Firm Dynamics and Random Search over the Business Cycle

Richard Audoly*

October 2020
(First version: January 2017)

Latest version available here.

Abstract

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.

*Department of Economics, Norwegian School of Economics (NHH). Email: richard.audoly@nhh.no. This paper is a revised version of the first chapter of my doctoral dissertation. I am thankful to my supervisors at UCL, Fabien Postel-Vinay and Jan Eeckhout, for their advice and support. For many helpful comments and suggestions, I also wish to thank Giuseppe Moscarini and Vincent Sterk as well as seminar participants at SaM 2017 (Barcelona), the Young Economists Conference 2018 (Belgrade), EEA 2018 (Cologne), the SaM Workshop 2019 (Bristol), SaM 2019 (Oslo), and the 3rd Dale T. Mortensen conference (Aarhus). Any errors are my own. *Data Disclaimer*: The UK Data Service agrees that the outputs are non-disclosive, and cannot be used to identify a person or organisation. The use of the Business Structure Database does not imply the endorsement of the data owner or the UK Data Service at the UK Data Archive in relation to the interpretation or analysis of the data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

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 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 gauge 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

[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.

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 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

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

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 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.

Two key contributions for the present paper are Moscarini and Postel-Vinay (2013) and Coles and Mortensen (2016). I maintain the assumption that firms can commit to state-contingent contracts from the former and build on the size independence result in the latter to characterize the optimal contract in the presence of firm-specific shocks. With respect to these papers, the key addition in my framework is to make firm entry and exit endogenous, which allows the selection of firms to evolve over the business cycle. Engbom (2019) and Gouin-Bonenfant (2019) also rely on a similar constant returns to scale assumption to simplify the firm's problem. They study, respectively, the impact of aging on dynamism in the labor market and of the dispersion in firm productivity on the labor share. Both focus on transition experiments between steady-states and abstract from aggregate shocks.

Lastly, two contemporaneous papers, Elsby and Gottfries (2019) and Bilal et al. (2019), present related environments, but relax the constant returns to scale assumption on the firm production technology. Both elegantly characterize the equilibrium as a job ladder in terms of some marginal value—respectively of production and joint surplus—a property closely related to the job ladder in terms of productivity in my framework. With respect to my paper, Elsby and Gottfries (2019) abstract from firm entry and exit, while Bilal et al. (2019) center on a steady-state environment to study the role of search frictions in explaining firm size and the life cycle of the firm. By contrast, my paper emphasizes the role of aggregate shocks in jointly determining the selection of firms and how workers move across these firms to derive the implications for aggregate labor productivity.

Outline. Section 2 documents novel facts on firm dynamics and labor productivity from British firm-level data. Section 3 introduces the model. Section 4 defines the equilibrium. Section 5 describes the calibration and numerical solution. Section 6 analyzes the reallocation properties of the model during a recession and Section 7 concludes.

2 Data

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 combines several administrative datasets from the UK to get a measure of labor productivity at the firm level for a large sample of British firms (Office for National Statistics, 2019, 2020a,b). Details on the construction of this sample are given in Appendix B.1.1. Importantly, these data cover several years before and after the Great Recession—officially starting in 2008Q2 and ending in 2009Q2 in the UK— thus allowing to decompose labor productivity before, during, and after the onset of this episode.

Aggregate labor productivity from firm-level data. I follow the definitions in Bartelsman et al. (2013) and construct aggregate labor productivity as $LP_t := \sum_i ES_{i,t} \cdot LP_{i,t}$, where the employment share, $ES_{i,t}$, and labor productivity measure, $LP_{i,t}$, at firm i in period t are given by

$$ES_{i,t} := \frac{\text{employment}_{i,t}}{\sum_i \text{employment}_{i,t}}, \quad LP_{i,t} := \ln\left(\frac{\text{value added}_{i,t}}{\text{employment}_{i,t}}\right). \tag{1}$$

Beyond consistency with Bartelsman et al. (2013), using the logarithm of the ratio of value added to employment is also convenient to get a unit-free productivity measure that can readily be compared to the model. I show in Appendix B.1.4 that the main results in this section hold when instead using a labor productivity measure in levels.

Macro-level changes in worker reallocation. To assess the role of worker reallocation in accounting for the overall change in labor productivity, I decompose LP_t as

$$LP_t = \sum_i ES_{i,t} \cdot LP_{i,t} = \underbrace{\overline{LP}_t}_{\text{average firm prod.}} + \underbrace{\sum_i \left(ES_{i,t} - \overline{ES}_t\right)\left(LP_{i,t} - \overline{LP}_t\right)}_{\text{OP misallocation measure}}, \tag{2}$$

an equality referred to in the literature as the "OP decomposition" (Olley and Pakes, 1996).[4] 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: it increases as more firms with above average productivity have a larger than average employment share. The use of this specific decomposition is guided by the fact that it admits an intuitive counterpart in the notation of the model introduced in subsequent sections.[5]

The evolution of each of these terms immediately around the time of the Great Recession is depicted in Figure 1a, expressed in deviation from their respective pre-recession linear trend. Figure 1a shows that both 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, the OP measure of misallocation has kept moving down since 2008. 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.

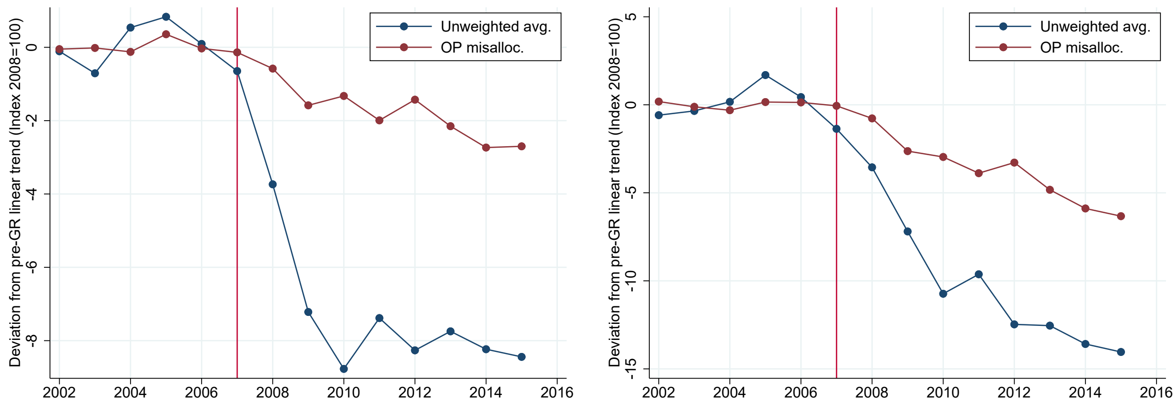
Impact on wages. A similar decomposition can be used to study changes in wages at the aggregate level. I define the following wage index

$$W_t := \sum_i ES_{i,t} \cdot W_{i,t}, \quad W_{i,t} := \ln\left(\frac{\text{total employment costs}_{i,t}}{\text{employment}_{i,t}}\right),$$

and break it down into an average firm wage term and a misallocation term using the identity given in Equation (2). The change in each of these terms is shown in Figure 1b. This figure shows that

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

[5] 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.



(a) Labor productivity

(b) Wages

Figure 1: OP decomposition in the aftermath of the Great Recession.

worker reallocation across high- and low-paying firms is also a channel accounting for the overall decrease in wages in the aftermath of the Great Recession, mirroring the labor productivity pattern.

Micro-level changes in worker reallocation. At the firm level, this aggregate pattern translates into a lower association between firm labor productivity or wages and their subsequent employment growth at the firm level. Table 1 shows regressions of the form

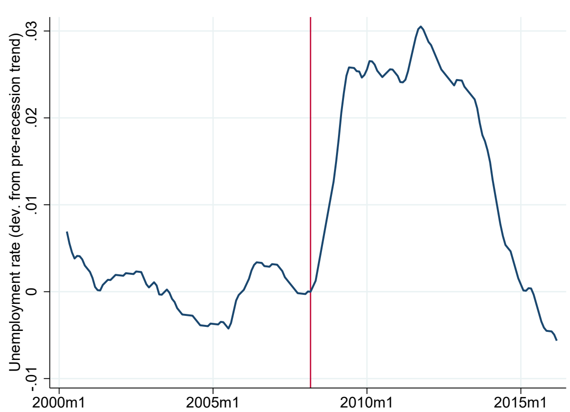
$$\Delta \ln n_{i,t+1} = \alpha y_{i,t} + \beta \left(\text{post}_t \cdot y_{i,t}\right) + \mu_{t,s} + \epsilon_{i,s,t}, \tag{3}$$

where $y_{i,t}$ is either firm labor productivity ($LP_{i,t}$) or wages ($W_{i,t}$), "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 $y_{i,t}$ 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 third post-recession. Consistent with the macro-level evidence, this finding suggests that employment growth is not as productivity enhancing as prior to the recession. A broadly similar pattern holds for wages.

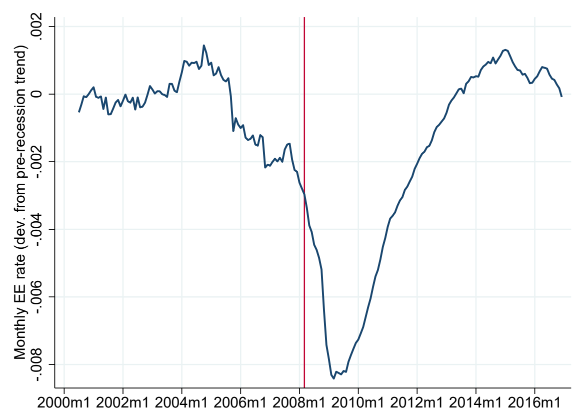
Labor market aggregates. The slow down in worker reallocation documented in Table 1 comes at the same time as some significant changes in the labor market, which are summarized in Figure 2. The unemployment rate rises by about three percentage points after March 2008 (left panel). The rate of direct transitions between employers (EE) is about one-third lower than prior to the start of the episode. The recovery period is then characterized both by a much larger pool of unemployed workers and a drop in, potentially productivity enhancing, employer-to-employer transitions.

	$y_{i,t} = LP_{i,t}$		$y_{i,t} = W_{i,t}$	
	(1)	(2)	(3)	(4)
$y_{i,t}$	0.102	0.102	0.136	0.136
	0.002	0.002	0.003	0.003
$y_{i,t} \cdot \text{post}_t$	-0.031		-0.012	
	0.003		0.003	
$y_{i,t} \cdot \text{recession}_t$		-0.039		-0.007
		0.004		0.006
$y_{i,t} \cdot \text{recovery}_t$		-0.029		-0.013
		0.003		0.003
Year-Industry FEs	Yes	Yes	Yes	Yes
Size Control ($\ln n_{i,t}$)	Yes	Yes	Yes	Yes
N	504,785	504,785	467,793	467,793

Table 1: Reallocation during the Great Recession at the firm level. The specification is given in Equation (3). The dependent variable is the (log-) change in employment in the next period, conditional on survival. "post" is an indicator for all years after 2007, "recession" is an indicator for 2008-2009, and "recovery" for all years after 2009. Robust standard errors.



(a) Unemployment rate



(b) Job-to-job transition rate

Figure 2: Aggregate labor market statistics during the UK Great Recession. Left: monthly unemployment rate (Office for National Statistics). Right: job-to-job monthly transition rate derived from the British Household Panel Survey (Postel-Vinay and Sepahsalari, 2019).

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 2: unemployment, job-to-job transitions, but also reallocation of workers across firms with heterogeneous productivities.

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 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.[6] 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 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 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.[7]

[6]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.

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

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

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

where c is assumed to satisfy $c' > 0,\ c'' > 0$ and $c(0) = 0$. As will become clear when writing down discounted 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 recruitment 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.[8]

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.

[8] 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.

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

$$\Pi_t\left(p_t, n_t, \overline{V}\right) = \max_{\substack{h_{t+1}\geq 0 \\ w_t \\ W_{t+1}}} \Bigg\{ (\omega_t p_t - w_t)n_t + \beta E_t\Big[\chi_{t+1}\Big(-c(h_{t+1})(1-\mu)(1-\delta)n_t + \Pi_{t+1}(p_{t+1}, n_{t+1}, W_{t+1})\Big)\Big]\Bigg\}, \quad (4)$$

where the firm's continuation decision is given by $\chi_{t+1} := \mathbb{1}\Big\{\big(W_{t+1} \geq U_{t+1}\big) \cap \big(\Pi_{t+1} \geq 0\big)\Big\}$, since the firm needs to offer at least U_{t+1} for its workers not to quit and I assume that the firm must make non-negative discounted profits to stay active. Anticipating on the results in Section 4, in equilibrium the firm's continuation decision can be expressed solely in terms of the firm's current idiosyncratic productivity, though this threshold changes with the business cycle.

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

$$n_{t+1} = \Big[1 - \underbrace{q_{t+1}(W_{t+1})}_{\text{quit rate}} + \underbrace{h_{t+1}}_{\text{hiring rate}}\Big] \underbrace{(1-\mu)(1-\delta)n_t}_{\text{remaining workers at search stage}}. \quad (5)$$

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.[9] Equation (5) 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.

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

[9] I define $\bar{F}_t := 1 - F_t$. Note that conditional on $q_{t+1}(W_{t+1})$ and h_{t+1} Equation (5) holds exactly by a Law of Large Number argument since n_t is the measure of workers employed at the firm.

the right-hand side of (4) and using the law of motion for employment, it can be shown that

$$\pi_t\left(p_t,\overline{V}\right) = \max_{\substack{h_{t+1}\geq 0\\ w_t\\ W_{t+1}}} \Bigg\{(\omega_t p_t - w_t)$$

$$+\beta E_t\Big[(1-\mu)(1-\delta)\chi_{t+1}\Big(-c(h_{t+1})+(1-q(W_{t+1})+h_{t+1})\pi_{t+1}(p_{t+1},W_{t+1})\Big)\Big]\Bigg\}, \quad (6)$$

still subject to the Promise-Keeping constraint. See Appendix A.1.1.

While this property is related to the result in Coles and Mortensen (2016), it is obtained under very different assumptions on the wage-setting protocol. I assume that firms can commit to a full wage schedule after each realization of the aggregate state, which is reflected in the Promise-Keeping constraint. They, by contrast, assume that workers form beliefs on the firm's productivity based on its offered wage. I give further details on the difference implied by these two wage-setting protocol in Appendix A.2.

It follows directly from Equation (6) 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 in the event 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 5.

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.[10] They then decide to enter if $U_t \leq S_t(p_0)$ for some

[10]Note that given the value of a firm is linear in n_t and given the equal treatment constraint, the surplus of the

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.[11] These firms then move on to the production stage, and become incumbent firms from the next period onward.

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

$$U_t = b + \beta E_t \Big\{ \mu Q_{t+1} + (1-\mu)\Big[(1-\lambda_{t+1})U_{t+1} + \lambda_{t+1} \int \max\big(W', U_{t+1}\big) dF_{t+1}(W')\Big]\Big\}. \tag{7}$$

Similarly to unemployed workers, employees are hit with probability μ 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

$$\begin{aligned} W_t = w_t + \beta E_t \Big\{ \mu Q_{t+1} + (1-\mu)\Big[\Big((1-\chi_{t+1}) + \delta\chi_{t+1}\Big)U_{t+1} \\ + \chi_{t+1}(1-\delta)\Big(\big(1-q_{t+1}(W_{t+1})\big)W_{t+1} + s\lambda_{t+1}\int \max\big(W', U_{t+1}\big)\, dF_{t+1}(W')\Big)\Big]\Big\}. \end{aligned} \tag{8}$$

firm and all of its workers is simply $\Pi_t(p_t, n_t, \overline{V}) + n_t\overline{V} = n_t S_t(p_t)$.

[11] 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."

3.5 Joint firm-worker surplus

The firm and worker problems can be summarized in a single expression combining Equations (6) and (8). I show in Appendix A.1.2 that the following expression for $S_t := \pi_t(p_t, \overline{V}) + \overline{V}$ can be obtained after rearranging these two equations

$$S_t(p) = p_t\omega_t + \beta E_t\Bigg\{\mu Q_{t+1} + (1-\mu)\Big[(1-\chi_{t+1}(p_{t+1}))U_{t+1} + \chi_{t+1}(p_{t+1})\Big(\delta U_{t+1} + (1-\delta)\psi_{t+1}(p_{t+1})\Big)\Big]\Bigg\}. \quad (9)$$

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

$$\psi_{t+1}(p_{t+1}) := \max_{\substack{h_{t+1}\geq 0 \\ W_{t+1}}}\Bigg\{-c(h_{t+1}) + (1-q_{t+1}(W_{t+1}))S_{t+1}(p_{t+1}) + h_{t+1}(S_{t+1}(p_{t+1}) - W_{t+1}) + s\lambda_{t+1}\int_{W_{t+1}}^{\infty} W'dF_{t+1}(W')\Bigg\}. \quad (10)$$

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 (10). Importantly, the resulting contract fully internalizes the Promise-Keeping constraint.

3.6 Wages

While wages do not appear from the firm-worker surplus, they are pinned down by 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 (10), the offered wage, w_t, is implicitly defined by

$$\overline{V} = w_t + \beta E_t\Bigg\{\mu Q_{t+1} + (1-\mu)\Big[\Big((1-\chi_{t+1}) + \delta\chi_{t+1}\Big)U_{t+1} + \chi_{t+1}(1-\delta)\Big(\big(1-q_{t+1}(V_{t+1})\big)V_{t+1} + s\lambda_{t+1}\int \max\big(V', U_{t+1}\big)\,dF_{t+1}(V')\Big)\Big]\Bigg\},$$

where V_{t+1} denotes the (state-contingent) solution to the joint-surplus maximization problem in Equation (10). Intuitively, given V_{t+1}, wages adjust so that the Promise-Keeping constraint is satisfied.

3.7 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,

$$Z_t := (1-\mu)\left[u_t + (1-\delta)s\int \chi_t(p)nd\nu_t\right], \tag{11}$$

where the unemployment rate is $u_t := 1 - \int nd\nu_t$. 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

$$A_t := \int \chi_t(p)a_t(p,n)d\nu_t. \tag{12}$$

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)

$$F_t(W) := A_t^{-1}\int \mathbb{1}\{W_t(p) \leq W\}\chi_t(p)a_t(p,n)d\nu_t. \tag{13}$$

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

$$h_t(p) \underbrace{(1-\mu)(1-\delta)n}_{\text{remaining workers at search stage}} = \underbrace{a_t(p,n)}_{\text{vacancies}} \underbrace{\eta_t}_{\text{contact rate}} \underbrace{Y_t(W_t(p)),}_{\text{acceptance rate}} \tag{14}$$

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.[12]

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

[12] I provide a full expression for $Y_t(W_t(p))$ in Appendix A.1.3.

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 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.

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 (6) 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* (11)-(14) *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* (10)*. The continuation decision is given by* $\chi(p, \omega, \nu) = \mathbb{1}\{V(p, \omega, \nu) \geq U(\omega, \nu)\}$*;*

3. S *and* U *solve, respectively,* (9) *and* (7)*.*

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.*[13]

[13] $\Gamma(.|p') \leq \Gamma(.|p)$ for $p' > p$ with strict inequality for some productivity level.

Then:

1. *The firm-worker surplus defined by Equation* (9) *is increasing in p;*
2. *Any equilibrium is Rank-Monotonic if* $\forall h \geq 0$,

$$\frac{c''(h)h}{c'(h)} \geq 1.$$

The proof is in Appendix A.1.4. 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 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.

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 A.1.5 shows that the optimal contract takes the following form

$$V(p,\omega,L) = \frac{uU(\omega,L) + s(1-\delta)\int_{p_E}^{p} S(\tilde{p},\omega,L)dL(\tilde{p})}{u + s(1-\delta)\Big(L(p) - L(p_E)\Big)}. \tag{15}$$

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 (10)

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

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

$$\lambda_t F_t\left(V(p)\right) = \int_{p_E}^{p} \frac{h_t(\tilde{p})}{u_t + s(1-\delta)\Big(L_t(\tilde{p}) - L_t(p_E)\Big)} dL_t(\tilde{p}). \tag{16}$$

The derivations can be found in Appendix A.1.5.

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

$$L_t^P(p) = \mu \int_{p_E}^{p} \chi_t(\tilde{p})d\Gamma_0(\tilde{p}) + (1-\mu)\chi_t(p)\Big[L_t(p)\rho_t(V_t(p)) + u_t\lambda_t F_t(V_t(p))\Big], \tag{17}$$

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.[14] 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.

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 its optimal 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* (17) *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 6.

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 $\epsilon^p_{t+1} \sim \mathcal{N}(0, 1)$. 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 condition in Result 1 provided $c_2 \geq 2$. I enforce this condition when searching over the parameter space. Taken together, these functional form assumptions yield the following vector of parameters to calibrate: $\big(\beta, \delta, c_1, c_2, s, \mu, b, \rho_p, \sigma_p, \mu_0, \sigma_0\big)$.

[14] The firm's policies can also be expressed as a function of the net surplus $\phi(p, \omega, L) := S(p, \omega, L) - U(\omega, L)$, which I do in practice when simulating the model. To economize on notation, all the corresponding expressions are relegated to Appendix A.1.6.

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) (log value added 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, an entry/exit productivity threshold, and a monthly hiring and quit rate for surviving firms at each productivity level. 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 $\mu(1 - \Gamma_0(p_E)) + (1 - \mu)\lambda$, the chance for a prospective entrepreneur to be successful at starting a business and for a searching worker to get an offer.

The moments relating to firm dynamics are derived from yearly data, and their model counterpart is obtained by simulating a panel of firms. I simulate a cohort of 150,000 entrants (roughly the size of a typical cohort of entrants in the British data) for forty-five years and aggregate the output from that simulation exactly like the data. Monthly value added 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 concept in the firm-level data and compute the labor productivity measure defined in Equation (1). Since this last productivity measure is in logs, the actual units of value-added are irrelevant to my calibration.

The model fit to the targeted moments is shown in Table 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.

The estimated parameters are listed in Table 3. The estimated job destruction rate is low, since the bulk of EU transitions come from firm exit in the model.[15] There is no clear benchmark in

[15] A potential strategy to further discipline this feature would be to get an estimate of the fraction of EU transitions

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 2: Targeted moments.

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 the economy, which is about half the value used in Shimer (2005).[16] 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.

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 3 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 repli-

coming from firm exit. This information is not readily available in UK data.

[16]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 ($\approx$5% annual)	0.996
Estimated		
δ	prob. job destruction ($\times$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 ($\times$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 3: Parameter estimates.

cate 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 developed by the literature on the emergence of power law distribution 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.[17] First, as shown in Section 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.

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 (Definition 2). This extra state

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

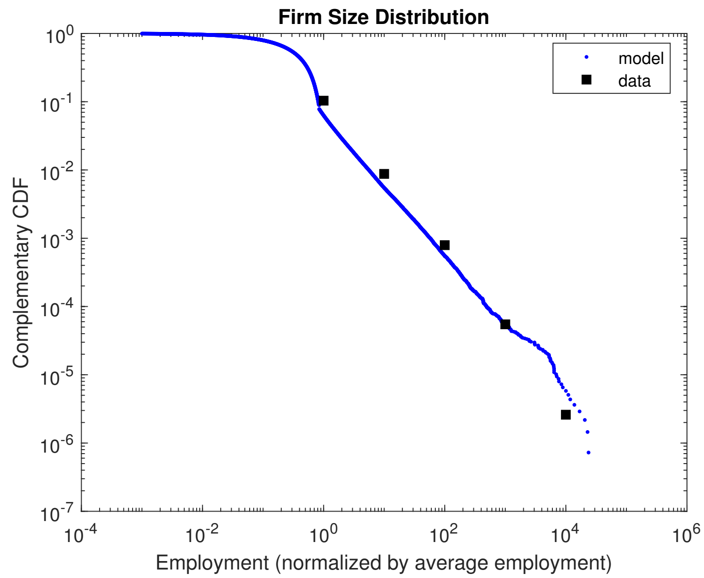


Figure 3: Firm size distribution at the calibrated parameters. Normalized employment is defined as employment at the firm divided by average firm employment in the economy. The complementary CDF at firm employment n is given by $\Pr(\text{employment}_{i,t} > n)$. The data series comes from the BSD and 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.

variable presents a technical difficulty 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. With the model solution in hand, I then proceed with a series of exercises highlighting the interplay of firm dynamics and search frictions in accounting for labor productivity changes following the Great Recession in the UK.

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 statistics. It 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 $L_t / \int dL_t(p)$.[18]

Given this 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.[19] For instance, the value function for workers in unemployment is approximated as

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

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

[19]I approximate the value functions and not the firm's policies directly since the latter are not smooth functions of the aggregate states due to the entry/exit threshold.

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

Table 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.

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 firm-productivity (p) node. An advantage of this modification of the Krusell and Smith (1998) procedure, which would posit an aggregate law of motion for the moments in $\boldsymbol{m}_t$, is that there is no need to specify a grid for them. They are generated directly as part of the simulation procedure. The solution algorithm proceeds by repeatedly simulating the model and regressing the value functions on the aggregate state vector until the coefficients converge. Additional details regarding the implementation of this algorithm, including the choice of the number of moments to be included in $\boldsymbol{m}_t$, can be found in Appendix B.2.3.

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 B.2.5.

To close the description of the numerical solution of the model with aggregate shocks, I make an assumption on the Markov process for aggregate productivity. These shocks are 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.[20] They are shown in Table 4.

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 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 stress that this aspect of the data is not targeted in the the calibration procedure, which does not target business

[20]I target fluctuations in the unemployment rate directly since, as is well understood in the literature, this type of model does not generate a lot of amplification from TFP shocks to the unemployment rate (Shimer, 2005; Hagedorn and Manovskii, 2008). The emphasis of the present paper is on the implications of labor market flows for the allocation of workers to firms.

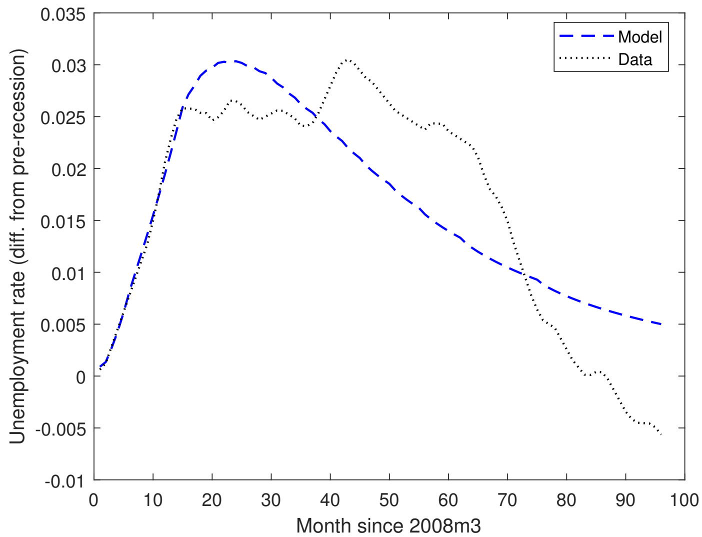


Figure 4: Simulated and actual unemployment rate during the Great Recession.

cycles statistics and centers on pre-recession features of 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 labor market response. Figure 4 depicts the Great Recession experiment I run in the model. I input a sequence of aggregate shocks designed to replicate the sharp increase in unemployment observed during the recession period (2008m3–2009m6). Aggregate productivity is then left to revert back to its steady-state level according to the persistence parameter given in Table 4.[21]

Table 5 further reports the peak to trough response (difference between the value in 2008m3 and 2009m6) of several labor market aggregates in the model and in the data. I consider the response in the various labor market transition rates, as well as in the number of (log) vacancies.[22] Though the drop in the EE rate appears smaller than in the data, the model does a good job at replicating the drop in these aggregates, keeping in mind that it is calibrated on the pre-recession period and based on its stationary solution.

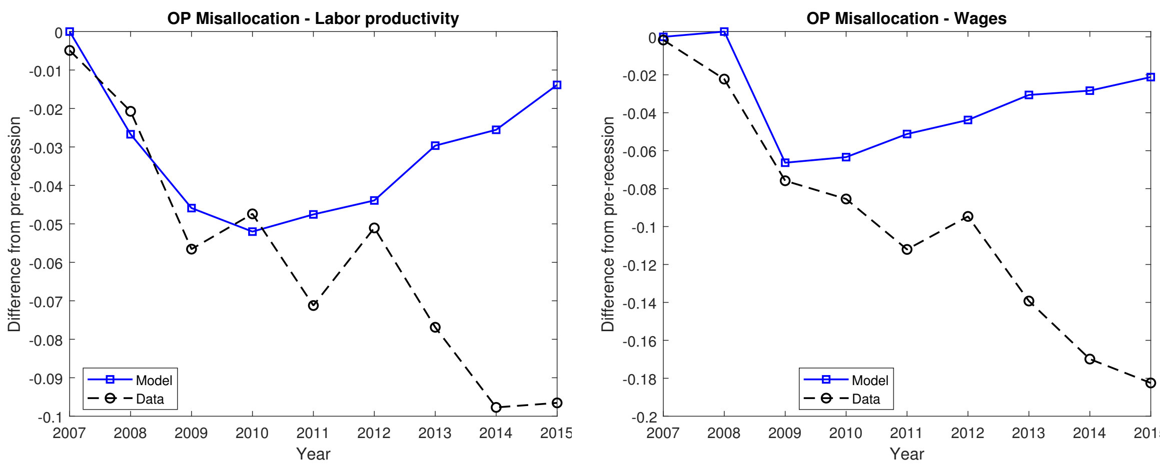
Macro-level worker reallocation. I study the reallocation of workers implied by the model in the simulated recession. I use the OP misallocation measure introduced in the empirical part of the paper to quantify changes in the allocation of workers to firms. This measure is given by $\sum_i \left(ES_{i,t} - \overline{ES}_t\right)\left(LP_{i,t} - \overline{LP}_t\right)$, which captures to which extent firms with a higher than average

[21]Figure 14 in Appendix B.2.6 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.

[22]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.

	Peak – Trough	
Aggregate	Model	Data
UE	0.011	0.009
EU	-0.002	-0.001
EE	0.002	0.005
ln Vac.	0.495	0.489

Table 5: Peak to trough response in labor market aggregates.



(a) Labor productivity **(b)** Employment costs

Figure 5: 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.

labor productivity also account for a higher than average employment share. Figure 5 (left panel) 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 similar in magnitude to that observed in the data. I also report results for the firm-level measure of wages, the log of employment costs per worker, defined in Section 2 (right panel). The model also implies a drop in the misallocation component of wages at the onset of the recession, though it is less persistent than in the data. By construction, the simulated recession cannot speak to the prolonged reduction in these measures observed towards the end of the sample period, since unemployment is left to return to its pre-recession steady-state in the model. This pattern suggests that, while the onset of the recession triggers the initial reduction, there are potentially some additional changes in the structure of the UK economy that account for the reduction observed in the medium term.

Micro-level worker reallocation. Table 6 further reports firm-level regressions similar to the ones shown in the empirical part of the paper. These regressions are given by

$$\Delta \ln n_{i,t+1} = \alpha y_{i,t} + \beta \left(\text{post}_t \cdot y_{i,t}\right) + \mu_{t,s} + \epsilon_{i,s,t},$$

	$y_{it} = LP_{it}$		$y_{it} = W_{it}$	
	Model	Data	Model	Data
y_{it}	0.318	0.102	0.338	0.136
$y_{it} \cdot \text{post}_t$	-0.031	-0.031	-0.028	-0.012

Table 6: Change in firm-level employment growth during the simulated recession.

where $y_{i,t}$ denotes either labor productivity $LP_{i,t}$ or wages $W_{i,t}$. Table 6 shows that the model can emulate the weakening of the relationship between the firm's current productivity (wages) and its subsequent growth found in the data following the Great Recession. This change is also not directly targeted in the calibration.

The OP decomposition through the lens of the model. Recall that the labor productivity index used in the empirical part of the paper is given by $LP_t = \sum_i ES_{i,t} \cdot LP_{i,t}$, 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

$$LP_t = \int \ln\left(\frac{\omega_t p n}{n}\right) dn\overline{\nu}_t^P(p,n) = \ln\omega_t + \int \ln(p) d\overline{L}_t^P(p),$$

where the superscript "P" denotes the production stage (end of period) and a bar denotes a normalized measure.[23] 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 component and a worker reallocation component. The equality

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

can be written

$$LP_t = \underbrace{\ln\omega_t}_{\text{aggregate shock}} + \underbrace{\int \ln(p) dK_t^P(p)}_{\text{firm selection}} + \underbrace{\int \ln(p) dL_t^P(p) - \int \ln(p) dK_t^P(p)}_{OP_t:=\text{measure of allocation}} \tag{18}$$

in the notation of the model. 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, while 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 evolving over the business cycle

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

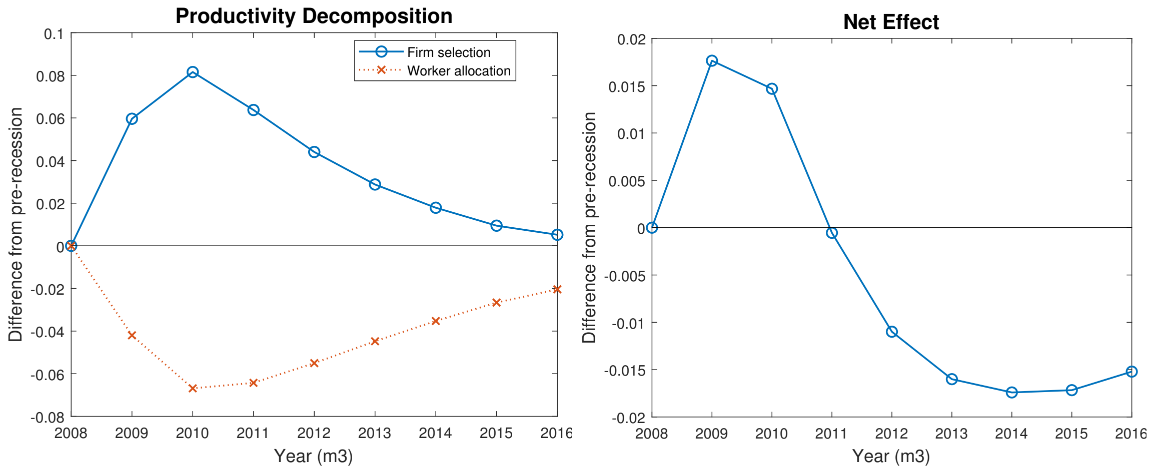


Figure 6: Labor productivity decomposition in the model.

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 18). I plot the evolution of these two components over the course of the simulated recession in Figure 6. It shows that they represent forces moving in opposite directions as they shape aggregate 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 6. It therefore acts as a negative force on labor productivity in the medium term.

Inspecting the worker reallocation mechanism. I illustrate the main worker reallocation mechanism at the micro level in Figure 7. On top of the firm selection effect, which shifts the entry threshold upward, how well labor is allocated to firms also depends on which firms grow faster following the shock. Figure 7 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 6, I find that the model can account for the fact that the relationship between firm-level labor productivity and employment growth becomes less positive in the aftermath of the recession. While this relationship

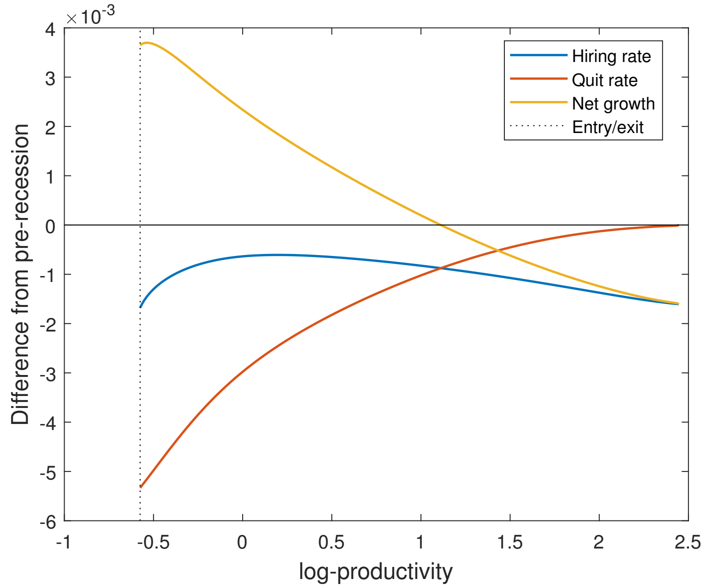


Figure 7: Firm policies during the simulated recessionary episode. Policies are averaged across simulation periods and presented in deviation from the stationary equilibrium. Idiosyncratic productivity is truncated to only include firms above the entry/exit threshold at all point in time after the shock.

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 transitions out of the bottom rung of the wage ladder—where firms are ranked based on wages and not productivity as in Figure 7—decline by eighty-five percent during the US Great Recession.

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.[24] Specifically, the value of non-employment is assumed to vary with unemployment benefits according to

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

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.

[24]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.

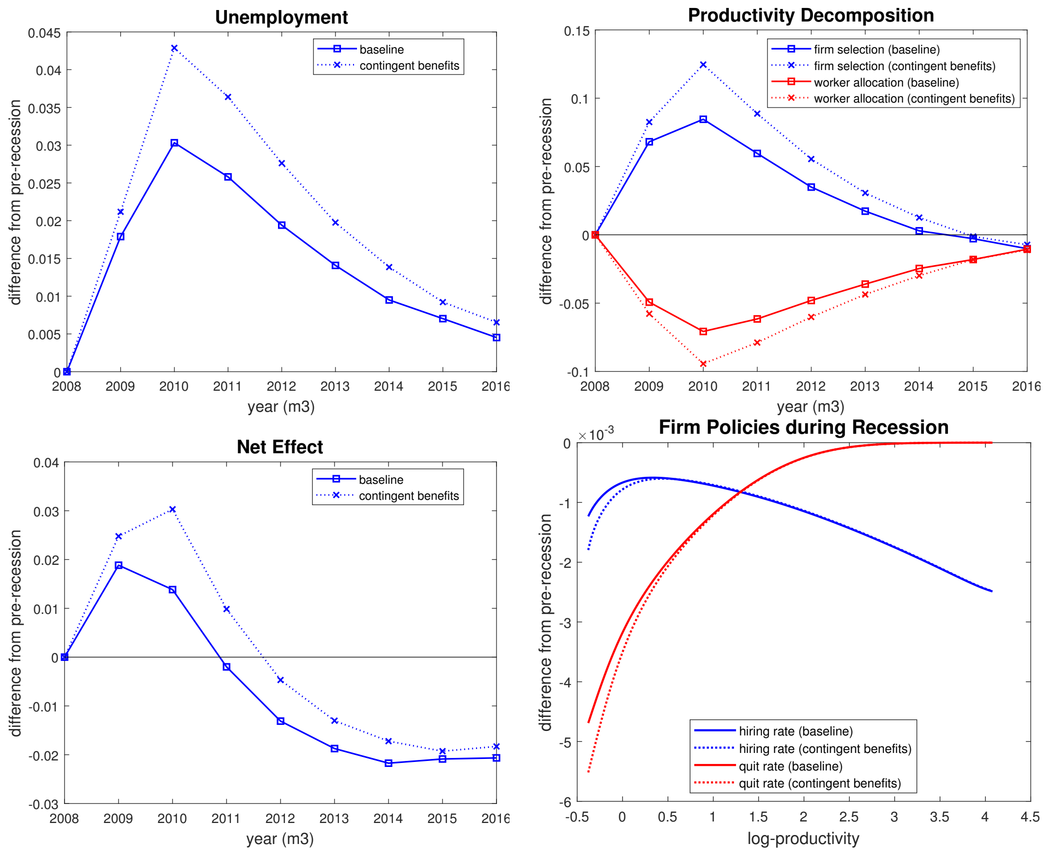


Figure 8: Baseline vs Unemployment-contingent benefits model. Model response to the simulated recession using the same sequence of aggregate shocks.

Figure 8 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.

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.

References

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

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013). Cross-country differences in productivity: The role of allocation and selection. *American Economic Review 103*(1), 305–34.

Bilal, A., N. Engbom, S. Mongey, and G. L. Violante (2019). Firm and worker dynamics in a frictional labor market.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 257–273.

Caballero, R. J. and M. L. Hammour (1994). The cleansing effect of recessions. *The American Economic Review*, 1350–1368.

Clementi, G. L. and B. Palazzo (2016). Entry, exit, firm dynamics, and aggregate fluctuations. *American Economic Journal: Macroeconomics 8*(3), 1–41.

Coles, M. G. and D. T. Mortensen (2016). Equilibrium labor turnover, firm growth, and unemployment. *Econometrica 84*(1), 347–363.

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

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

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

Elsby, M. W. and R. Michaels (2013). Marginal jobs, heterogeneous firms, and unemployment flows. *American Economic Journal: Macroeconomics 5*(1), 1–48.

Elsby, M. W. L. and A. Gottfries (2019). Firm Dynamics, On-the-Job Search and Labor Market Fluctuations. Working Paper.

Engbom, N. (2019). Firm and worker dynamics in an aging labor market. Technical report.

Fort, T. C., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2013). How firms respond to business cycles: The role of firm age and firm size. *IMF Economic Review 61*(3), 520–559.

Foster, L., C. Grim, and J. Haltiwanger (2016). Reallocation in the great recession: cleansing or not? *Journal of Labor Economics 34*(S1), S293–S331.

Foster, L., J. C. Haltiwanger, and C. J. Krizan (2001). Aggregate productivity growth: lessons from microeconomic evidence. In *New developments in productivity analysis*, pp. 303–372. University of Chicago Press.

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

Gavazza, A., S. Mongey, and G. L. Violante (2018). Aggregate recruiting intensity. *American Economic Review 108*(8), 2088–2127.

Gouin-Bonenfant, E. (2019). Productivity dispersion, between-firm competition, and the labor share. Technical report.

Griliches, Z. and H. Regev (1995). Firm productivity in israeli industry 1979–1988. *Journal of econometrics 65*(1), 175–203.

Gu, R. (2019). Specific capital, optimal firm insurance, and the dynamics of the postgraduate wage premium. Technical report.

Hagedorn, M. and I. Manovskii (2008). The cyclical behavior of equilibrium unemployment and vacancies revisited. *American Economic Review 98*(4), 1692–1706.

Haltiwanger, J. C., H. R. Hyatt, L. B. Kahn, and E. McEntarfer (2018). Cyclical job ladders by firm size and firm wage. *American Economic Journal: Macroeconomics 10*(2), 52–85.

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

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics 124*(4), 1403–1448.

Kaas, L. and P. Kircher (2015, Oct). Efficient firm dynamics in a frictional labor market †. *American Economic Review 105*(10), 3030–3060.

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

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

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

Lise, J. and J.-M. Robin (2017). The macrodynamics of sorting between workers and firms. *The American Economic Review 107*(4), 1104–1135.

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

Mortensen, D. T. and C. A. Pissarides (1994). Job creation and job destruction in the theory of unemployment. *The review of economic studies 61*(3), 397–415.

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

Moscarini, G. and F. Postel-Vinay (2013, October). Stochastic Search Equilibrium. *The Review of Economic Studies 80*(4), 1545–1581.

Moscarini, G. and F. Postel-Vinay (2016). Wage posting and business cycles: A quantitative exploration. *Review of Economic Dynamics 19*, 135–160.

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

Office for National Statistics (2020a). Annual Business Survey, 2008-2018: Secure Access. http://doi.org/10.5255/UKDA-SN-7451-13. 13th Edition. UK Data Service. SN:7451.

Office for National Statistics (2020b). Annual Respondents Database, 1973-2008: Secure Access. http://doi.org/10.5255/UKDA-SN-6644-5. 3rd Edition. UK Data Service. SN:6644.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

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

Postel-Vinay, F. and A. Sepahsalari (2019). Labour mobility and earnings in the uk, 1992-2016. Technical report.

Reed, W. J. (2001). The pareto, zipf and other power laws. *Economics letters 74*(1), 15–19.

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

Riley, R., C. Rosazza Bondibene, and G. Young (2015). The UK productivity puzzle 2008-2013: evidence from British businesses. *Bank of England Staff Working Paper* (531).

Rujiwattanapong, W. S. (2019). Unemployment dynamics and endogenous unemployment insurance extensions.

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

Sedláček, P. and V. Sterk (2017). The growth potential of startups over the business cycle. *American Economic Review 107*(10), 3182–3210.

Sepahsalari, A. (2016). Financial market imperfections and labour market outcomes.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American economic review*, 25–49.

Shimer, R. (2010). *Labor markets and business cycles.* Princeton, N.J: Princeton University Press.

Winberry, T. (2016). Lumpy investment, business cycles, and stimulus policy. *Job market paper, Princeton University/University of Chicago Booth School of Business*.

A Appendix

A.1 Derivations and proofs

A.1.1 Size-independent Discounted Profits

We want to guess and verify that a solution to the functional equation (4) has the form $n_t \pi_t(p_t, \overline{V})$. That is, we want to show that

$$\Pi_{t+1}(p_{t+1}, n_{t+1}, W_{t+1}) = n_{t+1}\pi_{t+1}(p_{t+1}, W_{t+1}) \implies \Pi_t(p_t, n_t, \overline{V}) = n_t \pi_t(p_t, \overline{V}).$$

Start from (4), still subject to the Promise-Keeping constraint and the law of motion for its workforce. Plugging in the guess in (4) gives

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

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 (6).

A.1.2 Firm-worker 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

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

Substituting w_t in the expression for firm profit per worker (6) gives

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

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

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

A.1.3 Definition Acceptance Rate

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

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

The acceptance rate at some offered W is then given by

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

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.

A.1.4 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, $\frac{c''(h)h}{c'(h)} \geq 1, \forall h \geq 0$ is sufficient to guarantee that V is increasing in p;
2. The firm-worker surplus mapping defined by (9) 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 (9) defines the optimal contract and hiring rate after all histories. At any interior maximum, the following first-order conditions are associated with (10)

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

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 (9) 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 (9) maps from the space of differentiable, bounded and increasing functions into itself. Differentiability in p follows directly from noting that the expectation in (9) 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 (9) 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 (10) gives

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

where V^*, h^* denote optimal policies. The term inside the expectation in the firm-worker surplus (9) 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

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

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

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

and

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

Finally, summing up the last two inequalities, we get

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

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

A.1.5 RME Contracts

This Appendix proves that the RME contract has the form given in (15). Before turning to the actual proof, I first show that the contract offer distribution, F_t, rewrites

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

where the substitution follows from the firm's vacancy posting position (14) and the equality $\eta_t A_t = \lambda_t Z_t$. Besides, in a RME, contracts are strictly increasing in p, so we have

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

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 (16)

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

To get (15), start from the first-order condition with respect to the optimal contract from (10) 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 (16) 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 (15) in the main text

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

A.1.6 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 (9), 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

$$\phi_t(p) := S_t(p) - U_t = p_t\omega_t - b + \beta(1-\mu)E_t\Bigg[\chi_{t+1}(p_{t+1})\Bigg\{\tilde{\psi}_{t+1}(p) - \lambda_{t+1}\int_0^{\infty} \theta d\tilde{F}_{t+1}(\theta)\Bigg\}\Bigg] \quad (19)$$

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

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

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 (15), which gives

$$V(p,\omega,L) - U(\omega,L) := \tilde{V}(p,\omega,L) = \frac{s(1-\delta)\int_{p_E}^{p} \phi(\hat{p},\omega,L)dL(\hat{p})}{u + s(1-\delta)\Big(L(p) - L(p_E)\Big)}, \tag{20}$$

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 $\chi(p,\omega,L) = \mathbb{1}\{\phi(p,\omega,L) \geq 0\}$.

A.2 Difference with Coles and Mortensen (2016)

The key 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 are precisely defined in the main text. 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,

$$\pi_t\left(p_t, \overline{V}\right) = \max_{\substack{h\geq 0\\ w\\ W}} \Bigg\{ \omega_t p_t - w + \beta E_t\Big[-c(h) + (1 - q_{t+1}(W) + h)\pi_{t+1}(p_{t+1}, W)\Big]\Bigg\}, \tag{21}$$

subject to the promise-keeping constraint

$$\overline{V} = w + \beta E_t\Big\{\delta U_{t+1} + (1-\delta)\Big[(1 - q_{t+1}(W))W + s\lambda_{t+1}\int \max\left(W', U_{t+1}\right) dF_{t+1}(W')\Big]\Big\}. \tag{22}$$

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 (21) using (22), and the optimal contract can be shown to be increasing in productivity and expressed as a function of the firm-worker surplus, as described in the main text.

The discrete time equivalent to (21) in Coles and Mortensen (2016) is given by

$$\pi_t\left(p_t, w_t\right) = \omega_t p_t - w_t + \beta E_t \max_{\substack{w \\ h \geq 0}} \left\{ -c(h) + (1 - q_{t+1}(w) + h)\pi_{t+1}(p_{t+1}, w) \right\}, \tag{23}$$

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.[25] 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 (21) and (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, by contrast, 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 9 shows the wage profile obtained by simulating a version of the model under each wage-setting protocol. I use a different calibration than in the main text 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.

[25] 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.

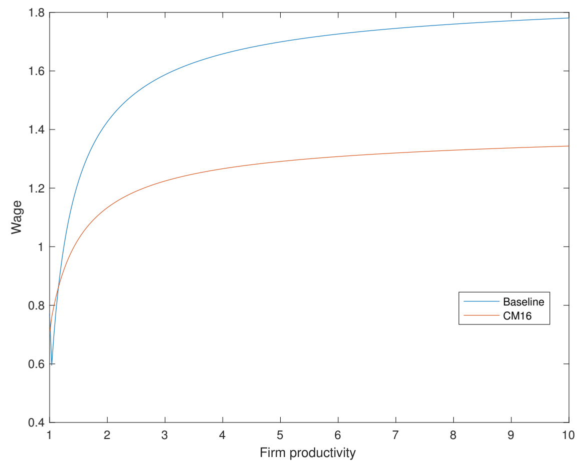


Figure 9: Wages in Baseline vs Coles and Mortensen (2016) model at the steady-state.

B Online appendix

B.1 Data

B.1.1 Firm-level data

I use three sources of firm-level data: the Annual Respondents Database (Office for National Statistics, 2020b), the Annual Business Survey (Office for National Statistics, 2020a), and the Business Structure Database (Office for National Statistics, 2019). The Annual Respondents Database (ARD) and the Annual Business Survey (ABS) are the same survey changing name over time. This survey requests detailed balance sheet information from the universe of large firms (with more than 250 employees) and a stratified random sample of smaller businesses (with less than 250 employees). The Business Structure Database is a snapshot from the registry of all British businesses (for all businesses paying VAT), but with only a handful of variables (employment, turnover, industry).

Since the Business Structure Database (BSD) does not have information on value added or employment costs, I follow the procedure in Riley et al. (2015)to obtain meaningful aggregates from the Annual Respondents Database (ARD)/Annual Business Survey (ABS). I use the "gross value added at factor costs" and "total employment costs" variables, which are harmonized across survey years by the data provider, as the relevant concepts for a firm's value added and wage bill. I deflate these measures using industry-level deflators provided by the Office for National Statistics. The employment variable is directly taken from the Business Structure Database.

To gross the data, I construct survey weights directly from the Business Structure Database, which represents the (near) universe of private sector employment. I define industry×firm-size cells and use the BSD employment counts as weights for the ARD/ABS. In constructing the analysis sample, I drop a few problematic sectors in the ARS/ABS: farming (A), mining & quarrying (B), energy supply (D), water (E), and real estate (L). All sectors dominated by public employment in the UK (education, health care, and social work) are also excluded. Finally, I also trim the top and bottom two percent of firms in the distribution of labor productivity, $LP_{i,t}$, in each industry×firm-size cell.

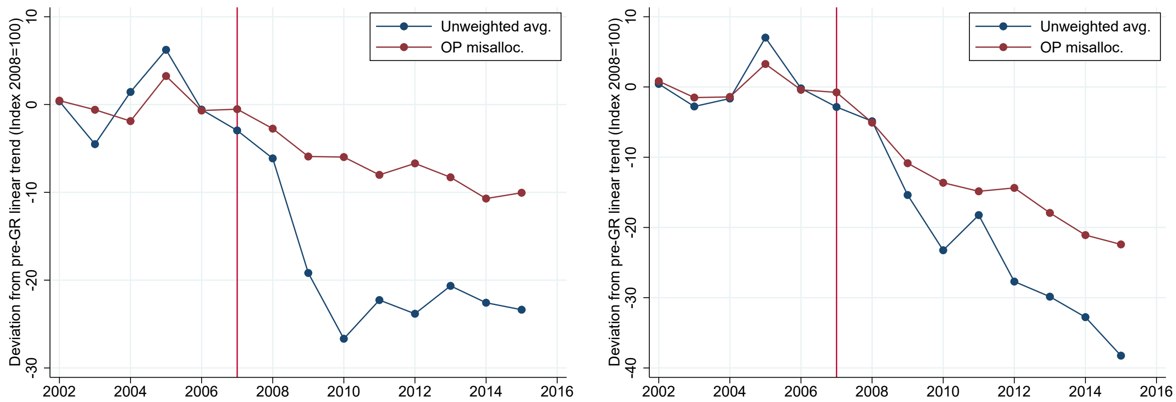
B.1.2 Labor market transitions

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.[26]

B.1.3 Additional macro series

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

[26] 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.



(a) Labor productivity

(b) Employment cost per worker

Figure 10: OP decomposition in the aftermath of the Great Recession.

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

B.1.4 Labor productivity and employment costs in levels

In the main text, firm-level labor productivity and wages are defined as

$$LP_{i,t} := \ln\left(\frac{\text{value added}_{i,t}}{\text{employment}_{i,t}}\right), \quad W_{i,t} := \ln\left(\frac{\text{total employment cost}_{i,t}}{\text{employment}_{i,t}}\right).$$

Figure 10 reports the OP decomposition shown in Figure 1, but with these measures in levels instead. Overall a similar pattern is found. The "misallocation" term contributes to lower both aggregate labor productivity and wages per worker, even to a larger extent, using these alternative measures.

B.2 Numerical solution

B.2.1 Stationary solution

As shown in Appendix A.1.6, the firm's policies can be expressed in terms of a single value function, the net surplus given in Equation (19). 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 (20) 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 (20).

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.

B.2.2 Estimation

The parameters are calibrated by targeting the moments listed in Table 2. In practice, I minimize the distance between the model generated moments and their empirical counterpart using the following

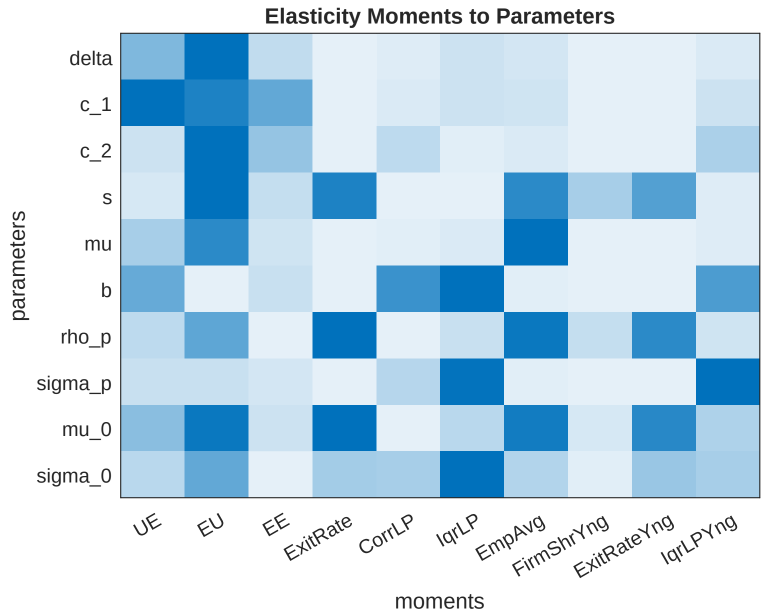


Figure 11: Elasticity of each moment to each parameter. Each cell corresponds to $\left|\frac{d \ln \text{moment}_j}{d \ln \text{parameter}_i}\right|$ for each parameter in row i and each moment in column j. A darker shade of blue indicates a larger absolute elasticity. The elasticities are computed by solving the model in a small neighborhood around the parameters and fitting a line through each parameter-moment series in logs.

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 12 further shows slices of the objective function around the estimated parameter values.

B.2.3 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

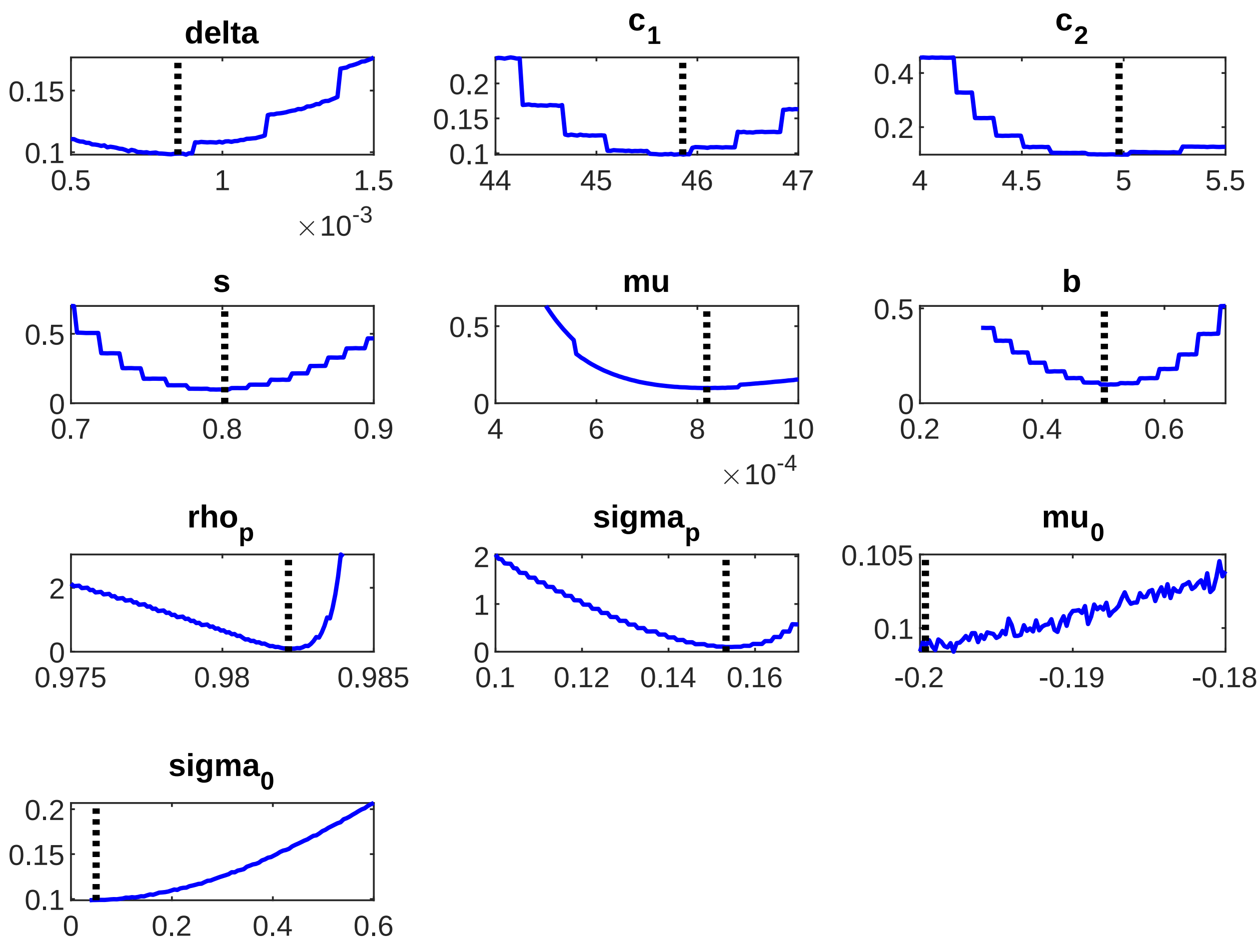


Figure 12: Slices of objective function for each parameter. Vertical dotted line denotes estimated parameter value.

summarized by a set of (un-centered) moments and the unemployment rate

$$\begin{aligned}
m_t^0 &:= u_t = 1 - \int_{\underline{p}}^{\overline{p}} dL_t(p) \\
m_t^1 &:= \int_{\underline{p}}^{\overline{p}} \ln p d\bar{L}_t(p) \\
m_t^2 &:= \int_{\underline{p}}^{\overline{p}} (\ln p)^2 d\bar{L}_t(p) \\
&\ldots
\end{aligned} \tag{24}$$

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 B.2.4.

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 be 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

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

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 gathering all moments in (24) in log-deviation from steady-state, while $\overline{S}$ and $\overline{U}$ stand for 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 use 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

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

I process similarly to find the coefficients for $\tilde{S}(p_i, \omega_t, \tilde{\boldsymbol{m}}_t; \theta_{p_i})$ at each productivity node p_i.

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.

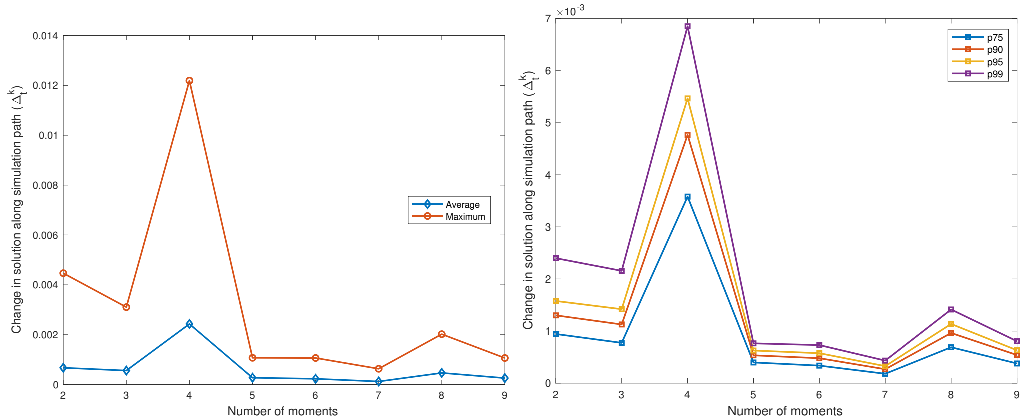
B.2.4 Robustness to number of moments

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 13a reports the average and maximum $\Delta_t^k(p)$ along the simulated sequence of shocks as more moments are included. This 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 13b confirms this pattern by showing several percentiles of $\Delta_t^k(p)$.



(a) Average and Maximum

(b) Percentiles

Figure 13: Robustness to number of moments included in approximation.

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 7: Accuracy Tests

B.2.5 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 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.

B.2.6 Non-linearities in shock size

Figure 14 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

Figure 14: 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.

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 the uncertainty around aggregate shocks matters in determining macroeconomic outcomes.