs as a stepping-stone to get better ones later. Lastly, firm dynamics allows the selection of firms to adjust over the business cycle through entry and exit, in line with the large drop in the number of active firms observed during recession.

In the model, firms make hiring and exit decisions and commit to a long term state-contingent wage contract. In designing these contracts, firms face a trade-off between making larger profits through offering lower wages and preventing their worker from getting poached by other employers with larger wage payments. I provide conditions on the primitives of the model such that the optimal wage contract is increasing in the firm's own productivity after all histories in equilibrium. This monotonicity property implies that workers move from less to more productive firms when making direct transitions between employers, since more productive firms offer larger wage contracts.

This property of the optimal wage contract is also central in retaining the tractability of the model. With on-the-job search, the optimal contract itself depends on the whole distribution of offered contracts through the rate at which workers quit firms to take better paying jobs, a daunting fixed-point problem. Instead, the fact that contracts are increasing in firm productivity makes the distribution of workers across firm productivity levels sufficient to characterize the firm's policies out of steady-state. I approximate this distribution with a set of its moments to numerically solve the full model with aggregate shocks.

I calibrate the model to match a set of labor market and firm dynamics moments from pre-recession British data. In doing so, I specifically include moments capturing workers' transition rates in and out of unemployment and between employers, as well as moments disciplining the selection of firms upon entry. These moments include the firm exit rate, as well as the persistence and dispersion of labor productivity at the firm level, which I obtain from the firm-level data. While not being targeted directly in the calibration, the model does a very good job at replicating the large concentration of employment in the largest firms observed in the data. This is important since any measure of aggregate productivity derived from firm data is shaped by this high level of employment concentration.

Given the calibrated model, I feed in a sequence of aggregate shocks triggering a sharp and prolonged increase in unemployment, akin to the UK experience during the Great Recession. The model generates firm dynamics and labor market aggregates in line with the data, a set of series not targeted as part of the calibration procedure. Importantly, it also replicates the drop in worker allocation across firms measured in the firm-level data after 2008. In the simulated recession, the model captures the magnitude and a large part of the persistence of this effect: it accounts for about sixty percent of the overall reduction in labor productivity that can be attributed to the allocation of workers to firms found in the data by 2015. Search frictions in the labor market then represent a compelling explanation for the drop in the worker allocation measure found in the data.

To understand the respective contribution of firm dynamics and search frictions, I leverage the model to decompose labor productivity into three components: (i) aggregate shock, (ii) firm selection, (iii) worker reallocation. This model-based decomposition allows to isolate the effects of firm selection and worker reallocation from the direct impact of the aggregate shock, which is subsumed in the empirical decomposition implemented on firm data. I can then assess the role of each endogenous component in driving aggregate labor productivity in the simulated recession. While firm selection has a large positive effect on labor productivity in the short run, I find that the worker reallocation component has a medium-term negative impact on labor productivity. On net, it consistently dominates the firm selection effect three years after the start of the recession.

The reason the allocation of workers to firms is downgraded following the shock comes from on-the-job search. Firms have two margins to control the rate at which they adjust their workforce in the model: the rate at which they hire and the rate at which workers quit their job to work at more productive firms. While the hiring rate drops everywhere in the productivity distribution, the rate at which workers quit their job decreases primarily on the lower part of the firm productivity distribution, as these workers now compete for good jobs with a larger pool of unemployed workers. This second effect dominates in the calibrated model. As a result, low productivity firms do not shrink as fast in the aftermath of the recession as in normal times.

This mechanism finds empirical support in the data in three dimensions. First, the fact that low productivity firms do not shrink as fast in the aftermath of the shock is in line with the empirical finding that the relationship between employment growth and labor productivity is still positive but not as strong following the Great Recession in the UK firm data. Second, the lower rate of voluntary quits implies a lower aggregate rate of job-to-job transitions. These direct transitions between employers drop sharply at the time of the Great Recession in the UK. Finally, recent evidence described in Haltiwanger et al. (2018) for the United States points to a substantial reduction in the rate of job-to-job transitions out of low-paying firms during the last recession. While I cannot investigate this channel directly in the British firm-level data, this last finding is consistent with the model prediction that voluntary quits from low productivity firms fall after a negative shock.

Related literature. This work is first related to the literature stressing the connection between the dispersion of firm productivity within narrowly defined industries and the allocation of production factors in determining aggregate productivity. Hsieh and Klenow (2009) study the misallocation of production factors across countries in a model with monopolistic competition and heterogeneous firms. Lentz and Mortensen (2008) quantify the importance of input reallocation for aggregate growth using a model of product innovation. My paper similarly starts from the existence of a large dispersion in firm productivity but instead emphasizes the interplay between search frictions in the labor market and the business cycle as a key driver of the reallocation of labor across heterogeneous firms.

There is a large empirical literature that studies the reallocation effects of recessions using micro-level firm data. Starting with Davis and Haltiwanger (1992), economic downturns have been recognized as periods of changing job reallocation. Moscarini and Postel-Vinay (2012) and Fort et al. (2013) further point out the existence of heterogeneous responses in employment growth respectively by firm size and firm age. Because of data limitations, these papers focus on employment changes, but do not link them to productivity explicitly. A contribution of this paper is to use data with a concept of output at the firm level to directly tie employment changes to firm-level productivity. These data allow me extend the analysis of the US manufacturing sector in Foster et al. (2016) to the entire private sector in the UK. My finding that worker reallocation has a negative impact on labor productivity is in line with their result that the reallocation of jobs triggered by the US Great Recession has been less productivity enhancing than in previous contractions. In addition, the central contribution of this paper is to describe a tractable model with both a meaningful definition of firm productivity and endogenous worker flows, thus providing a rich framework to analyze these empirical regularities.

Second, this paper contributes to the growing literature that combines firm dynamics with search frictions in the labor market. Models of firm dynamics traditionally center on adjustment costs in capital inputs (Khan and Thomas, 2013; Clementi and Palazzo, 2016) or in gaining customers (Sedláček and Sterk, 2017), but they maintain the assumption that labor markets clear. Conversely, the macro labor literature stresses the role of search frictions to account for the evolution of labor market aggregates over the business cycle, but these models center on the notion of jobs – a one worker-one firm match with idiosyncratic productivity (Mortensen and Pissarides, 1994; Shimer, 2005; Lise and Robin, 2017). As such, they do not suggest a natural way to aggregate these jobs into a meaningful definition of a firm.

My work adds to the recent papers integrating firm dynamics and search friction in the labor market by merging three unique features: firm dynamics, random search with on-the-job search, and a global solution with aggregate shocks. Gavazza et al. (2018), Kaas and Kircher (2015), and Sepahsalari (2016) abstract from job-to-job flows and center on heterogeneity in the efficiency of hiring over the business cycle. Elsby and Michaels (2013) describe a rich random search environment with heterogeneous firms but abstract from on-the-job search. Schaal (2017), finally, focuses on the impact of uncertainty shocks in a related model cast in a directed search framework. With respect to the random search environment considered in my model, his framework implies that firms are indifferent between contracts in equilibrium, and as such job-to-job transitions need not be productivity enhancing. In my model, by contrast, more productive firms poach workers from less productive firms in equilibrium. This equilibrium property offers a clear channel through which recessions can affect the allocation of workers to firms: by slowing down the reallocation of workers from less to more productive firms, as these workers compete with a larger pool of unemployed workers during downturns.

My work is most closely related to several recent papers describing random search environments with both firm dynamics and on-the-job search. In each case, I stress the main differences with my model.

Moscarini and Postel-Vinay (2013) extend the standard Burdett and Mortensen (1998) model to an environment with aggregate shocks. My model uses a similar contract structure, in which the firm can commit to state-contingent wage payments going forward. I improve on their framework by embedding a proper notion of firm dynamics: firms enter, are hit by idiosyncratic shocks, which eventually lead them to exit. This was not possible with their definition of equilibrium, which requires that more productive firms have larger employment over the business cycle—what they term a ‘’Rank-preserving” equilibrium. This characterization rules out, for example, a very productive firm entering, as it would have low employment by definition. I show that under a specific recruitment technology, the equilibrium can be characterized in terms of a single productivity ladder, independently of a firm's size.

Coles and Mortensen (2016) present a model with both firm dynamics and on-the-job search. I rely on their hiring cost formulation to establish size-independence in the firm's policies. They use a different wage-setting protocol to mine, and I expand on this difference in Appendix 1.A.1. My paper also diverges from theirs in that firm entry and exit is endogenous, I do not restrict search effort to be the same for employed and unemployed workers, and thoroughly analyze the business cycle properties of my model, implementing a full numerical solution for the equilibrium out of steady-state.

Elsby and Gottfries (2019) develop a model with firm dynamics and on-the-job search in which firms operate a decreasing returns-to-scale production function. They show under two bargaining protocols that the job ladder can be characterized in terms of a single variable, the marginal product of labor. A key difference with my framework is that they do not allow for firm entry and exit. So all reallocation occurs through worker transitions, contrary to my framework where firm selection also plays a role. Another relates to their exploration of the cyclical property of their model, which is limited to a comparison of steady-states and transitional dynamics.[4]

Bilal et al. (2019), finally, describe a model with firm dynamics (including firm entry and exit), on-the-job search, and a production function with decreasing returns to scale. They build on the framework developed in Lentz and Mortensen (2012) to characterize the job ladder in terms of the marginal value of the firm-workers surplus. Similarly to Elsby and Gottfries (2019), the main difference with my paper is that they restrict themselves to studying the transitional dynamics property of their model, staying away from studying the full dynamics out of steady-state.

Outline. Section 1.2 documents novel facts on firm dynamics and labor productivity from British firm-level data. Section 1.3 introduces the model. Section 1.4 defines the equilibrium. Section 1.5 describes the calibration and numerical solution. Section 1.6 analyzes the reallocation properties of the model during a recession and Section 1.7 concludes.

[4] At least in the current working paper version.

1.2 Firm Productivity and Labor Productivity during the Great Recession

To document the interaction between firm productivity at the micro level and labor productivity at the macro level, I construct an index of labor productivity aggregating from the ground up, starting from firm-level data. This paper uses the Business Structure Database (Office for National Statistics, 2019), a dataset with yearly information on the universe of British firms between 1998 and 2016. Importantly, these data cover about a decade before and after the Great Recession – officially starting in 2008Q2 and ending in 2009Q2 in the UK – thus allowing to decompose labor productivity before the onset of the recession and during the recovery period.

Aggregate labor productivity from firm-level data. The Business Structure Database (BSD) combines several administrative sources to derive the employment, sales, and industry for each active firm in a given year. In the subsequent analysis, I exclude Health and Education, which are mostly public in the UK, as well as a few industries whose aggregates do not line up with official UK statistics (Finance and Insurance, Mining and Quarrying). The final sample is made up of more than 33 million firm-year observations. Further details regarding sample selection, the construction of these variables, and the validation of the aggregate series derived from the BSD against official UK statistics can be found in Appendix 1.B.1.

I follow Bartelsman et al. (2013) in defining the following industry-level labor productivity index

$$LP_t := \sum_i \underbrace{ES_{i,t}}_{\text{employment share}} \times \underbrace{LP_{i,t}}_{\text{firm labor productivity}} \tag{1.1}$$

with the employment share, $ES_{i,t}$, and labor productivity measure, $LP_{i,t}$, at firm i in period t given by

$$ES_{i,t} := \frac{\text{employment}_{i,t}}{\sum_i \text{employment}_{i,t}}, \quad LP_{i,t} := \ln\left(\frac{\text{sales}_{i,t}}{\text{employment}_{i,t}}\right). \tag{1.2}$$

While the Business Structure Database reports sales, and not value-added as more conventionally used to define labor productivity, its key advantage is to make this information available throughout the firm's life span and for all firms in the economy, independently of their size or sector.[5] Decker et al. (2018), who also study the relation between firm productivity and worker reallocation, define labor productivity at the firm-level similarly. In addition, I compare the sales-based labor productivity measure to a value added one from a different dataset of British firms in Appendix 1.B.2. I find that the two measures are very strongly correlated (correlation coefficient of .925), in line with findings based on US data (Foster et al., 2001).

The analysis is carried out within industries to account for price differences (in product or input) across sectors.[6] To further abstract from trends in industry shares over the sample period, these industry-level measures are aggregated using time-invariant industry weights. These weights are defined as the labor share of each industry in the pooled sample.[7]

Worker reallocation and productivity during the UK Great Recession. Figure 1.1 briefly outlines the UK experience during the Great Recession. It shows that the sharp and prolonged increase in unemployment comes at the same time as a persistent reduction in aggregate labor productivity growth, a pattern known in the UK as the "labour productivity puzzle." While previous work has analyzed the role of cross-sector reallocation in accounting for the economy-wide pattern (Patterson et al., 2016), these data show that labor productivity growth also markedly slows down within sectors.

To assess the role of worker reallocation in accounting for the overall drop in labor productivity, I decompose LP_t as

$$\begin{aligned} LP_t &= \sum_i ES_{i,t} \times LP_{i,t} \\ &= \underbrace{\overline{LP}_t}_{\text{average firm productivity}} + \underbrace{\sum_i \left(ES_{i,t} - \overline{ES}_t\right)\left(LP_{i,t} - \overline{LP}_t\right)}_{\text{OP allocative efficiency measure}}, \end{aligned}$$

an equality referred to in the literature as the "OP decomposition" (Olley

[5] To be precise, the BSD covers all firms either above the VAT registration threshold or with at least one employee liable for income tax. These thresholds are not overly restrictive as they represent low levels of economic activity.

[6] In my baseline analysis, I use the "division" level of the British Standard Industrial Classification (SIC). With about seventy categories, this level of decomposition is roughly equivalent to three-digits sub-sectors in the NAICS system. The graphs shown in this section are similar when using a thinner SIC level.

[7] The decline in manufacturing is the most noticeable trend in the UK over the period. The share of manufacturing in employment falls from 16 to 8 percent between 1998 and 2016.

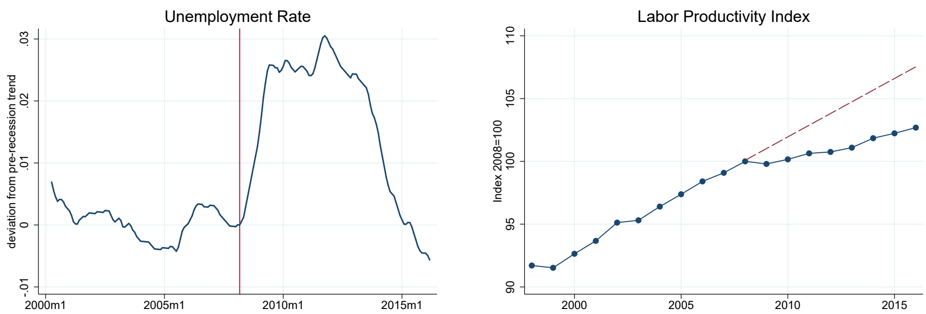


Figure 1.1: The Great Recession in the UK. Left: monthly unemployment rate (Office for National Statistics). Right: Labor productivity index, as defined in Equation (1.1). The index is computed separately for each industry and aggregated using time-invariant industry weights (see main text for details).

and Pakes, 1996).[8] In this last expression, the first term is the average (unweighted) productivity of firms in the economy. The second term measures how well labor is allocated to firms, as it increases as more firms with above average productivity also have a larger than average employment share.

The evolution of each of these terms immediately before and after the Great Recession is depicted in Figure 1.2. They are both expressed in deviation from their respective pre-recession linear trend. Figure 1.2 shows that average firm productivity and the allocation of labor to firms have contributed to lower labor productivity growth in the aftermath of the recession. In particular, after a small increase right after the onset of the recession, the OP measure of allocation has kept moving down since 2010. By the end of the sample period, it represents about a fourth of the overall reduction with respect to the pre-recession trend in labor productivity.[9]

Firm-level evidence. At the firm level, this aggregate pattern shows up as a lower association between firm labor productivity and their subsequent employment growth at the firm level. Table 1.1 shows regressions of the

[8] This equality follows directly from expanding the second term and noting that, by definition, $\sum_i ES_{i,t} = 1$.

[9] Many different decompositions of productivity have been proposed in the literature (e.g., Griliches and Regev, 1995; Foster et al., 2001; Diewert and Fox, 2010). Most closely related to the exercise in this paper, Riley et al. (2015) similarly find using a battery of dynamic productivity decomposition that the "external" – between firms – component of labor productivity changes tends to increase over time as a share of the overall productivity drop following the Great Recession in the UK. The use of the OP decomposition is motivated by the fact that it admits an intuitive counterpart in the notation of the model to be introduced in Section 1.3.

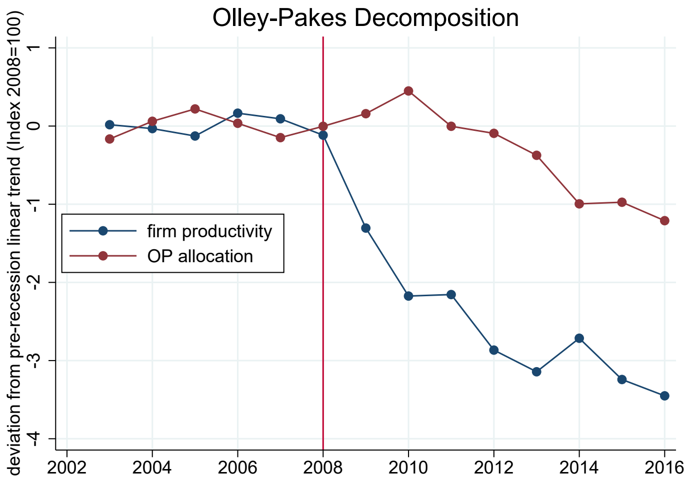


Figure 1.2: OP decomposition in the aftermath of the Great Recession. Each series is shown in deviation from its pre-recession linear trend.

form

$$\Delta \ln n_{i,t+1} = \alpha LP_{i,t} + \beta \text{post}_t \times LP_{i,t} + \mu_{t,s} + \epsilon_{i,s,t}, \tag{1.3}$$

where $LP_{i,t}$ is firm labor productivity, "post" is an indicator for the years following the Great Recession, and $\mu_{s,t}$ a set of industry-year fixed effects. The coefficients (α, β) then measure the strength of the relationship between labor productivity and employment growth within an industry-year cell. The coefficient β shows that the positive association between firm labor productivity and employment growth drops by about a fourth post-recession, implying the average growth between a firm one standard deviation above and below the mean is about 2.3 percent lower. This finding suggests that employment growth is not as productivity enhancing as prior to the recession. The model developed in subsequent sections offers a rationale in terms of search frictions for this firm behavior at the micro level. (Table 1.1 also reports results on the productivity of entering and exiting firms, showing minor differences between the pre- and post-recession periods.)

Job-to-job transitions. The slow down in worker reallocation documented in Table 1.1 comes at the same time as a net reduction in job-to-job transitions, as depicted in Figure 1.3. In the aftermath of the downturn, the rate of direct transitions between employers is about one-third lower than prior to the start of the episode. The recovery period is then characterized by both a much larger pool of unemployed workers and a drop in,

	$\Delta \ln n_{i,t+1}$ (continuers)	Firm Entry	Firm Exit
$LP_{i,t}$	0.065	-0.004	-0.020
	(0.000)	(0.000)	(0.000)
$LP_{i,t} \times \text{post}_t$	-0.017	-0.004	0.003
	(0.000)	(0.000)	(0.000)
Year-Industry FEs	Yes	Yes	Yes
Size Control ($\ln n_{i,t}$)	Yes	Yes	Yes
N (million)	22.01	27.08	27.78

Table 1.1: Reallocation during the Great Recession at the firm level. The specification is given in Equation (1.3). The first column is the change in employment in the next period (restricted to surviving firms). The second is a linear probability model for firms exiting in the next period (restricted to surviving and exiting firms). The third is a linear probability model for entering firms (restricted to firms entering in the current period and incumbents). Robust standard errors in parentheses.

potentially productivity enhancing, employer-to-employer transitions.

1.3 A Model of Firm Dynamics with On-the-Job Search

The section describes a model of firm dynamics in which the transitions of workers in-and-out of unemployment and between employers are constrained by search frictions in the labor market. The model thus offers a counterpart to the various reallocation patterns documented in Section 1.2: unemployment, job-to-job transitions, but also reallocation of workers across firms with heterogeneous productivities.

1.3.1 Environment

Time is discrete and the horizon is infinite. Aggregate productivity is driven by an economy-wide shock, ω_t, which follows a stationary first-order Markov process, $Q(\omega_{t+1}|\omega_t)$.

Agents. There are two types of agents in the economy: workers and firms. Both are risk-neutral, infinitely-lived, and maximize their pay-offs discounted with factor β. The labor force is represented by a continuum of working age individuals with measure one. These workers are ex-ante identical and supply one unit of labor in-elastically. There is an endogenously

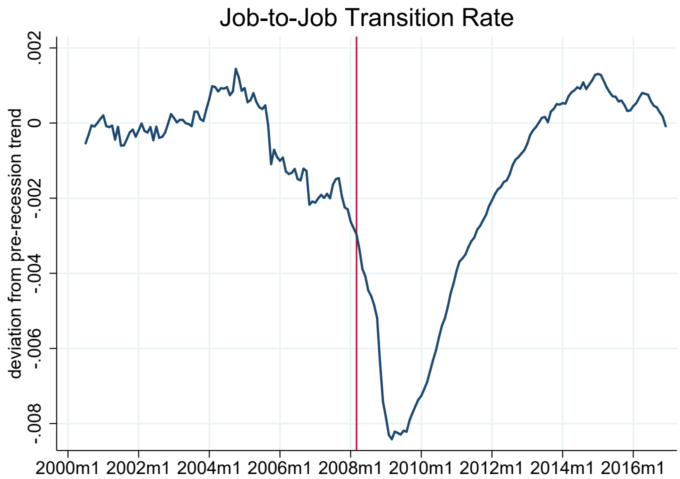


Figure 1.3: Job-to-job monthly transition rate derived from the British Household Panel Survey (Postel-Vinay and Sepahsalari, 2019). See Appendix 1.B.3 for additional details on the construction of this series.

evolving measure of firms shaped by firm entry and exit. These firms face an idiosyncratic productivity shock evolving according to a distinct first-order Markov process denoted by $\Gamma(p_{t+1}|p_t)$.

Timing. Each period t can be divided into the six following phases:

1. Productivity shocks. Aggregate productivity, ω_t, and firm-specific productivity p_t are realized.

2. Entrepreneurial shock. With probability μ, workers become potential entrepreneurs. They draw an initial idea with productivity $p_0 \sim \Gamma_0$ and decide whether to enter.

3. Firm exit. Firms decide whether to stay on or discontinue their operations based on the realization of the productivity shocks. If they exit, all of their workers become unemployed.

4. Exogenous separations. Employees at continuing firms lose their jobs with exogenous probability δ.

5. Search. Recruitment at incumbent firms takes place. Firms post vacancies to hire. Both unemployed and employed workers search for jobs.

6. Production and payments. Unemployed workers have home production b. Firms produce with their employees after the search stage.

Wages accrue to employed workers. Newly created businesses start producing with a single worker, the entrepreneur.

It is assumed that workers becoming unemployed due to firm exit or a δ-shock start searching in the next period. Similarly, potential entrepreneurs (workers hit by a μ-shock) quit their job and do not search in the current period. They become unemployed should they choose not to pursue their business opportunity.

A recursive formulation is used throughout the paper. All value functions in subsequent sections are written from the production and payments stage onward, taking expectation over the events occurring in period $t+1$, conditional on the information available at the end of period t. The measure of unemployed workers and incumbent firms, which are formally defined below, are recorded at the very start of the period, before the entrepreneurial shock occurs.

Contracts. Each firm designs and commits to an employment contract. This agreement between a firm and a worker specifies a wage payment contingent on the realization of some state variable, which is made precise once the agents' problems are formally introduced. The firm chooses this contract to maximize its long-run profits, taking other firms' contracts as given. In addition, it is assumed that firms are bound by an equal treatment constraint, which restricts them to offering the same contract to all of their employees, independently of when they are hired.[10] With full commitment, the discounted sum of future wage payments can be summarized by a contract value W_t, where t denotes the realization of the contractible state in the current period. In Section 1.4, I derive a closed-form expression for the optimal contract that makes the model straightforward to simulate.

Workers, on the other hand, cannot commit to a firm and are free to walk away at any point. Outside offers are their private information and are therefore not contractible. Given the equal treatment constraint, if the realization of the state entails a contract value below the value of unemployment, the firm loses its entire workforce and is forced to exit. This can equally be interpreted as the employment contract specifying firm exit after certain realizations of the state.

Search and matching technology. Search is random. The probability that a vacancy reaches a worker is denoted η_t. The probability that an

[10] Because all workers are ex-ante identical and since there is no learning on the job, this constraint can be interpreted as a non-discrimination rule.

unemployed worker draws an offer is denoted λ_t. Employed workers have less time to search; their probability to draw an offer is given by $s\lambda_t$, for some exogenous search intensity parameter $s < 1$. Denoting A_t the stock of vacancies and Z_t aggregate search effort (from employed and unemployed workers), accounting for contacts between workers and vacancies in each period directly gives the equality

$$\lambda_t Z_t = \eta_t A_t, \qquad \lambda_t, \eta_t \leq 1.$$

Following Burdett and Mortensen (1998), there is no bargaining: workers draw a take-it-or-leave-it offer from a distribution, F_t, which is endogenously determined in equilibrium. Since workers can only accept or decline these offers, their decision boils down to accepting better contracts.[11]

1.3.2 Incumbent firms

Production. Firms operate a constant returns to scale technology with labor as its only input. n_t denotes the measure of workers currently employed at the firm. The productivity factor is given by $\omega_t p_t$. ω_t stands for the aggregate component of productivity, which is common to all firms, while p_t represents the firm's idiosyncratic productivity. ω_t and p_t follow independent first-order Markov processes and are positive by assumption.

Hiring technology. Following Merz and Yashiv (2007) and Coles and Mortensen (2016), hiring is modeled as an adjustment cost, where the cost of hiring is spread equally amongst current firm employees. A firm of current size n_t hiring a total of H_t workers has a total recruitment bill of

$$n_t c\left(\frac{H_t}{n_t}\right) = n_t c(h_t), \qquad h_t := \frac{H_t}{n_t},$$

where c is assumed to satisfy $c' > 0,\ c'' > 0$ and $c(0) = 0$. As will become clear when writing down profits, a linear recruitment technology in the firm's employment at the time of hiring implies that the firm's problem is linear in n_t. This simplification makes the model more tractable, as the firm's policy functions do not depend on n_t.

In economic terms, this formulation of the firm's hiring cost should be seen as a screening and training cost for new hires. Similarly to the model developed in Shimer (2010), current employees are an input in the recruit-

[11]Burdett and Mortensen (1998) use the term “wages” instead of “contract”, since, in their stationary environment, a contract is a constant, non-renegotiable wage.

ment process, with additional hires decreasing the revenue from each worker all else equal. The fact that this cost is convex in hires per current employees reflect the smaller disruption to production of recruiting, say, five new employees at a business with 100 workers than at a ten-worker one. Empirical evidence suggests that these training costs can be substantial: Gu (2019) finds that job adaptation, as assessed by employers, takes 22.5 weeks on average for US non-college workers.[12]

Once the firm has chosen its target number of hires, I assume that the actual vacancies corresponding to these hires, which I formally define once the notation for aggregates is introduced, are posted at no extra cost.

Discounted Profits. Because the firm fully commits to a contract upon entry, its profits can be written in recursive form by requiring that the firm offers at least the value of the current contract in equilibrium (Promise-Keeping constraint). Let $\overline{V}$ denote the value of this contract given the realization of the states this period. Let χ_t further denote the firm's decision to continue given the realization of the shocks at the start of period t.

A firm with current productivity p_t employing n_t workers has discounted profits

$$\Pi_t\left(p_t, n_t, \overline{V}\right) = \max_{\substack{h_{t+1}\geq 0 \\ w_t \\ W_{t+1}}} \Bigg\{ (\omega_t p_t - w_t) n_t \\ + \beta E_t \Big[\chi_{t+1} \Big(- c(h_{t+1})(1-\mu)(1-\delta) n_t + \Pi_{t+1}(p_{t+1}, n_{t+1}, W_{t+1}) \Big) \Big] \Bigg\}, \tag{1.4}$$

where the firm's continuation decision is given by

$$\chi_{t+1} := \mathbb{1}\Big\{ \big(W_{t+1} \geq U_{t+1}\big) \cap \big(\Pi_{t+1} \geq 0\big) \Big\}, \tag{1.5}$$

since the firm needs to offer at least U_{t+1} for its workers not to quit and I assume that it must make non-negative discounted profits. Anticipating on the results in Section 1.4, in equilibrium the firm's continuation decision can be expressed solely in terms of the firm's current idiosyncratic productivity, though this threshold evolves with the business cycle.

In addition, the firm's maximization problem is subject to the two fol-

[12]The exact question in the Multi-City Study of Urban Equality asks about the time it takes a typical employee in an occupation to become fully competent. See Gu (2019) for details.

lowing constraints. First, full commitment implies a Promise-Keeping constraint in the sequential form problem. The firm's choice of wages, w_t, and contract values in the next period, W_{t+1}, has to give workers a value of at least $\overline{V}$ in expectation. Second, the size of its workforce, conditional on the firm surviving, is defined as

$$n_{t+1} = \left[1 - \underbrace{q_{t+1}(W_{t+1})}_{\text{quit rate}} + \underbrace{h_{t+1}}_{\text{hiring rate}}\right] \underbrace{(1-\mu)(1-\delta)n_t.}_{\text{remaining workers at search stage}} \tag{1.6}$$

The measure of workers employed at the search stage is $(1-\mu)(1-\delta)n_t$, those not leaving to become potential entrepreneurs (rate μ) or exogenously becoming unemployed (rate δ). $q_{t+1}(W_{t+1})$ denotes the rate at which workers still employed at the search stage leave the firm to take on better jobs, conditional on the firm offering value W_{t+1}. The quit rate is given by $q_{t+1}(W_{t+1}) := s\lambda_{t+1}\overline{F}_{t+1}(W_{t+1})$, the rate at which workers employed at the firm find better jobs in the current period.[13] Equation (1.6) makes the firm's trade-off in controlling the growth of n_t explicit. It can either offer better contracts, thus limiting poaching, or intensify its hiring effort through h at a higher recruitment cost.

Coles and Mortensen (2016) describe an alternative wage determination protocol, in which firms cannot commit to a wage payment schedule in the future and workers do not observe the firm's productivity, only its offered wages. Their characterization, however, is obtained under stronger assumptions made for tractability. I provide a detailed discussion of the implied difference between the two wage determination mechanisms in Appendix 1.A.1.

Linearity of Discounted Profits. It can be guessed and verified that discounted profits are linear in n_t. Define profit per worker as $n_t\pi_t(p_t, \overline{V}) := \Pi_t(p_t, n_t, \overline{V})$. By substituting this guess in the right-hand side of (1.4) and

[13] I define $\bar{F}_t := 1 - F_t$. Note that conditional on $q_{t+1}(W_{t+1})$ and h_{t+1} Equation (1.6) holds exactly by a Law of Large Number argument since n_t is the measure of workers employed at the firm.

using the law of motion for employment, it can be shown that

$$
\begin{aligned}
\pi_t\left(p_t, \overline{V}\right) = \max_{\substack{h_{t+1}\geq 0\\ w_t\\ W_{t+1}}} \Bigg\{ & \omega_t p_t - w_t \\
& + \beta E_t \bigg[(1-\mu)(1-\delta)\chi_{t+1}\Big(- c(h_{t+1}) \\
& + (1 - q(W_{t+1}) + h_{t+1})\pi_{t+1}(p_{t+1}, W_{t+1})\Big)\bigg]\Bigg\}, \qquad (1.7)
\end{aligned}
$$

still subject to the Promise-Keeping constraint. See Appendix 1.A.2.

It follows directly from Equation (1.7) that the firm's optimal policies do not depend on its current size n_t. In particular, there is no partial layoff in the model, since the continuation decision χ_t is the same at all n_t. Jobs are only terminated in the following four cases: (i) exogenous entrepreneurial shocks at rate μ, (ii) exogenous separations at rate δ, (iii) voluntary quits for better jobs at rate $s\lambda_t\overline{F}_t(W_t)$, (iv) firm exit.

To sum up, firms are defined in the model by a recruitment technology – the cost function c – and a "contract policy" – the state-contingent contract W_{t+1} it offers to all its workers. While firm size does not enter directly the firm's policy functions, it is still well-defined in the model. This is because these policies pin down, conditional on survival, the growth rate of employment. Even if two firms with the same idiosyncratic productivity in a given period grow at the same rate, the accumulation of firm-specific shocks generates a firm-size distribution in the cross-section. The model actually replicates the Pareto tail of the empirical firm size distribution very well. I return to this point when calibrating the model in Section 1.5.

1.3.3 Firm Entry

Firm entry is governed by the decision of workers to become entrepreneurs. I assume that unemployed and employed workers draw a business idea with probability μ from an exogenous distribution Γ_0 at the start of each period t. This distribution gives the initial (firm-specific) productivity of entering businesses. I further make the assumption that employed workers cannot go back to their previous job when hit by such an "entrepreneurial shock". They must either enter the market with their new idea or become unemployed.

The decision of potential entrepreneurs to start a new business then weighs the value of starting up a firm against the value of unemployment. Entering entrepreneurs are assumed to get the full surplus $S_t(p_t) := \pi_t(p_t, \overline{V}) + \overline{V}$ of the match.[14] They then decide to enter if $U_t \leq S_t(p_0)$ for some initial draw p_0 from Γ_0. If they choose not to take this business opportunity, they fall back into unemployment until next period (they do not search in the current period). If they choose to enter, it is assumed that entrepreneurs have their business purchased by some outside investors (not modeled), and become the first workers at these firms.

Similarly to Gavazza et al. (2018), firms need to have positive employment to operate the recruiting technology. There is no meaningful notion of a firm with zero worker in this framework, and entrepreneurs therefore become the first workers at newborn firms. With a continuum of workers, the interpretation of this entry process is that a measure μ of workers, both employed and unemployed, becomes potential entrepreneurs in each period. They then create firms at which they become the first workers, and these firms have employment n_0. I normalize $n_0 = 1$, so that the measure of entering firms is equal to that of starting entrepreneurs.[15] These firms then move on to the production stage, and become incumbent firms from the next period onward.

1.3.4 Value of Employment and Unemployment

First, let Q_t denote the value of a potential entrepreneur, a worker hit by a μ-shock,

$$Q_t := \int \max\big(S_t(p), U_t\big) d\Gamma_0(p).$$

An unemployed worker has home production b and receives job offers with probability λ_{t+1}, conditional on not being hit by an entrepreneurial shock, μ. The value of being unemployed is then

$$\begin{aligned} U_t = b + \beta E_t \Big\{ & \mu Q_{t+1} \\ & + (1-\mu)\Big[(1-\lambda_{t+1})U_{t+1} + \lambda_{t+1} \int \max\big(W', U_{t+1}\big) dF_{t+1}(W')\Big]\Big\}. \end{aligned} \tag{1.8}$$

Similarly to unemployed workers, employees are hit with probability μ

[14]Note that given the value of a firm is linear in n_t and given the equal treatment constraint, the surplus of the firm and all of its workers is simply $\Pi_t(p_t, n_t, \overline{V}) + n_t\overline{V} = n_t S_t(p_t)$.

[15]In the British firm data, more than three quarters of entering firms report employment equals to one, where employment is defined as "employees and working proprietors."

by an "entrepreneurial shock", in which case they leave their present job to explore this idea. Otherwise, employed workers can search on the job with exogenous relative intensity $s < 1$. They separate with exogenous probability δ. Employed workers earn wages w_t in the current period, and are promised a state-contingent value W_{t+1} in the next period. Recall that due to the commitment structure, the firm exits and all of its workers become unemployed after some realizations, when it cannot offer its workers more than their reservation value, U_{t+1}, summarized by the indicator $\chi_{t+1}(W_{t+1})$.

Taken together, these shocks give rise to the following value function for the employed worker

$$\begin{aligned} W_t = w_t + \beta E_t \Bigg\{ \mu Q_{t+1} + (1-\mu) \Bigg[\Big((1-\chi_{t+1}) + \delta \chi_{t+1} \Big) U_{t+1} \\ + \chi_{t+1}(1-\delta) \Big(\big(1 - q_{t+1}(W_{t+1})\big) W_{t+1} \\ + s\lambda_{t+1} \int \max\left(W', U_{t+1}\right) dF_{t+1}(W') \Big) \Bigg] \Bigg\}. \end{aligned} \tag{1.9}$$

1.3.5 Joint Firm-Worker Surplus

The firm and worker problems can be summarized in a single expression combining Equations (1.7) and (1.9). I show in Appendix 1.A.3 that the following expression for $S_t := \pi_t(p_t, \overline{V}) + \overline{V}$ can be obtained after rearranging these two equations

$$\begin{aligned} S_t(p) = p_t \omega_t + \beta E_t \Bigg\{ \mu Q_{t+1} + (1-\mu) \Bigg[(1 - \chi_{t+1}(p_{t+1})) U_{t+1} \\ + \chi_{t+1}(p_{t+1}) \Big(\delta U_{t+1} + (1-\delta) \psi_{t+1}(p_{t+1}) \Big) \Bigg] \Bigg\}. \end{aligned} \tag{1.10}$$

In this last expression, $\psi_{t+1}(p_{t+1})$ denotes the joint value of a firm-worker pair, conditional on the firm not exiting, which writes

$$\begin{aligned} \psi_{t+1}(p_{t+1}) := \max_{\substack{h_{t+1} \geq 0 \\ W_{t+1}}} \Bigg\{ - c(h_{t+1}) + (1 - q_{t+1}(W_{t+1})) S_{t+1}(p_{t+1}) \\ + h_{t+1}(S_{t+1}(p_{t+1}) - W_{t+1}) + (1-\delta) s \lambda_{t+1} \int_{W_{t+1}}^{\infty} W' dF_{t+1}(W') \Bigg\}. \end{aligned} \tag{1.11}$$

This simplification directly follows from the assumptions that the firm fully commits to its workers and that utility is transferable, since both firms and workers are risk-neutral. Conditional on survival, the optimal contract and hiring rate maximize Equation (1.11), and, importantly, the resulting contract fully internalizes the Promise-Keeping constraint.

While wages disappear from this last expression, they can still be computed from the Promise-Keeping constraint. Given that the firm guarantees at least $\overline{V}$ in expectation to its workers and that future contract values are defined as the solution to Equation (1.11), the offered wage is implicitly defined by

$$\overline{V} = w_t + \beta E_t \Bigg\{ \mu Q_{t+1} + (1-\mu) \Bigg[\Big((1-\chi_{t+1}) + \delta \chi_{t+1} \Big) U_{t+1}$$

$$+ \chi_{t+1}(1-\delta) \Big(\big(1-q_{t+1}(V_{t+1})\big) V_{t+1} + s\lambda_{t+1} \int \max\left(V', U_{t+1}\right) dF_{t+1}(V') \Big) \Bigg] \Bigg\},$$

where V_{t+1} denotes the (state-contingent) solution to the joint-surplus maximization problem in Equation (1.11).

1.3.6 Aggregation

Search Effort, Vacancies, and Offer Distribution. Let $\nu_t(p, n)$ denote the cumulative measure of firms with productivity less than p and workforce less than n at the start of period t, before workers are hit by "entrepreneurial" shocks and firm exit takes place. Aggregate search effort is the measure of searching workers, both unemployed and employed,

$$Z_t := (1-\mu) \left[u_t + s(1-\delta) \int \chi_t(p) n d\nu_t \right], \tag{1.12}$$

where the unemployment rate is $u_t := 1 - \int n d\nu_t$. This expression excludes potential entrepreneurs and displaced workers, who do not search in period t by assumption.

Let $a_t(p, n)$ denote the vacancies posted by a continuing firm with productivity p and workforce n. Total vacancy posting aggregates the vacancies of all active firms in the economy

$$A_t := \int \chi_t(p) a_t(p, n) d\nu_t. \tag{1.13}$$

Finally, the cumulative density of offered contracts is the sum of vacancies offering a contract less than some contract value W over the total posted vacancies (since search is random)

$$F_t(W) := A_t^{-1} \int \mathbb{1}\left\{W_t(p) \leq W\right\} \chi_t(p) a_t(p,n) d\nu_t. \tag{1.14}$$

Firm Vacancy Posting. To close the model, we need to specify vacancy posting by firms, $a_t(p,n)$. Since there is no cost of posting vacancies by assumption, the firm simply posts as many as required by its target hiring rate, $h_t(p)$. $a_t(p,n)$ is then implicitly defined by the accounting equation

$$h_t(p) \underbrace{(1-\mu)(1-\delta)n}_{\text{remaining workers at search stage}} = \underbrace{a_t(p,n)}_{\text{vacancies}} \underbrace{\eta_t}_{\text{contact rate}} \underbrace{Y_t(W_t(p)),}_{\text{acceptance rate}} \tag{1.15}$$

where η_t is the probability that this vacancy reaches a worker and $Y_t(W_t(p))$ is the chance it is accepted. This probability is determined by whether the worker reached by the vacancy is currently employed at a firm offering less than $W_t(p)$ in the current period.[16]

1.4 Rank-Monotonic Equilibrium

This section formalizes the definition of equilibrium used in the remainder of the paper. I provide conditions on the cost of hiring function such that the optimal contract is increasing in the current realization of idiosyncratic productivity after all histories. I label these equilibria as "Rank-Monotonic" in the rest of the paper. This characterization is similar in spirit to the "Rank-Preserving" property defined in Moscarini and Postel-Vinay (2013) in the sense that the optimal contract is increasing in firm-specific productivity in both cases. However, while in their framework with constant productivity this property implies that more productive firms are always larger along the equilibrium path – it preserves the rank of firms in the firm-size distribution – in Moscarini and Postel-Vinay (2013), idiosyncratic productivity shocks break the direct link between a firm's rank in the productivity distribution and firm size in my framework. Though more productive firms are still growing faster and therefore more likely to be large in equilibrium – contracts are monotonic in a firm's productivity – my model also allows for new, fast-growing entering start-ups. These firms show up in the model as firms entering near the top of the productivity distribution, which will grow fast while being initially small.

This property drastically simplifies the numerical solution of the model

[16] I provide a full expression for $Y_t(W_t(p))$ in Appendix 1.A.4.

since, (i) there is no need to compute the full distribution of offered contracts, a daunting fixed-point problem as the optimal contract itself depends on this distribution, (ii) the optimal contract has a closed-form solution.

1.4.1 Recursive Equilibrium

Given the Markov structure of the shocks, attention can be restricted to recursive equilibria in which the state-space relevant to the firm's decision is made of the two shocks and the measure of firms in the (n, p)-space, ν, the latter being sufficient to compute all aggregates in the model. In addition, Equation (1.7) makes clear that the firm's current size is not part of this state-space. More formally:

Definition 1 *A Recursive Equilibrium is a triple of policy functions (V,h,χ) and a pair of value functions (S,U) that depend on the current realization of aggregate productivity, the current realization of idiosyncratic productivity, and the measure of firms at the start of the period. Given that all firms follow the policies given by (V,h,χ), these functions satisfy:*

1. *Equations* (1.12)-(1.15) *hold with* $\chi_{t+1}(p) = \chi(p, \omega, \nu)$*,* $h_{t+1}(p) = h(p, \omega, \nu)$*, and* $V_{t+1}(p) = V(p, \omega, \nu)$*;*

2. *The contract and hiring functions solve the maximization problem in* (1.11)*. The continuation decision is given by*

$$\chi(p, \omega, \nu) = \mathbb{1}\left\{V(p, \omega, \nu) \geq U(\omega, \nu)\right\};$$

3. *S and U solve, respectively,* (1.10) *and* (1.8)*.*

1.4.2 Rank-Monotonic Equilibrium

A Rank-Monotonic Equilibrium (RME) adds the following requirement to the optimum contract:

Definition 2 *A Rank-Monotonic Equilibrium is a Recursive Equilibrium such that the optimal contract,* $V(p, \omega, \nu)$*, is weakly increasing in p for all* ω *and* ν*.*

Result 1 further provides sufficient conditions on the cost of hiring function such that a Recursive Equilibrium is in fact Rank-Monotonic.

Result 1 *Assume that the hiring cost function is twice differentiable, increasing and convex. Assume the Markov process for idiosyncratic productivity satisfies first-order stochastic dominance ($\Gamma(.|p') \leq \Gamma(.|p)$ for $p' > p$ with strict inequality for some productivity level). Then:*

1. *The firm-worker surplus defined by Equation* (1.10) *is increasing in* p*;*

2. *Any equilibrium is Rank-Monotonic if*

$$\frac{c''(h)h}{c'(h)} \geq 1, \quad \forall h \geq 0.$$

The proof is in Appendix 1.A.5. Similarly to the result in Moscarini and Postel-Vinay (2013), Result 1 is not an existence statement, but a characterization of the properties of the optimal contract conditional on the existence of such an equilibrium.

The condition on the cost function in Result 1 is an additional convexity requirement. Firms use the retention margin – through offering better contracts – only in the extent the hiring technology is sufficiently costly. With identical workers and no learning on the job, the model could potentially generate a large amount of churning at the top of the productivity distribution if employers have little incentives to promise their worker higher values to retain them. Given the conditions in Result 1, hiring costs become so high for larger h that firms find it optimal to use both the retention and hiring margins to control their optimal growth rate.

The rest of the paper centers on Rank-Monotonic equilibria. When taking the model to the data, I restrict the parameter space to ensure that the convexity requirement on the cost of hiring function in Result 1 is satisfied.

1.4.3 Additional Characterization of RMEs

Because the optimal contract is increasing in p after every history in a Rank-Monotonic Equilibrium, several aggregates can be recast as functions of p, which allows to further characterize the optimal contract. I start by defining the measure of workers employed at firms less productive than p at the start of period t

$$L_t(p) := \int_{\tilde{p} \leq p} n d\nu_t(\tilde{p}, n).$$

Note that this last measure fully summarizes acceptance/quit decisions at each level of productivity in a Rank-Monotonic Equilibrium since the optimal contract is increasing in p. Firms will poach workers from firms with productivity below them and lose workers to firms with productivity above them.

Optimal policies. First, since both the firm-worker surplus and the optimal contract are increasing in p in a Rank-Monotonic Equilibrium, the entry and exit thresholds coincide. The firm's continuation policy can be written $\chi(p,\omega,L) = \mathbb{1}\{S(p,\omega,L) \geq U(\omega,L)\}$. I denote $p_E(\omega,L)$ the corresponding entry and exit productivity threshold, which is implicitly defined by $S(p_E,\omega,L) = U(\omega,L)$.

Second, Appendix 1.A.6 shows that the optimal contract takes the following form

$$V(p,\omega,L) = \frac{uU(\omega,L) + s(1-\delta)\displaystyle\int_{p_E}^{p} S(\tilde{p},\omega,L)dL(\tilde{p})}{u + s(1-\delta)\Big(L(p) - L(p_E)\Big)}. \tag{1.16}$$

The optimal contract is therefore a weighted average between the value of unemployment and the firm-worker surplus, where the weights are given by, respectively, the measure of workers in unemployment and the measure of workers searching this period at firms with productivity less than p. This expression is reminiscent of the Nash-Bargaining solution used in classic search models, which breaks down the firm-worker surplus between each party with a constant exogenous weight (e.g., Mortensen and Pissarides, 1994). The difference in my setting is that the weights are fully endogenous and evolve with the distribution of workers over the business cycle.

Third, the optimal hiring rate follows directly from inverting the derivative of the cost function in the firm's corresponding first-order condition from Equation (1.11)

$$c'(h(p,\omega,L)) = S(p,\omega,L) - V(p,\omega,L).$$

Wages. Given the expression for the optimal contract in Equation (1.16), wages are straightforward to solve for. They can be solved from the value for the employed worker (1.9) in which the contract value is now known and given by Equation (1.16).

Distribution of offered contracts. In a RME, the acceptance rate for a firm with current productivity p can be expressed as a function of the measure of workers employed at firms with current productivity below p. The distribution of offered contracts can then be simplified as

$$\lambda_t F_t\left(V(p)\right) = \int_{p_E}^{p} \frac{h_t(\tilde{p})}{u_t + s(1-\delta)\Big(L_t(\tilde{p}) - L_t(p_E)\Big)} dL_t(\tilde{p}). \tag{1.17}$$

The derivations can be found in Appendix 1.A.6.

Employment Law of Motion. Taken together, these policies imply the following law of motion for the measure of employed worker

$$L_t^P(p) = \\ \mu \int_{p_E}^{p} \chi_t(\tilde{p}) d\Gamma_0(\tilde{p}) + (1-\mu)\chi_t(p)\Big[L_t(p)\rho_t(V_t(p)) + u_t\lambda_t F_t(V_t(p))\Big], \tag{1.18}$$

where L_t^P denotes the measure of workers at firms with productivity less than p at the end of period t (at the production stage). The first term corresponds to entering entrepreneurs with initial draws less than p. The two terms in the square brackets give, first, the fraction of workers retained at firms less than p and the inflow from unemployment. The end of period and beginning of next period measures are directly linked by

$$\frac{dL_{t+1}(p)}{dp} = \int_{\underline{p}}^{\overline{p}} \frac{dL_t^P(\tilde{p})}{d\tilde{p}} d\Gamma(p|\tilde{p}),$$

which corresponds to the "re-shuffling" of workers across productivity levels due to the firm-specific shocks.

To sum up, knowing the value functions S and U for all values of the aggregate shock and the measure of employment across firm productivity is enough to simulate the model in the presence of aggregate shocks.[17] The firm's optimal policies admit closed-form solutions conditional on these value functions, and these policies in turn determine the law of motion for workers across firm productivity.

[17] It is also possible to define the net surplus as $\phi(p,\omega,L) := S(p,\omega,L) - U(\omega,L)$ and express the firm's policies as a function of this single value function, which I do in practice when simulating the model. To economize on notation, all the corresponding expressions are relegated to Appendix 1.A.7.

1.5 Calibration

This section presents the calibration and simulation procedure. Though the size independence and Rank-Monotonic equilibrium results simplify the firm's problem, solving for the firm's policies still requires to keep track of the measure of workers across firm idiosyncratic productivity levels, L_t. I then proceed in two steps to calibrate the model.

I start by solving the model without aggregate shocks, and target some key labor market and firm dynamics moments from British data to calibrate the main parameters. In doing so, I focus on a Stationary Rank-Monotonic Equilibrium. Formally:

Definition 3 *A Stationary Rank-Monotonic Equilibrium is a triple of policy functions (V,h,χ), a pair of value functions (S,U), and a measure of workers across firm productivity L, that depend on the current realization of the firm's productivity p. These functions satisfy the following requirements:*

1. *The conditions for a Rank-Monotonic Equilibrium in Definition 2 are satisfied;*

2. *The law of motion for the measure of worker induced by the firm's optimal policies* (1.18) *is constant and equal to L.*

I return to the full model with aggregate shocks in a second step, and describe how the measure of workers is approximated out of steady-state in Section 1.6.

1.5.1 Parametrization

A period t is set to a month. I specify the Markov processes for idiosyncratic productivity shocks as $\ln p_{t+1} = \rho_p \ln p_t + \sigma_p \epsilon^p_{t+1}$ with ϵ^p_{t+1}. Such a process satisfies first-order stochastic dominance conditional on past realizations, which is required for the equilibrium to be Rank-Monotonic (Result 1). The productivity of initial ideas, Γ_0 is assumed to follow a log-normal distribution with mean μ_0 and standard deviation σ_0. The functional form for the cost of hiring function is guided by the conditions derived in Result 1. I calibrate the parameters in the following cost function $c(h) = {c_2}^{-1}\big(c_1 h\big)^{c_2}$, which satisfies the convexity requirements in Result 1 provided $c_2 \geq 2$. I enforce this condition when searching over the parameter space. Taken together, these functional form assumptions give the following vector of eleven parameters to calibrate $\{\beta, \delta, c_1, c_2, s, \mu, b, \rho_p, \sigma_p, \mu_0, \sigma_0\ \}$.

1.5.2 Calibration strategy

The discount factor, β, is set in line with a 5% annual discount rate. This leaves ten parameters to calibrate, which I pin down by targeting an equal number of moments from the data. My choice of moment targets reflects both the search and firm dynamics components of the model. To discipline worker transitions in and out of unemployment and between employers, I target the unemployment to employment (UE), employment to unemployment (EU), and job-to-job (EE) average monthly transition rates in the UK over the pre-crisis period (2000-2007). These series are derived from the British Household Panel Survey (BHPS) following the methodology described in Postel-Vinay and Sepahsalari (2019).

To discipline the life cycle of firms, I target the firm exit rate, as well as the auto-correlation and inter-quartile range of labor productivity. I use the measure of labor productivity defined in Equation (1.2) (log sales over employment), deviated from year-industry averages. These moments are computed directly from the Business Structure Database (BSD), and are therefore yearly measures. In addition, I also include moments that relate specifically to the dynamics of young firms. Firms are labeled as "young" if they are less than ten years old, since this cut-off implies an equal share of young and old firms on average. These moments are the share of young firms, the share of workers employed by young firms, and the exit rate and inter-quartile range of labor productivity at young businesses. They are also derived from the BSD.

To compute the moments implied by the model, I solve for a Stationary Rank-Monotonic Equilibrium, given a vector of candidate parameters. This yields a distribution of firms and workers across productivity levels, as well as an entry/exit productivity threshold and a monthly employment growth rate for surviving firms. Figure 1.4 depicts the obtained distributions and employment growth rate at the estimated parameters. The monthly transition rates can then be computed directly based on this equilibrium. For instance, the monthly probability to find a job when unemployed implied by the model is given by

$$\underbrace{\mu \int_{\tilde{p} \geq p_E} d\Gamma_0(\tilde{p})}_{\text{successful entrepreneurs}} \quad + \quad \underbrace{(1-\mu)\lambda}_{\text{successful job search}}.$$

However, because the moments relating to firm dynamics are derived from yearly data, their model counterpart are obtained by simulating a panel

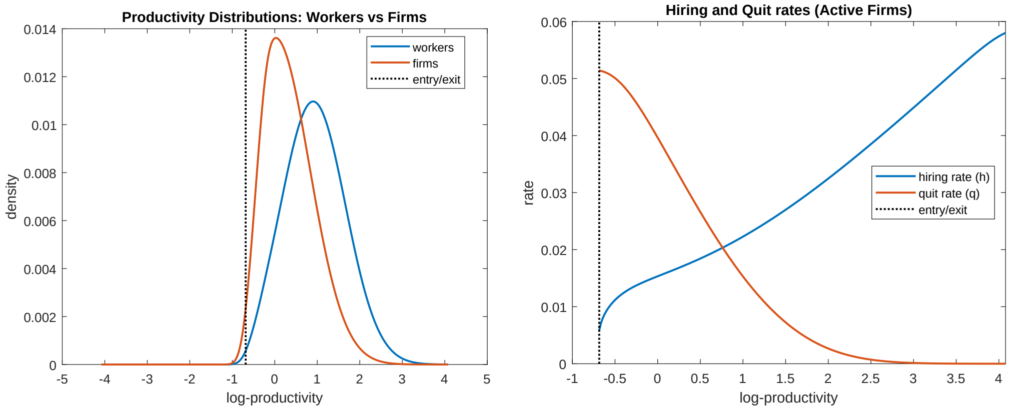


Figure 1.4: Properties of Stationary Rank-Monotonic Equilibrium at estimated parameters. Left panel: beginning of period productivity distributions of firms and workers. Right panel: monthly employment growth rate implied by the firm's policies.

of firms. I simulate a cohort of 150,000 entrants (roughly the size of a typical cohort in the Business Structure Database) for twenty years and aggregate the output from that simulation exactly like the data. Note that monthly turnover at firm i and in month t is defined as $p_{i,t}n_{i,t}$ and summed over a year to get a model equivalent to the turnover concept in the BSD and compute the labor productivity measure defined in Equation (1.2). Since this last productivity measure is in logs, the actual units of sales are irrelevant to my calibration.

The model fit to the targeted moments is shown in Table 1.2. Overall, the model replicates these statistics well, with the exception of the exit rate at young firms and the persistence of labor productivity, which are both slightly lower in the model than their empirical counterpart. The model can still account for about half of the difference in firm exit between young and old businesses.

I show how each parameter is related to each moment in Figure 1.5.[18] The figure depicts the absolute elasticity of each moment to each parameter around its estimated value. Though each parameter drives more than one moment, the following broad groups can be derived from the figure. The main parameters determining the transition rates are the job destruction rate (δ), the cost of hiring function parameters (c_1, c_2), and relative search effort s. The exit rate moments are primarily driven by the persistence of idiosyncratic productivity (ρ_p) and the mean relative productivity of entrants (μ_0). The dispersion and correlation of labor productivity are mainly determined through the flow-value of unemployment (b) and the

[18] I also report slices of the objective function for each parameter in Appendix 1.C.2.

Moment	Model	Data
Worker transitions (monthly)		
UE	0.067	0.069
EU	0.004	0.004
EE	0.020	0.020
Firm dynamics(yearly)		
Exit Rate	0.133	0.128
$\rho(LP_{i,t}, LP_{i,t-1})$	0.632	0.798
$IQR(LP_{i,t})$	0.669	0.678
Average Employment	12.8	12.8
Firm Share Young	0.621	0.560
Exit Rate Young	0.141	0.176
$IQR(LP_{i,t})$ Young	0.645	0.616

Table 1.2: Targeted moments.

standard deviation of the shocks (σ_p, σ_0). Finally, the share of young firms and employment at young firms are primarily responding to the rate of arrival of business ideas (μ), with the relative search intensity (s) and persistence of idiosyncratic productivity (ρ_p) also playing a role.

The estimated parameters are listed in Table 1.3. The estimated job destruction rate is low, since the bulk of EU transitions come from firm exit in the model.[19] There is no clear benchmark in the literature for the hiring cost function parameters because this functional form has seldom been used. I find that the implied average hiring cost as a fraction of monthly sales is 5.3%. Among the studies using a related specification, Merz and Yashiv (2007) estimates the exponent to be approximately cubic, but in a pure adjustment cost model without search frictions, while Moscarini and Postel-Vinay (2016) use a highly convex function (exponent = 50) in their baseline calibration, but with this cost applying to the number of actual hires and not the hiring rate. The relative search effort (s) of employed worker is large compared to traditional estimates obtained from US data. This reflects the fact the EE transition rate is much larger relative to the UE transition rate in British data (respectively .02 and .07 monthly in the British Household Panel Survey) than in US data (respectively .02 and .21 monthly in the Survey of Income and Program Participation). The flow-value of unemployment represents 19% of the average wage in

[19] A potential strategy to further discipline this feature would be to get an estimate of the fraction of EU transitions coming from firm exit. This information is not readily available in UK data.

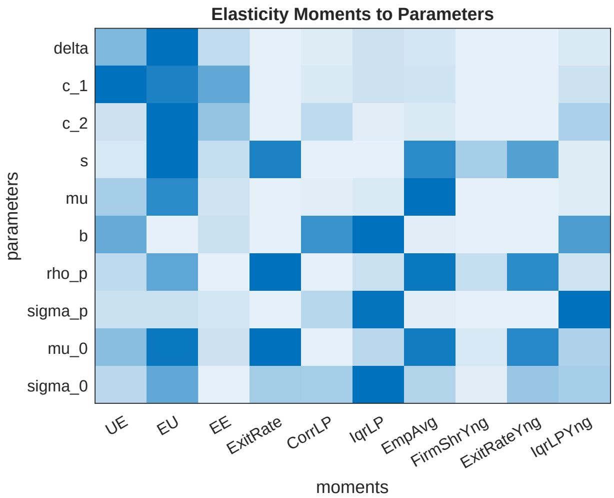


Figure 1.5: Elasticity of each moment to each parameter. Each cell corresponds to $\left|\frac{d \ln \text{moment}_j}{d \ln \text{parameter}_i}\right|$ for each parameter in row i and each moment in column j. A darker shade of blue indicates a larger absolute elasticity. The elasticities are computed by solving the model in a small neighborhood around the parameters and fitting a line through each parameter-moment series in logs.

the economy, which is about half the value used in Shimer (2005).[20] The idiosyncratic shock parameters, finally, imply a large degree of persistence of idiosyncratic productivity and a much larger dispersion of idiosyncratic productivity post-entry than pre-entry. As such post-entry shocks are a key driver of the life-cycle of the firm in the model.

1.5.3 Firm size distribution

Though the firm size distribution is not included in the set of targeted moments, the model still generates the large concentration of employment in the largest firms observed in the data. Figure 1.6 displays the normalized employment size (employment at the firm divided by average firm employment in the economy) and the associated complementary CDF (the firm's rank in terms of employment size) in the model and the data on a log-log scale. It shows that the model can replicate the log-linear relationship between firm employment and tail probability, a well-documented empirical feature of the firm size distribution. The resulting Pareto coefficient, estimated for the sample of firms larger than average size, is 1.066 in the data and 1.03 in the model.

This feature of the model can be rationalized within the framework de-

[20]Hornstein et al. (2011) show that lower values of b in the Burdett and Mortensen (1998) model yield a mean to min wage ratio more in line with the data.

Parameter	Description	Value
Pre-calibrated		
β	discount factor (≈5% annual)	0.996
Estimated		
δ	prob. job destruction (×100)	0.085
c_1	hiring cost:	45.855
c_2	$c(h) = (c_1 h)^{c_2}/c_2$	4.977
s	relative search effort	0.802
μ	prob. of start-up (×100)	0.082
b	flow value of unemployment	0.502
ρ_p	firm productivity process:	0.982
σ_p	$\ln p_{t+1} = \rho_p \ln p_t + \sigma_p \epsilon^p_{t+1}$	0.153
μ_0	initial productivity draw:	-0.200
σ_0	$\ln p_0 \sim \mathcal{N}(\mu_0, \sigma_0)$	0.050

Table 1.3: Parameter estimates.

veloped by the literature on power laws in economics (e.g., Gabaix, 1999). This line of research stresses several characteristics of the underlying process driving the size of individual units – firm employment in my setting, but typically the population of cities – that lead to a Pareto tail in steady-state. First, the growth rate of individual units is modeled through an evolving, but size independent growth rate (Gibrat's Law). Second, these individual units must be exposed to a birth-death process (Reed, 2001).

Without going into the technical details underpinning these results, I note that the evolution of firm size in my model is consistent with these requirements.[21] First, as shown in Section 1.3, the constant returns to scale assumption implies that the firm's policies are independent of employment. Conditional on survival, the growth rate at a firm with current productivity realization p is given by $(1-\mu)(1-\delta)\big(1-q(V(p))+h(p)\big)$, irrespective of its current employment. Second, the entry-exit threshold naturally generates firm birth and death.

1.6 Business Cycle

In a Stationary Equilibrium, the distribution of workers across firm productivity is stationary and consistent with the firm's optimal policies by definition. But in the presence of aggregate shocks, this distribution evolves over time and enters the firm's state space (see Definition 2). This extra

[21] Gouin-Bonenfant (2019) also gets a similarly good fit to the firm size distribution in a search model with similar properties.

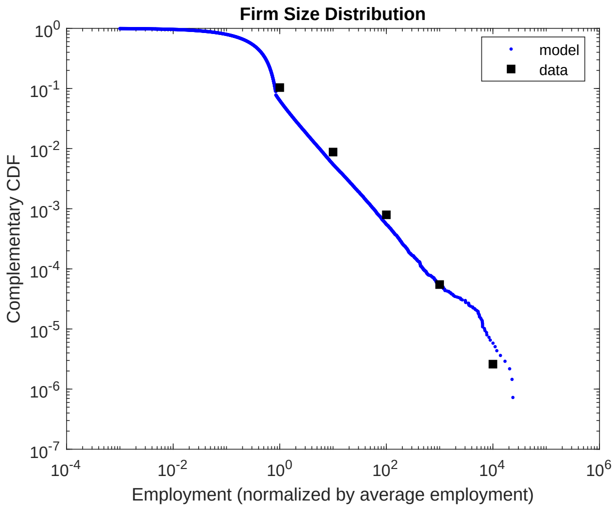


Figure 1.6: Firm size distribution at the calibrated parameters. Normalized employment is defined as employment at the firm divided by average firm employment in the economy. The complementary CDF at firm employment n is given by $\Pr(\text{employment}_{i,t} > n)$. The data series comes from the BSD and is computed separately for each year and averaged across years. The model series is obtained by simulating a cohort of entrants for one hundred years.

state variable comes with a technical hurdle since the distribution of workers across firm productivity is an infinitely dimensional object.

In this section, I start by describing the approximation used to solve the model out of steady-state. I then proceed with a series of exercises highlighting the interplay of firm dynamics and search frictions in accounting for labor productivity following the Great Recession in the UK.

1.6.1 Solving the Model with Aggregate Shocks

I now reintroduce aggregate shocks in the model. In the spirit of Krusell and Smith (1998), the measure of workers out of steady-state is approximated by a set of its moments. This measure is summarized by the unemployment rate, $u_t = 1 - \int dL_t(p)$, and a vector of moments $\boldsymbol{m}_t$ from the normalized measure of workers $L_t / \int dL_t(p)$.[22]

In addition, simulating the full model with aggregate shocks requires to solve for the firm's policy functions for all values of the aggregate shock and the distribution of workers, L_t. Given the approximation of L_t, the state-space relevant to the firm now reduces to ω_t, u_t, and $\boldsymbol{m}_t$. I then approximate the firm-worker surplus and the unemployed worker's value function out of steady-state with a polynomial.[23] For instance, the value

[22]Recall that there is a measure one of workers, so $u_t + \int dL_t(p) = 1$ by definition.

[23]I approximate the value functions and not the firm's policies directly since the latter

ω parameters		$\hat{u}_t$ targets	Data	Model
ρ_ω	0.965	$\mathrm{corr}(\hat{u}_t, \hat{u}_{t-1})$	0.937	0.920
σ_ω	0.110	$\mathrm{sd}(\hat{u}_t)$	0.081	0.084

Table 1.4: Parameters aggregate shock (ω_t). Persistence and volatility are, respectively, the first autocorrelation and standard deviation of HP-filtered (log) unemployment in the UK between 1971Q1 and 2018Q4. The model series are obtained from simulating the model and aggregating and filtering its output similarly to the data.

function for workers in unemployment is approximated as

$$\ln U(\omega_t, L_t) - \ln \bar{U} \approx \tilde{U}(\omega_t, \tilde{u}_t, \tilde{\boldsymbol{m}}_t; \theta_U)$$

where $\tilde{x}_t$ denotes a variable in log-deviation from steady-state and θ_U is a vector of coefficients to be solved for. The firm-worker surplus is similarly approximated, using a separate polynomial at each idiosyncratic productivity node. The solution algorithm proceeds by repeatedly simulating the model until the coefficients converge. Additional details regarding the implementation of this algorithm can be found in Appendix 1.C.3.

An alternative approach to simulate heterogeneous agents models with aggregate shocks is to use the perturbation method proposed by Reiter (2009). Such linearization techniques have been successfully applied to firm dynamics models (Sedláček and Sterk, 2017; Winberry, 2016). However, my simulations suggest that this first-order approximation is highly inaccurate in the context of my model due to the discontinuity implied by the firm's entry and exit threshold. I therefore choose the simulation-based approach outlined here and report accuracy tests for my proposed algorithm in Appendix 1.C.5.

Finally, the Markov process for aggregate productivity shocks is assumed to follow $\ln \omega_{t+1} = \rho_\omega \ln \omega_t + \sigma_\omega \epsilon^\omega_{t+1}$ with $\epsilon^\omega_{t+1} \sim \mathcal{N}(0, 1)$. The parameters in this process $(\rho_\omega, \sigma_\omega)$ are chosen to replicate the model-simulated persistence and volatility of unemployment in the UK between 1971 and 2018. They are shown in Table 1.4.

1.6.2 The Great Recession in the Model

To understand the reallocation effects of a large recession in the model, I input a sequence of aggregate shocks that triggers a sharp rise in unemployment, akin to the UK experience during the Great Recession. I show

are not smooth functions of the aggregate states due to the entry/exit threshold.

that the model can generate a reduction in the OP measure of allocation of workers to firms that is in line with the patterns documented in the data. I then leverage the model to further account for changes in labor productivity following the recession in terms of firm selection and worker reallocation.

Aggregate responses. Figure 1.7 displays key model generated aggregates along with the respective data series. The top left panel shows the Great Recession counterfactual I run in the model: I input a sequence of aggregate shocks designed to replicate the sharp increase in unemployment observed at the onset of the recession. Aggregate productivity is then left to revert back to its steady-state level according to the persistence parameter given in Table 1.4.[24]

The remaining plots benchmark the implied responses in the number of active firms, the labor market transition rates, and vacancies against the corresponding data series.[25] All data series are shown in deviation from their pre-recession linear trend. Overall, the model does a decent job at replicating the overall pattern of these aggregates, keeping in mind that these series are not targeted and that the model is calibrated on the pre-recession period based on its stationary solution.

As an additional validation, I study the reallocation of workers implied by the model in the simulated recession. The measure of worker reallocation used is similar to the empirical part of the paper and given by

$$\sum_i \left(ES_{i,t} - \overline{ES}_t\right)\left(LP_{i,t} - \overline{LP}_t\right),$$

an expression that increases with a higher share of workers employed at the most productive firms. Figure 1.8 benchmarks the model response against the data in deviation from their pre-recession linear trend. It shows that the model generates a drop in this measure that is, overall, similar in magnitude to that observed in the data. Though it recovers more quickly than the data, it still accounts for more than fifty percent of the overall response by 2015, seven years after the start of the recession.

[24]Figure 1.19 in Appendix 1.C.6 shows that unemployment exhibits marked non-linearities as a response to aggregate shocks in the model. These non-linearities justify fully solving the model with aggregate shocks so that agents appropriately incorporate future uncertainty in their decisions over the course of the simulated recession.

[25]While it is not necessary to specify a matching function to solve the model since it can be solved using the identity $\lambda_t Z_t = \eta_t A_t$, a functional from assumption is required to back out vacancies. I use the standard Cobb-Douglas form $\xi A_t^{\alpha} Z_t^{1-\alpha}$ where I normalize $\xi = 1$ and set the elasticity of matches to vacancies to .5.

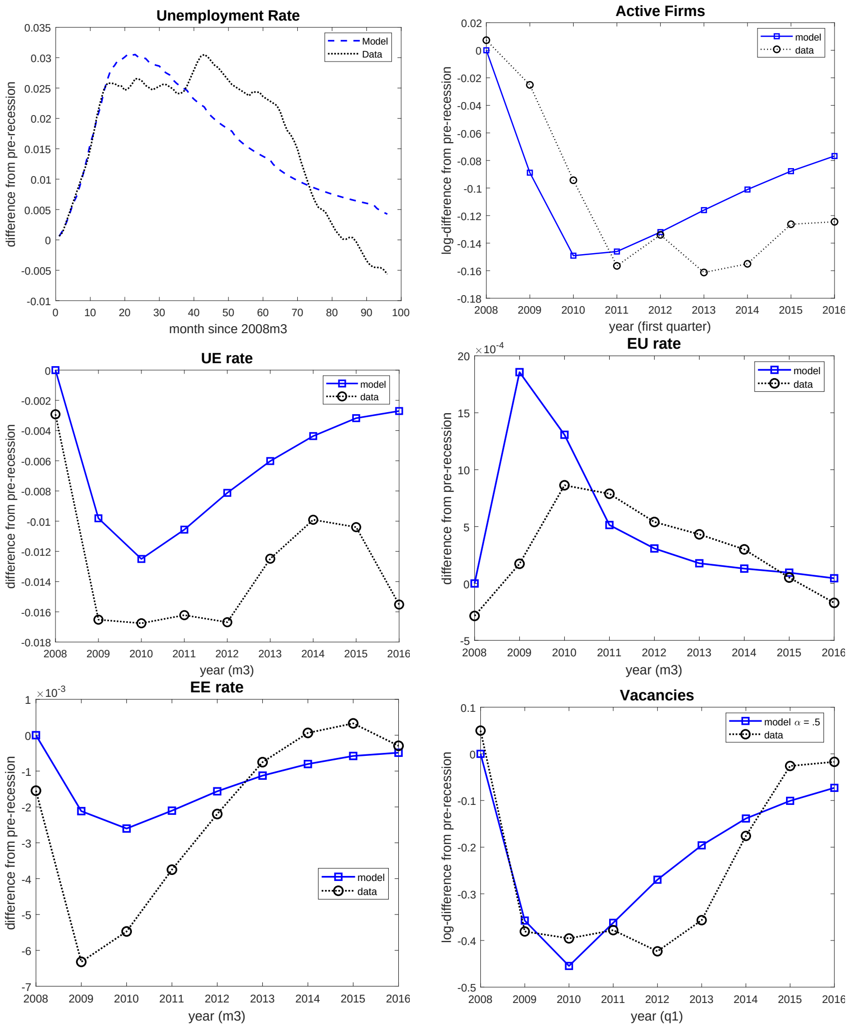


Figure 1.7: Aggregate model responses to a sequence of productivity shocks triggering the increase in unemployment depicted in the top left panel. See Appendix 1.B.4 for the source of additional series.

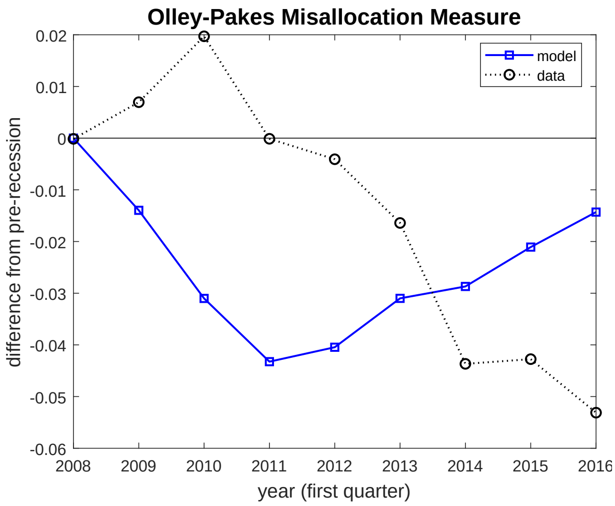


Figure 1.8: Olley-Pakes misallocation measure during the simulated recession. The data is computed in deviation from their pre-recession linear trend. The model series is computed simulating a cohort of firms over the course of the recessionary episode and aggregating its output similarly to the data.

The Olley-Pakes decomposition through the lens of the model. Recall that the labor productivity index used in the empirical part of the paper is given by

$$LP_t = \sum_i ES_{i,t} \times LP_{i,t},$$

where $ES_{i,t}$ and $LP_{i,t}$ denote, respectively, the employment share and labor productivity at firm i in period t. In the notation of the model, this expression rewrites

$$LP_t = \int \underbrace{\ln\left(\frac{\omega_t pn}{n}\right)}_{\text{firm productivity}} \underbrace{d n\overline{v}_t^P(p,n)}_{\text{employment ``share''}} = \ln \omega_t + \int \ln(p) d\overline{L}_t^P(p),$$

where the superscript "P" denotes the production stage (end of period) and a bar denotes a normalized measure.[26] This last equality makes clear that aggregate labor productivity is determined by the aggregate shock and the employment-weighted distribution of firm productivity, $\overline{L}_t^P$, an object shaped by firm dynamics and search frictions in equilibrium.

Again LP_t can be further decomposed into a firm productivity compo-

[26]So $\overline{L_t}^P(p) := \int_{\tilde{p}\leq p} dL_t^P(\tilde{p}) / \int dL_t^P(p)$. Besides, because a model period is a month, this expression is monthly labor productivity.

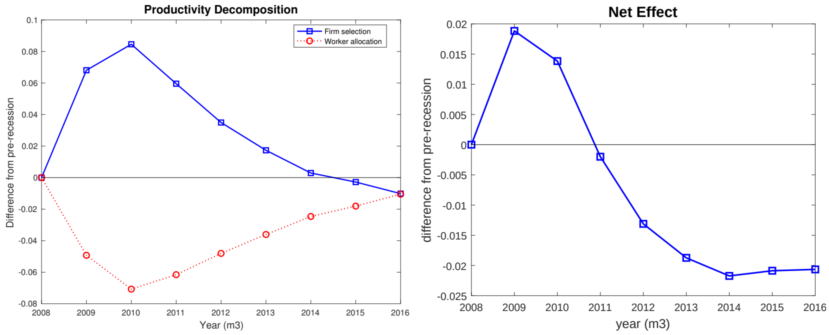


Figure 1.9: Labor productivity decomposition in the model.

nent and a worker reallocation component. The equality

$$LP_t = \sum_i ES_{i,t} \times LP_{i,t} = \overline{LP}_t + \sum_i \left(ES_{i,t} - \overline{ES}_t\right)\left(LP_{i,t} - \overline{LP}_t\right)$$

can be written in the model as

$$LP_t = \underbrace{\ln \omega_t}_{\text{aggregate shock}} + \underbrace{\int \ln(p) dK_t^P(p)}_{\text{firm selection}} + \underbrace{\int \ln(p) dL_t^P(p) - \int \ln(p) dK_t^P(p)}_{OP_t:=\text{measure of allocation}}. \tag{1.19}$$

In this expression, the first term gives the direct impact of the aggregate shock, the second term captures changes in the distribution of firms across productivity level. Finally, the last term corresponds to the allocation of worker to firms. In the model, it relates directly to the difference between the distribution of workers across firm productivity (L_t^P) and the distribution of firms across productivity (K_t^P), two objects jointly determined in equilibrium in the model.

Through the lens of the model, aggregate labor productivity is then made up of two endogenous terms: firm selection and worker reallocation (the last two terms in Equation 1.19). I plot the evolution of these two components over the course of the simulated recession in Figure 1.9. It shows that they represent opposite forces shaping labor productivity. However, while they are initially of the same magnitude, the worker reallocation term exhibits more persistence. It is still negative eight years after the start of the recession. On net, the worker reallocation effect dominates in the medium term, as shown on the right panel of Figure 1.9.

Inspecting the worker reallocation mechanism. I illustrate the main worker reallocation mechanism at the micro level in Figure 1.10. On top of the firm selection effect, which shifts the entry threshold upward, how

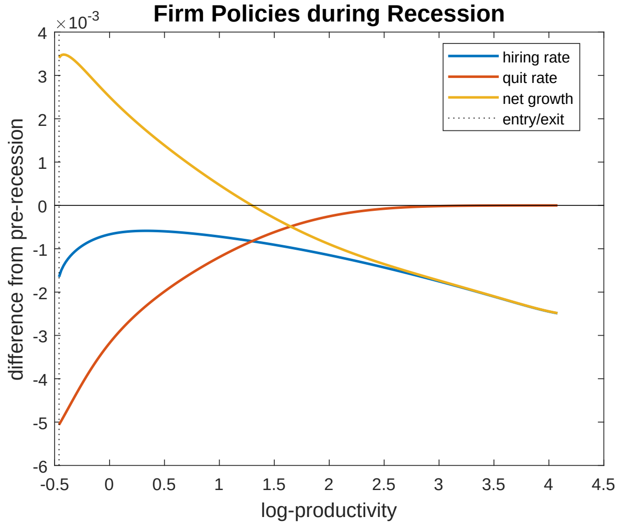


Figure 1.10: Firm policies during the simulated recessionary episode. Policies are averaged across simulation periods and presented in deviation from the stationary equilibrium. Idiosyncratic productivity is truncated to only include firms above the entry/exit threshold at all point in time after the shock.

well labor is allocated to firms also depends on which firms grow faster following the shock. Figure 1.10 shows changes in the quit rate, hiring rate, and net employment rate with respect to their pre-recession level along the firm-specific productivity dimension.

The figure shows that while the hiring rate drops at all productivity levels with respect to the pre-recession period, the quit rate drops even more at the bottom of the productivity distribution. This is because, in a random search environment, the probability for workers to draw an offer from a high-productivity firm is reduced as they compete with more unemployed workers. Since voluntary quits are always productivity enhancing in equilibrium, this reduction in the quit rate contributes to dampening labor productivity.

The fact that the resulting net employment growth rate increases – in relative terms, since these firms are still shrinking, but not as fast as they would in normal times – at the bottom of the productivity distribution during the shock is consistent with the firm-level data. In Table 1.1, I find that the relationship between labor productivity at the firm level and employment growth becomes less positive in the aftermath of the recession. While this relationship cannot be decomposed further into hires, quits, and layoffs without matched employer-employee data, the drop in the quit rate at low quality firms is consistent with the evidence described in Haltiwanger et al. (2018) for the United States. These authors find that job-to-job

A. Log-deviation from pre-recession (2015)	
Data	-0.207
Model	-0.133
B. Implied decomposition	
Average Firm Wage	0.517
OP wage term	0.483

Table 1.5: Wage decomposition in the simulated model.

transitions out of the bottom rung of the wage ladder – where firms are ranked based on wages and not productivity as in Figure 1.10 – decline by eighty-five percent during the Great Recession.

Implications for wages. It has been documented that the Great Recession resulted in substantial downward pressure on wages.[27] A potential channel to account for this fall in average wages is the reallocation of workers across low- and high-wage employers induced by aggregate shocks. The average wage paid at each point in time can be written, similarly to Equation (1.19), as

$$\underbrace{\int w_t(p) dK_t^P(p)}_{\text{average wage paid by firm}} + \underbrace{\int w_t(p) dL_t^P(p) - \int w_t(p) dK_t^P(p)}_{\text{allocation of workers across wages}}. \tag{1.20}$$

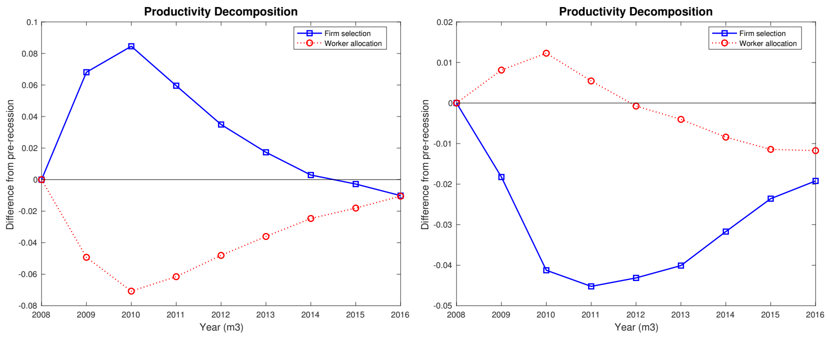
In Equation (1.20), the first term captures the reduction of labor productivity passed down to wages by firms, while the second term relates to the re-allocation of workers across the distribution of wages paid by firms.

Table 1.5 reports this decomposition for the year 2015 in the simulated model, seven years after the start of the recession.[28] It first shows that the model accounts for a substantial share of the overall drop in the average wage following the onset of the recession. Second, through the lens of the model, about half of this drop stems from the reallocation of workers to employers offering different wages – the second term in Equation (1.20). It then suggests that the changing patterns of worker reallocation over the course of the recession can act as potential channel downgrading wages.

Alternative aggregate shock. It is assumed so far that a single aggregate productivity shock drives the business cycle. I report results for

[27] See Blundell et al. (2014) for UK evidence.

[28] This decomposition cannot be decomposed directly from the data, as the Business Structure Database has no information on pay.



(a) Productivity shock **(b)** Job destruction shock

Figure 1.11: Labor productivity decomposition in the model.

an alternative type of aggregate shock, whereby the business cycle is now driven by a job destruction shock instead. I assume that the job destruction rate—the rate at which workers are exogenously separated into unemployment independently of a firm's characteristics—follows $\ln \delta_{t+1} = \rho_\delta \ln \delta_t + \sigma_\delta \epsilon^\delta_{t+1}$, $\epsilon^\delta_{t+1} \sim \mathcal{N}(0,1)$, where ρ_δ and σ_δ are set to replicate the persistence and volatility of unemployment in the UK, similarly to the calibration of the aggregate productivity shock.

I benchmark the model's reallocation properties in response to each type of aggregate shock. I find a sequence of $\{\epsilon^\delta_\tau\}^{t_1}_{\tau=t_0}$ shocks that replicate the sharp increase in the unemployment rate observed in the UK at the start of the Great Recession, the equivalent to the main experiment run with aggregate productivity shocks. ($[t_0, t_1]$ corresponds to the period when GDP is contracting in the UK, from 2008m3 to 2009m6.)

Figure 1.11 shows the productivity decomposition from Equation (1.19) in each case. A striking feature is that the implied decompositions look very different. In particular, while an aggregate productivity shock creates a standard cleansing effect, making firms at the bottom of the productivity distribution exit, a destruction shock has the opposite effect. It has no direct impact on firm output, and lowers the rate of voluntary quits, as there are now more unemployed workers that can be reached by more productive firms. Also a destruction shock can have a positive impact on firm profits, as the optimal contract directly depends (negatively) on δ, as seen in (1.16). This appears to dominate at the lower end of the productivity distribution, so that the exit threshold becomes less stringent for this type of shock. As a consequence, both firm selection and worker reallocation contribute to pushing productivity down.

1.6.3 Policy Experiment: Unemployment-contingent Benefits

The trade-off between firm selection and worker reallocation during a recession can be further illustrated in the following policy experiment. In the spirit of unemployment insurance extensions in the US, I allow the value of non-employment, b, to depend on the unemployment rate.[29] Specifically, the value of non-employment is assumed to vary with unemployment benefits according to

$$\ln b_t - \ln \overline{b} = \kappa \times (\ln u_t - \ln \overline{u}),$$

where $\overline{b}$ and $\overline{u}$ denote, respectively, the value of non-employment and the unemployment rate in the stationary equilibrium and $\kappa \geq 0$ is the elasticity of unemployment benefits to the unemployment rate. I tentatively set $\kappa = .3$ and solve the model again using the same sequence of aggregate shocks as in the benchmark economy.

Figure 1.12 compares the model response under the unemployment-contingent benefit policy ($\kappa > 0$) to the baseline model with constant b ($\kappa = 0$) over the course of the simulated Great Recession. With respect to labor productivity, such a policy has two opposite effects. First, it makes the selection effect more stringent. Unemployment increases by three percentage points at its peak in the baseline model and by almost four and a half points under the alternative policy. This effect is reflected in the firm selection term, which is also more positive since the entry threshold is higher.

Second, unemployment-contingent benefits magnify the worker reallocation effect resulting from search frictions. As can be seen from the firm's policies, the quit rate drops even more at the bottom of the distribution in this case: workers employed at these firms must compete with more unemployed workers to climb up the contract-productivity ladder. While the net effect of the policy is still positive in my calibration, the model does suggest that such policies can also have negative consequences on labor productivity by decreasing the pace of worker reallocation to more productive units.

[29] The actual policy makes the duration, and not the level, of unemployment benefits contingent on the unemployment rate. I focus on the level of these benefits to avoid the need to introduce an extra state variable for unemployed workers off and on benefits. See Rujiwattanapong (2019) for a model fully capturing the unemployment insurance extension mechanism.

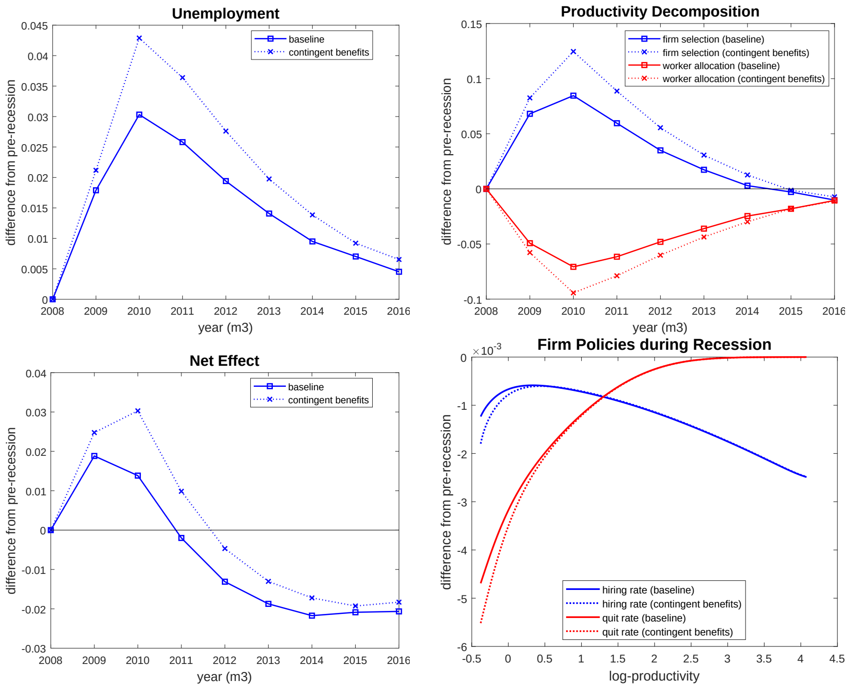


Figure 1.12: Baseline vs Unemployment-contingent benefits model. Model response to the simulated recession using the same sequence of aggregate shocks.

1.7 Conclusion

I develop a model with three key features: (i) on-the-job search, (ii) firm dynamics, (iii) aggregate shocks. Firms with heterogeneous productivities compete to attract and retain workers in a frictional labor market. In equilibrium, job-to-job transitions are always productivity enhancing, as more productive firms offer better contracts. I use the model to analyze how firms' recruiting behaviors at the micro level drive the evolution of aggregate labor productivity at the macro level in the aftermath of a recession.

The central insight of the model is that search frictions dampen labor productivity following a large aggregate shock. On-the-job search causes the quit rate – the rate at which workers voluntary leave their current job to take a better one – to drop on the lower part of the productivity distribution after a recession. Search frictions then hamper the reallocation of workers from less to more productive firms.

In an experiment designed to replicate the increase in unemployment observed during the UK Great Recession, I find that this channel can account for a large portion of the drop in the allocation of workers to firms measured in British firm-level data. Through the lens of the model, this negative worker reallocation effect dominates the positive firm selection effect implied by the aggregate shocks in the medium term.

Appendix 1.A Theory Appendix

1.A.1 Difference with Coles and Mortensen (2016)

The central difference between my approach and the model developed in Coles and Mortensen (2016) is in the wage setting protocol. Similarly to Moscarini and Postel-Vinay (2013), I assume that firms can fully commit to delivering a state-contingent wage after each future realization of some firm-specific and aggregate states, which I precisely define in the paper. Coles and Mortensen (2016) assume firms cannot commit to such a wage plan, but instead that workers do not observe firm-level productivity and form beliefs on that productivity from the wage offered by the firm.

To make this difference explicit, I rewrite the firm's problem under each set of assumptions on the wage-setting protocol. A result common to both papers is that the present value of profits is linear in firm employment n. I therefore focus on the present value of profits per worker π_t. In my model,

$$\pi_t\left(p_t, \overline{V}\right) = \max_{\substack{h\geq 0 \\ w \\ W}} \left\{ \omega_t p_t - w + \beta E_t \Big[-c(h) + (1 - q(W) + h)\pi_{t+1}(p_{t+1}, W) \Big] \right\}, \quad (1.21)$$

subject to the promise-keeping constraint

$$\overline{V} = w + \beta E_t \Big\{ \delta U_{t+1} + (1-\delta)\Big[(1 - q(W))W + s\lambda_{t+1} \int \max\left(W', U_{t+1}\right) dF_{t+1}(W')\Big] \Big\}. \quad (1.22)$$

In the recursive formulation, full commitment on the firm's side implies that it must deliver, in expectation, $\overline{V}$ when choosing the wage rate w and continuation values W. With risk-neutral workers, wages can be substituted out from (1.21) using (1.22), and the optimal contract can be shown to be increasing in productivity and expressed as a function of the firm-worker surplus.

The discrete time equivalent to (1.21) in Coles and Mortensen (2016) is given by

$$\pi_t\left(p_t, \overline{w}\right) = \omega_t p_t - \overline{w} + \beta E_t \max_{\substack{w \\ h\geq 0}} \Big\{ -c(h) + (1 - q(w) + h)\pi_{t+1}(p_{t+1}, w) \Big\}, \quad (1.23)$$

where there is no commitment to a wage plan across periods, though I assume for simplicity that the firm can commit to pay at the production stage the wage it announces at the search stage.[30] They then describe an equilibrium in which workers form beliefs on the productivity of firms and show that it is optimal for more productive firms to offer higher wages.

Equations (1.21) and (1.23) make clear that the firm trades off hiring new workers and retain existing employees in controlling its rate of employment growth. And both characterizations of equilibrium entail that workers move towards more productive firms, as they offer higher wages, when making a job-to-job move. But the characterization in Coles and Mortensen (2016) is obtained under stronger assumptions:

1. No wedge in search effort, so $s = 1$ in the notation of my model.

2. No endogenous firm entry and exit, which requires $\underline{p} > b$.

While these restrictions can potentially be relaxed, as their main purpose is to have the reservation wage equal to the value of non-employment (b), numerically solving for the reservation wage in the general case could be demanding, as it involves an intricate fixed-point problem. My model can readily accommodate endogenous entry and exit, as well as a different level of search effort for workers on and off the job.

To gage the quantitative difference implied by each set of assumption on wage determination, Figure 1.13 shows the wage profile obtained by simulating a version of the model under each wage-setting protocol. Note that I use a different calibration than in the thesis to accommodate the extra restrictions imposed by Coles and Mortensen (2016). This exercise suggests that, at least for this choice of parameters, workers are able to extract more from the production flow in the bargaining protocol with firm commitment.

1.A.2 Size-independent Discounted Profits

We want to guess and verify that a solution to the functional equation (1.4) has the form $n_t \pi_t(p_t, \overline{V})$. That is, we want to show that

$$\Pi_{t+1}(p_{t+1}, n_{t+1}, W_{t+1}) = n_{t+1}\pi_{t+1}(p_{t+1}, W_{t+1})$$

[30] Coles and Mortensen (2016) do not have to deal with this complication as their model is set in continuous time. I translate their model to discrete time to make the comparison sharper.

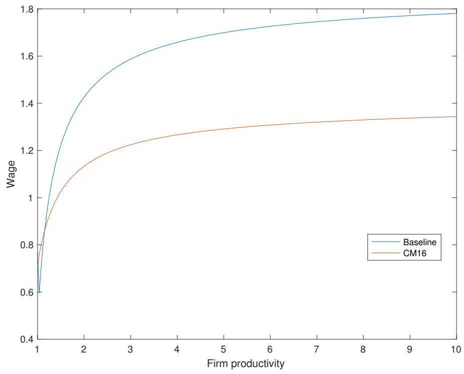


Figure 1.13: Wages in Baseline vs Coles and Mortensen (2016) model at the steady-state.

entails that

$$\Pi_t(p_t, n_t, \overline{V}) = n_t \pi_t(p_t, \overline{V}).$$

Start from (1.4), still subject to the Promise-Keeping constraint and the law of motion for its workforce. Plugging in the guess in (1.4) gives

$$\begin{aligned} &E_t\Big[\chi_{t+1}(p_{t+1})\Big(-c(h_{t+1})(1-\mu)n_t + \Pi_{t+1}(p_{t+1}, n_{t+1}, W_{t+1})\Big)\Big] \\ =&E_t\Big[\chi_{t+1}(p_{t+1})\Big(-c(h_{t+1})(1-\mu)n_t + n_{t+1}\pi_{t+1}(p_{t+1}, W_{t+1})\Big)\Big]. \end{aligned}$$

Now substitute the law of motion for the firm's workforce in the last expression. Note that with a continuum of workers, it is assumed to hold exactly condition on the firm surviving and $\rho_{t+1}(W_{t+1})$, h_{t+1}. This substitution would still work with a discrete number of workers as long as the law of motion holds in expectation, so the Law of Iterated Expectations can be applied conditioning on the realization of the shocks at the start of the period. Substituting n_{t+1} then yields

$$n_t E_t\Big[\ \chi_{t+1}(p_{t+1})\Big(-c(h_{t+1}) + (\rho_{t+1}(W_{t+1}) + h_{t+1})\pi_{t+1}(p_{t+1}, W_{t+1})\Big)\Big].$$

Using this last expression in the main profit equation, it follows directly that firm profits are linear in n_t, as shown in (1.7).

1.A.3 Derivation Match Surplus

Recall that the joint value of a match is defined as $S_t(p_t) := \pi_t(\overline{V}) + \overline{V}$. Rearranging the Promise-Keeping constraint gives an expression for w_t

$$\begin{aligned} w_t = \overline{V} - \beta E_t \Bigg\{ & \mu Q_{t+1} \\ & + (1-\mu)\Bigg[(1-\chi_{t+1}(p_{t+1}))U_{t+1} \\ & + \chi_{t+1}(p_{t+1})\Big(\delta_{t+1}U_{t+1} + (1-\delta_{t+1})(1-s\lambda_{t+1}\bar{F}_{t+1}(W_{t+1}))W_{t+1} \\ & + s\lambda_{t+1}\int_{W_{t+1}}^{\infty}\theta dF_{t+1}(\theta)\Big)\Bigg]\Bigg\}. \end{aligned}$$

Substituting w_t in the expression for firm profit per worker (1.7) gives

$$\begin{aligned} S_t(p) &:= \pi_t(p, \overline{V}) + \overline{V} \\ &= -\overline{V} + \max_{\substack{h_{t+1}\geq 0 \\ W_{t+1}}} \Bigg\{ p_t\omega_t + \beta E_t\Bigg[\mu Q_{t+1} \\ &+ (1-\mu)\Big((1-\chi_{t+1}(p_{t+1}))U_{t+1} \\ &+ \chi_{t+1}(p_{t+1})\big(\delta U_{t+1} + (1-\delta)((1-s\lambda_{t+1}\bar{F}_{t+1}(W_{t+1}))W_{t+1} \\ &+ s\lambda_{t+1}\int_{W_{t+1}}^{\infty}\theta dF_{t+1}(\theta)) \\ &- c(h_{t+1}) + (\rho_{t+1}(W_{t+1}) + h_{t+1})\pi_t(p_{t+1}, W_{t+1}))\Big)\Bigg]\Bigg\} + \overline{V}. \end{aligned}$$

Finally, taking the max operator inside the expectation and grouping terms yields

$$\begin{aligned} S_t(p) = p_t\omega_t + \beta E_t\Bigg[& \mu Q_{t+1} \\ & + (1-\mu)\Bigg((1-\chi_{t+1}(p_{t+1}))U_{t+1} \\ & + \chi_{t+1}(p_{t+1}) \max_{\substack{h_{t+1}\geq 0 \\ W_{t+1}}} \big\{ -c(h_{t+1}) \\ & + \rho_{t+1}(W_{t+1})S_{t+1}(p_{t+1}) + h_{t+1}(S_{t+1}(p_{t+1}) - W_{t+1}) \\ & + (1-\delta)s\lambda_{t+1}\int_{W_{t+1}}^{\infty}\theta dF_{t+1}(\theta)\big\}\Bigg)\Bigg]. \end{aligned}$$

1.A.4 Definition Acceptance Rate

Define G_t the share of workers employed at firms offering contract value less than W in the current period

$$G_t(W) := \frac{\displaystyle\int \mathbb{1}\left\{W_t(p) \le W\right\} \chi_t(p) n d\nu_t}{\displaystyle\int \chi_t(p) n d\nu_t}.$$

The acceptance rate at some offered W is then given by

$$Y_t(W) := \frac{u_t + s(1-\delta) G_t(W) \displaystyle\int \chi_t(p) n d\nu_t}{u_t + s(1-\delta) \displaystyle\int \chi_t(p) n d\nu_t},$$

where the numerator is the (intensity-weighted) measure of workers currently employed at firms offering contracts less than W and the denominator is the total measure of such workers.

1.A.5 Proof Rank-Monotonic Equilibrium

The outline of the proof is similar to that in Moscarini and Postel-Vinay (2013, 2016). The key difference is that the firm's problem can be considered separately for each worker since $\Pi_t\left(p, n, \overline{V}\right) = n\pi_t\left(p, \overline{V}\right)$. There is therefore no need to show super-modularity of the firm-worker surplus in its productivity and own size. It is enough to show that the firm-worker surplus is increasing in p, which implies that the optimal contract is also increasing in p, conditional on some convexity requirements of the cost of hiring function. We want to prove the two following statements:

1. Conditional on S being increasing in p, $\frac{c''(h)h}{c'(h)} \ge 1, \forall h \ge 0$ is sufficient to guarantee that V is increasing in p;

2. The firm-worker surplus mapping defined by (1.10) implies that S is increasing in p.

Taking each point in order:

1. Sufficient conditions on c for a RME Conditional on the firm surviving, the maximization problem associated with (1.10) defines the optimal contract and hiring rate after all histories. At any interior maximum, the following first-order conditions are associated with (1.11)

$$\begin{aligned}[h]: &\quad c'(h) = S(p) - W \\ [W]: &\quad \rho'(W)(S(p) - W) = h,\end{aligned}$$

where I have dropped the time subscripts, but S and ρ implicitly depend on calendar time in what follows. In addition, at any maximum, the associated Hessian matrix, H, is negative-definite, which requires

$$\det(H) = -c''(h)\Big(\rho''(W)\big(S(p) - W\big) - \rho'(W)\Big) - 1 > 0.$$

The two FOCs can be combined to give the following expression in W

$$-c'\big(\rho'(W)\big(S(p) - W\big)\big) + S(p) - W = 0$$

and totally differentiating that last expression with respect to p gives

$$\frac{dW}{dp} = \frac{\frac{\partial S(p)}{\partial p}\big(c''(h)\rho'(W) - 1\big)}{\det(H)}.$$

In this last expression, the denominator is positive at any maximum. By assumption, the firm-worker surplus is increasing in p, so $\frac{\partial S(p)}{\partial p} \geq 0$. Noting that the two FOCs can be combined to give $\rho'(w) = \frac{h}{c'(h)}$, it follows that

$$\frac{dW}{dp} \geq 0 \iff c''(h)\rho'(W) \geq 1 \iff \frac{c''(h)h}{c'(h)} \geq 1.$$

2. Firm-worker surplus increasing in p In this part of the proof, we want to show that S is increasing in p, which was assumed in the previous part. I follow the proof strategy outlined in (Moscarini and Postel-Vinay, 2013, Appendix A) and start by showing that the mapping defined by (1.10) maps from the space of differentiable, bounded and increasing functions into itself, conditional on a constant measure of firms $\nu(p, n)$. With this condition, the Continuous Mapping Theorem can be applied, so the net-surplus defined by the mapping exists, is unique, and increasing in p.

In a second step, the condition on ν is relaxed. In this case, the Continuous Mapping Theorem cannot be applied, as S is no longer defined on $\mathcal{R}^N$. But, since it is known that S is increasing in p in the restrictive case and that this solution is unique, we know that every candidate solution of the unrestricted mapping should have the property as well.

In the remainder of the proof, we then fix the beginning of period measure of firms to some value. We want to show that the mapping defined by (1.10) maps from the space of differentiable, bounded and increasing functions into itself. Differentiability in p follows directly from noting that the expectation in (1.10) is differentiable in p as long as the conditional probability density of future productivity is. This can be assumed. Since the support of p is convex and closed, it also follows that the mapping defined in (1.10) maps into the set of bounded functions.

Finally, to show that the mapping is increasing in p, first note that, for continuing firms, the envelope condition on the firm's optimization problem (1.11) gives

$$\begin{aligned}\frac{d\psi_{t+1}(p)}{dp} &= \frac{\partial\psi_{t+1}(p)}{\partial p}\\ &= \Big(\rho_{t+1}(V^*) + h^*\Big)\frac{\partial S_{t+1}(p)}{\partial p} \geq 0,\end{aligned}$$

where V^*, h^* denote optimal policies. The term inside the expectation in the firm-worker surplus (1.10) is then weakly increasing in p: constant on the part of the support of p_{t+1} where the firm exits, and weakly increasing otherwise.

To complete the proof, an additional assumption is needed on the idiosyncratic productivity shock. Namely, it has to be assumed that given a higher realization of productivity in the current period, the conditional Cumulative Distribution Function of future productivity satisfies first-order stochastic dominance.

With this assumption, conditional on any two distinct previous realizations of p, the conditional densities of future idiosyncratic productivity satisfy a single-crossing property. Let p_0 denote this crossing point and let $p_1,\ p_2$ be two values in $[\underline{p}, \overline{p}]$ such that $p_2 > p_1$, then

$$\begin{aligned}S_t(p_2) - S_t(p_1) &= \omega_t(p_2 - p_1)\\ &\quad + \beta(1-\mu)\bigg(E_t\Big[\kappa_{t+1}(p)\ |p_2\Big] - E_t\Big[\kappa_{t+1}(p)\ |p_1\Big]\bigg),\end{aligned}$$

where $\kappa_{t+1}(p)$ is a notation for the terms inside the expectation

$$\kappa_{t+1}(p) := (1-\chi_{t+1}(p))U_{t+1} + \chi_{t+1}(p)\Big(\delta U_{t+1} + \psi_{t+1}(p)\Big).$$

(The μQ_{t+1} terms are independent of the previous value of p, so they cancel.) Showing that S_t is increasing in p now amounts to show that the difference in expectation in the last expression is non-negative. This difference can be rewritten

$$\int_{\underline{p}}^{\overline{p}} E_t\Big[\kappa_{t+1}(p)\Big]\Big(\gamma(p|p_2) - \gamma(p|p_1)\Big)dp,$$

where $\gamma(p|p_i)$ is the density of the p-shock conditional on p_i.

Now, given the crossing-point p_0, we can rewrite

$$\begin{aligned}
&\int_{\underline{p}}^{\overline{p}} E_t\Big[\kappa_{t+1}(p)\Big]\Big(\gamma(p|p_2) - \gamma(p|p_1)\Big)dp \\
&= \int_{\underline{p}}^{p_0} E_t\Big[\kappa_{t+1}(p)\Big]\Big(\gamma(p|p_2) - \gamma(p|p_1)\Big)dp \\
&+ \int_{p_0}^{\overline{p}} E_t\Big[\kappa_{t+1}(p)\Big]\Big(\gamma(p|p_2) - \gamma(p|p_1)\Big)dp
\end{aligned}$$

and, since $E_t\Big[\kappa_{t+1}(p)\Big]$ is weakly increasing in p, we get the following inequalities

$$\begin{aligned}
&\int_{\underline{p}}^{p_0} E_t\Big[\kappa_{t+1}(p)\Big]\Big(\gamma(p|p_2) - \gamma(p|p_1)\Big)dp \\
&\geq E_t\Big[\kappa_{t+1}(p)\Big]\int_{\underline{p}}^{p_0} \Big(\gamma(p|p_2) - \gamma(p|p_1)\Big)dp
\end{aligned}$$

and

$$\begin{aligned}
&\int_{p_0}^{\overline{p}} E_t\Big[\kappa_{t+1}(p_0)\Big]\Big(\gamma(p|p_2) - \gamma(p|p_1)\Big)dp \\
&\geq E_t\Big[\kappa_{t+1}(p_0)\Big]\int_{p_0}^{\overline{p}} \Big(\gamma(p|p_2) - \gamma(p|p_1)\Big)dp.
\end{aligned}$$

Finally, summing up the last two inequalities, we get

$$\begin{aligned}
&E_t\Big[\kappa_{t+1}(p)\ |p_2\Big] - E_t\Big[\kappa_{t+1}(p)\ |p_1\Big] \\
&= \int_{\underline{p}}^{\overline{p}} E_t\Big[\kappa_{t+1}(p)\Big]\Big(\gamma(p|p_2) - \gamma(p|p_1)\Big)dp \geq 0,
\end{aligned}$$

which shows that $S_t(p_2) \geq S_t(p_1)$ for $p_2 > p_1$.

1.A.6 RME Contracts

This Appendix proves that the RME contract has the form given in (1.16). Before turning to the actual proof, I first show that the contract offer distribution, F_t, rewrites

$$\begin{aligned}F_t(W) &:= A_t^{-1} \int \mathbb{1}\left\{W_t(p) \leq W\right\} \chi_t(p) a_t(p,n) d\nu_t \\ &= \int \mathbb{1}\left\{W_t(p) \leq W(p)\right\} \frac{\chi_t(p)(1-\mu)n}{Z_t \lambda_t Y_t(W)} d\nu_t,\end{aligned}$$

where the substitution follows from the firm's vacancy posting position (1.15) and the equality $\eta_t A_t = \lambda_t Z_t$. Besides, in a RME, contracts are strictly increasing in p, so we have

$$G_t(W_t(p)) = \frac{\displaystyle\int_{\underline{p}}^{p} \chi_t(p') dL_t(p')}{\displaystyle\int_{\underline{p}}^{\overline{p}} \chi_t(p') dL_t(p')} = \frac{L_t(p) - L_t(p_E)}{L_t(\overline{p}) - L_t(p_E)},$$

where p_E denotes firm's entry/exit threshold and the acceptance rate can now be simplified as

$$Y_t(V_t(p)) = \frac{u_t + s(1-\delta)\Big(L_t(p) - L_t(p_E)\Big)}{u_t + s(1-\delta)\Big(L_t(\overline{p}) - L_t(p_E)\Big)}.$$

Finally, plugging this last expression into the contract offer distribution evaluated at $V(p)$ gives Equation (1.17)

$$\lambda_t F_t\left(V(p)\right) = \int_{p_E}^{p} \frac{h_t(p')}{u_t + s(1-\delta)\Big(L_t(p') - L_t(p_E)\Big)} dL_t(p').$$

To get (1.16), start from the first-order condition with respect to the optimal contract from (1.11) for active firms at some productivity level p

$$[W]: \quad \rho'(W)(S(p) - W) = h,$$

where I drop the time subscripts on ρ, S, but these functions depend implicitly on ω and L. The derivative of the retention rate is given by

$$\rho'(W) = s(1-\delta)\lambda \frac{dF(W)}{dW},$$

and, in a Rank-Monotonic Equilibrium, the derivative of the offer function can be expressed from (1.17) as

$$\lambda \frac{dF(W)}{dW}\frac{dW}{dp} = \frac{hl(p)}{u + s(1-\delta)\big(L(p) - L(p_E)\big)}.$$

Combining these three expressions yields the following first-order differential equation in W

$$\frac{dW}{dp} + \frac{s(1-\delta)l(p)}{u + s(1-\delta)\big(L(p) - L(p_E)\big)} W = \frac{s(1-\delta)l(p)}{u + s(1-\delta)\big(L(p) - L(p_E)\big)} S(p)$$

with boundary condition $W(p_E) = U$. Noting that

$$\frac{d \ln \Big(u + s(1-\delta)\big(L(p) - L(p_E)\big)\Big)}{dp} = \frac{s(1-\delta)l(p)}{u + s(1-\delta)\big(L(p) - L(p_E)\big)},$$

the corresponding integrating factor is then

$$\exp \int \frac{s(1-\delta)l(p)}{u + s(1-\delta)\big(L(p) - L(p_E)\big)} dp = u + s(1-\delta)\big(L(p) - L(p_E)\big).$$

Along with the boundary condition, this yields (1.16) in the main text

$$W(p) = \frac{uU + s(1-\delta)\displaystyle\int_{p_E}^{p} S(p')dL(p')}{u + s(1-\delta)\big(L(p) - L(p_E)\big)}.$$

1.A.7 Derivations Net Surplus

This Appendix shows that the model can be recast in a single value function by subtracting the unemployed worker's value function to the firm-worker surplus. I omit it from the main text not to clutter the description of the model. However, this more compact formulation is used in solving and simulating the model since the firm's policies can all be expressed as a function of the net surplus.

Net Surplus Equation The net firm-worker surplus is defined as $\phi_t(p) := \pi_t + \overline{V} - U_t := S_t(p) - U_t$. Adding and subtracting U_{t+1} in (1.10), the firm-worker surplus can be rewritten

$$\begin{aligned} S_t(p) = p_t\omega_t + \beta E_t\Bigg[& U_{t+1} + \mu Q_{t+1} \\ & + (1-\mu)\Bigg(\chi_{t+1}(p_{t+1}) \max_{\substack{h_{t+1}\geq 0 \\ W_{t+1}}} \Big\{ - c(h_{t+1}) + \rho_{t+1}(W_{t+1})\phi_{t+1}(p_{t+1}) \\ & + h_{t+1}(\phi_{t+1}(p_{t+1}) - (W_{t+1} - U_{t+1})) \\ & + (1-\delta)s\lambda_{t+1} \int_{W_{t+1}}^{\infty} \theta - U_{t+1} dF_{t+1}(\theta) \Big\} \Bigg) \Bigg]. \end{aligned}$$

Using the same strategy, the unemployed worker's value can similarly be rearranged as

$$U_t = b + \beta E_t \Bigg[U_{t+1} + \mu Q_{t+1} + (1-\mu)\lambda_{t+1} \int \max\left\{\theta - U_{t+1}, 0\right\} dF_{t+1}(\theta) \Bigg].$$

The net surplus can then be expressed as

$$\begin{aligned} \phi_t(p) &:= S_t(p) - U_t \\ &= p_t\omega_t - b \\ &+ \beta(1-\mu)E_t\Bigg[\chi_{t+1}(p_{t+1})\Bigg\{\tilde{\psi}_{t+1}(p) - \lambda_{t+1}\int_0^{\infty} \theta d\tilde{F}_{t+1}(\theta)\Bigg\}\Bigg] \end{aligned} \tag{1.24}$$

where $\tilde{F}_{t+1}$ defines the offer distribution for the firm's contract net of the value of unemployment, and $\tilde{\psi}_t(p)$ is the firm's optimization problem in net surplus form

$$\begin{aligned} \tilde{\psi}_t(p) := \max_{\substack{h\geq 0 \\ V}} \Bigg\{ & - c(h) + \rho_t(V)\phi_t(p) \\ & + h\Big(\phi_t(p) - V\Big) \\ & + (1-\delta)s\lambda_t \int_V^{\infty} \theta d\tilde{F}_t(\theta) \Bigg\} \end{aligned}$$

where, the firm now picks a contract V net of the value of unemployment.

Firm policies as a function of ϕ in a RME Since $\phi = S - U$ and U does not depend on p, ϕ is also increasing in p for every candidate equilibrium. In a Rank-Monotonic Equilibrium, the corresponding net contract follows by subtracting $U(\omega, L)$ in (1.16), which gives

$$V(p,\omega,L) - U(\omega,L) := \tilde{V}(p,\omega,L) = \frac{s(1-\delta)\displaystyle\int_{p_E}^{p} \phi(\hat{p},\omega,L)dL(\hat{p})}{u + s(1-\delta)\Big(L(p) - L(p_E)\Big)}, \tag{1.25}$$

The optimal hiring rate can also be expressed as solving

$$c'(h(p,\omega,L)) = \phi(p,\omega,L) - \tilde{V}(p,\omega,L),$$

and the entry/exit decision as $\chi(p,\omega,L) = \mathbb{1}\{\phi(p,\omega,L) \geq 0\}$.

Appendix 1.B Data Appendix

1.B.1 Firm Data: The Business Structure Database

Variables. I gather the definitions of the main analysis variables here. Note that a given variable is potentially drawn from multiple sources depending on whether the enterprise is selected to be part of a survey in the last year. [31]

- Employment: Sum of employees and working proprietors. This variable comes from different sources, but, for the majority of firms, employment is derived from income tax data – which is deduced directly from pay in the UK. For these firms, the employment figure corresponds to either the last or four last available quarters when the snapshot is taken, between March and April each year.

- sales: Income from the "sale of good or services to third parties". These figures are net of VAT, but include other taxes (alcohol, tobacco). For the majority of businesses, these sales figures are drawn from VAT returns for the past financial year, which ends in early April each year.

- Industry: The Standard Industry Classification is updated twice over the sample period, in 2003 and 2007. Since this classification is given, in most cases, in the contemporaneous vintage, I convert all industries to the 2007 classification by i) directly assigning their SIC07 industry

[31] I am grateful to Davide Melcangi and the research support team at ONS for clarifying the timing of some of these variables.

code for firms that survive until then, ii) creating crosswalks from the SIC92 to SIC03 and SIC03 to SIC07, based on the firms surviving across present both before and after each respective update comes into effect.

- Age: I follow Fort et al. (2013) in deriving a firm's age from establishment data. Each establishment has a birth year which corresponds to when they first appear on the registers. I define firm age as the age of the firm's oldest establishment when the firm first becomes active. It then ages naturally from this point onward, for each extra year in the sample. An advantage of this definition is that it avoids artificially classifying as "young" firms appearing in the data as a result of mergers, changes of ownership, etc.

- Labor Productivity: As discussed in the main text, labor productivity is defined as the logarithm of sales over employment.

Validation with national statistics aggregates. To assess the accuracy of the aggregates derived from the Business Structure Database, I compare some of these aggregates series with the closest official series from the Office for National Statistics. Figure 1.14 reports two such benchmarks: employment and sales. The corresponding ONS series are, respectively, workforce jobs and "domestic output at basic prices".[32] As shown in Figure 1.14, some sectors have trends at odds with the official series, especially for the sales variable. I proceed by excluding the following aggregate sectors: B (mining and quarrying), K (finance and insurance), M (professional and technical services), R (arts and entertainment). I also drop sectors O-Q (public administration, education, and health), the last two being mostly public in the UK.

Analysis sample. I finally apply a set of restrictions to construct a panel of firms over available survey years. I drop all firms that report sales or employment zero in any given year. I also drop firms which do not report hiring anyone over all survey year. Figure 1.15 shows aggregate employment and sales for the analysis sample and the official aggregates from the Office for National Statistics.

[32] "domestic output at basic prices" relates to sales in the BSD, as it corresponds to an industry's gross output (not net of intermediary consumptions).

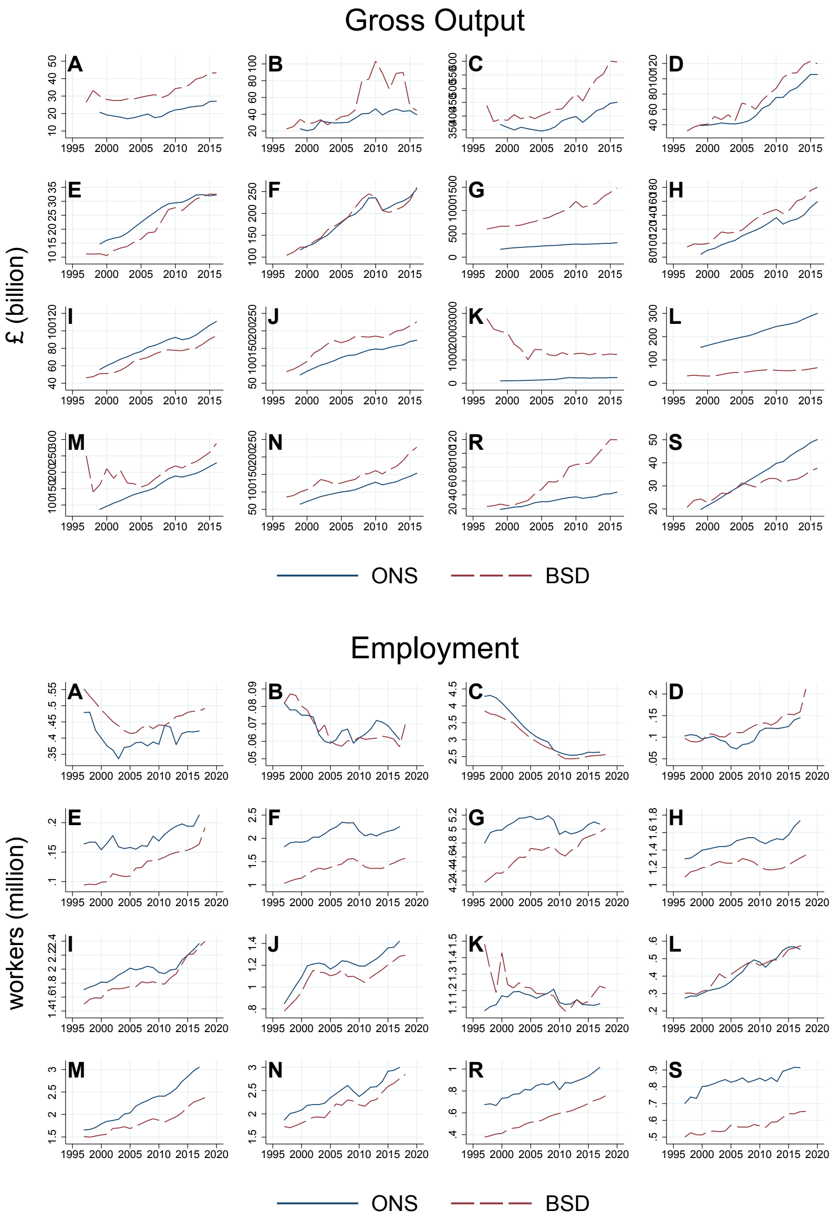


Figure 1.14: Benchmark with official statistics by broad industry. Comparison with official aggregate series by broad industry in the SIC07 classification (in bold, top left corner). The corresponding SIC07 industries are given in Table 1.6. See main text for details on the definition of these series.

SIC07 Section	Description
A	Agriculture, Forestry and Fishing
B	Mining and Quarrying
C	Manufacturing
D	Electricity, Gas, Steam and Air Conditioning Supply
E	Water Supply; Sewerage, Waste Management
F	Construction
G	Wholesale and Retail Trade; Repair of Motor Vehicles
H	Transportation and Storage
I	Accomodation and Food Service Activities
J	Information and Communication
K	Financial and Insurance Activities
L	Real Estate Activities
M	Professional, Scientific and Technical Activities
N	Administrative and Support Service Activities
O	Public Administration and Defence; Compulsory Social Security
P	Education
Q	Human Health and Social Work Activities
R	Arts, Entertainment and Recreation
S	Other Service Activities

Table 1.6: Description of SIC07 broad industries.

Figure 1.15: Aggregates from analysis sample.

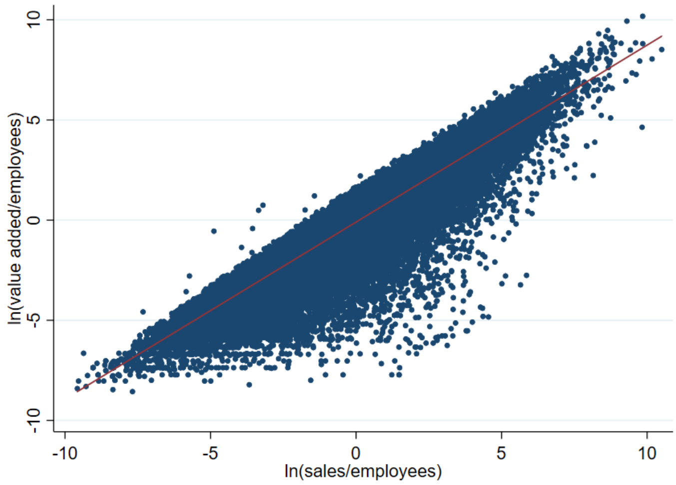


Figure 1.16: Correlation between turnover and value added labor productivity measures. These measures are defined, respectively, as the log of turnover per employee and turnover minus cost of sales per employee, deviated from industry means.

1.B.2 Labor Productivity Measure

As discussed in the main text, the Business Structure Database only makes turnover available for each firm. Figure 1.16 benchmarks the turnover-based labor productivity measure against a value-added based labor productivity measure. The data is from Fame, a commercial database of company information, including detailed balance sheet data, for the UK and Ireland. The figure shows that, within industries, these two labor productivity measures are strongly associated.

1.B.3 Labor Market Transitions

The labor market transition rates are taken from Postel-Vinay and Sepahsalari (2019). They are derived from the British Household Panel Survey (BHPS) and its successor Understanding Society (UKHLS). Note that because of the transition from the BHPS to UKHLS, there is a gap in the series between August 2008 and December 2009, which is smoothed over using moving averages.[33]

1.B.4 Additional Macro Series

Several additional series are taken directly from the Office for National Statistics website:

[33] I am grateful to the authors for sharing these series, and to Pete Spittal for explaining how the transition between the two surveys affects them.

- Unemployment rate (aged 16 and over, seasonally adjusted): MGSX
- UK vacancies - total: AP2Y

Appendix 1.C Numerical Solution

1.C.1 Stationary solution

As shown in Appendix 1.A.7, the firm's policies can be expressed in terms of a single value function, the net surplus given in Equation (1.24). A Stationary Rank-Monotonic Equilibrium (see Definition 3) can similarly be defined as a fixed-point in the net surplus, ϕ and the measure of workers, L. The algorithm below is given in terms of net firm-worker surplus for concision.

Discretization. In a Rank-Monotonic Equilibrium, all heterogeneity in the model arises through p. I discretize idiosyncratic productivity using Tauchen's procedure with $N_p = 400$ points. This yields a $\{p_1, \ldots, p_{N_p}\}$ grid and the associated transition matrix for p.

This discretization can seen as the relevant policy or value function being constant on some (small) half-open interval. This provides an intuitive way to integrate against the measure of workers, Lt, by replacing the integral by the appropriate employment share weighted sum. For instance, the net optimal contract (1.25) at some productivity node p_k can be approximated as

$$
\begin{aligned}
\tilde{V}(p_k) &= \frac{s(1-\delta)\displaystyle\int_{p_1}^{p_k} \chi(p')\phi(p')dL(p')}{u + s(1-\delta)\Big(L(p_k) - L(p_E)\Big)} \\
&= \frac{s(1-\delta)\displaystyle\sum_{i=2}^{k}\int_{p_{i-1}}^{p_i} \chi(p')\phi(p')dL(p')}{u + s(1-\delta)\Big(L(p_k) - L(p_E)\Big)} \\
&\approx \frac{s(1-\delta)\displaystyle\sum_{i=2}^{k}\chi(p_{i-1})\phi(p_{i-1})\int_{p_{i-1}}^{p_i} dL(p')}{u + s(1-\delta)\Big(L(p_k) - L(p_E)\Big)},
\end{aligned}
$$

where the last integral in the approximation is simply the fraction of workers employed at firms in the interval between p_{i-1} and p_i.

Algorithm stationary equilibrium. Given this discretization, I iterate on the following steps:

1. Guess initial values for ϕ and L on the grid for idiosyncratic productivity. In line with the RME result, I start with some increasing function for the net surplus. In practice, I set $L = 0$ (all workers initially unemployed) as a first step.

2. Conditional on values for ϕ and L, the agents' optimal policies can be computed. For example, the activity threshold, p_E, is the point at which ϕ becomes positive. The optimal contract can be computed from (1.25).

3. The net surplus equation and the law of motion for employment shares imply new values for ϕ and L on the grid. Note that the net surplus equation gives an update for ϕ in the previous period, while that for the employment mass yields next period's employment for each productivity level. But this does not matter since the algorithm solves for a stationary equilibrium.

4. The final step consists in computing the Euclidean norm to check the convergence of L and ϕ. If this is the case, the pair (ϕ, L) represents a stationary equilibrium. Otherwise, go back to point 2 with the updated values until convergence.

1.C.2 Estimation

The parameters are calibrated by targeting the moments listed in Table 1.2. In practice, I minimize the distance between the model generated moments and their empirical counterpart using the following objective function

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

where θ denotes the parameter vector, $\mathbf{m}_{\text{data}}$ 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: $\mathbf{\Lambda} = \text{diag}(1/\mathbf{m}^2_{\text{data}})$. Figure 1.17 further shows slices of the objective function around the estimated parameter values.

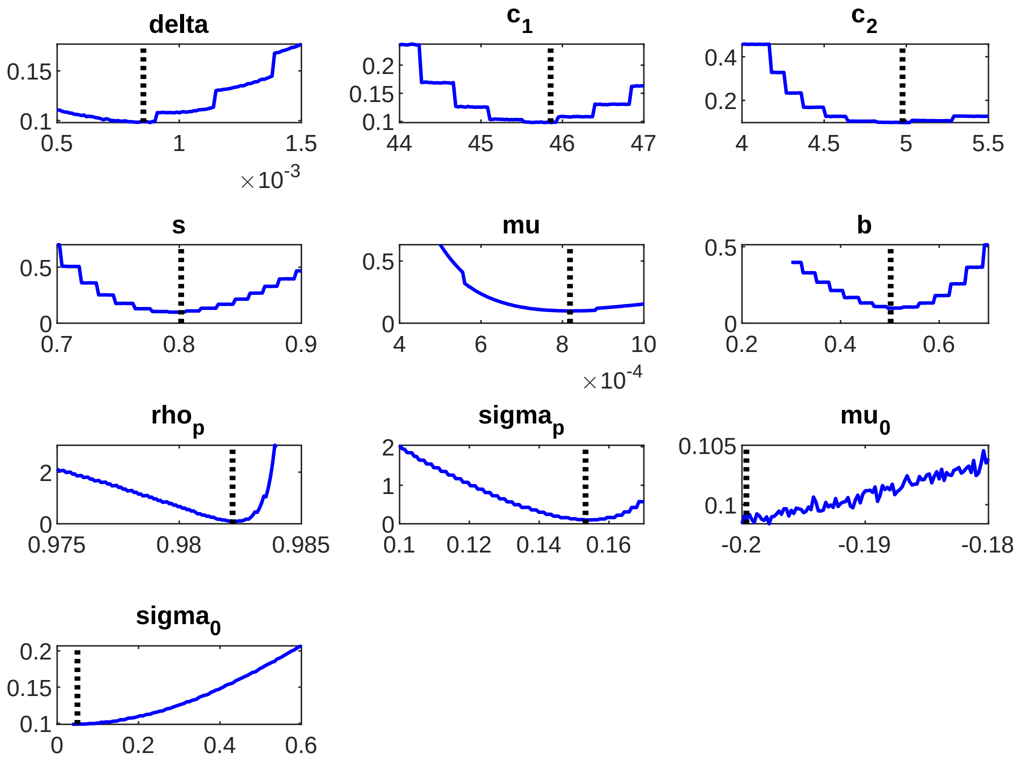


Figure 1.17: Slices of objective function for each parameter. Vertical dotted line denotes estimated parameter value.

1.C.3 Aggregate shocks solution

As explained in the main text, the simulation algorithm in the presence of aggregate shocks relies on two approximations. First, the measure of employment at firms of different productivity is summarized by a set of (un-centered) moments and the unemployment rate:

$$\begin{aligned} u_t &= 1 - \int_{\underline{p}}^{\overline{p}} dL_t(p) \\ m_t^1 &= \int_{\underline{p}}^{\overline{p}} \ln p d\bar{L}_t(p) \\ m_t^2 &= \int_{\underline{p}}^{\overline{p}} (\ln p)^2 d\bar{L}_t(p) \\ &\dots \end{aligned} \tag{1.26}$$

where $\bar{L}_t$ denotes the cumulative density associated with the cumulative measure of workers on p, $\bar{L}_t(p) = \frac{L_t(p)}{\int dL_t(p)}$. I report some robustness checks on the number of moments included in the approximation in Appendix 1.C.4.

Second I parameterize the value functions for the firm-worker surplus, S_t, and the unemployed worker, U_t, with a polynomial. I choose to parameterize these value functions separately instead of the net surplus since they are positive by definition, so they can expressed in log-deviation from steady-state.

Because preserving the monotonicity of S_t (especially around the entry threshold) is central to the procedure, I use a separate polynomial for each productivity node p_i. The value functions are approximated outside of steady-state as

$$\ln S(p_i, \omega_t, L_t) - \ln \overline{S}(p_i) \approx \tilde{S}(p, \omega_t, \tilde{\boldsymbol{m}}_t; \theta_{p_i}) \quad p_i \in \{p_1, ..., p_{N_p}\}$$

and

$$\ln U(\omega_t, L_t) - \ln \overline{U} \approx \tilde{U}(\omega_t, \tilde{\boldsymbol{m}}_t; \theta_U)$$

where $\tilde{\boldsymbol{m}}_t$ denotes the vector of moments in (1.26) in log-deviation from steady-state, while $\overline{S}$ and $\overline{U}$ are the firm and worker surplus and value of unemployment at the steady-state.

The algorithm then solves for the coefficients by iterating on the four following steps:

1. Draw a sequence of aggregate productivity shocks and guess an initial value for the coefficients of $\tilde{S}$ and $\tilde{U}$. I initialize them at zero in practice.

2. Simulate the measure of employment forward, starting from the stationary solution. Conditional on the current value of $\boldsymbol{\theta}$, agents make optimal hiring and contract offer decisions given the current states, which induces a law of motion for employment at each productivity level. The simulated measure of workers is approximated by a set of moments as described above.

3. Update $\tilde{S}$ and $\tilde{U}$, conditional on the simulation of L_t obtained in the previous step. This requires to take an expectation over future realizations of the aggregate shock. The aggregate shocks is discretized using Tauchen procedure with $N_\omega = 19$ nodes in practice.

4. Run a regression of $\tilde{S}$ and $\tilde{U}$ on the state variables to update the coefficients. Go back to step 2 and iterate until convergence.

I find the coefficients by running separate regressions for the firm-worker surplus at each p-node on the variables in the state-space. I omit the constant, thus imposing that the steady-state holds exactly at each node. Since these regressors are sometimes close to collinear, I make the procedure more robust by using ridge regression to regularize the problem. For instance, the coefficients for the unemployed worker's value function are found by solving

$$\min_{\theta_U} \sum_t (\ln U_t - \ln \overline{U} - \tilde{U}(\omega_t, \tilde{\boldsymbol{m}}_t; \theta_U))^2 + \zeta \sum_i \theta_{U_i}^2$$

where θ_{U_i} denotes individual elements of θ_U, $\zeta > 0$ is the associated regularization parameter, and I assume

$$\tilde{U}(\omega_t, \tilde{\boldsymbol{m}}_t; \theta_U) = \\ (\ln \omega_t - 0)\, \theta_U^{\omega} + \left(\ln u_t^1 - \ln \bar{u}^1\right) \theta_U^{u} + \sum_{k=1}^{N_m} \left(\ln m_t^k - \ln \bar{m}^k\right) \theta_U^{m_k},$$

with N_m the number of moments of $\bar{L}_t$ included in the approximation.

The regularization parameter, $\zeta > 0$, ensures that the matrix of regressors is invertible by adding to it a ζ-diagonal matrix. I finally allow for less than full updating by appropriately dampening the obtained coefficients. I proceed similarly for each polynomial of the firm-worker surplus. Note that with these parametric assumptions, the coefficients $\{\theta_U, \theta_{p_1}, \ldots, \theta_{p_{N_p}}\}$ are elasticities, which gives some intuition about the appropriate convergence condition.

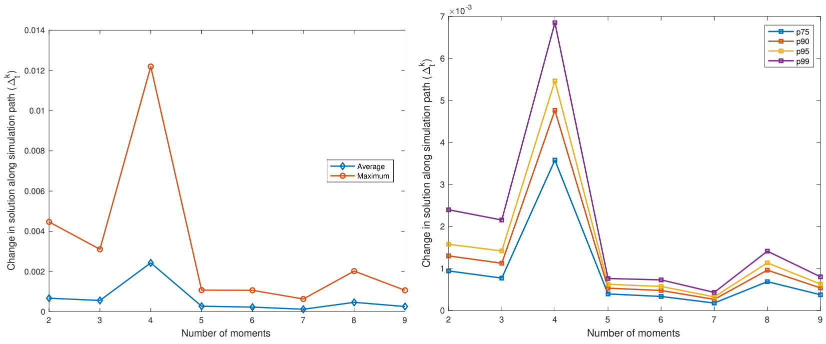
1.C.4 Importance of higher order moments

To assess the sensitivity of this solution method to the number of moments used in approximating L_t, I perform the following test. I incrementally introduce up to $N_m = 9$ moments to summarize L_t, and solve the model using the same sequence of aggregate shocks each time. I can then compute a solution for $\tilde{S}^k(p, \omega_t, \tilde{\boldsymbol{m}}_t; \theta_{p_i})$ and $\tilde{U}^k(\omega_t, \tilde{\boldsymbol{m}}_t; \theta_U)$ along the same sequence of aggregate shocks, where $k = 1, \ldots, N_m$ indexes the number of moments included in the approximation.

I proceed by defining the following measure of sensitivity of the global solution to the inclusion of extra moments

$$\Delta_t^k(p) := \left|\tilde{S}_t^k(p) - \tilde{S}_t^{k-1}(p)\right| = \left|\ln S_t^k(p) - \ln S_t^{k-1}(p)\right|$$

and similarly for $\tilde{U}_t^k$. Figure 1.18a reports the average and maximum $\Delta_t^k(p)$ along the simulated sequence of shocks as more moments are included. This



(a) Average and Maximum **(b)** Percentiles

Figure 1.18: Robustness to number of moments included in approximation.

test suggests that at least up to the 4th moment should be included as there is a pronounced change in sensitivity at this point, as shown by the large spike in the picture. To check that this is not purely driven by outliers, Figure 1.18b confirms this pattern by showing several percentiles of $\Delta_t^k(p)$.

1.C.5 Accuracy tests

The accuracy of the procedure is assessed through the tests proposed in den Haan (2010), adapted to the current setting. I compute the firm-worker surplus, $S_t(p)$ and unemployment value, U_t in two different ways. Given a sequence of aggregate shocks $\{\omega_s\}_{s=1}^T$, $S_t(p)$ and U_t can be obtained either using their respective approximation based on θ_p and θ_u, or computed directly solving the model backward in time and explicitly taking an expectation over ω_{t+1} in each period.

Table 1.7 reports these statistics for an alternative sequence of shocks, different to the one used to solve for the coefficients. I report the average and maximum absolute percent error between the approximation and explicit solutions, i.e. $100(y_t^{\text{approx.}} - y_t^{\text{explicit}})$, taken at each point in time and each node, where $y_t^{\text{approx.}}$ denotes $\tilde{S}(p, \omega_t, \hat{\boldsymbol{m}}_t; \theta_p)$ or $\tilde{U}(\omega_t, \hat{\boldsymbol{m}}_t; \theta_U)$ as appropriate.

1.C.6 Non-linearities in shock size

Figure 1.19 shows the response in unemployment to several one-time negative productivity shocks of different magnitudes. It illustrates that there are substantial non-linearities in the response of unemployment to aggregate shocks. These non-linearities justify the need for a full solution of the model with aggregate shocks, and not merely a transition experiment, since

Variable	Absolute Error (in %) Mean	Max
Value Functions		
S_t	0.060	0.284
U_t	0.027	0.128
Moments L_t $(\boldsymbol{m}_t)$		
u_t	1.177	4.274
m_t^1	0.048	0.358
m_t^2	0.034	0.238

Table 1.7: Accuracy tests

Figure 1.19: Unemployment increase in response to negative shocks of different sizes. Series are normalized by the response to a one standard deviation negative ω_t-shock.

the uncertainty around aggregate shocks matters in determining macroeconomic outcomes.

Chapter 2

Self-employment and Unemployment Risk

2.1 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.

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 be-

[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 (2018a). In terms of private unemployment insurance, anecdotal evidence suggests that when a private UI market does exist, the self-employed cannot buy these policies.

tween involuntary 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 involuntary 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 fifty 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. (2019a) 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, n.d.); 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 2.3 describes the model. Section 2.4 details the procedure to discretize worker heterogeneity. Section 2.5 shows the model fit, and Section 2.6 analyzes the effects of extending UI benefits to the self-employed.

2.2 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

[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 2.1: Summary statistics by labor form for the selected sample.

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.

2.2.1 The self-employed in the SIPP data

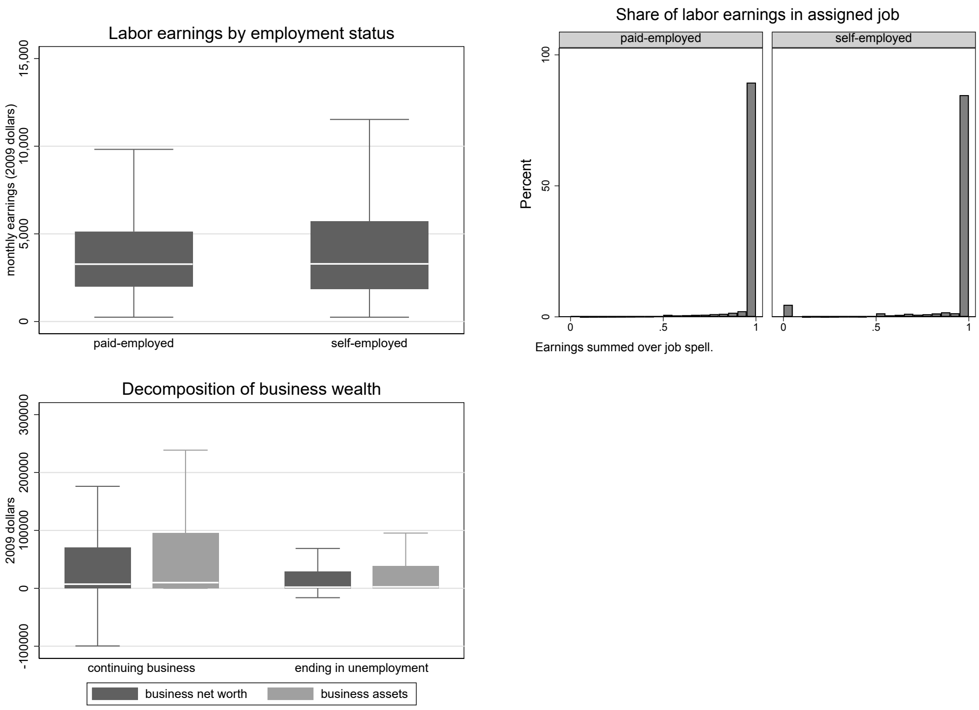
The sample is restricted to the individuals with the largest labor earnings in each household over the duration of the survey. These workers are assigned to one of three labor market states depending on their primary source of labor income over the duration of each job or business: paid-employed (P), self-employed (S), or unemployed (U).[3]

Table 2.1 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 2.1 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).[4] Figure 2.1 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.

[3] Appendix 2.A provides a complete description of the SIPP and the procedure to ascribe workers to each labor market state.

[4] Appendix Table 2.11 further shows this breakdown for income growth.

Figure 2.1: 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 2.2: 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

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.[5] The distribution of business wealth, displayed in the bottom left panel of Figure 2.1, 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.

Table 2.2 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.

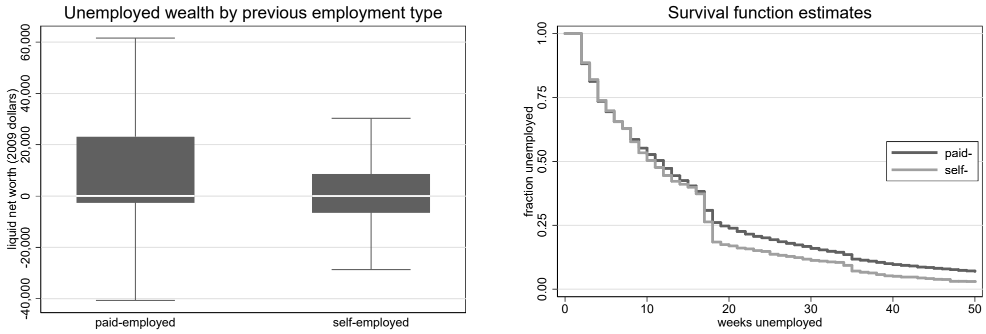
2.2.2 Unemployment risk

Table 2.2 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.[6]

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

[6] Appendix Table 2.11 also suggests that the capacity of these households to access basic consumer credit markets is not overly restricted by the main earner's labor form status. The distribution of "Insecured Debt" shows that the self-employed are able to accumulate at least as much unsecured debt.

Figure 2.2: 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.

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 2.2, 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 2.2, right panel, plots the survival curve in unemployment for each type of worker.[7] Workers previously in self-employment appear to exit unemployment slightly sooner.[8] I confirm this finding in a series of proportional hazard models. Table 2.3 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 this pattern of earlier exit is robust to controlling for standard demographics, as well as previous industry and occupation effects.

To sum up, the empirical evidence outlined here points to the self-employed—defined as workers deriving their primary income from a business—

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

[8] A Wilcoxon test of equality gives $p = .13$, implying no rejection 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 2.3: 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.

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.

2.3 A search model with self-employment and savings

Based on the empirical evidence outlined in Section 2.2, 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.[9]

2.3.1 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.[10]

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

[9]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).

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

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.

2.3.2 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 separated, 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'.

2.3.3 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

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

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

$$R_k^S(a,y) = \max\Big\{ V^U(a), \\ \delta_S V^U(a) + (1-\delta_S)\Big[V_k^S(a,y) + \lambda_{SP}\mu_k^{SP}(a,y) + \lambda_{SS}\mu_k^{SS}(a,y) \Big]\Big\}, \quad (2.1)$$

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

$$R_k^P(a,w) = \max\Big\{ V^U(a), \\ \delta_P V^B(a,w) + (1-\delta_P)\Big[V_k^P(a,w) + \lambda_{PP}\mu_k^{PP}(a,w) + \lambda_{PS}\mu_k^{PS}(a,w) \Big]\Big\}. \quad (2.2)$$

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

$$\begin{aligned} V_k^s(a,y) &= \max_{c,\tilde{a}} \Big\{ u(c) + \beta R_k^s(\tilde{a},y) \Big\} \\ \text{s.t} \quad c + \frac{\tilde{a}}{1+r} &= y + a + y_k^{HH} \\ \tilde{a} &\geq \underline{a}. \end{aligned} \quad (2.3)$$

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.[11] When hit by a δ_P-shock, workers get UI payments $UI(w)$. Their present value at the consumption stage is then given by

$$\begin{aligned} V_k^B(a,w) &= \max_{c,\tilde{a}} \left\{ u(c) + \beta R_k^U(\tilde{a}) \right\} \\ \text{s.t} \quad c + \frac{\tilde{a}}{1+r} &= UI^P(w) + a + y_k^{HH} \\ \tilde{a} &\geq \underline{a}. \end{aligned} \tag{2.4}$$

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

$$R_k^U(a) = \lambda_{UP}\mu_k^{UP}(a) + \lambda_{US}\mu_k^{US}(a), \tag{2.5}$$

where when drawing a potential labor income, workers' outside option is $V_k^U(a)$, the counterpart to (2.4) in the absence of unemployment insurance payments.

2.3.4 Stationary Equilibrium

Taken together, optimal savings decisions and the reservation incomes across the different states imply a stationary distribution of workers over labor force states (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 (2.1)-(2.5);

[11] 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.

2. The asset and consumption choices, $\hat{a}_k^s$ and $\hat{c}_k^s$, solve equations (2.3) and (2.4). The reservation wage functions $\rho_k^{ss'}$ are defined by the solutions to $\mu_k^{ss'}$ in equations (2.1), (2.2), and (2.5);

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 $\Gamma^*_k = Q_k \circ \Gamma^*_k$.

2.4 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.

2.4.1 Using k-means to classify workers

Following Manresa et al. (2017) and Bonhomme et al. (2019a), 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

$$\underset{\tilde{h},k_1,\ldots,k_N}{\arg\min} \sum_{i=1}^{N} \left\| \hat{h}_i - \tilde{h}(k_i) \right\|^2, \tag{2.6}$$

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 (2.6) 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.[12]

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

The logic behind this classification step is closest to that described in Bonhomme et al. (2019a). 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. (2019a); they similarly capture heterogeneity in labor earning potentials that can be correlated with both observed and unobserved characteristics.

2.4.2 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.[13] The vector of features $\hat{h}_i$ includes first a measure of 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,

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

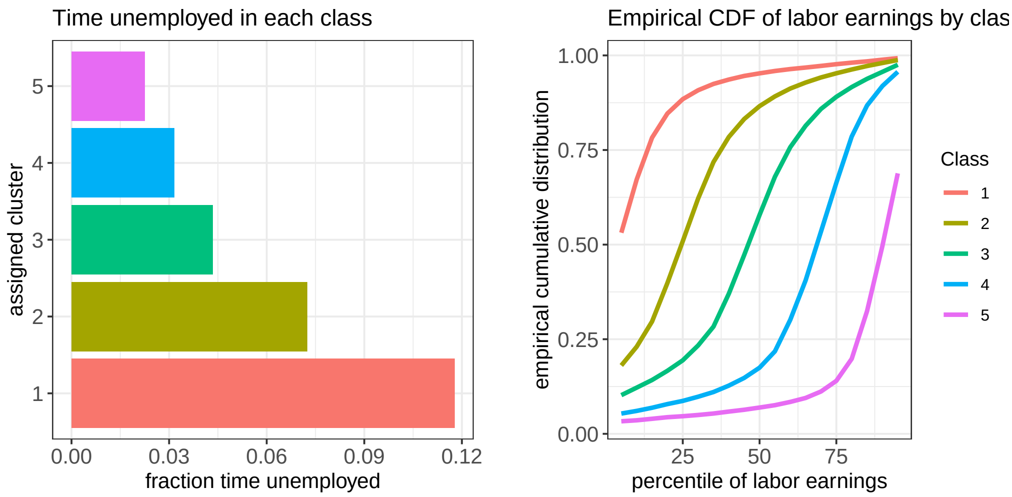
with y_{it} denoting labor income for i in period t. I compute this empirical CDF for twenty quantiles of the empirical distribution of labor earnings in the sample.

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

algorithm in the base R function "kmeans".

[13]Figure 2.7 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 2.3: 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.

for labor market attachment (left) and the empirical CDF of labor market earnings (right) are shown in Figure 2.3. It suggests that a clear partition exists, where low earners also have a weaker labor market attachment.

2.5 Calibration

2.5.1 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.

Table 2.4: Externally calibrated parameters

Parameter	Description	Value	Target/Source
σ	CRRA utility parameter	2	Standard
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

2.5.2 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.[14]

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.

2.5.3 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 2.4, which also reports the corresponding targets.

2.5.4 Internally calibrated parameters

Table 2.5 displays the seventeen remaining parameters that are calibrated internally by minimizing the distance between a set of model simulated

[14]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 2.5: 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	

moments and their empirical counterpart.[15] 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).

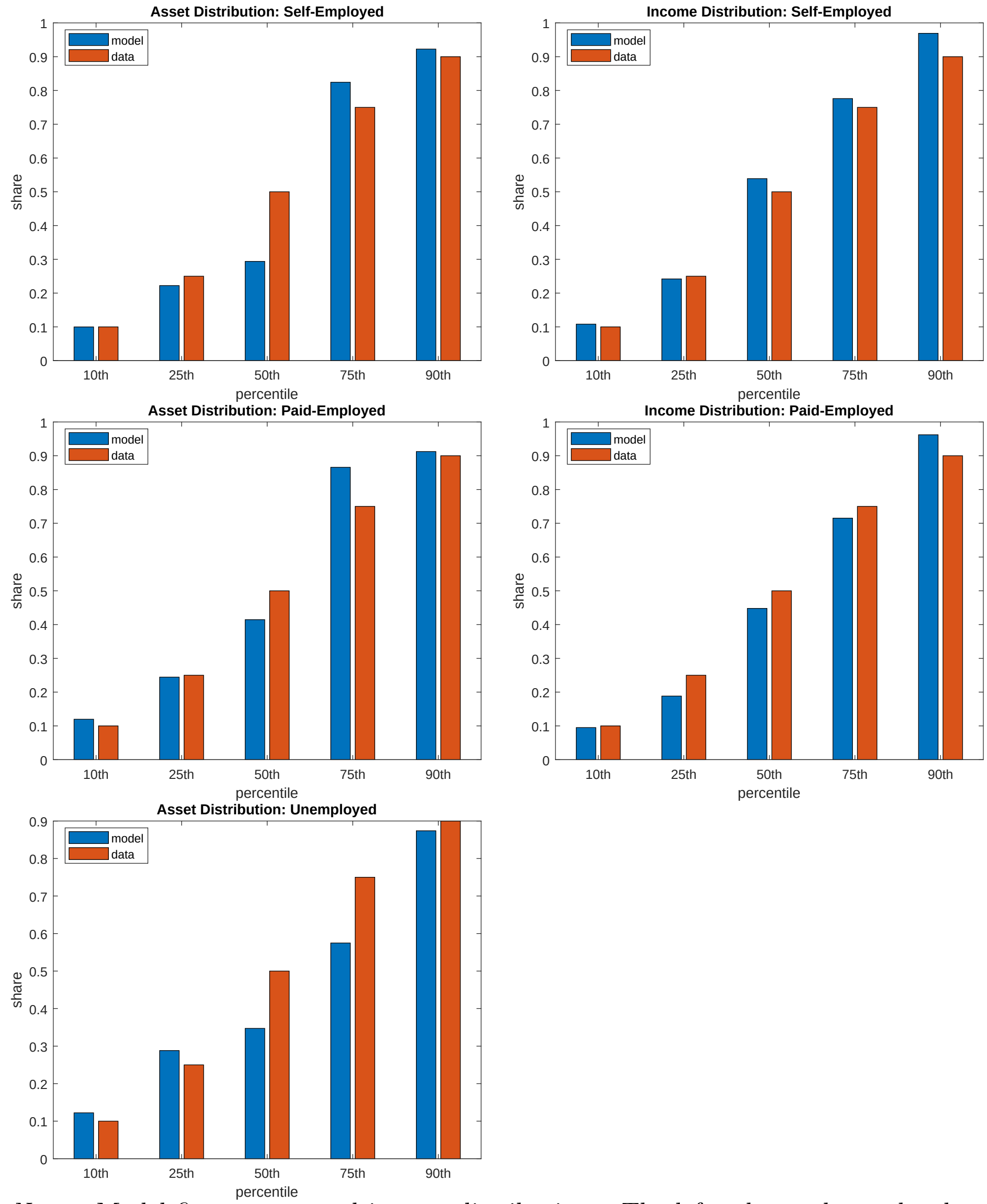
Table 2.6 and Figure 2.4 report the model fit to the targeted moments. Table 2.6 displays the fit to workers' transition rates across the three labor forms. Figure 2.4 shows the net liquid asset (left column) and labor income (right column) distributions in the model and the data. Overall, the model fit is good. It can replicate the transition patterns observed in the SIPP well. Regarding the asset and income distributions, the model does not fully replicate the upper tail of the distribution of wealth and income for individuals in employment (paid- or self-employed). The opposite is true for the net liquid wealth distribution of workers in unemployment, with slightly more workers near the top.

[15]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: $\mathbf{\Lambda} = \text{diag}(1/\mathbf{m}^2_{\text{data}})$.

Figure 2.4: 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.

Table 2.6: 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

2.6 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.

2.6.1 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. This system is made of a replacement rate, b, such that workers get

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

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

$$\tau \int y d\Gamma^S_{\tau,b}(a,y) = (1-\tau)\int UI^S(y) d\Gamma^C_{\tau,b}(a,y) \tag{2.7}$$

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

Worker class (\$$y_k^{HH}$)	E(contributions)	E(benefits)	Ratio ben. to cont.
1208	3.7	6.2	1.69
1628	8.1	8.5	1.05
2068	5.9	5.3	0.89
2588	21.4	16.1	0.76
2879	13.4	10.9	0.81

Table 2.7: 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.

activity. Going back to the value of self-employment at the beginning of the search stage (2.2), this expression now reads

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

where V_k^C 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).

2.6.2 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 (2.7).

Table 2.7 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 $\Delta_{b,\tau}^{\text{comp}}$

$\$y_k^{HH}$	**All workers**		$s = S$		$s = C$	
	$\%\ \Delta_{b,\tau}^{\text{comp}}$	$\Delta_{b,\tau}^{\text{transfer}}$	$\%\ \Delta_{b,\tau}^{\text{comp}}$	$\Delta_{b,\tau}^{\text{transfer}}$	$\%\ \Delta_{b,\tau}^{\text{comp}}$	$\Delta_{b,\tau}^{\text{transfer}}$
1208	0.04	720	0.2	1350	0.4	2626
1628	0.13	1520	0.4	3972	0.9	6790
2068	0.07	1383	0.3	4889	0.8	9063
2588	0.01	695	0.2	4471	0.8	10589
2879	0.10	2517	0.5	10603	0.8	15426
all	0.07	1379	0.3	5138	0.8	9099

Table 2.8: Insurance value of the self-employment UI scheme by worker type. See main text for the definitions of $\Delta_{b,\tau}^{\text{comp}}$ and $\Delta_{b,\tau}^{\text{transfer}}$. Worker class is indicated by y_k^{HH}, the median household income when the worker is in unemployment. "All workers" columns give averages over all states in the model; $s = S$ is for workers in self-employment; $s = C$ is for formerly self-employed on benefits.

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 2.8 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).

2.6.3 Sensitivity to stochastic discount factor

The baseline interest rate is exogenously set to 4 percent a year in the calibration. The discount factor, β, is internally calibrated to try to match, among other moments, the distribution of net liquid wealth found in the data. In my baseline analysis, I obtain a large discount factor, which translates into a yearly subjective discount rate of $\rho = 0.012$. As a result, agents are markedly more patient than the market in my analysis, and inclined to accumulate precautionary savings.

	$\rho = 0.012$ (baseline)			$\rho = .025$			$\rho = .04$		
Earn. Cluster	All	$s = S$	$s = C$	All	$s = S$	$s = C$	All	$s = S$	$s = C$
861	288	288	425	174	175	396	146	146	419
1772	708	708	917	454	457	841	385	388	880
3048	1702	1708	1928	1051	1059	1634	921	932	1663
5791	2752	2771	3280	1733	1749	2677	1435	1452	2567
All	1399	1405	1669	875	882	1418	744	753	1419

Table 2.9: Discount factor experiments. The numbers are the cash grant (in \$2009) required to make the agents indifferent between being in the baseline economy instead of the economy with benefits for the self-employed. Averages of these cash grants are reported for the worker types in rows and the labor market states in columns.

In Table 2.9, I report the value of extending UI benefits to the self-employed rates under several assumptions on the subjective discount rate: the baseline calibration (annual discount rate $\rho = 0.012$), an intermediary scenario (annual discount rate $\rho = .025$), and a scenario in which it is the same as the market rate (annual discount rate $r = \rho = .04$). To measure how different groups of agents value this public insurance policy, I again compute the cash transfer, $\Delta_{b,\tau}^{\text{transfer}}$, required to make the agents indifferent between being in the baseline economy instead of the economy with benefits for the self-employed. The table displays averages conditional on a worker's unobserved earnings type (rows) and their current labor market state s (columns).

This exercise shows that the value of implementing the policy drops sharply as the discount rate increases.[16] Moving from the economy in the baseline scenario of $\rho = .012$ to $\rho = .04$ implies a compensating cash grant roughly cut in half. There is, however, a tension entailed by a higher ρ for the group of workers currently on benefits and previously self-employed ($s = C$). The compensating cash grant plateaus, or even increases depending on the earnings cluster, as ρ increases. It reflects the lower level of savings with a higher ρ, which bites particularly hard for households at the bottom of the earnings distribution who have to terminate their self-employment activity.

2.6.4 Optimal policies

It is also possible to derive the socially desirable policy (b^*, τ^*) given some social welfare function.[17] In the model's notations, the utilitarian welfare

[16] The numbers given in the baseline scenario differs from those reported in Table 2.8, as they are obtained matching the data for the subsample of low-educated households.

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

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 2.10: Optimal policies by worker type. The replacement rate (b^*) and social contribution rate (τ^*) are the solution to (2.8) 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 (2.8) for the whole economy. "equal treatment" is the case where b is aligned with the paid-employed.

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

$$\max_{b,\tau} \Omega(b,\tau) \quad \text{s.t.} \quad \tau \int y d\Gamma_{\tau,b}^S(a,y) = (1-\tau)\int UI^S(y)d\Gamma_{\tau,b}^C(a,y). \tag{2.8}$$

Table 2.10 displays optimal policies, defined as the solution to (2.8), 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 values of b.

2.7 Conclusion

This paper investigates the labor market risks faced by workers in self-employment. I use monthly panel data from the SIPP to show that, much like regular wage workers, the self-employed go through unemployment spells. I also show that, prior to the start of these spells, a large portion of these workers do not have substantial savings to self-insure, suggesting that there are potential welfare gains in providing them with some form of public insurance.

To evaluate the welfare implications of such policies, I proceed by setting up a search model in which risk-averse workers move across three labor market states (paid-employment, self-employment, and unemployment) and have the possibility to borrow and save to self-insure. I use the model to analyze the welfare effects of extending a mandatory public insurance scheme to this group. My preliminary results suggest modest welfare gains from such a policy. They also point to the importance of properly accounting for the vast dispersion in income amongst the self-employed in assessing the welfare impacts of such policies.

Appendix 2.A Data

The main data source is the Survey of Income and Program Participation (SIPP), a survey centered on how Americans use welfare programs.[18] This Appendix describes the sample selection and expands on the definition of paid-employment and self-employment.

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.[19] 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-

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

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

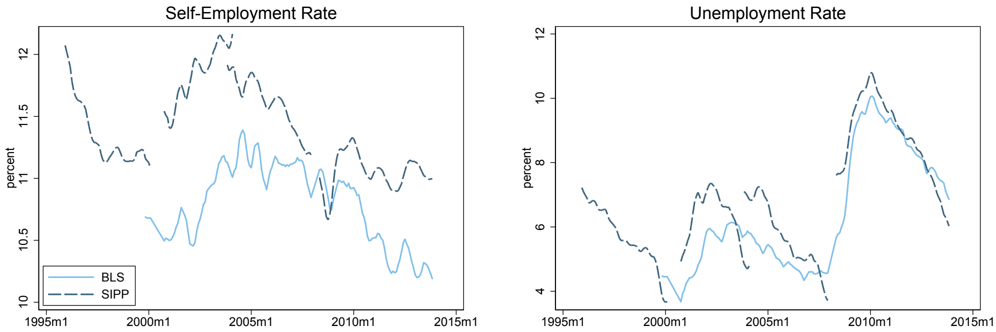


Figure 2.5: Benchmark of aggregates derived from the SIPP with the corresponding Bureau of Labor Statistics series. The unemployment and self-employment rates in the SIPP are based on the definitions of labor force status given in the main text. Observations are weighted using the provided cross-sectional weights.

, 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.[20] 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.

Appendix 2.B Additional tables and figures

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

			Percentile				
	Labor market status	Avg.	p10	p25	p50	p75	p90
Labor earnings (monthly, $2009)							
main earner	never SE	3,944	1,341	2,093	3,296	5,033	7,328
	some SE	4,284	857	1,634	3,065	5,539	9,406
other earners	never SE	1,299	0	0	0	2,180	3,861
	some SE	1,239	0	0	0	1,892	3,780
Wealth ($2009)							
Net liquid wealth	never SE	64,409	-12,572	-885	3,296	49,866	185,555
	some SE	84,826	-16,248	-1,322	2,989	66,972	255,998
Illiquid wealth	never SE	92,219	0	3,836	35,549	117,781	248,603
	some SE	242,218	0	14,208	97,860	298,143	661,911
Unsecured debt	never SE	9,434	0	0	1,472	8,887	23,983
	some SE	12,930	0	0	1,404	9,834	27,660
Earnings growth							
6-month	never SE	0.096	-0.237	-0.019	-0.008	0.071	0.390
	some SE	0.364	-0.523	-0.194	-0.008	0.256	1.155
12-month	never SE	0.132	-0.268	-0.035	-0.012	0.126	0.498
	some SE	0.465	-0.594	-0.248	-0.014	0.391	1.483

Table 2.11: Wealth and earnings: workers with at least some self-employment vs workers never self-employed.

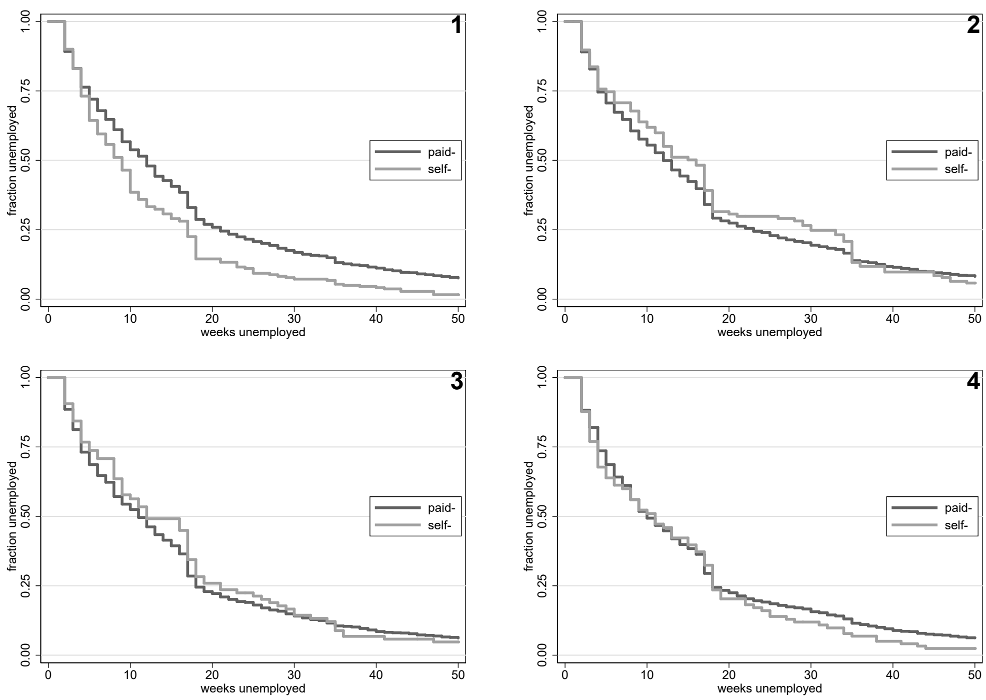


Figure 2.6: Survival function by net liquid wealth quartile. Kaplan-Meier estimates by net liquid wealth quartile and previous employment status (paid- and self-). The digit in the top-right corner corresponds to the individual's household's net liquid wealth at or before the start of the unemployment 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 2.12: 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.

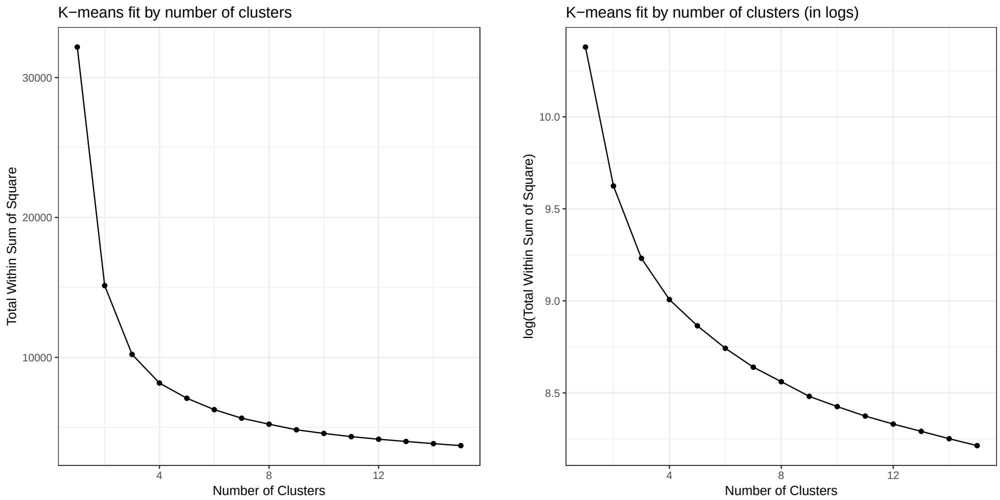


Figure 2.7: Change in kmeans fit with number of clusters. The left panel shows a measure of fit for k-means, the total within sum of squares, as the number of clusters increases. The right panel shows the same figure with the total sum of squares in logs.

Chapter 3

Sorting in Knowledge Production: Evidence from Patent Data

3.1 Introduction

A large literature studies the determinants of knowledge production. This literature has emphasized, for instance, the role of the accumulation of team-specific human capital (Jaravel et al., 2018), the geographic concentration of inventors and firms in "tech clusters" (Moretti, 2019), high-quality interactions (Akcigit et al., 2018), or social dynamics in a field dominated by a star scientist (Azoulay et al., 2019), in explaining the output of researchers and inventors.

A common challenge in these empirical studies is to hold constant the sorting of inventors in teams, research institutions, or regions based on unobserved characteristics related to their ability. This is typically done by centering on exogenous events reshuffling the inputs to knowledge production, such as the death of a star scientist, or by controlling for a large number of fixed-effects. Sorting, however, can also be seen as a a potential determinant of knowledge production to be measured.

In this paper, I propose to quantify inventor sorting in patent data. I focus on one dimension of this process: the association between the ability of inventors and the quality of the institutions—private corporations, universities, public research institutes which I term "firms" for short in what follows—at which they work. I translate econometric methods developed to measure sorting in matched employer-employee data sets to patent data, using the panel dimension to recover the contribution of inventors, firms, and inventor-firm sorting to patent quality. My main finding is that inventor-firm sorting is positive and makes up for on average five percent of the total variance in inventor output found in patent data.

I use data on the universe of patents granted by the United States Patent Trading Office between 1975 and 2009 consolidated as part of the PatentsView project (USPTO, 2019). These data allow to track inventors, their firms, and their locations over time. I construct measures of inventor output based on the quality of their patents, which are defined as the number of citations these patents receive, a well-established measure in the literature (Hall et al., 2001). Because a salient feature of patent data is that most inventors have very few patents, I focus on a sub-sample of top inventors for whom the panel dimension is relevant, as they are repeatedly observed patenting for extended periods of time.

Given this panel of top inventors, I fit a series of two-way fixed effects models, following the approach introduced by Abowd et al. (1999) in their study of French matched employer-employee data. This framework allows to recover estimates of inventor and firm fixed effects and decompose inventor output into three components: inventor heterogeneity, firm heterogeneity, and sorting. A concern with this estimation technique is that this decomposition is potentially biased by the very low number of inventors moving across firms. I show that clustering firms in a pre-estimation step alleviates this concern by substantially increasing the number or movers per firm cluster. Following Bonhomme et al. (2019b), I use a kmeans algorithm to group firms prior to estimation based on the quality of their patents.

My estimates suggest that there is positive sorting between inventors and firms. As such this form of sorting represents another micro-level mechanism contributing to knowledge production. Across periods, sorting accounts for about five percent of the total variance in inventor output. The correlation coefficient between inventor and firm effects ranges from .10 to .24, a value in the upper middle range of those typically found in matched employer-employee data. Finally, these estimates further point to inventor-firm regional collocation—good inventors and good firms locating in the same geographical areas—as a factor accounting for regional disparities in inventor output.

Related literature. This paper relates to two main strands of literature. There is first a substantial literature embedding a stylized notion of knowledge acquisition/diffusion in economic growth model to account for technological progress (Lucas, 1988; Aghion and Howitt, 1992; Eeckhout and Jovanovic, 2002). All the papers cited at the start of the introduction can be seen as investigating potential micro-level foundations for this process of knowledge acquisition (Moretti, 2019; Jaravel et al., 2018; Azoulay et al., 2019). Akcigit et al. (2018) makes this connection explicit by using patent level data from the European Patent Office to discipline the knowledge creation process in a growth model. My paper investigates the role of another micro-level mechanism underlying knowledge production by documenting the prevalence of inventor-firm sorting.

Second, there is an active line of research using econometric models with two-sided unobserved heterogeneity to investigate the determinant of wage dispersion using matched employer-employee data. Following Abowd et al. (1999), these models have been estimated on data from a number of countries (see Card et al., 2018, for an overview), and they also have been used to discipline job search models (Lopes de Melo, 2018). This paper is, to the best of my knowledge, the first to leverage this framework to analyze patent data.

Outline. Section 3.2 describes the data and sample selection. Section 3.3 outlines the econometric framework. Section 3.4 reports the results, and Section 3.5 concludes.

3.2 Data

This study uses data from PatentsView, a project developed by the United States Patent and Trademark Office (USPTO) to promote the use of their data on granted patents (USPTO, 2019). PatentsView contains data on the universe of patents granted by USPTO over the period 1976-2019.[1] The key advantage of these data is to thoroughly disambiguate the names of inventors and make available the resulting inventor identifiers, thus allowing to track inventors over the course of their career.

Disambiguation of inventor names. Since USPTO does not issue a unique identifier for inventors the first time they are listed on a patent, there is no direct way to track an inventor over time. Besides, due to homonyms (two distinct patents listing "John Smith") or to variations in the spelling of an inventor's name ("John Smith" vs "J. Smith"), identical or similar

[1] PatentsView also makes data on patent applications available for the period after 2001. Because of the way I define patent quality, I focus on granted patents.

names across patents cannot be assumed to identify the same person. In the PatentsView data, the names of inventors are disambiguated using a Discriminative Hierarchical Coreference algorithm to yield unique inventor identifiers (Monath and McCallum, 2015). This algorithm was chosen by USPTO as it outperformed other algorithms in an open competition to disambiguate inventor names in patent data. The accuracy of the resulting inventor identifiers should therefore be higher than with alternative algorithms proposed in the literature (Li et al., 2014; Morrison et al., 2017). I take this disambiguation step and the resulting inventor identifiers as given in what follows.

Definitions. I follow Akcigit et al. (2018) and Moretti (2019) in defining the relevant inventor level concepts of time, output, geography, and "firm". These definitions are motivated by the need to capture the context of an inventor's work on the patented innovation, as this context is interpreted as an input in patent production.

Time t is defined as the year when the inventor applied for the patent, not when the patent is ultimately granted. The application year is likely to be close to the time when the underlying work leading to the patent is undertaken. There is, on average, a 2.6 year lag between the application date and granting date.

An inventor's output is measured by the quality of her patents. Patent quality is defined as the number of forward citations from published patents, potentially truncated, that a patent p receives. The quality of a patent is then given by $q_{p,t} := \sum_{t'=t}^{t+\tau} \text{citations}_{p,t'}$, where the time index refers to the application date for both citing and cited patents. τ is the truncation date for forward citations, which I set to five years in my baseline specifications to avoid artificially inflating the quality of older patents, as those have had more time to accumulate citations. I also experiment with measures accounting for team size, by dividing $q_{i,t}$ by the number of inventors listed on the patent, and for the state of technology by regressing these quality measures on a set of year-technology category dummies.[2]

To define an inventor-level measure of output, I sum up the quality of all patents on which an inventor is listed in a given time period. With my baseline measure of quality, inventor i's output in year t is then simply the sum of all forward citations the patents she produces in year t receive: $y_{i,t} := \sum_p q_{p,t}$. The same definition of output can be readily extended to

[2]Technology classes are defined as the 38 subcategories in the NBER classification of patents.

alternative measures of patent quality.

Location is defined as the thinnest level (TL3) in the OECD's territorial level classification of regions (see OECD, 2018b, Annex A for details).[3] In the case of the US, this territorial level corresponds to the 179 Economic Areas defined by the Bureau of Economic Analysis (BEA). For instance, "Los Angeles-Long Beach-Riverside" and "San Diego-Carlsbad-San Marcos" correspond to two distinct TL3 regions in that classification. I use the modal location if an inventor lists several addresses on her patents within a time period.

A "firm", finally, is defined as the assignee of the patent. The assignee is the legal owner of the patent, and can be an individual or an organization.[4] The overwhelming majority—more than 99 percent—of patents are owned by an organization. In addition, 98.3 percent of patents are owned by a private entity, company or corporation, including private universities.[5] I therefore use the term "firm" interchangeably with "assignee" in the remainder of the paper and see these entities as an input in the patent production function. This input potentially captures the intellectual environment, processes, research amenities that these firms offer to their inventors, though I do not attempt to characterize its nature further for lack of data.

Sample selection. As detailed in the next section, the econometric approach in this study is to rely on two-way fixed effect models to decompose inventor output into their ability and their research environment. This approach puts some constraints on the number of observations for each inventors if individual fixed effects are to be consistently estimated. In matched employer-employee datasets, on which two-way fixed effects models are traditionally estimated, full-time workers are usually observed every year, as their earnings are recorded for tax purposes. This differs from patent data where most inventors only have a single patent over the course of their career. I therefore proceed by imposing a set of restrictions to build a panel of inventors with repeated observations.

First, I focus on inventors listing an address in the US. These inventors represent roughly fifty percent of all inventors mentioned on granted patents

[3]A virtue of the OECD's classification is to have full geographical coverage for all OECD member countries.

[4]If there is more than one assignee, which is the case for 3 percent of patents, I use the first one listed on the patent.

[5]The names of assignees are subject to the same limitations as that of inventors in the extent their spelling could vary across patent, and are therefore disambiguated in the PatentViews database using string matching.

in the raw data. This restriction is made to be in line with the inventor mobility requirements underlying two-way fixed-effects models, which I detail in Section 3.3. Inventors are way more likely to move across firms into an integrated labor market, such as the US.

Second, since the data cover more than four decades, I consider a set of ten-year overlapping periods separately: 1975-1984, 1980-1989, and so forth by five-year increments until 2000-2009.[6] Patents are included in each time interval based on their application dates. This choice balances the need to have a long enough panel of inventors with the possibility that the production function underlying inventor output may be changing over time.

Third, within each sub-period, I select inventors with the largest number of granted patents, in the spirit of Moretti (2019). I focus on inventors in the top five percent in the distribution of the number of patents in each ten-year period. Table 3.1 shows how these inventors differ from the whole sample.[7] By construction, the selected inventors have a markedly larger number of patents within each sub-period, which implies that the length of the panel ranges, on average, from 5.3 to 7.1 years. In addition, while the average size of the team (the number of inventors listed on the patent) are comparable for both groups, the average quality of patents produced by the group of selected inventors is clearly greater, as their patents attract more citations on average.[8] In all, Table 3.1 makes clear that the selected sample is made of star inventors.

3.3 Econometric methods

Following the large literature on two-way fixed effects models pioneered by Abowd et al. (1999), I propose to analyze inventor output in a linear framework with unobserved heterogeneity in inventors and firms. Because, to the best of my knowledge, this model has not been previously applied to individual-level patent data, I also document the mobility patterns of indi-

[6] I exclude the last decade available in the PatentsView data, as these patents could not accumulate citations from future patents, which makes the measure of patent quality inaccurate. This problem is made starker by the lag between application and granting dates.

[7] I also assess the robustness of my results to focusing on inventors in the top twenty and ten percent in the distribution of the number of patents in each period. Table 3.7 in Appendix 3.A shows similar statistics for these alternative sample criteria.

[8] Comparing the technology fields of patents in each group, I find no salient differences, though selected inventors do appear to have more patents in "Chemical" for earlier periods.

Period	Sample	Inventors	Avg. patents	Avg. years	Team size	Patent Quality ($\overline{q}_i$) Avg.	p25	p50	p75
1975-84	all	209,586	2.43	1.72	2.17	2.31	0.50	1.50	3.00
	top 5%	12,946	12.52	5.31	2.10	2.46	1.24	2.00	3.15
1980-89	all	244,475	2.39	1.72	2.48	3.25	1.00	2.00	4.00
	top 5%	14,894	12.09	5.23	2.37	3.49	1.67	2.75	4.43
1985-94	all	329,935	2.58	1.80	2.90	5.30	1.00	3.00	6.00
	top 5%	18,860	13.79	5.65	2.84	5.87	2.56	4.27	7.16
1990-99	all	464,638	3.10	1.94	3.28	7.72	2.00	4.00	9.00
	top 5%	26,384	18.19	6.03	3.40	9.14	3.80	6.63	11.36
1995-04	all	601,179	3.62	2.06	3.65	6.71	1.11	3.60	8.00
	top 5%	33,980	23.18	6.56	3.80	8.98	3.99	6.60	11.00
2000-09	all	656,964	3.97	2.17	3.90	4.42	0.83	2.00	5.00
	top 5%	34,543	27.96	7.10	4.06	6.57	2.78	4.56	7.58

Table 3.1: Inventors selected in analysis sample vs all inventors. The selected inventors are in the top five percent in the distribution of the number of patents in each period. The table shows inventor-level statistics for each sample periods. “Patents” and “Years” are averages of inventor-level totals. “Team Size” and “Patent Quality” are, respectively, for averages and summary statistics of inventor-level averages.

viduals (“inventors”) across production units (“assignees”) in this context.

Econometric specification. Within each ten-year sample, inventors i are observed producing output y_{it} in year t. Let $j(i,t)$ denote the assignee (”firm”) of inventor i in year t. I assume the following linear model for inventor’s i output in year t

$$y_{it} = \alpha_i + \psi_{j(i,t)} + \epsilon_{it}, \tag{3.1}$$

where α_i captures the contribution of inventor i and ψ_j the contribution of firms to output. It is assumed that the error term satisfies the strict exogeneity condition $E[\epsilon_{it}|\alpha_i, \psi_{j(i,t)}] = 0$. As noted in Card et al. (2013), this restriction does not rule out mobility patterns related to the inventor or firm effects. It does, however, rule out mobility based on the realization of the idiosyncratic component of output, ϵ_{it}.

		Share in largest component		
Period	Components	Observations	Inventors	Assignees
1975-1984	1329	0.857	0.848	0.539
1980-1989	1479	0.855	0.846	0.587
1985-1994	1638	0.871	0.863	0.627
1990-1999	1536	0.911	0.909	0.726
1995-2004	1449	0.937	0.932	0.799
2000-2009	1278	0.946	0.941	0.823

Table 3.2: Component statistics in Inventor-Assignee network.

Variance decomposition. Given the model specified in Equation (3.1), the variance of inventor output admits the following decomposition

$$\mathrm{Var}(y_{it}) = \underbrace{\mathrm{Var}(\alpha_i)}_{\text{inventor comp.}} + \underbrace{\mathrm{Var}(\psi_{j(i,t)})}_{\text{firm comp.}} + \underbrace{2\ \mathrm{Cov}(\alpha_i, \psi_{j(i,t)})}_{\text{inventor-firm sorting}} + \underbrace{\mathrm{Var}(\epsilon_{it})}_{\text{error comp.}}. \tag{3.2}$$

I use this decomposition to quantify the contribution of different factors to inventor output. Of particular interest are the variance of the firm effects $\psi_{j(i,t)}$, capturing the contribution of firms to inventor output, and the sorting term, $2 \times \mathrm{Cov}(\alpha_i, \psi_{j(i,t)})$, related to the amount of sorting between the quality of inventors and the quality of their firm.

Low mobility bias. It is well-documented in the literature on two-way fixed effects models that Equation (3.2) yields biased estimates of sorting because many firms have very few movers (Andrews et al., 2008; Lamadon et al., 2019). It has been shown that this low mobility bias tends to underestimate the sorting component of the variance of wages in favor of the firm component.

This problem of low mobility translates directly to inventor-assignee data. First, as shown in Abowd et al. (2002), the fixed effects are only identified within the connected components defined by inventor mobility. Table 3.2 shows summary statistics on the implied components for each period. While the number of components is large, never less than a thousand in each period, the largest component covers the vast majority of observations, around 90 percent across periods.[9] Incidentally, this share is similar to that found in Lamadon et al. (2019) for the same period in regular US matched employer-employee data.

Second, even within the largest component, the number of movers is

[9]The figure is similar when looking at the number of inventors in the largest component, but markedly smaller when consider assignees instead.

	Number of movers per firm					
Clustering	Avg.	min	p25	p50	p75	max
No	4.52	1	1	1	3	875
$K = 10$	4096.70	1723	2132	3804	5502	7834
$K = 20$	2289.40	879	1215	1966	2633	5393
$K = 30$	1571.77	532	735	1038	1943	5084
$K = 40$	1202.98	218	568	973	1588	3397

Table 3.3: Summary statistics on number of movers per firm (2000-2009).

small for the majority of firms. Table 3.3 shows that 50 percent of assignees have a single inventor moving in the period 2000-2009. To address this issue, I follow the procedure outlined in Bonhomme et al. (2019b) and cluster assignees into groups in a first, pre-estimation step. I use a kmeans clustering algorithm and partition assignees based on the empirical CDF of inventor output into K groups.[10] Table 3.3 further shows that clustering firms makes them more strongly connected by substantially increasing the number of movers. Even with $K = 40$ assignee clusters, the number of movers within a cluster is never less than three hundred. I subsequently set the number of clusters to $K = 30$ and check the robustness of my results to varying this parameter.

3.4 Results

Baseline results. Table 3.4 presents variance decomposition results based on Equation (3.2). The results are broken down by period. A first observation is that the fit of such a linear model is far from perfect with as skewed an outcome as patent quality. Across periods, about fifty percent of the variance falls in the error term. This suggests that fully specifying the likelihood as a function of inventor and assignee latent heterogeneity, an approach taken by Bonhomme et al. (2019b) and Lentz et al. (2018) with matched-employee data, is a potentially promising extension.

Second, turning to the part of the variance that is explained by the fixed effects, the decomposition suggests that the main factor driving differences in inventor output is the inventor fixed effect, in line with what is traditionally found with matched employer-employee data. Though it varies slightly across periods, the share of inventor output that is explained by inventor fixed effects does not exhibit a clear trend and oscillates between

[10] This procedure is now standard (Dauth et al., 2018; Lamadon et al., 2019). The details of the implementation can be found in Bonhomme et al. (2019b).

	Share of Var(y_it) explained by:				
Period	Var(α_i)	Var(ψ_j)	2 Cov(α_i,ψ_j)	Var(ϵ_{it})	Cor(α_i,ψ_j)
1975-1984	0.329	0.045	0.029	0.597	0.119
1980-1989	0.319	0.065	0.043	0.574	0.148
1985-1994	0.340	0.094	0.072	0.494	0.202
1990-1999	0.343	0.101	0.085	0.471	0.230
1995-2004	0.333	0.079	0.052	0.536	0.159
2000-2009	0.369	0.048	0.046	0.537	0.173

Table 3.4: Decomposition of inventor output. The table reports the decomposition of Var($y_{i,t}$) across sample period as given by (3.2). All terms are computed using their sample analog, where the fixed effects are first estimated in the linear regression given by (3.1). Rows may not sum up to one due to rounding.

thirty and thirty-five percent.

Third, the fraction of the variance of inventor output respectively explained by the firm and sorting component is of similar magnitude, though the sorting component remains slightly smaller across the period. This suggests that firms do matter for inventor output and that there is a fair amount of positive sorting between inventors and assignees. Surprisingly, there is also no clear upward trend in these two components. They both appear to rise up to the 1990s, but subsequently fall in the 1995-2004 and 2000-2009 intervals. A tentative explanation could be that these two periods are plagued by recessions—one mild in the early 2000s, but centered on the high-tech sector, the other much more severe—which could have impaired the quality of matches.

The correlation between the inventor and firm fixed-effect varies from .11 to .24. In matched employer-employee datasets, studies have typically found a very wide range of values for the correlation between worker and firm fixed effects (see Lentz et al., 2018, Table 5 for a summary). By contrast, the correlations found in Table 3.4 are (i) clearly positive and (ii) slightly on the upper end of the estimates found in the matched employer-employee literature, though this might reflect methodological differences since some of these other studies are not necessarily addressing the bias entailed by low worker mobility.

Robustness checks. Table 3.5 presents a series of robustness checks: using a different number of firm clusters, varying the measure of inventor output, and adding covariates in Equation (3.1). I focus on the sample period 2000-2009 and report the same variance decomposition as in Table

3.4.

Overall, the main message of Table 3.5 is that the decomposition outlined above is robust to these alternative specifications. Both the overall breakdown between the different terms and the amount of sorting remain roughly identical. Surprisingly, the explained part of inventor output does not markedly increase when additional covariates are included.

Finally, I also test the robustness of the decomposition to alternative definitions of the sample of inventors. The final rows in Table 3.5 shows estimates of Equation 3.2 when the sample includes inventors in the top ten and top twenty percent in the distribution of the number of patents per inventor, instead of the five percent threshold used in the baseline estimation. Though these samples imply a slightly lower amount of inventor-firm sorting, the overall breakdown between the different components is robust to the definition of "top inventors". Table 3.8 and 3.9 in Appendix 3.A further establish that this breakdown remains robust across periods for the alternative samples.

Technology clusters and sorting. The role of technology clusters as a salient vector of knowledge production has received a lot of attention in the literature (see the dicussion in Lucas, 1988, for an early example). I revisit this role through the lens of the two-way fixed effect model in Equation (3.1). Similarly to the exercise in Dauth et al. (2018) for the earnings of workers in West Germany, I decompose the variance of average inventor output between regions as

$$\mathrm{Var}(E_r(q_{it})) = \mathrm{Var}(E_r(\alpha_i)) + \mathrm{Var}(E_r(\psi_{j(i,t)})) + 2\ \mathrm{Cov}(E_r(\alpha_i), E_r(\psi_{j(i,t)})), \tag{3.3}$$

where $E_r(.)$ denotes the within-region mean. This expression gives a decomposition of the dispersion in average patent quality across regions in an inventor component, a firm component, and a term capturing sorting at the region level.

Recall that regions are defined according the OECD's nomenclature, which in the US corresponds to the "Economic Areas" defined by the Bureau of Economic Analysis. Importantly, this classification aims at generating economically meaningful geographical units, and therefore can cut across traditional administrative units, as is the case for the "New York-Newark-Bridgeport" region. The decomposition in Equation (3.3) therefore gives some insight into where differences in the average inventor output of

	Share of Var(y_{it}) explained by:				
	Var(α_i)	Var(ψ_j)	2 Cov	Var(ϵ_{it})	Cor
Baseline specification					
$y_{it} := \sum_p q_{pt}$, $K = 30$	0.369	0.048	0.046	0.537	0.173
Number of clusters					
$K = 10$	0.390	0.033	0.036	0.541	0.158
$K = 20$	0.377	0.043	0.041	0.539	0.158
$K = 40$	0.360	0.056	0.048	0.535	0.170
Alternative outcomes					
Adj. for year-cat	0.371	0.040	0.045	0.544	0.187
9-year forward cit.	0.403	0.042	0.046	0.509	0.176
Adj. team size	0.370	0.047	0.046	0.538	0.175
$y_{it} := \sum_{p'} \operatorname{asinh}(q_{p't})$	0.355	0.066	0.040	0.538	0.130
$y_{it} := \sum_{p'} \log\big(1 + q_{p't}\big)$	0.360	0.066	0.037	0.537	0.119
# patents	0.347	0.052	0.030	0.571	0.112
Adding regressors					
Age cubic	0.414	0.043	0.044	0.533	0.164
Year FEs	0.373	0.043	0.047	0.532	0.185
Age & Year	0.394	0.043	0.045	0.532	0.173
Alternative samples					
Top 20% inventors	0.401	0.038	0.037	0.524	0.150
Top 10% inventors	0.388	0.041	0.039	0.533	0.154

Table 3.5: Robustness checks. The variance decomposition is the same as in Table 3.4 for the period 2000-2009. The different specifications are listed in the first column. "Adj. for year-cat" refers to residual inventor output in a regression on application year × patent category dummies. "Adj. team size"refers to forward citations divided by team size. The additional regressors considered are year fixed-effects and a quadratic in age, where age=0 in the inventor's first recorded application year. Alternative samples are for inventors in the top twenty and ten percent in the distribution of the number of patents (baseline is five percent). Rows may not sum up to one due to rounding. For the specifications with additional covariates, rows do not sum up to one because these additional covariates are not shown in the variance decomposition.

	Share of Var($E_r(y_{it})$) explained by:				
Period	Var($E_r(\alpha_i)$)	Var($E_r(\psi_j)$)	2 Cov	Regions	std($E_r(y_{it})$)
1975-1984	0.450	0.303	0.246	110	1.749
1980-1989	0.367	0.355	0.279	115	2.403
1985-1994	0.335	0.229	0.436	116	8.135
1990-1999	0.341	0.327	0.332	126	11.571
1995-2004	0.416	0.238	0.346	125	16.895
2000-2009	0.582	0.241	0.177	129	13.850

Table 3.6: Decomposition of variance between regions. The table reports the decomposition of Var($E_r(y_{it})$) given in (3.3). All elements are computed using their sample analog. Var($E_r(y_{it})$) is computed based on the fitted value of y_{it} in (3.1). Regions from the bottom quartile in terms of number of observations are excluded from the computation.

these areas stem from.

Table 3.6 shows this decomposition for all sample periods. It first makes clear that the dispersion in average inventor output across regions is sizable. For example, the standard deviation of this measure is eight citations (per inventor-year) in the 1985-1994 interval.

It further shows that all three terms in Equation (3.3) contribute substantially to the variation in average inventor output between regions. Though it evolves across periods, all terms account for roughly a third of the total variance. As a consequence, there appears to be a lot of regional inventor-firm sorting, in the sense that good inventors appear to locate into regions with better firms. A substantial part of the regional differences in patent quality can therefore be ascribed to a form of collocation between inventors and firms.

3.5 Conclusion

This paper analyzes a new channel to explain variation in the quality of inventor output: inventor-firm sorting. I translate econometric methods traditionally used in the study of matched employer-employee data to patent data. This methodology allows me to recover the contributions of inventors, firms, and inventor-firm sorting to the total variance of inventor output.

I find evidence of positive inventor-firm sorting. This mechanism accounts on average for five percent of the total variance of inventor output in the US between 1975 and 2010. My analysis further suggests that the geographical sorting of inventors and firms is a key channel to explain regional disparities in inventor output.

			Avg.	Avg.	Team	Patent Quality ($\overline{q}_i$)			
Period	Sample	Inventors	patents	years	size	Avg.	p25	p50	p75
1975-84	all	209,586	2.43	1.72	2.17	2.31	0.50	1.50	3.00
	top 20%	51,413	6.08	3.41	2.12	2.41	1.00	1.86	3.14
	top 10%	23,346	9.35	4.51	2.10	2.45	1.17	2.00	3.17
	top 5%	12,946	12.52	5.31	2.10	2.46	1.24	2.00	3.15
1980-89	all	244,475	2.39	1.72	2.48	3.25	1.00	2.00	4.00
	top 20%	59,659	5.96	3.39	2.42	3.42	1.38	2.60	4.33
	top 10%	27,137	9.07	4.45	2.38	3.48	1.60	2.67	4.40
	top 5%	14,894	12.09	5.23	2.37	3.49	1.67	2.75	4.43
1985-94	all	329,935	2.58	1.80	2.90	5.30	1.00	3.00	6.00
	top 20%	87,873	6.23	3.51	2.90	5.62	2.00	3.75	6.75
	top 10%	41,910	9.37	4.57	2.87	5.77	2.38	4.00	7.00
	top 5%	18,860	13.79	5.65	2.84	5.87	2.56	4.27	7.16
1990-99	all	464,638	3.10	1.94	3.28	7.72	2.00	4.00	9.00
	top 20%	102,972	8.81	4.11	3.39	8.49	3.00	5.59	10.25
	top 10%	47,133	13.62	5.26	3.41	8.91	3.54	6.22	10.89
	top 5%	26,384	18.19	6.03	3.40	9.14	3.80	6.63	11.36
1995-04	all	601,179	3.62	2.06	3.65	6.71	1.11	3.60	8.00
	top 20%	151,398	9.93	4.25	3.75	7.82	2.83	5.25	9.50
	top 10%	62,901	16.76	5.69	3.78	8.51	3.56	6.08	10.38
	top 5%	33,980	23.18	6.56	3.80	8.98	3.99	6.60	11.00
2000-09	all	656,964	3.97	2.17	3.90	4.42	0.83	2.00	5.00
	top 20%	136,598	12.69	4.90	4.05	5.62	2.00	3.70	6.63
	top 10%	76,813	18.04	5.92	4.07	5.99	2.35	4.10	7.00
	top 5%	34,543	27.96	7.10	4.06	6.57	2.78	4.56	7.58

Table 3.7: Summary statistics on alternative samples on inventors. "top 5%" is the benchmark used in the main text. The table shows inventor-level statistics for each sample periods. "Patents" and "Years" are averages of inventor-level totals. "Team Size" and "Patent Quality" are, respectively, for averages and summary statistics of inventor-level averages.

Appendix 3.A Additional tables

Period	Share of Var(y_it) explained by: Var(α_i)	Var(ψ_j)	2 Cov(α_i,ψ_j)	Var(ϵ_{it})	Cor(α_i,ψ_j)
1975-1984	0.388	0.035	0.027	0.550	0.116
1980-1989	0.391	0.070	0.050	0.489	0.152
1985-1994	0.412	0.069	0.067	0.452	0.198
1990-1999	0.412	0.077	0.079	0.432	0.221
1995-2004	0.376	0.060	0.050	0.515	0.166
2000-2009	0.401	0.038	0.037	0.524	0.150

Table 3.8: Decomposition of inventor output with inventors in the top 20% in the distribution of the number of patents. The table reports the decomposition of Var($y_{i,t}$) across sample period as given by (3.2). All terms are computed using their sample analog, where the fixed effects are first estimated in the linear regression given by (3.1). Rows may not sum up to one due to rounding.

Period	Share of Var(y_it) explained by: Var(α_i)	Var(ψ_j)	2 Cov(α_i,ψ_j)	Var(ϵ_{it})	Cor(α_i,ψ_j)
1975-1984	0.352	0.043	0.033	0.572	0.132
1980-1989	0.346	0.069	0.049	0.537	0.158
1985-1994	0.374	0.072	0.069	0.485	0.210
1990-1999	0.372	0.094	0.085	0.449	0.226
1995-2004	0.348	0.068	0.053	0.531	0.174
2000-2009	0.388	0.041	0.039	0.533	0.154

Table 3.9: Decomposition of inventor output with inventors in the top 10% in the distribution of the number of patents. The table reports the decomposition of Var($y_{i,t}$) across sample period as given by (3.2). All terms are computed using their sample analog, where the fixed effects are first estimated in the linear regression given by (3.1). Rows may not sum up to one due to rounding.

Bibliography

Abowd, John M, Francis Kramarz, and David N Margolis, "High wage workers and high wage firms," *Econometrica*, 1999, *67* (2), 251–333. 120, 121, 124

_ **, Robert H Creecy, Francis Kramarz et al.**, "Computing person and firm effects using linked longitudinal employer-employee data," Technical Report, Center for Economic Studies, US Census Bureau 2002. 126

Acemoglu, Daron and Robert Shimer, "Efficient Unemployment Insurance," *Journal of Political Economy*, October 1999, *107* (5), 893–928. 89, 91

Aghion, Philippe and Peter Howitt, "A model of growth through creative destruction," *Econometrica*, 1992, *51.* 121

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi, "Dancing with the stars: Innovation through interactions," Technical Report, National Bureau of Economic Research 2018. 119, 121, 122

Andrews, Martyn J, Len Gill, Thorsten Schank, and Richard Upward, "High wage workers and low wage firms: negative assortative matching or limited mobility bias?," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2008, *171* (3), 673–697. 126

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S Graff Zivin, "Does science advance one funeral at a time?," *American Economic Review*, 2019, *109* (8), 2889–2920. 119, 121

Barlevy, Gadi, "The sullying effect of recessions," *The Review of Economic Studies*, 2002, *69* (1), 65–96. 23, 24

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, "Cross-country differences in productivity: The role of allocation and selection," *American Economic Review*, 2013, *103* (1), 305–34. 30

Bilal, Adrien, Niklas Engbom, Simon Mongey, and Giovanni L Violante, "Firm and Worker Dynamics in a Frictional Labor Market," 2019. 7, 29

Blundell, Richard, Claire Crawford, and Wenchao Jin, "What can wages and employment tell us about the UK's productivity puzzle?," *The Economic Journal*, 2014, *124* (576), 377–407. 62

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa, "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 2019, *87* (3), 699–739. 90, 102, 103

_ , _ , and _ , "A distributional framework for matched employer employee data," *Econometrica*, 2019, *87* (3), 699–739. 120, 127

Bradley, Jake, "Self-Employment in an Equilibrium Model of the Labor Market," *IZA Journal of Labor Economics*, June 2016, *5* (1), 6. 92

_ , Fabien Postel-Vinay, and Hélene Turon, "Public Sector Wage Policy and Labor Market Equilibrium: A Structural Model," *Journal of the European Economic Association*, 2017, p. jvw026. 92, 98

Burdett, Kenneth and Dale T Mortensen, "Wage differentials, employer size, and unemployment," *International Economic Review*, 1998, pp. 257–273. 24, 28, 37, 53

Caballero, Ricardo J and Mohamad L Hammour, "The Cleansing Effect of Recessions," *The American Economic Review*, 1994, pp. 1350–1368. 23

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline, "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 2018, *36* (S1), S13–S70. 121

_ , Jörg Heining, and Patrick Kline, "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly journal of economics*, 2013, *128* (3), 967–1015. 125

Catherine, Sylvain, "Keeping Options Open: What Motivates Entrepreneurs?," 2017, p. 57. 92

Chaumont, Gaston Rene and Shouyong Shi, "Wealth Accumulation, on the Job Search and Inequality," SSRN Scholarly Paper ID 2893777, Social Science Research Network, Rochester, NY January 2017. 92

Chetty, Raj, "A General Formula for the Optimal Level of Social Insurance," *Journal of Public Economics*, November 2006, *90* (10–11), 1879–1901. 89, 91

_ , "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, April 2008, *116* (2), 173–234. 91, 96

Clementi, Gian Luca and Berardino Palazzo, "Entry, exit, firm dynamics, and aggregate fluctuations," *American Economic Journal: Macroeconomics*, 2016, *8* (3), 1–41. 28

Coles, Melvyn G and Dale T Mortensen, "Equilibrium Labor Turnover, Firm Growth, and Unemployment," *Econometrica*, 2016, *84* (1), 347–363. 12, 15, 29, 37, 39, 66, 67, 68

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum, "Matching in cities," Technical Report, National Bureau of Economic Research 2018. 127, 129

Davis, Steven J. and John Haltiwanger, "Gross Job Creation, Gross Job Destruction, and Employment Reallocation*," *The Quarterly Journal of Economics*, 08 1992, *107* (3), 819–863. 27

de Melo, Rafael Lopes, "Firm wage differentials and labor market sorting: Reconciling theory and evidence," *Journal of Political Economy*, 2018, *126* (1), 313–346. 121

Decker, Ryan A, John C Haltiwanger, Ron S Jarmin, and Javier Miranda, "Changing business dynamism and productivity: Shocks vs. responsiveness," Technical Report, National Bureau of Economic Research 2018. 31

den Haan, Wouter J., "Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents," *Journal of Economic Dynamics and Control*, 2010, *34* (1), 79 – 99. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty. 87

Diewert, W Erwin and K Fox, "On measuring the contribution of entering and exiting firms to aggregate productivity growth," *Price and productivity measurement*, 2010, *6*, 41–66. 32

Eeckhout, Jan and Boyan Jovanovic, "Knowledge spillovers and inequality," *American Economic Review*, 2002, *92* (5), 1290–1307. 121

Elsby, Michael W. L. and Axel Gottfries, "Firm Dynamics, On-the-Job Search and Labor Market Fluctuations," Working Paper 2019. 7, 29

Elsby, Michael WL and Ryan Michaels, "Marginal jobs, heterogeneous firms, and unemployment flows," *American Economic Journal: Macroeconomics*, 2013, *5* (1), 1–48. 28

Fort, Teresa C, John Haltiwanger, Ron S Jarmin, and Javier Miranda, "How firms respond to business cycles: The role of firm age and firm size," *IMF Economic Review*, 2013, *61* (3), 520–559. 27, 78

Foster, Lucia, Cheryl Grim, and John Haltiwanger, "Reallocation in the Great Recession: cleansing or not?," *Journal of Labor Economics*, 2016, *34* (S1), S293–S331. 23, 27

_ , John C Haltiwanger, and Cornell John Krizan, "Aggregate productivity growth: lessons from microeconomic evidence," in "New developments in productivity analysis," University of Chicago Press, 2001, pp. 303–372. 31, 32

Fujita, Shigeru and Giuseppe Moscarini, "Recall and Unemployment," *American Economic Review*, December 2017, *107* (12), 3875–3916. 114

Gabaix, Xavier, "Zipf's law for cities: an explanation," *The Quarterly journal of economics*, 1999, *114* (3), 739–767. 54

Gavazza, Alessandro, Simon Mongey, and Giovanni L Violante, "Aggregate recruiting intensity," *American Economic Review*, 2018, *108* (8), 2088–2127. 28, 41

Gouin-Bonenfant, Emilien, "Productivity Dispersion, Between-Firm Competition, and the Labor Share," Technical Report 2019. 54

Griliches, Zvi and Haim Regev, "Firm productivity in Israeli industry 1979–1988," *Journal of econometrics*, 1995, *65* (1), 175–203. 32

Gu, Ran, "Specific Capital, Optimal Firm Insurance, and the Dynamics of the Postgraduate Wage Premium," Technical Report 2019. 38

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, "The NBER patent citation data file: Lessons, insights and methodological tools," Technical Report, National Bureau of Economic Research 2001. 120

Haltiwanger, John C, Henry R Hyatt, Lisa B Kahn, and Erika McEntarfer, "Cyclical job ladders by firm size and firm wage," *American Economic Journal: Macroeconomics*, 2018, *10* (2), 52–85. 24, 27, 61

Hamilton, Barton H., "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment," *Journal of Political Economy*, June 2000, *108* (3), 604. 92

Hansen, Gary D. and Ayşe Imrohoroglu, "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard," *Journal of Political Economy*, February 1992, *100* (1), 118. 91

Hornstein, Andreas, Per Krusell, and Giovanni L Violante, "Frictional wage dispersion in search models: A quantitative assessment," *American Economic Review*, 2011, *101* (7), 2873–98. 53

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 2009, *124* (4), 1403–1448. 27

Humphries, John Eric, "The Causes and Consequences of Self-Employment over the Life Cycle," 2018, p. 83. 92

Hurst, Erik and Benjamin Wild Pugsley, "What Do Small Businesses Do?," Working Paper 17041, National Bureau of Economic Research May 2011. 95

Jaravel, Xavier, Neviana Petkova, and Alex Bell, "Team-specific capital and innovation," *American Economic Review*, 2018, *108* (4-5), 1034–73. 119, 121

Kaas, Leo and Philipp Kircher, "Efficient Firm Dynamics in a Frictional Labor Market †," *American Economic Review*, Oct 2015, *105* (10), 3030–3060. 28

Khan, Aubhik and Julia K Thomas, "Credit shocks and aggregate fluctuations in an economy with production heterogeneity," *Journal of Political Economy*, 2013, *121* (6), 1055–1107. 28

Kolsrud, J., Camille Landais, P. Nilsson, and Johannes Spinnewijn, "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 2015. 91

Krusell, Per and Anthony A Smith Jr, "Income and wealth heterogeneity in the macroeconomy," *Journal of political Economy*, 1998, *106* (5), 867–896. 55

_, Toshihiko Mukoyama, and Ayşegül Şahin, "Labour-Market Matching with Precautionary Savings and Aggregate Fluctuations," *Review of Economic Studies*, October 2010, *77* (4), 1477–1507. 91, 109

Lamadon, Thibaut, "Productivity Shocks, Long-Term Contracts and Earnings Dynamics," p. 86. 92

_, Magne Mogstad, and Bradley Setzler, "Imperfect competition, compensating differentials and rent sharing in the US labor market," Technical Report, National Bureau of Economic Research 2019. 126, 127

Lentz, Rasmus and Dale T Mortensen, "An empirical model of growth through product innovation," *Econometrica*, 2008, *76* (6), 1317–1373. 27

_ and _, "Labor market friction, firm heterogeneity, and aggregate employment and productivity," Working Paper 2012. 29

_, Suphanit Piyapromdee, and Jean-Marc Robin, "On worker and firm heterogeneity in wages and employment mobility: Evidence from danish register data," *PIER Discussion Papers*, 2018, *91.* 127, 128

Li, Guan-Cheng, Ronald Lai, Alexander D'Amour, David M Doolin, Ye Sun, Vetle I Torvik, Z Yu Amy, and Lee Fleming, "Disambiguation and co-authorship networks of the US patent inventor database (1975–2010)," *Research Policy*, 2014, *43* (6), 941–955. 122

Lise, Jeremy, "On-the-Job Search and Precautionary Savings," *The Review of Economic Studies*, January 2013, *80* (3), 1086–1113. 92, 98, 104

_ and Jean-Marc Robin, "The Macrodynamics of Sorting between Workers and Firms," *The American Economic Review*, 2017, *107* (4), 1104–1135. 28

Lucas, Robert E, "On the mechanics of economic development," *Journal of monetary economics*, 1988. 121, 129

Manresa, Elena, Thibaut Lamadon, and Stéphane Bonhomme, "Discretizing Unobserved Heterogeneity," Technical Report, The IFS March 2017. 90, 102

McCall, J. J., "Economics of Information and Job Search," *The Quarterly Journal of Economics*, 1970, *84* (1), 113–126. 99

Meghir, Costas, Renata Narita, and Jean-Marc Robin, "Wages and Informality in Developing Countries†," *American Economic Review*, April 2015, *105* (4), 1509–1546. 92, 98

Merz, M and E Yashiv, "Labor and the market value of the firm," *American Economic Review*, 2007, *97* (4), 141. 37, 52

Monath, N and A McCallum, "Discriminative hierarchical coreference for inventor disambiguation," in "PatentsView Inventor Disambiguation Technical Workshop, USPTO, Alexandria, VA" 2015. 122

Moretti, Enrico, "The effect of high-tech clusters on the productivity of top inventors," Technical Report, National Bureau of Economic Research 2019. 119, 121, 122, 124

Morrison, Greg, Massimo Riccaboni, and Fabio Pammolli, "Disambiguation of patent inventors and assignees using high-resolution geolocation data," *Scientific data*, 2017, *4*, 170064. 122

Mortensen, Dale T and Christopher A Pissarides, "Job creation and job destruction in the theory of unemployment," *The review of economic studies*, 1994, *61* (3), 397–415. 23, 28, 47

Moscarini, Giuseppe and Fabien Postel-Vinay, "The contribution of large and small employers to job creation in times of high and low unemployment," *The American Economic Review*, 2012, *102* (6), 2509–2539. 27

_ **and** _ , "Stochastic Search Equilibrium," *The Review of Economic Studies*, October 2013, *80* (4), 1545–1581. 28, 44, 46, 66, 70, 71

_ **and** _ , "Wage posting and business cycles: A quantitative exploration," *Review of Economic Dynamics*, 2016, *19*, 135–160. 52, 70

OECD, *The Future of Social Protection* 2018. 20, 89

__ , *OECD Regions and Cities at a Glance 2018* 2018. 123

Office for National Statistics, "Business Structure Database, 1997-2018: Secure Access," `http://doi.org/10.5255/UKDA-SN-6697-10` 2019. 10th Edition, SN:6697. 30

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1297. 31

Ouyang, Min, "The scarring effect of recessions," *Journal of Monetary Economics*, 2009, *56* (2), 184–199. 23

Patterson, Christina, Ayşegül Şahin, Giorgio Topa, and Giovanni L Violante, "Working hard in the wrong place: A mismatch-based explanation to the UK productivity puzzle," *European Economic Review*, 2016, *84*, 42–56. 31

Postel-Vinay, Fabien and Alireza Sepahsalari, "Labour Mobility and Earnings in the UK, 1992-2016," Technical Report 2019. 35, 50, 81

Reed, William J, "The Pareto, Zipf and other power laws," *Economics letters*, 2001, *74* (1), 15–19. 54

Reiter, Michael, "Solving heterogeneous-agent models by projection and perturbation," *Journal of Economic Dynamics and Control*, 2009, *33* (3), 649–665. 56

Riley, Rebecca, Chiara Rosazza Bondibene, and Garry Young, "The UK productivity puzzle 2008-2013: evidence from British businesses," *Bank of England Staff Working Paper*, 2015, (531). 32

Rujiwattanapong, W Similan, "Unemployment dynamics and endogenous unemployment insurance extensions," 2019. 64

Saporta-Eksten, Itay, "Job Loss, Consumption and Unemployment Insurance," *Unpublished manuscript*, 2014, pp. 1–50. 104, 105

Schaal, Edouard, "Uncertainty and unemployment," *Econometrica*, 2017, *85* (6), 1675–1721. 28

Sedláček, Petr and Vincent Sterk, "The growth potential of startups over the business cycle," *American Economic Review*, 2017, *107* (10), 3182–3210. 28, 56

Sepahsalari, Alireza, "Financial market imperfections and labour market outcomes," 2016. 28

Shi, Shouyong, "Directed Search for Equilibrium Wage–Tenure Contracts," *Econometrica*, March 2009, *77* (2), 561–584. 92

Shimer, Robert, "The cyclical behavior of equilibrium unemployment and vacancies," *American economic review*, 2005, pp. 25–49. 28, 53

_ , *Labor markets and business cycles*, Princeton, N.J: Princeton University Press, 2010. 37

_ **and Ivan Werning**, "Liquidity and Insurance for the Unemployed," *American Economic Review*, December 2008, *98* (5), 1922–1942. 89

USPTO, "PatentsView," `https://www.patentsview.org/download/` 2019. Updated December 31, 2019. 120, 121

Visschers, Ludo, Ana Millan, and Matthias Kredler, "Great Opportunities or Poor Alternatives: Self-Employment, Unemployment and Paid Employment over the Business Cycle," Technical Report 597, Society for Economic Dynamics 2014. 92, 98

Winberry, Thomas, "Lumpy investment, business cycles, and stimulus policy," *Job market paper, Princeton University/University of Chicago Booth School of Business*, 2016. 56