Firm Dynamics and Random Search over the Business Cycle

Richard Audoly*

This version: December 2020
First version: January 2017

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 British firm-level data 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), Helleveien 30, 5045 Bergen. 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 University College London, Fabien Postel-Vinay and Jan Eeckhout, for their advice and support. For many helpful comments and suggestions, I also wish to thank Wei Cui and Rasmus Lentz, as well as seminar participants at SaM 2017 (Barcelona), the Young Economists Conference 2018 (Belgrade), EEA 2018 (Cologne), the SaM Workshop 2019 (Bristol), and the Dale T. Mortensen conference 2019 (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, Annual Respondents Database, and Annual Business Survey 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 in the presence of search frictions in the labor market.

This framework is motivated by a series of empirical regularities on the evolution of firm and worker flows during recessions. On the firm side, the number of active businesses tends to substantially drop during an economic contraction.[1] On the worker side, recessions both markedly increase flows into unemployment 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.[2] To which extent these changes in firm and worker flows reallocate workers to more productive firms is an open question.[3]

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 because 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. In an application to the Great Recession in the UK, I use the model to assess the quantitative significance of this mechanism in holding back labor productivity. I find that worker reallocation has sizable and persistent negative effects on aggregate labor productivity, more than canceling out the positive effects resulting from firm selection.

My analysis is anchored in firm-level data. I use administrative data for a large sample of British firms over the period 2002-2016 and construct a labor productivity index from the ground up, starting from a measure of labor productivity at the firm level. These data cover several years before and after the Great Recession (officially starting in 2008q2 in the UK), one of the largest post-war economic contractions 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, during, and after the Great Recession. I find that the allocation of workers to firms is significantly downgraded following the Great Recession. Firm-level regressions confirm that the positive relationship between the labor productivity of firms and the growth of their workforce 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.

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

[2] See, for example, Blanchard et al. (1990) for unemployment inflows and outflows and Fujita et al. (2020) for job-to-job transitions.

[3] 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.

A key contribution of this paper is to develop a tractable model of firm and worker reallocation to study these empirical regularities. My framework combines the three following features: aggregate shocks, search frictions, and firm dynamics. Aggregate shocks are a pre-requisite to studying the evolution of labor productivity over the business cycle. Search frictions in the labor market constrain the transition of workers out of unemployment. In the spirit of the random search framework with on-the-job search proposed by Burdett and Mortensen (1998), I also allow workers to search while employed. While it complicates the solution of the model, this addition is central since (i) about half of gross job creation and destruction flows originate in direct employer-to-employer transitions in the data;[4] (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. Firm dynamics, lastly, is key to obtain a model counterpart to productivity at the micro level, which is only measured for firms in the data. This feature also allows the selection of firms to adjust over the business cycle through entry and exit, in line with the volatility of the number of active businesses found in the data at business cycle frequency.

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, by offering lower wages, and preventing their worker from getting poached by other employers, by instead increasing 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. This monotonicity property implies that workers move from less to more productive firms when making direct transitions between employers, since more productive firms pay better in equilibrium.

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 productivity levels sufficient to characterize the firm's optimal choices at every point over the business cycle. I approximate this distribution with a vector 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 how labor productivity and employment evolve at the firm level. While not being targeted directly in the calibration, the model does a 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 degree of employment concentration.

Given the calibrated model, I feed in a sequence of aggregate shocks triggering a sharp and prolonged decrease in the job-finding rate, akin to the UK experience during the Great Recession. Though not directly targeted, the model replicates well the empirical findings on the interaction be-

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

tween firm productivity and worker reallocation during this episode. It tracks closely the reduction in the measure of allocation of workers to firms over the post-recession period found in the data. It is also in line with the regression results showing a weaker association between labor productivity and employment growth at the firm level post-recession.

To understand the forces at play, 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 firm selection and worker reallocation 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 an equally large, but more persistent, negative impact in the simulated recession. On net, this second effect therefore tends to matter more quantitatively in the medium term.

The reason the allocation of workers to firms is downgraded in the model following a negative shock comes from on-the-job search. Firms have two margins to control the rate at which they adjust their workforce: 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 after a negative shock, the rate at which workers quit their job decreases primarily on the lower part of this distribution, as these workers now compete for good-paying 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 after the shock is in line with the empirical finding that the relationship between employment growth and labor productivity is still positive but weaker following the Great Recession in the British 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 US Great Recession. While I cannot investigate this channel directly in the British firm 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 paper contributes to the growing literature that combines firm dynamics with search frictions in the labor market. This literature brings together firm dynamics models in the tradition of Hopenhayn and Rogerson (1993), which maintain the assumption that labor markets clear (Khan and Thomas, 2013; Clementi and Palazzo, 2016; Sedláček and Sterk, 2017), and search and matching models in the tradition of Mortensen and Pissarides (1994), which emphasize firm-worker matches, but with no meaningful way to aggregate these matches into firms. These multi-worker matches are a pre-requisite to jointly study firm-level concepts, such as productivity, and labor market flows, such as transitions in and out of unemployment and between employers. My work adds to the recent papers integrating firm dynamics and search frictions in the labor market by merging three unique features: (i) firm dynamics, (ii) random search with on-the-job search, and (iii) business cycle fluctuations.

Two broad approaches have been followed to combine firm dynamics and search frictions in the existing literature. There is first a series of papers building on the theoretical results in Menzio and Shi (2011) to introduce firm dynamics in an environment where workers can direct their search to specific job offers (Kaas and Kircher, 2015; Schaal, 2017). This approach is highly tractable in the presence of aggregate shocks. Given the appropriate free-entry condition, all distributions vanish from the state-space and the model can be solved numerically using standard methods. While Kaas and Kircher (2015) abstract from on-the-job search, Schaal (2017) presents an elegant model combining firm dynamics, on-the-job search, and aggregate shocks. A limitation of his setting to analyze the reallocation of workers across firms over the business cycle is that free-entry requires firms to be indifferent as to which type of jobs they offer. As a result, hires cannot be further broken down in hires from unemployment and workers poached from other firms, so to which extent workers are reallocated across firm types over the business cycle is not clear in theory.

By contrast, models with on-the-job search cast in a random search environment all feature a characterization of job-to-job transitions in terms of some job ladder, a ranking of which jobs workers find more valuable which makes clear across which firm types workers move when changing employer. My work is most closely related to that second broad approach to combining firm dynamics and search frictions when workers can search on the job. I highlight below how my model differs from several recent studies also set in a random search environment.

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

Coles and Mortensen (2016) develop a model with both firm dynamics and on-the-job search. I rely on their hiring cost technology to establish size-independence in the firm's policies. My paper diverges from theirs in that I use a different wage-setting protocol: I allow firms to fully commit to future wage payments.[5] This departure allows me to relax some of the restrictions they impose on their environment. Firm entry and exit is endogenous in my framework, and I need not restrict search effort to be the same for employed and unemployed workers. I show that allowing

[5]I return to this point when introducing the model in Section 3.

for endogenous firm selection plays an important role in quantifying the reallocation properties of the model.

Lastly, two contemporaneous papers, Elsby and Gottfries (2019) and Bilal et al. (2019), consider similar environments, but relax the constant returns to scale assumption on the firm's production technology. Elsby and Gottfries (2019) show under two wage-setting protocols that the job ladder can be characterized in terms of a single variable, the marginal product of labor. Contrary to my framework, however, they do not allow for firm entry and exit. As a result, all reallocation arises through worker transitions in their setting, in contrast to my framework where firm selection contributes to determining aggregate labor productivity. Bilal et al. (2019) describe a model related to Elsby and Gottfries (2019) that allows for firm entry and exit. They build on the framework introduced in Lentz and Mortensen (2012) to characterize the job ladder in terms of the marginal joint value of a firm and its workers. A potential limitation of this approach is that wages are not defined, contrary to my framework, where there is a natural way to retrieve them and study their evolution over the cycle.

The key difference between my paper and these two important contributions is that their analysis of aggregate shocks is restricted to a comparison of steady-states and transitional dynamics. Here I propose a full solution of the model outside of steady-state, which allows me to study to which extent firm selection and worker reallocation shape aggregate labor productivity over the business cycle. I also depart from these studies in that I thoroughly benchmark the business cycle properties of the model against firm-level data on productivity covering a major recession episode.

Outline. Section 2 documents several facts on worker reallocation 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 empirically the interaction between firm productivity at the micro level, worker reallocation, 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 obtain 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. Importantly, these data cover several years before and after the Great Recession (officially lasting for five quarters between 2008q2 and 2009q2 in the UK) thus allowing to decompose labor productivity before, during, and after the onset of this episode. I return to the data patterns documented here in Section 6, where I study the quantitative properties of the model for the reallocation of workers across the firm productivity distribution.

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 convenient to get a unit-free productivity measure that can readily be compared to the model. I show in Appendix B.4 that the main empirical patterns documented here still 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 around the time of the Great Recession, I decompose LP_t as

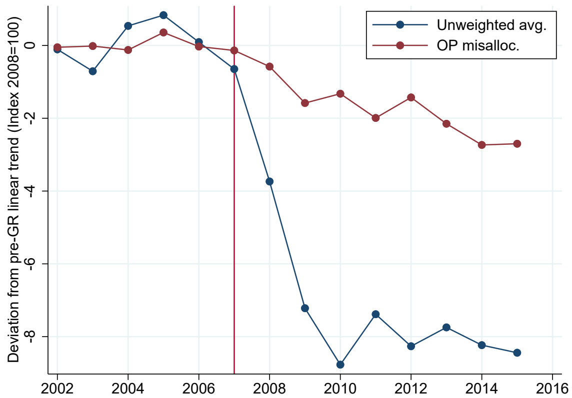
$$LP_t = \sum_i ES_{i,t} \cdot LP_{i,t} = \underbrace{\overline{LP}_t}_{\text{average firm prod.}} + \underbrace{\sum_i \left(ES_{i,t} - \overline{ES}_t\right)\left(LP_{i,t} - \overline{LP}_t\right)}_{\text{OP misallocation measure}}, \tag{2}$$

an equality referred to in the literature as the "OP decomposition" (Olley and Pakes, 1996).[6] 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 to be introduced in subsequent sections.[7]

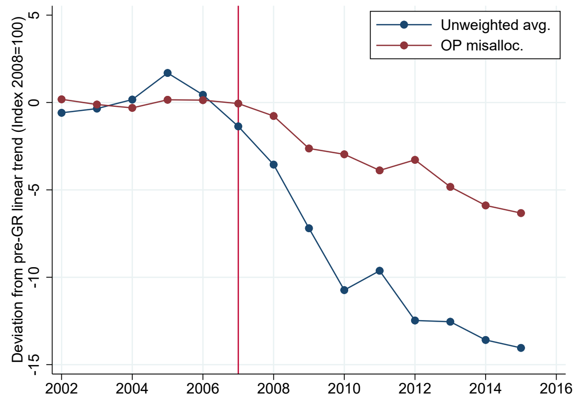
The evolution of each of these terms immediately around the time of the Great Recession is depicted in Figure 1a, expressed in deviation from their respective pre-recession linear trend. Figure 1a shows that both average firm productivity and the allocation of labor to firms have contributed to lower labor productivity growth in the aftermath of the recession. In particular, the OP measure of misallocation keeps moving down after 2008. By the end of the sample period, it represents about a fourth of the overall reduction relative to the pre-recession trend of labor productivity.

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

[7] 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) Wage payments

Figure 1: OP decomposition during the UK Great Recession.

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

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

and break it down into an average firm wage term and a misallocation term using the same identity as in Equation (2). The change in each of these terms is shown in Figure 1b. This figure suggests that 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 firm wage payments and their subsequent employment growth post-recession. Table 1 shows regressions of the form

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

where $y_{i,t}$ is either firm labor productivity ($LP_{i,t}$) or wage payments ($WP_{i,t}$), "post" is an indicator for the years following the Great Recession, and $\mu_{s,t}$ a set of industry-year fixed effects.[8] By design, the sample is restricted to continuing firms, for which employment is observed in the next period. 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 similar pattern holds for wages.

[8]See Foster et al. (2016) for a similar empirical strategy in a US context.

	$y_{i,t} = LP_{i,t}$		$y_{i,t} = WP_{i,t}$	
	(1)	(2)	(3)	(4)
$y_{i,t}$	0.102	0.102	0.136	0.136
	(0.002)	(0.002)	(0.003)	(0.003)
$y_{i,t} \times \text{post}_t$	-0.031		-0.012	
	(0.003)		(0.003)	
$y_{i,t} \times \text{recession}_t$		-0.039		-0.007
		(0.004)		(0.006)
$y_{i,t} \times \text{recovery}_t$		-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 in parenthesis.

Labor market aggregates. The slow down in worker reallocation documented in Table 1 comes at the same time as some significant cyclical changes in the labor market, which are summarized in Figure 2. The unemployment rate rises by about three percentage points after March 2008 (Figure 2a). The job-to-job transition rate is about one-third lower than prior to the start of the episode (Figure 2b). In the labor market, 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.

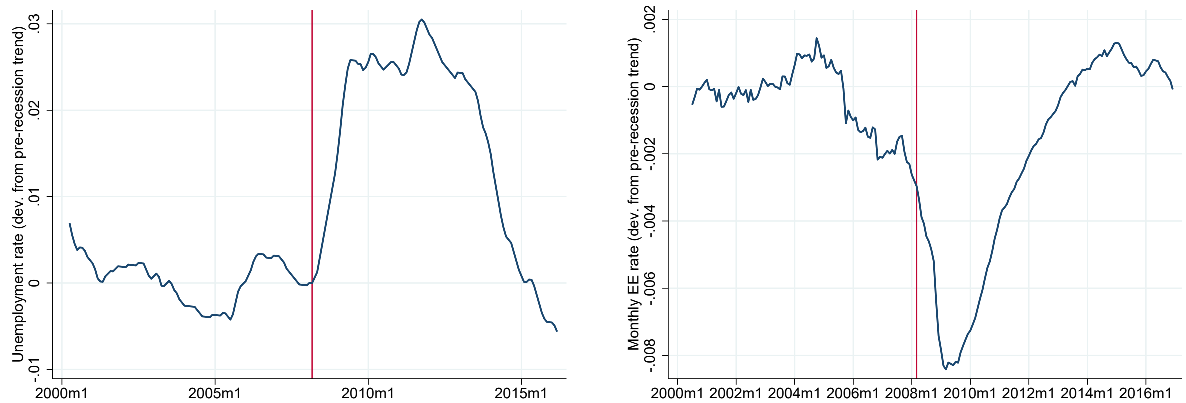
3 A model of firm dynamics with on-the-job search

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 framework to study the various reallocation patterns documented in Section 2: worker flows map into employment changes at 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.

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



(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).

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 stationary first-order Markov process denoted by $\Gamma(p_t|p_{t-1})$. Realizations of p_t are assumed to lie in some positive interval $[\underline{p}, \overline{p}]$.

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 shock. 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.

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, conditional on the information available at the end of period $t-1$. 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 variables, which are 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.[9] 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, which makes the model straightforward to simulate.

Conversely, workers 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 the contacts between workers and vacancies in each period gives

$$\lambda_t Z_t = \eta_t A_t = m(Z_t, A_t) \leq \min\{A_t, Z_t\},$$

where $m(.)$ is a matching function, increasing and concave in each argument, and the inequality ensures that $\lambda_t,\ \eta_t$ are always less than one.

Following Burdett and Mortensen (1998), there is no bargaining: workers draw a take-it-or-leave-it offer from the distribution of offered contracts, F_t, which is endogenously determined in

[9] 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.

equilibrium. Since workers can only accept or decline these offers, their decision boils down to accepting better contracts.[10]

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 hiring H workers has a total recruitment bill of

$$n \cdot c\left(\frac{H}{n}\right) = n \cdot c(h),$$

where c is assumed to satisfy $c' > 0,\ c'' > 0$ and $c(0) = 0$ and I define the gross hiring rate $h := H/n$. As will become clear when writing down discounted profits, such a linear recruitment technology makes the model substantially more tractable, as it can be shown that the firm's policy functions are linear in its employment size n.

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. 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.[11]

Once the firm has chosen its gross hiring rate, h, I assume that the actual vacancies corresponding to these hires are posted at no extra cost. I provide an expression for these vacancies once the notation for the labor market aggregates is introduced, in Section 3.7 below.

Discounted profits. Because the firm fully commits to a contract upon entry, its profits can be written in recursive form subject to a Promise-Keeping constraint: the firm is bound to deliver the value of the current contract to its workers in expectation. Let W_{t-1} denote the value of this contract given the realization of the states in period $t-1$. Let χ_t further denote the firm's decision to continue given the realization of the shocks at the start of period t.

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

[11] 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 idiosyncratic productivity p_{t-1} employing n_{t-1} workers has discounted profits

$$\Pi_{t-1}\left(p_{t-1}, n_{t-1}, W_{t-1}\right) = \max_{\substack{w,W \\ h\geq 0}} \left\{ (\omega_{t-1}p_{t-1} - w)n_{t-1} + \beta E_t \Big[\chi_t \Big(- c(h)(1-\mu)(1-\delta)n_t + \Pi_t(p_t, n_t, W) \Big) \Big] \right\}, \quad (4)$$

where the firm's continuation decision is given by $\chi_t := \mathbb{1}\Big\{ \big(W \geq U_t\big) \cap \big(\Pi_t \geq 0\big) \Big\}$, since the firm needs to offer at least U_t 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, Promise-Keeping requires that the firm's choice of wages, w, and (state-contingent) contract values in the next period, W, delivers workers a value of at least W_{t-1} in expectation. Second, the size of its workforce, conditional on the firm surviving, is evolving according to the law of motion

$$n_t = \big[1 - q_t(W) + h\big](1-\mu)(1-\delta)n_{t-1}. \quad (5)$$

In this last expression, $(1-\mu)(1-\delta)n_t$ is the measure of workers still employed at the search stage, those not leaving to become potential entrepreneurs (rate μ) or exogenously becoming unemployed (rate δ). Workers still employed at the search stage leave the firm to take on better jobs at rate $q_t(W)$, where the quit rate is given by $q_t(W) := s\lambda_t(1-F_t(W))$, the rate at which workers employed at the firm find better jobs in the current period. Finally, new workers join the firm at rate h.[12] Equation (5) makes the firm's trade-off in controlling the growth of n_{t-1} explicit. It can either offer better contracts, thus limiting poaching, or intensify its hiring efforts 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 profits per worker as $n_t\pi_t(p_t, W_{t-1}) := \Pi_t(p_t, n_t, W_{t-1})$. By substituting this guess in the right-hand side of Equation (4) and using the law of motion for employment, it can be shown that

$$\pi_{t-1}\left(p_{t-1}, W_{t-1}\right) = \max_{\substack{w,W \\ h\geq 0}} \left\{ (\omega_{t-1}p_{t-1} - w) + \beta E_{t-1} \Big[(1-\mu)(1-\delta)\chi_t \Big(- c(h) + (1 - q_t(W) + h)\pi_t(p_t, W) \Big) \Big] \right\}, \quad (6)$$

[12] Conditional on $q_t(W)$ and h, Equation (5) holds exactly by a Law of Large Number argument since n_t is the measure of workers employed at the firm.

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

While this property of the value of profits is similar 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 the current wage it offers. With the firm committing to future wages, I am able to relax some of the restrictions they impose on their environment, thus making the firm's entry and exit decisions endogenous. I give further details on the difference between these two wage-setting protocols in Appendix A.7.

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 $q_t(W)$, and (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 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 the choice of hiring effort and contracts pin down, conditional on survival, the growth rate of employment. Even though 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 productivity of prospective entrants. 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 a firm against the value of unemployment. Entering entrepreneurs are assumed to get the full surplus $S_t(p_t) := \pi_t(p_t, W_t) + W_t$ of the match.[13] They then decide to enter if $U_t \leq S_t(p_0)$ for some initial draw p_0 from Γ_0. If they choose not to take this business opportunity, they fall back into unemployment until next period (they do not search in the current period). If they choose to enter, it is assumed that entrepreneurs have their business purchased by some outside investors (not modeled), and become the first workers at these firms.

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

Similarly to Gavazza et al. (2018), firms need to have positive employment to operate the recruitment technology c. 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, become 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. These firms then move on to the production stage, and become incumbent firms from the next period onward.

3.4 Value of unemployment and employment

Let Q_t denote the value of an entrepreneur, an employed or unemployed worker hit by a μ-shock, $Q_t := \int \max\big(S_t(p), U_t\big) d\Gamma_0(p)$. Let $\overline{W}_t$ denote the contract with the largest value offered by firms given the realization of the states in period t.[14]

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

$$U_{t-1} = b + \beta E_{t-1}\Big\{\mu Q_t + (1-\mu)\Big[(1-\lambda_t)U_t + \lambda_t \int_{U_t}^{\overline{W}_t} W' dF_t(W')\Big]\Big\}. \tag{7}$$

Similarly to unemployed workers, employees have a chance μ to get a business idea, in which case they leave their present job to explore this opportunity. 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-1}, and are promised a state-contingent value W_t. Recall that due to the commitment structure, the firm needs to promise its workers a contract value of at least U_t to stay in business. This is summarized by the indicator χ_t. Taken together, these shocks give rise to the following value function for the employed worker

$$\begin{aligned} W_{t-1} = w_{t-1} + \beta E_{t-1}\Big\{ & \mu Q_t \\ & + (1-\mu)\Big[\Big((1-\chi_t) + \delta\chi_t\Big)U_t + \chi_t(1-\delta)\Big(\big(1-q_t(W_t)\big)W_t + s\lambda_t \int_{W_t}^{\overline{W}_t} W' dF_t(W')\Big)\Big]\Big\}. \end{aligned} \tag{8}$$

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.2 that the following expression for $S_{t-1}(p) := \pi_{t-1}(p_{t-1}, W_{t-1}) +$

[14] This upper bound must exist as $\overline{\omega p}/(1-\beta)$ represents a natural upper limit on the value of offered contracts.

W_{t-1} can be obtained after rearranging these two equations

$$S_{t-1}(p_{t-1}) = \\ p_{t-1}\omega_{t-1} + \beta E_{t-1}\Big\{\mu Q_t + (1-\mu)\Big[(1-\chi_t(p_t))U_t + \chi_t(p_t)\Big(\delta U_t + (1-\delta)\psi_t(p_t)\Big)\Big]\Big\}. \quad (9)$$

In this last expression, $\psi_t(p_t)$ denotes the joint value of a firm-worker pair, conditional on the firm not exiting, which writes

$$\psi_t(p_t) := \max_{W,h\geq 0}\Big\{-c(h) + \big(1-q_t(W)\big)S_t(p_t) + \big(S_t(p_t) - W\big)h + s\lambda_t \int_{W}^{\overline{W}_t} W' dF_t(W')\Big\}. \quad (10)$$

This formulation of the firm-worker surplus 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 in the expression for the firm-worker surplus, they are still pinned down by the Promise-Keeping constraint. Given that the firm guarantees at least W_{t-1} in expectation to its workers and that future contract values are defined as the solution to Equation (10), the offered wage, w_{t-1}, is implicitly defined by

$$W_{t-1} = w_{t-1} + \beta E_{t-1}\Big\{\mu Q_t \\ + (1-\mu)\Big[\Big((1-\chi_t) + \delta\chi_t\Big)U_t + \chi_t(1-\delta)\Big(\big(1-q_t(W_t)\big)W_t + s\lambda_t\int_{W_t}^{\overline{W}_t} W' dF_{t+1}(W')\Big)\Big]\Big\},$$

where W_t denotes the (state-contingent) solution to the joint-surplus maximization problem in Equation (10). Intuitively, given W_t, wages can freely adjust to satisfy the Promise-Keeping constraint.

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,

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

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

Let $a_t(p,n)$ denote the vacancies posted by a continuing firm with productivity p and workforce n. Total vacancy posting is simply the sum of all vacancies posted by active firms in the economy $A_t := \int \chi_t(p) a_t(p,n) d\nu_t$. Because search is random, the cumulative density of offered contracts is then given by the sum of vacancies offering a contract less than some contract value W over all posted vacancies

$$F_t(W) := A_t^{-1} \int \mathbb{1}\{W_t \leq W\} \chi_t(p) a_t(p,n) d\nu_t. \tag{12}$$

Firm vacancy posting. To close the model, I specify how vacancy posting at the firm level, $a_t(p,n)$, is pinned down by its hiring choice. Since there is no cost of posting vacancies by assumption, the firm simply posts as many as required by its optimal hiring rate, $h_t(p)$, given the current aggregate conditions in the labor market. $a_t(p,n)$ is then implicitly defined by the accounting equation

$$h_t(p)(1-\mu)(1-\delta)n = a_t(p,n)\eta_t Y_t\big(W_t(p)\big). \tag{13}$$

In Equation (13), the term on the left-hand side gives the firm's gross hires in period t. $a_t(p,n)$ is then chosen taking into account the rate at which these vacancies reach workers, the contact rate η_t, and the chance $Y_t(W_t(p))$ that the firm's offer is accepted. This probability is determined by whether the worker reached by the vacancy is on a contract offering less than the firm's choice of contract $W_t(p)$ in the current period. The expression for Y_t is given in Appendix A.3.

4 Rank-Monotonic Equilibrium

This section formalizes the definition of equilibrium used in the remainder of the paper. I derive sufficient conditions on the cost of hiring function that make the optimal contract offered by firms increasing in their current realization of idiosyncratic productivity, after all histories. I label these equilibria as "Rank-Monotonic" in the rest of the paper.

This characterization is related to the "Rank-Preserving" property conceptualized in Moscarini and Postel-Vinay (2013) in the sense that the optimal contract is increasing in firm-specific productivity in both cases. However, while in their framework with constant firm-specific productivity this property implies that more productive firms are always larger along the equilibrium path (aggregate shocks preserve 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 and firm-size distribution 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 allows for richer firm dynamics. Young firms can enter near the top of the productivity distribution and quickly increase their workforce though they are initially small. Old and large firms can be led to exit following a bad realization of idiosyncratic productivity.

The "Rank-Monotonic" 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 (ω, p) 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 policies are linear in its current employment n. More formally:

Definition 1 *A Recursive Equilibrium is a triple of policy functions (W,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 ν at the start of the period. Given that all firms follow the policies given by (W,h,χ), these functions satisfy:*

1. *Equations* (11)-(13) *hold with firms' optimal choices $\chi_t(p) = \chi(p, \omega, \nu)$, $h_t(p) = h(p, \omega, \nu)$, and $W_t(p) = W(p, \omega, \nu)$;*

2. *The contract and hiring functions solve the maximization problem in* (10)*. The continuation decision is given by $\chi(p, \omega, \nu) = \mathbb{1}\{W(p, \omega, \nu) \geq U(\omega, \nu)\}$;*

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 optimal contract:

Definition 2 *A Rank-Monotonic Equilibrium is a Recursive Equilibrium such that the optimal contract, $W(., \omega, \nu)$, 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.*[15]

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

Then, for any Recursive Equilibrium such that the distribution of offered contract $F(.,\omega,\nu)$ is everywhere differentiable:

1. *The firm-worker surplus defined by Equation* (9) *is differentiable and increasing in p;*
2. *The equilibrium is Rank-Monotonic provided $hc''(h)/c'(h) \geq 1$ at all $h \geq 0$.*

The proof is in Appendix A.4. I stress that Result 1 is not an existence statement, but a characterization of the properties of the optimal contract conditional on the existence of a such a Recursive Equilibrium.

The condition on the cost function in Result 1 is an additional convexity requirement. Firms use the retention margin, by offering better contracts to limit quits, 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 requirement that $F(.,\omega,\nu)$ is everywhere differentiable rules out Recursive Equilibria in which all firms offer the same contract, irrespective of their idiosyncratic productivity. The classic argument in random search model with on-the-job search to show that the distribution of offers is not degenerate is that firms can increase profits by offering jobs paying slightly more, thus poaching workers from other firms (see, for example, Burdett and Mortensen, 1998). This argument does not translate to this framework because the quit and hiring margins are separately controlled by the firm, respectively through contracts and hiring effort. For instance, if all firms offer the value of unemployment at all realizations of the aggregate states, a marginal increase in wage payments decreases profits with no employment gains, since the firm independently chooses hiring. I note that such equilibria are counter-factual, as they do not give rise to job-to-job transitions.

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. I further assume that firms believe there are voluntary worker quits when designing the optimal contract, so the worker retention margin is used, which in turn ensures the distribution of contracts is non-degenerate.

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 these equilibria. I start by defining the measure of workers employed at firms less productive than p at the start of period t, $L_t(p) := \int_{\tilde{p}\leq p} n d\nu_t(\tilde{p}, n)$. This last measure fully summarizes quit decisions at each level of productivity in a Rank-Monotonic Equilibrium since the optimal contract is increasing in p. Firms 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.5 shows that the optimal contract takes the following form

$$W(p,\omega,L) = \frac{uU(\omega,L) + s(1-\delta)\int_{p_E}^{p} S(\tilde{p},\omega,L)dL(\tilde{p})}{u + s(1-\delta)\Big(L(p) - L(p_E)\Big)}. \tag{14}$$

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 many standard search and matching 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 first-order condition for their choice of hiring rate in Equation (10)

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

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

$$\lambda_t F_t\left(W_t(p)\right) = \int_{p_E}^{p} \frac{(1-\delta)h_t(\tilde{p})}{u_t + s(1-\delta)\Big(L_t(\tilde{p}) - L_t(p_E)\Big)} dL_t(\tilde{p}). \tag{15}$$

The derivations can be found in Appendix A.5.

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

$$L_t^P(p) = \mu\Big[\Gamma_0(p) - \Gamma_0(p_E)\Big] + (1-\mu)\Big[L_t(p)(1-\delta)\big(1 - q_t(W_t(p))\big) + u_t\lambda_t F_t(W_t(p))\Big], \tag{16}$$

where L_t^P denotes the cumulative 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 square brackets give, first, the fraction of workers retained at firms less than p and, second, the inflow of workers from unemployment. Finally, the end of period and beginning of next period measures are directly linked through $dL_{t+1}(p)/dp = \int_{\underline{p}}^{\overline{p}} dL_t^P(\tilde{p})/d\tilde{p} d\Gamma(p|\tilde{p})$, which corresponds to the "re-shuffling" of workers across productivity levels due to the firm-specific shocks.

Summary. In a Rank-Monotonic Equilibrium, knowledge of 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.[16] 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, an infinitely dimensional object. 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 Rank-Monotonic Equilibrium in which the measure of workers is stationary. Formally:

Definition 3 *A Stationary Rank-Monotonic Equilibrium is a triple of policy functions (W,h,χ), a pair of value functions (S,U), and a measure of workers across firm productivity L, that only depend on the current realization of firm productivity p ($\omega_t = \omega > 0$ for all t). 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* (16) *is constant and equal to L.*

In a second step, I return to the full model with aggregate shocks 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 process for idiosyncratic productivity shocks as $\ln p_t = \rho_p \ln p_{t-1} + \sigma_p \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0,1)$. Such a process satisfies first-order stochastic dominance conditional on past realizations, which is required for the equilibrium to be Rank-Monotonic (Result 1). Draws from Γ_0, the distribution of productivity for initial ideas, are assumed to come from the stationary distribution implied by the process for idiosyncratic productivity. 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

[16] 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 space, all the corresponding expressions are relegated to Appendix A.6.

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: $\xi := \left(\beta, \delta, c_1, c_2, s, \mu, \rho_p, \sigma_p, b\right)^{\intercal}$.

5.2 Calibration strategy

The discount factor, β, is set in line with a 5% annual discount rate. This leaves eight parameters to calibrate, which I pin down by targeting an equal number of moments from the data (see Appendix C.2 for details). My choice of moment targets reflects both the search and firm dynamics components of the model. All data moments are computed for the pre-crisis period.

To discipline the search component of the model, I target the unemployment to employment (UE), employment to unemployment (EU), and job-to-job (EE) average monthly transition rates in the UK. These series are derived from the British Household Panel Survey (BHPS) following the methodology described in Postel-Vinay and Sepahsalari (2019). To discipline how employment and labor productivity jointly evolve at individual firms, I target the inter-decile range of labor productivity ($LP_{i,t}$), the autocorrelation of log-employment, the average employment of firms, and the share of job destruction stemming from firm exit.[17] I also include the coefficient from regressing a firm's employment growth on its current labor productivity, controlling for industry-year fixed effects. These moments are computed directly from the administrative firm-level data described in Appendix B.1, and are therefore yearly measures.

Given a candidate vector of parameters, I solve for a Stationary Rank-Monotonic Equilibrium to derive the moments implied by the model. 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 p. The monthly worker 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 come from yearly data. Their model counterpart is therefore obtained by simulating a panel of firms. I simulate a cohort of 60,000 entrants (roughly forty percent of the number of firms in a typical cohort of entrants in the British data) 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 counterpart to the concept available in the firm-level data and compute the labor productivity measure defined by Equation (1). Since this measure is in logs, the actual units of value-added are irrelevant to my calibration.

All parameters and moments are related, as even parameters more directly linked to the worker side of the model alter the decisions of firms. For instance, the relative search intensity of employed workers, s, also impacts the entry and exit threshold, as it changes the value of employment relative to unemployment, and in turn determines the span of labor productivity. In Appendix Figure 9,

[17] I target the autocorrelation of employment and not labor productivity directly, as labor productivity is only available for a repeated cross-section of firms and not for a panel in the data.

Moment	Data	Model
A. Targeted moments		
UE	0.069	0.068
EU	0.004	0.004
EE	0.020	0.018
Job dest. from exit	0.526	0.528
Average employment	12.113	11.944
corr($\ln n_{i,t}, \ln n_{i,t-1}$)	0.949	0.989
Interdecile range($LP_{i,t}$)	2.241	2.233
Reg. $\Delta \ln n_{i,t+1}$ on $LP_{i,t}$	0.136	0.135
B. Non-targeted moments		
Share of young (age<5) firms	0.365	0.371
Share of empl. at young firms	0.116	0.100
Firm exit rate	0.130	0.097
Firm size tail parameter	1.066	1.073
Interdecile range($W_{i,t}$)	2.594	1.741
Reg. $W_{i,t}$ on $LP_{i,t}$	0.704	0.740

Table 2: Model fit.

I report slices of the objective function for each parameter that confirm that the parameters in ξ are well-informed by the targeted moments.

5.3 Model fit

The model fit to the targeted moments is shown in Table 2. Overall, the model replicates these statistics well, with the exception of the autocorrelation of employment, which the model tends to overestimate with respect to the data.

Table 2 further shows that the model performs well along a number of dimensions that are not directly targeted as part of the calibration procedure. First, the model implies a reasonable selection and growth of firms following entry. Both the share of young firms (less than five year old) and the share of total employment at those firms are in line with the data. The magnitude of the firm exit rate is also broadly similar to the actual value.

Second, the wages generated by the model imply a reasonable degree of wage dispersion with respect to the data. Again, because the data are at the firm-level, the wage concept is the logarithm of total employment costs over total employment at the firm ($WP_{i,t}$), as defined in Section 2. The model replicates about seventy percent of the between-firm dispersion in wages, as measured by the inter-decile range. The extent to which wages ($WP_{i,t}$) correlate with labor productivity ($LP_{i,t}$), which I measure by regressing $WP_{i,t}$ on $LP_{i,t}$ again controlling for industry-year effects, is also similar in the model and the data.

Lastly, the model implies a firm-size distribution with a long tail. I measure the concentration of employment at the largest firms by estimating the Pareto coefficient associated with this

Parameter	Description	Value
Externally set		
β	Discount factor (≈5% annual)	0.996
Calibrated		
δ	Prob. job destruction (×100)	0.003
c_1	Hiring cost:	52.506
c_2	$c(h) = (c_1 h)^{c_2}/c_2$	5.841
s	Relative search effort	0.735
μ	Prob. of start-up (×100)	0.075
ρ_p	Process for firm productivity:	0.978
σ_p	$\ln p_{t+1} = \rho_p \ln p_t + \sigma_p \epsilon_{t+1}$	0.269
b	Flow value of unemployment	0.308

Table 3: Parameter estimates.

distribution, where I restrict the sample to firms with employment at or above the average firm employment. This coefficient is 1.066 in the data and 1.073 in the model. This feature arises in the model because (i) the size-independence result implies that the growth rate of employment at a firm with current productivity realization p is given by $(1-\mu)(1-\delta)\big(1-q(W(p))+h(p)\big)$, which does not depend on its current size; (ii) this is a birth-death process, due to firm entry and exit. These two conditions on the process underlying firm employment are precisely those identified by the literature on the emergence of power law distributions in economics (e.g., Gabaix, 1999; Reed, 2001).[18]

5.4 Parameters

The calibrated parameters are listed in Table 3. I briefly comment on some key parameter values. There is no clear benchmark in the literature for the hiring cost function parameters because this functional form has seldom been used. I find that the implied average hiring cost as a fraction of monthly output is 8.4%. 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 16.6% of the average wage in the economy, which

[18]Gouin-Bonenfant (2019) reports similar findings in a related model.

is slightly under half the value used in Shimer (2005). In line with the results in Hornstein et al. (2011), lower values of b in this class of models tend to help matching the degree of wage dispersion observed in the data, as shown in Table 2.

6 Business cycle

In a Stationary Rank-Monotonic 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 it is therefore relevant to the firm's decisions (Definition 2). This extra state variable represents a technical hurdle since the distribution of workers across firm productivity is an infinitely dimensional object.

In this section, I start by describing the approximation used to solve the model with aggregate shocks. With the model solution in hand, I then proceed with a series of exercises highlighting the interplay between firm dynamics and search frictions in accounting for aggregate labor productivity changes over the business cycle.

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 vector of moments, which I denote by $\boldsymbol{m}_t$. This vector includes the unemployment rate, $m_t^0 := u_t = 1 - \int dL_t(p)$, and N_m moments m_t^1, $\ldots$, $m_t^{N_m}$ from the distribution of workers $L_t / \int dL_t(p)$.[19]

Given this approximation of L_t, the state-space relevant to the firm now reduces to ω_t and $\boldsymbol{m}_t$. I then parametrize the firm-worker surplus and the unemployed worker's value function out of steady-state with a polynomial. For instance, the value function for workers in unemployment is approximated as $\ln U(\omega_t, L_t) - \ln \bar{U} \approx \tilde{U}(\omega_t, \tilde{\boldsymbol{m}}_t; \theta_U)$, where $\tilde{x}_t$ denotes a variable in log-deviation from steady-state and θ_U is a vector of coefficients to be solved for. The firm-worker surplus is similarly approximated, using a separate polynomial at each grid point for firm productivity. 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 it does not require to specify a grid for these moments. 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 how to choose the number of moments to be included in $\boldsymbol{m}_t$, can be found in Appendix C.3.

An alternative approach to simulate heterogeneous agents models with aggregate shocks is to use the perturbation method proposed by Reiter (2009). Such linearization techniques have been successfully applied to firm dynamics models (Sedláček and Sterk, 2017; Winberry, 2020). However, my simulations suggest that taking a derivative around the steady-state is highly inaccurate in the

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

Parameters $\ln \omega_t$		Cyclicality stat.	Data	Model
ρ_ω	0.967	$\text{corr}(\widehat{UE}_t, \widehat{UE}_{t-1})$	0.823	0.823
σ_ω	0.147	$\text{std}(\widehat{UE}_t)$	0.052	0.047

Table 4: Parameters aggregate shock process.

context of my model due to the discontinuity implied by the firm's entry and exit threshold. I therefore rely on the simulation-based approach outlined here and report accuracy tests for my proposed algorithm in Appendix C.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 $\ln \omega_{t+1} = \rho_\omega \ln \omega_t + \sigma_\omega \epsilon_{t+1}$ with $\epsilon_{t+1} \sim \mathcal{N}(0, 1)$. The parameters in this process $(\rho_\omega, \sigma_\omega)$ are chosen to replicate the model-simulated persistence and volatility of the job-finding rate in the UK between 1992 and 2016. The persistence and volatility are, respectively, the first autocorrelation and standard deviation of the HP-filtered (log) job-finding rate $(\widehat{UE}_t)$ in the UK between 1992q1 and 2016q4, where the monthly series are first converted to quarterly by time averaging. The calibrated values for $(\rho_\omega, \sigma_\omega)$ are shown in Table 4. I target fluctuations in the job-finding rate directly since, as is well understood in the literature, this class of model typically does not generate a lot of amplification from aggregate productivity shocks to labor market variables (Shimer, 2005; Hagedorn and Manovskii, 2008), and this observation applies to my framework. Since the emphasis is on understanding how changes in labor flows arising at business cycle frequency contribute to changes in aggregate labor productivity, I choose to target these flows directly to calibrate the process for aggregate shocks. Equation (17) below makes clear that the direct impact of these shocks on labor productivity can be taken out in the context of this model.

6.2 Validation: a Great Recession episode

To understand the reallocation effects of a large recession in the model, I input a sequence of aggregate shocks that trigger a sharp drop in the job-finding rate, 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 empirical patterns documented in Section 2. I stress that this aspect of the data is not targeted in the calibration procedure outlined in Section 5.

Aggregate labor market response. Figure 3 depicts the Great Recession experiment I run in the model. I input a sequence of aggregate shocks designed to replicate the sharp drop in the job-finding rate observed in the UK following the start of the episode. The aggregate shock is then left to revert back to its steady-state level according to the persistence parameter given in Table 4.

Table 5 further reports the peak to trough response, defined as the difference between the values at the official start (2008m3) and end (2009m6) of the UK Great Recession, of several labor market

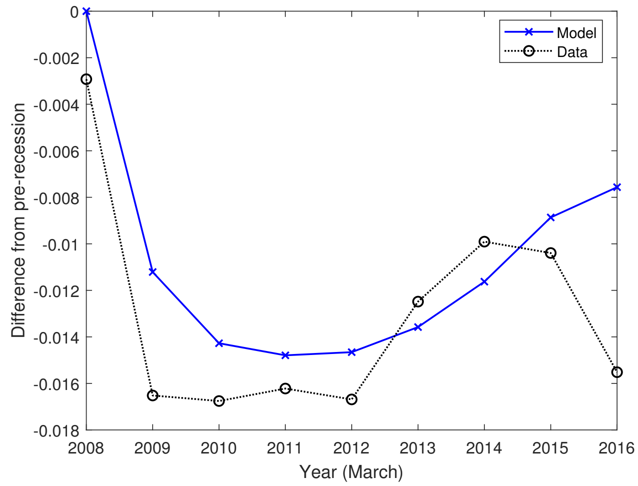


Figure 3: Job finding rate in Great Recession experiment.

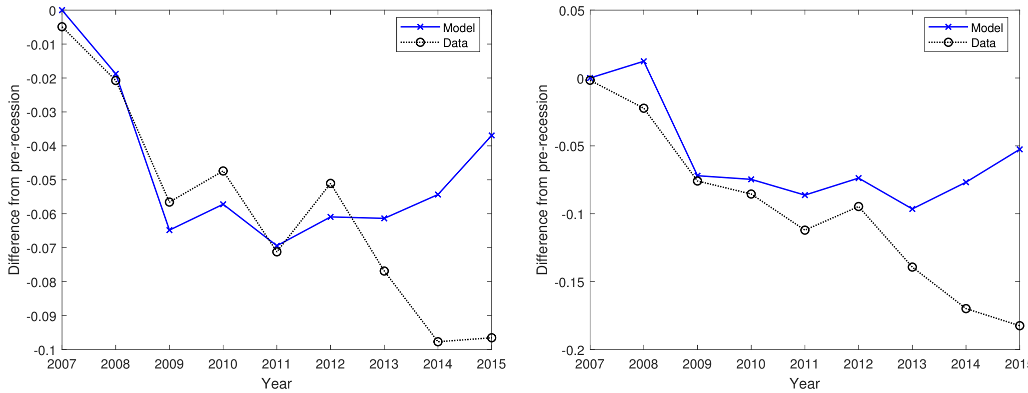
	Peak – Trough	
Aggregate	Model	Data
UE	0.012	0.009
EU	-0.001	-0.001
EE	0.003	0.005
ln Vac.	0.523	0.489

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

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.[20] Overall, though the drop in the EE rate appears slightly smaller than in the data, the model does a good job at replicating the changes in these aggregates.

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 $\sum_i \left(ES_{i,t} - \overline{ES}_t\right)\left(LP_{i,t} - \overline{LP}_t\right)$ and captures to which extent firms with a higher than average labor productivity also account for a higher than average employment share. To obtain the corresponding model series, I simulate a cohort of firms through the recession and aggregate the output of this simulation at yearly frequency. Figure 4a benchmarks the model response against the data in deviation from their pre-recession linear trend. It shows that the model tracks the drop in the OP misallocation measure closely. I also report results for the contribution of worker misallocation to changes in the aggregate wage index defined in Section 2 (Figure 4b). The model also does well at

[20]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 $m(A_t, Z_t) = aA_t^{\alpha} Z_t^{1-\alpha}$ where I normalize $a = 1$ and set the elasticity of matches to vacancies to .5.



(a) Labor productivity (LP_{it}) **(b)** Wage payments (WP_{it})

Figure 4: Changes in Olley-Pakes misallocation measure: model vs data

	$y_{it} = LP_{it}$		$y_{it} = W_{it}$	
	Model	Data	Model	Data
Coeff. on y_{it} (α)	0.162	0.102	0.222	0.136
Coeff. on $y_{it} \times \text{post}_t$ (β)	-0.031	-0.031	-0.028	-0.012
As fraction of pre-GR ($-\beta/\alpha$)	0.194	0.299	0.126	0.088

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

replicating this aspect of the data. The fact that the series diverge towards the end of the period suggests that other changes are potentially taking place in the structure of the British economy, which would materialize in the model as an additional shift in some parameters, to explain the persistence of the reduction towards the end of the period.

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

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

where $y_{i,t}$ denotes either labor productivity $LP_{i,t}$ or firm-level wage payments $WP_{i,t}$. Table 6 shows that the model can emulate how the association between the firm's current productivity (current wages paid) and their subsequent employment growth drops following the Great Recession. In the $LP_{i,t}$ regression, in particular, the baseline coefficient drops by 30% in the data and 20% in the model.

6.3 A model-based productivity decomposition

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, monthly aggregate labor productivity can be rewritten

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

where the superscript "P" denotes the production stage (end of period) and a bar denotes a normalized measure.[21] 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.

LP_t can be further decomposed into a firm productivity component and a worker reallocation component. The OP decomposition $LP_t = \overline{LP}_t + \sum_i \left(ES_{i,t} - \overline{ES}_t\right)\left(LP_{i,t} - \overline{LP}_t\right)$ can be written

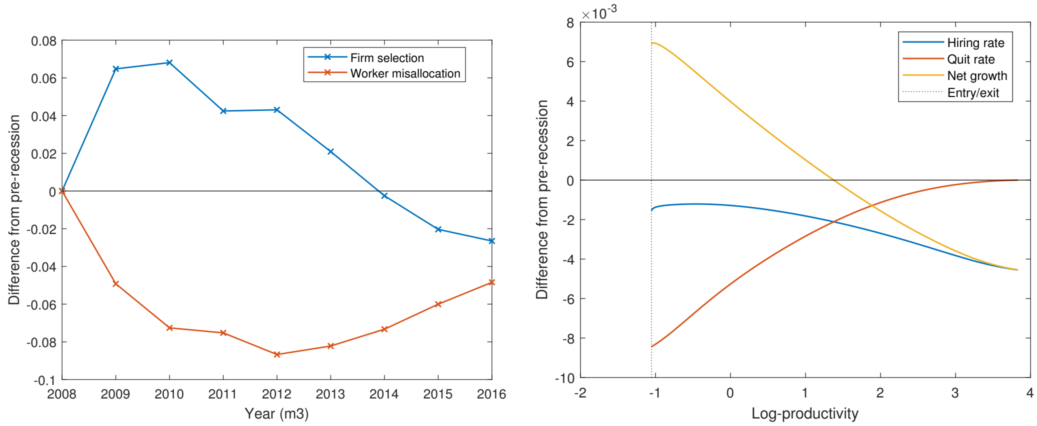
$$LP_t = \underbrace{\ln \omega_t}_{\text{aggregate shock}} + \underbrace{\int \ln(p) dK_t^P(p)}_{\text{firm selection}} + \underbrace{\int \ln(p) dL_t^P(p) - \int \ln(p) dK_t^P(p)}_{\text{measure of misallocation}} \tag{17}$$

in the notation of the model. In this expression, the first two terms correspond to average firm productivity ($\overline{LP}_t$): the first term gives the direct impact of the aggregate shock and the second term captures changes in the distribution of firms across productivity level. The last term in Equation (17) measures the misallocation of workers to firms. In the model, the OP measure of misallocation therefore 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 in the model.

Through the lens of the model, aggregate labor productivity can then be further unpacked in two endogenous terms on top of the direct impact of the shock: firm selection and worker reallocation (the last two terms in Equation (17)). I plot the evolution of these two components over the course of the simulated recession in Figure 5a. It shows that they contribute to aggregate labor productivity in mostly opposite directions. However, while they are initially of about the same magnitude, the worker reallocation term exhibits more persistence. It is still negative eight years after the start of the recession. It therefore acts as a force dragging down labor productivity in the medium term. On top of this negative effect, the firm selection term is also shaped by the overall labor market conditions: with a lot of unemployment, firms at the lower end of the productivity distribution find it easier to retain workers, which also alter their entry and exit decision. At the start of the recession, the direct impact of the aggregate productivity shock dominates and make firm selection more stringent. But towards the end of the episode, as the shock reverts back to its pre-recession level, the model actually implies this general equilibrium effect eventually becomes larger, creating an additional force pushing labor productivity down through firm selection.

I confirm the relative importance of the worker reallocation channel by decomposing labor productivity over the business cycle using Equation (17). I average the firm selection and worker misallocation terms in deviation from steady-state over the business cycle in two recession scenarios:

[21] So $\overline{L_t}^P(p) := \int_{\tilde{p}\leq p} dL_t^P(\tilde{p}) / \int dL_t^P(p)$.



(a) Firm selection and worker reallocation

(b) Firm policies

Figure 5: Labor productivity changes in the model.

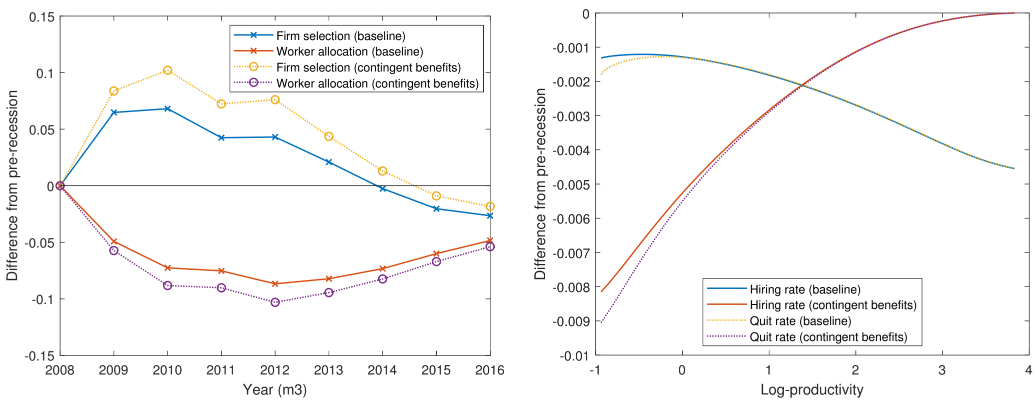
"small" denotes negative aggregate shocks between one and two standard deviations in the ergodic distribution of $\ln \omega_t$; "large" corresponds to aggregate shocks larger than two standard deviations in the same distribution. Table 7 shows that the negative worker reallocation effect also dominates the firm selection effect over the business cycle, both for small and large negative shocks.

I illustrate the main worker reallocation mechanism at the micro level in Figure 5b. On top of the firm selection effect, which shifts up the entry threshold, how well labor is allocated to firms also depends on which firms grow faster following the shock. Figure 5b 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 weakening of the relationship between firm-level labor productivity and employment growth in the aftermath of the recession. While this relationship cannot be decomposed further into hires, quits, and layoffs without matched employer-employee data, the drop in the quit rate at low productivity 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 5b) decline by eighty-five percent during the US Great Recession.

	Baseline		Policy	
Size of negative shock	Small	Large	Small	Large
Firm selection	0.018	0.033	0.030	0.055
Worker misallocation	-0.040	-0.059	-0.046	-0.069
Net effect	-0.023	-0.025	-0.017	-0.014

Table 7: OP decomposition over the business cycle.



(a) Productivity decomposition

(b) Optimal policies

Figure 6: Baseline vs unemployment-contingent benefits model.

6.4 Policy experiment: unemployment-contingent benefits

The tension between firm selection and worker reallocation as determinants of aggregate labor productivity can be further assessed in the following policy experiment. In the spirit of unemployment insurance extension programs in the United States, I allow the value of non-employment, b, to depend on the unemployment rate.[22] Specifically, the value of non-employment is assumed to vary with the unemployment rate according to

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

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.

As an example, I set $\kappa = .3$ and solve the model again using the same sequence of aggregate shocks as in the baseline model. I then compare the model response under the unemployment-contingent benefit policy ($\kappa > 0$) to the baseline model with constant b ($\kappa = 0$) over the course of

[22]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.

the simulated recession. With respect to labor productivity, Figure 6a shows that such a policy has two opposite effects. First, it makes the selection effect more stringent. Workers' outside option is better under the unemployment-contingent policy, which tightens the conditions for firms to enter and stay in business. This represents an extra push to the cleansing force in the model.

Second, Figure 6a also shows that unemployment-contingent benefits magnify the worker reallocation effect resulting from search frictions. More unemployment further slows down the transition of employed workers to more productive firms. As can be seen from the firm's policies in Figure 6b, 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 slowing down the pace of worker reallocation to more productive units. On balance, the additional firm selection effect seems to dominate in the simulated recession. I also find a similar effect over the business cycle, again classifying recessions as "small" or "large" according to the size of the aggregate shocks. As shown in Table 7, the firm selection effect seems to react more strongly to unemployment-contingent benefits, which alleviates the negative impact from the worker reallocation channel.

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' entry and exit decisions and recruitment behaviors at the micro level drive the evolution of aggregate labor productivity at the macro level over the business cycle.

The central insight of the model is that search frictions dampen labor productivity following a large aggregate shock. A larger pool of unemployed workers 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 employed workers from less to more productive firms.

In an experiment designed to replicate the reduction in the job-finding rate observed during the Great Recession in the UK, I find that this mechanism can account for most of the drop in the allocation of workers to firms measured in British firm-level data during this episode. Through the lens of the model, I further find that, in the medium term, this negative worker reallocation effect dominates the positive effect arising due to a more stringent selection of firms when accounting for the evolution of labor productivity in the aftermath of the recession. While the strength of these two forces is bound to differ for different countries or time periods, this paper identifies search frictions in the labor market as an important drag on the recovery of labor productivity.

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. Working Paper.

Blanchard, O. J., P. Diamond, R. E. Hall, and K. Murphy (1990). The cyclical behavior of the gross flows of us workers. *Brookings Papers on Economic Activity 1990*(2), 85–155.

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.

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

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.

Fujita, S., G. Moscarini, and F. Postel-Vinay (2020). Measuring employer-to-employer reallocation. Working Paper, National Bureau of Economic Research.

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. Working Paper.

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. Working Paper.

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.

Hopenhayn, H. and R. Rogerson (1993). Job turnover and policy evaluation: A general equilibrium analysis. *Journal of Political Economy*, 915–938.

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.

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 (2012). Labor market friction, firm heterogeneity, and aggregate employment and productivity. Working Paper.

Menzio, G. and S. Shi (2011). Efficient search on the job and the business cycle. *Journal of Political Economy 119*(3), 468–510.

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 (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. Working Paper.

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.

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. (2020). Lumpy investment, business cycles, and stimulus policy. Working Paper.

A Theory Appendix

A.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

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

Starting from Equation (4), still subject to the Promise-Keeping constraint and the law of motion for its workforce, the expectation on the right-hand side rewrites

$$\begin{aligned}
&E_{t-1}\Big[\chi_t(p_t)\Big(-c(h_t)(1-\mu)n_t+\Pi_t(p_t,n_t,W_t)\Big)\Big]\\
=&E_{t-1}\Big[\chi_t(p_t)\Big(-c(h_t)(1-\mu)n_t+n_t\pi_t(p_t,W_t)\Big)\Big]\\
=&n_{t-1}E_{t-1}\Big[\ \chi_t(p_t)\Big(-c(h_t)+(\rho_t(W_t)+h_t)\pi_t(p_t,W_t)\Big)\Big].
\end{aligned}$$

The second line substitutes in the guess. The last line uses the law of motion for the firm's workforce. Note that with a continuum of workers, it is assumed to hold exactly conditional on the firm surviving and $q_t(W_t),\ h_t$. 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. 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.2 Firm-worker match surplus

Recall that the joint value of a match is defined as $S_{t-1}(p_{t-1}):=\pi_{t-1}(\overline{V})+\overline{V}$. Rearranging the Promise-Keeping constraint gives an expression for w_{t-1}

$$\begin{aligned}
w_{t-1}=\overline{V}-\beta E_{t-1}\Bigg\{&\mu Q_t\\
&+(1-\mu)\Big[\big(1-\chi_t(p_t)\big)U_t+\chi_t(p_t)\Big(\delta_tU_t+(1-\delta_t)\big(1-q_t(W_t)\big)W_t+s\lambda_t\int_{W_t}^{\overline{W}_t}W'dF_t(W')\Big)\Big]\Bigg\}.
\end{aligned}$$

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

$$
\begin{aligned}
S_{t-1}(p) &:= \pi_{t-1}(p, \overline{V}) + \overline{V} \\
&= -\overline{V} + \max_{\substack{W_t \\ h_t \geq 0}} \Bigg\{ p_{t-1}\omega_{t-1} \\
&+ \beta E_{t-1}\Bigg[\mu Q_t + (1-\mu)\Big((1-\chi_t(p_t))U_t \\
&+ \chi_t(p_t)\big(\delta U_t + (1-\delta)((1-q_t(W_t))W_t + s\lambda_t \int_{W_t}^{\overline{W}_t} W' dF_t(W')) \\
&- c(h_t) + (1-q_t(W_t)) + h_t)\pi_t(p_t, W_t)\big)\Big)\Bigg]\Bigg\} + \overline{V}.
\end{aligned}
$$

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

$$
\begin{aligned}
S_{t-1}(p) = p_{t-1}\omega_{t-1} + \beta E_{t-1}\Bigg[\mu Q_t + (1-\mu)\Bigg((1-\chi_t(p_t))U_t \\
+ \chi_t(p_t) \max_{\substack{W_t \\ h_t \geq 0}} \Big\{ - c(h_t) + (1-q_t(W_t))S_t(p_t) + h_t(S_t(p_t) - W_t) \\
+ (1-\delta)s\lambda_t \int_{W_t}^{\overline{W}_t} W' dF_{t+1}(W') \Big\}\Bigg)\Bigg].
\end{aligned}
$$

A.3 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.4 Proof Rank-Monotonic Equilibrium

The proof is similar in spirit to the ones in Moscarini and Postel-Vinay (2013, 2016): the goal is to show that the optimal contract is increasing in the firm's current realization of productivity, p, assuming the existence of a recursive equilibrium (Definition 1). The key difference with Moscarini and Postel-Vinay (2013, 2016) is that the firm's problem can be considered separately for each worker since $\Pi_t\left(p, n, \overline{V}\right) = n\pi_t\left(p, \overline{V}\right)$.[23] There is therefore no need to show super-modularity of the optimal contract in productivity and firm 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 a convexity requirement on the cost of hiring function. We want to prove the two following statements:

1. Conditional on S_t being increasing in p_t, $hc''(h)/c'(h) \geq 1$ for all $h \geq 0$ is sufficient to guarantee that the optimal contract W_t is increasing in p_t;
2. The firm-worker surplus defined by Equation (9) maps differentiable and increasing functions of p into differentiable and increasing functions of p.

Taking each point in turn:

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. This maximization problem (Equation (10) in the main text) is given by

$$\psi_t(p_t) := \max_{\substack{W \\ h \geq 0}} \Big\{ -c(h) + \big(1 - q_t(W)\big) S_t(p_t) + h\big(S_t(p_t) - W\big) + s\lambda_t \int_W^{\overline{W}_t} W' dF_t(W') \Big\}.$$

At any interior maximum, the firm's optimal choice (W_t, h_t) must satisfy the following first-order conditions

$$\begin{aligned} 0 &= -c'(h_t) + S_t(p_t) - W_t \\ 0 &= -q_t'(W_t)(S_t(p_t) - W_t) - h_t. \end{aligned}$$

In addition, the associated Hessian matrix (H_t^{ψ}) must be negative-definite, which requires

$$\det\Big(H_t^{\psi}\Big) = -c''(h_t)\Big(q_t''(W_t)\big(S_t(p_t) - W_t\big) + q_t'(W_t)\Big) - 1 > 0.$$

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

$$-c'(-q_t'(W_t)\big(S_t(p_t) - W_t)\big) + S_t(p_t) - W_t = 0$$

and totally differentiating that last expression with respect to p gives

$$\frac{dW_t}{dp_t} = \frac{\partial S_t(p_t)}{\partial p_t} \cdot \frac{-q_t'(W_t)c''(h_t) - 1}{\det\Big(H_t^{\psi}\Big)}.$$

In this last expression, $\det\Big(H_t^{\psi}\Big)$ is positive at any maximum. By assumption, the firm-worker surplus is increasing in p, so $\partial S_t(p_t)/\partial p_t \geq 0$. Noting that the two FOCs can be combined to give

[23] I keep subsuming the aggregate states (ω, ν) into the time subscript t to streamline the notation throughout.

$-q_t'(W_t)c'(h_t) = h_t$, it follows that

$$\frac{dW_t}{dp} \geq 0 \iff -q_t'(W_t)c''(h_t) - 1 \geq 0 \iff \frac{h_t c''(h_t)}{c'(h_t)} \geq 1.$$

2. Firm-worker surplus increasing in p Assume that S_t is differentiable and increasing in p_t, we want to show that

$$S_{t-1}(p_{t-1}) = p_{t-1}\omega_{t-1} + \beta E_{t-1}\left\{\mu Q_t + (1-\mu)\Big[(1-\chi_t(p_t))U_t + \chi_t(p_t)\Big(\delta U_t + (1-\delta)\psi_t(p_t)\Big)\Big]\right\}.$$

is differentiable and increasing in p_{t-1}, where again ψ_t is given by

$$\psi_t(p_t) := \max_{\substack{W \\ h\geq 0}} \Big\{ -c(h) + \big(1-q_t(W)\big)S_t(p_t) + h\big(S_t(p_t) - W\big) + s\lambda_t \int_W^{\overline{W}_t} W' dF_t(W') \Big\}.$$

Differentiability of S_{t-1} in p_{t-1} follows directly from noting that the expectation in this last expression is differentiable in p_{t-1} as long as the conditional probability density of future productivity is. This is true by assumption.

To show that S_{t-1} is increasing in its first argument, first note that, for continuing firms, the envelope condition on the firm's optimization problem (10) gives

$$\frac{d\psi_t(p_t)}{dp_t} = \frac{\partial \psi_t(p_t)}{\partial p_t} = \Big(1 - q_t(W_t(p_t)) + h_t(p_t)\Big)\frac{\partial S_t(p_t)}{\partial p_t} \geq 0,$$

where $\big(W_t(p_t), h_t(p_t)\big)$ denote optimal policies in the firm's optimization problem. The term inside the expectation in the firm-worker surplus is then weakly increasing in p: constant on the part of the support of p where the firm exits, and weakly increasing otherwise.

To complete the proof, an additional assumption is needed on the idiosyncratic productivity shock. It has to be assumed that given a higher realization of productivity in the current period, the distribution of future productivity satisfies first-order stochastic dominance $\Gamma(.|p'') \leq \Gamma(.|p')$ for $p'' > p'$ with strict inequality for some productivity level. 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 and p_2 be two productivity values such that $p_2 > p_1$, then

$$S_{t-1}(p_2) - S_{t-1}(p_1) = \omega_{t-1}(p_2 - p_1) + \beta(1-\mu)\bigg(E_{t-1}\Big[\kappa_t(p_t)\ |p_2\Big] - E_{t-1}\Big[\kappa_t(p_t)\ |p_1\Big]\bigg),$$

where $\kappa_t(p_t)$ is a notation for the terms inside the expectation

$$\kappa_t(p_t) := (1-\chi_t(p_t))U_t + \chi_t(p_t)\Big(\delta U_t + \psi_t(p_t)\Big)$$

and I now explicitly condition on the current realization of productivity in the expectation operator.

Showing that S_{t-1} is increasing in its first argument now amounts to show that the difference in expectation in the last expression is non-negative. This difference can be rewritten

$$\int_{\underline{p}}^{\overline{p}} E_{t-1}\Big[\kappa_t(p_t)\Big]\Big(\gamma(p_t|p_2)-\gamma(p_t|p_1)\Big)dp_t,$$

denoting $\gamma(p_t|p_{t-1})$ the density of p_t conditional on p_{t-1} and the expectation is now taken over the aggregate states. Now, given the crossing-point p_0, we can rewrite

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

and, since $E_{t-1}\Big[\kappa_t(p)\Big]$ is weakly increasing in p, we can bound the terms in this last expression as

$$\int_{\underline{p}}^{p_0} E_{t-1}\Big[\kappa_t(p_t)\Big]\Big(\gamma(p_t|p_2)-\gamma(p_t|p_1)\Big)dp_t\geq E_{t-1}\Big[\kappa_t(p0)\Big]\int_{\underline{p}}^{p_0}\Big(\gamma(p_t|p_2)-\gamma(p_t|p_1)\Big)dp_t$$

and

$$\int_{p_0}^{\overline{p}} E_{t-1}\Big[\kappa_t(p_t)\Big]\Big(\gamma(p_t|p_2)-\gamma(p_t|p_1)\Big)dp_t\geq E_{t-1}\Big[\kappa_t(p_0)\Big]\int_{p_0}^{\overline{p}}\Big(\gamma(p_t|p_2)-\gamma(p_t|p_1)\Big)dp_t,$$

where I use that, by the single-crossing property, $\gamma(p_t|p_2)-\gamma(p_t|p_1)\leq 0$ for $p\in[\underline{p},p_0]$ and $\gamma(p_t|p_2)-\gamma(p_t|p_1)\geq 0$ for $p\in[p_0,\overline{p}]$. Finally, summing up the last two inequalities, we get

$$E_{t-1}\Big[\kappa_t(p_t)\ |p_2\Big]-E_{t-1}\Big[\kappa_t(p_t)\ |p_1\Big]=\int_{\underline{p}}^{\overline{p}} E_{t-1}\Big[\kappa_t(p_t)\Big]\Big(\gamma(p_t|p_2)-\gamma(p_t|p_1)\Big)dp_t\geq 0,$$

where I make use of $\int_{\underline{p}}^{\overline{p}}\gamma(p_t|p_2)-\gamma(p_t|p_1)dp_t=0$ by single-crossing. This last inequality shows that $S_{t-1}(p_2)\geq S_{t-1}(p_1)$ for $p_2>p_1$.

A.5 RME Contracts

This Appendix proves that the RME contract has the form given in (14). Before turning to the actual proof, I first show that the contract offer distribution, F_t, rewrites

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

where the substitution follows from the firm's vacancy posting Equation (13) 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

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

Finally, plugging this last expression into the contract offer distribution evaluated at $V(p)$ gives Equation (15)

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

To get (14), start from the first-order condition with respect to the optimal contract from (10) for active firms at some productivity level p

$$-q_t'(W_t(p))(S_t(p) - W_t(p)) = h_t(p).$$

The derivative of the quit rate is given by $q_t'(W) = -s\lambda_t F_t'(W)$ and, in a Rank-Monotonic Equilibrium, the derivative of the distribution of offers can be expressed from (15) as

$$\lambda_t F_t'(W_t(p))\frac{dW_t}{dp} = \frac{(1-\delta)h_t(p)l_t(p)}{u_t + s(1-\delta)\big(L_t(p) - L_t(p_E)\big)},$$

with $l_t(p) = dL_t(p)/dp$. Combining the last three expressions yields the following first-order differential equation in W_t

$$\frac{dW_t}{dp} + \frac{s(1-\delta)l_t(p)}{u_t + s(1-\delta)\big(L_t(p) - L_t(p_E)\big)} W_t = \frac{s(1-\delta)l_t(p)}{u_t + s(1-\delta)\big(L_t(p) - L_t(p_E)\big)} S_t(p)$$

with boundary condition $W_t(p_E) = U_t$. Noting that

$$\frac{d\ln\Big(u_t + s(1-\delta)\big(L_t(p) - L_t(p_E)\big)\Big)}{dp} = \frac{s(1-\delta)l_t(p)}{u_t + s(1-\delta)\big(L_t(p) - L_t(p_E)\big)},$$

the corresponding integrating factor is then

$$\exp \int \frac{s(1-\delta)l_t(p)}{u_t + s(1-\delta)\big(L_t(p) - L_t(p_E)\big)} dp = u_t + s(1-\delta)\big(L_t(p) - L_t(p_E)\big).$$

Along with the boundary condition, this yields (14) in the main text

$$W_t(p) = \frac{u_t U_t + s(1-\delta)\int_{p_E}^{p} S_t(p')dL_t(p')}{u_t + s(1-\delta)\big(L_t(p) - L_t(p_E)\big)}.$$

A.6 Derivations Net Surplus

This Appendix shows that the model can be recast in terms of a single value function by subtracting the unemployed worker's value function to the firm-worker surplus. I omit this notation from the main text not to clutter the description of the model. 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 $\phi_{t-1}(p) := \pi_{t-1}(p) + \overline{V} - U_{t-1} := S_{t-1}(p) - U_{t-1}$. Adding and subtracting U_t in (9), the firm-worker surplus can be rewritten

$$\begin{aligned} S_{t-1}(p) = p_{t-1}\omega_{t-1} + \beta E_{t-1}\Bigg[& U_t + \mu Q_t \\ & + (1-\mu)\Bigg(\chi_t(p_t)(1-\delta)\max_{\substack{W\\ h\geq 0}}\Big\{ - c(h) + \big(1 - q_t(W)\big)\phi_t(p_t) \\ & + h(\phi_t(p_t) - (W - U_t)) + s\lambda_t\int_{W}^{\overline{W}_t} W' - U_t dF_t(W')\Big\}\Bigg)\Bigg]. \end{aligned}$$

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

$$U_{t-1} = b + \beta E_{t-1}\Bigg[U_t + \mu Q_t + (1-\mu)\lambda_t\int_{U_t}^{\overline{W}_t} W' - U_t dF_t(W')\Bigg].$$

Let $V := W - U_t$ denote the value of the offered contract, net of the value of unemployment. Let $\tilde{F}_t$ be the corresponding distribution of contracts, so $F_t(W) = F_t(V + U_t) := \tilde{F}_t(V)$. Finally, let $\overline{V}_t := \overline{W}_t - U_t$ be the net contract with the largest value given the current realization of the states in period t. The net surplus can then be expressed as

$$\begin{aligned} \phi_{t-1}(p) &:= S_{t-1}(p) - U_{t-1} \\ &= p_{t-1}\omega_{t-1} - b + \beta(1-\mu)E_{t-1}\Bigg[\chi_t(p_t)\Bigg\{(1-\delta)\tilde{\psi}_t(p) - \lambda_t\int_0^{\overline{V}_t} V'd\tilde{F}_t(V')\Bigg\}\Bigg] \end{aligned} \tag{18}$$

where $\tilde{\psi}_t(p)$ is the firm's optimization problem in net surplus form

$$\tilde{\psi}_t(p) := \max_{\substack{V\\ h\geq 0}}\Bigg\{ - c(h) + \Big(1 - \tilde{q}_t(V)\Big)\phi_t(p) + h\Big(\phi_t(p) - V\Big) + s\lambda_t\int_{V}^{\overline{V}_t} V'd\tilde{F}_t(V')\Bigg\}.$$

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 (14), which gives

$$W(p, \omega, L) - U(\omega, L) := V(p, \omega, L) = \frac{s(1-\delta)\int_{p_E}^{p} \phi(p', \omega, L) dL(p')}{u + s(1-\delta)\Big(L(p) - L(p_E)\Big)}, \tag{19}$$

The optimal hiring rate can also be expressed as solving

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

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

A.7 Difference with Coles and Mortensen (2016)

The main difference between my approach and the model developed in Coles and Mortensen (2016) is in the wage setting protocol. I assume that firms can fully commit to delivering a state-contingent wage after each future realization of firm productivity and some aggregate outcomes, 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 models 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,

$$\pi_{t-1}\left(p_{t-1}, W_{t-1}\right) = \max_{\substack{w,W \\ h\geq 0}} \left\{ \omega_{t-1} p_{t-1} - w + \beta E_{t-1}\Big[- c(h) + (1 - q_t(W) + h)\pi_t(p_t, W) \Big] \right\}, \tag{20}$$

subject to the promise-keeping constraint

$$W_{t-1} = w + \beta E_{t-1} \left\{ \delta U_t + (1-\delta)\Big[(1 - q_t(W))W + s\lambda_t \int \max\left(W', U_t\right) dF_t(W')\Big] \right\}. \tag{21}$$

In this recursive formulation, full commitment on the firm side implies that it must deliver, in expectation, W_{t-1} when choosing the wage rate w and continuation values W. With risk-neutral workers, wages can be substituted out from (20) using (21), 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 Equation (20) in Coles and Mortensen (2016) is given by

$$\pi_{t-1}\left(p_{t-1}, w_{t-1}\right) = \omega_{t-1} p_{t-1} - w_{t-1} + \beta E_{t-1} \max_{\substack{w \\ h \geq 0}} \left\{ - c(h) + (1 - q_t(w) + h)\pi_t(p_t, w) \right\}, \qquad (22)$$

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.[24] They proceed by describing an equilibrium in which workers form beliefs on the productivity of firms based on the wage they offer and show that it is an optimal strategy for more productive firms to offer higher wages.

Equations (20) and (22) make clear that the firm trades off hiring new workers and retaining current employees in controlling its rate of employment growth. And both characterizations of equilibrium entail that workers move towards more productive firms, as they offer better jobs, when making a job-to-job move. But the characterization in Coles and Mortensen (2016) is obtained under stronger assumptions: (i) No difference in search effort between employed and unemployed workers, so $s = 1$ in the notation of my model; (ii) No endogenous firm entry and exit, which requires $\underline{p} > b$. While these restrictions can be relaxed, as their main purpose is to make the reservation wage equal the value of non-employment, b, numerically solving for the corresponding reservation wage in the general case with aggregate shocks is potentially very difficult, 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 7 shows the wage profile obtained by numerically solving a stationary 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 value of output in the bargaining protocol with firm commitment.

B Data Appendix

B.1 Firm-level data

I rely on three main sources of firm-level data: the Annual Respondents Database (Office for National Statistics, 2020b), it successor, 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) request 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, but it only makes available a

[24] 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 clearer.

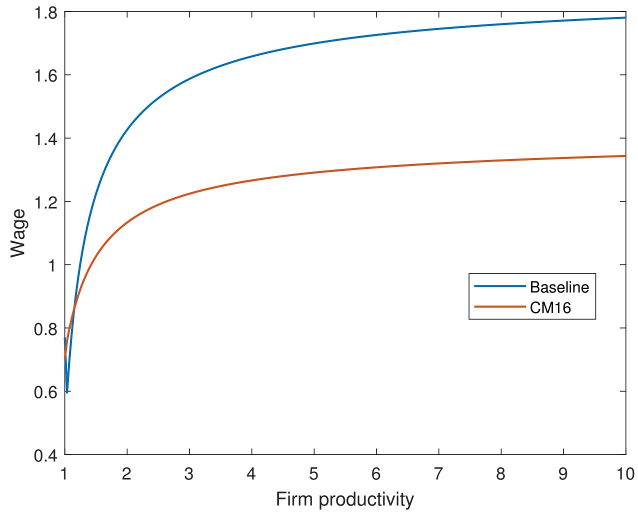


Figure 7: Wages in Baseline vs Coles and Mortensen (2016) model at the steady-state.

handful of variables (employment, estimated turnover, industry).[25]

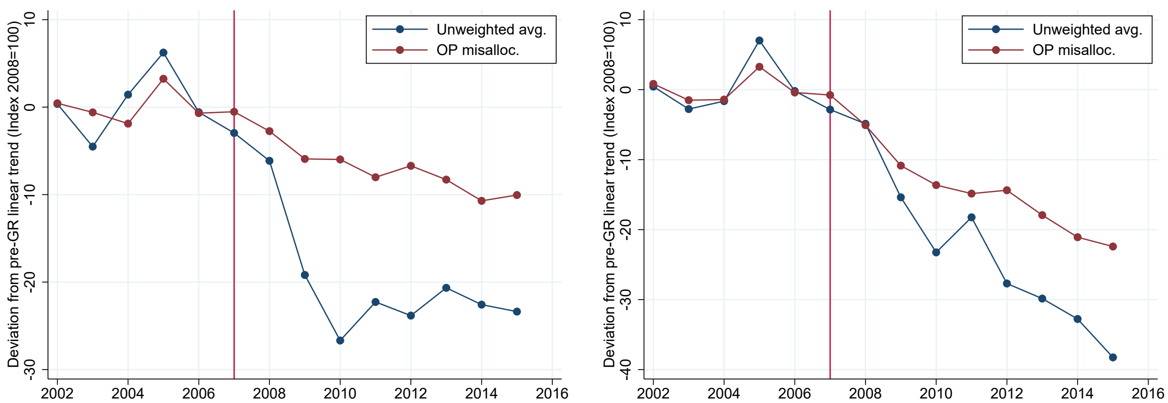
Since the Business Structure Database (BSD) does not have information on value added or employment costs, I follow the procedure in Riley et al. (2015) to obtain meaningful aggregates from the Annual Respondents Database (ARD)/Annual Business Survey (ABS). I use the "gross value added at factor costs" and "total employment costs" variables, which are harmonized across survey year 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.2 Labor market data

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). Because of the transition from the BHPS to UKHLS, there is a gap in the series between

[25] Businesses must satisfy one of two conditions to be included in these data. They must either have a sales turnover above the VAT registration threshold or have at least one employee. In practice, these restrictions imply that all but the smallest businesses and the self-employed are included in these data.



(a) Labor productivity　　**(b)** Employment cost per worker

Figure 8: OP decomposition in the aftermath of the Great Recession.

August 2008 and December 2009, which is smoothed over using moving averages.[26]

B.3 Additional macro series

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

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

B.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 8 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.

C Numerical Appendix

C.1 Stationary solution

As shown in Appendix A.6, the firm's policies can be expressed in terms of a single value function, the net surplus given in Equation (18). A Stationary Rank-Monotonic Equilibrium (see Definition

[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.

3) can similarly be defined as a fixed-point in the net surplus, $\phi = S - U$ and the measure of workers, L. The algorithm below is expressed in terms of the net firm-worker surplus formulation for concision.

Discretization. In a Rank-Monotonic Equilibrium, all heterogeneity in the model arises through p. I discretize the process for idiosyncratic log-productivity using Tauchen's procedure with $N_p = 401$ points. This yields a grid $\{p_1, \ldots, p_{N_p}\}$ grid and the associated transition matrix.

This discretization can be 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, L, by replacing the integral by the appropriate employment share weighted sum. For instance, the net optimal contract (19) at some productivity node p_k can be approximated as

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

where the integral in the last expression 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 of p for the net surplus. In practice, I also 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 Equation (19) and the firm's choice of hiring intensity from the corresponding FOC.

3. The net surplus equation and the law of motion for the measure of employed workers 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.

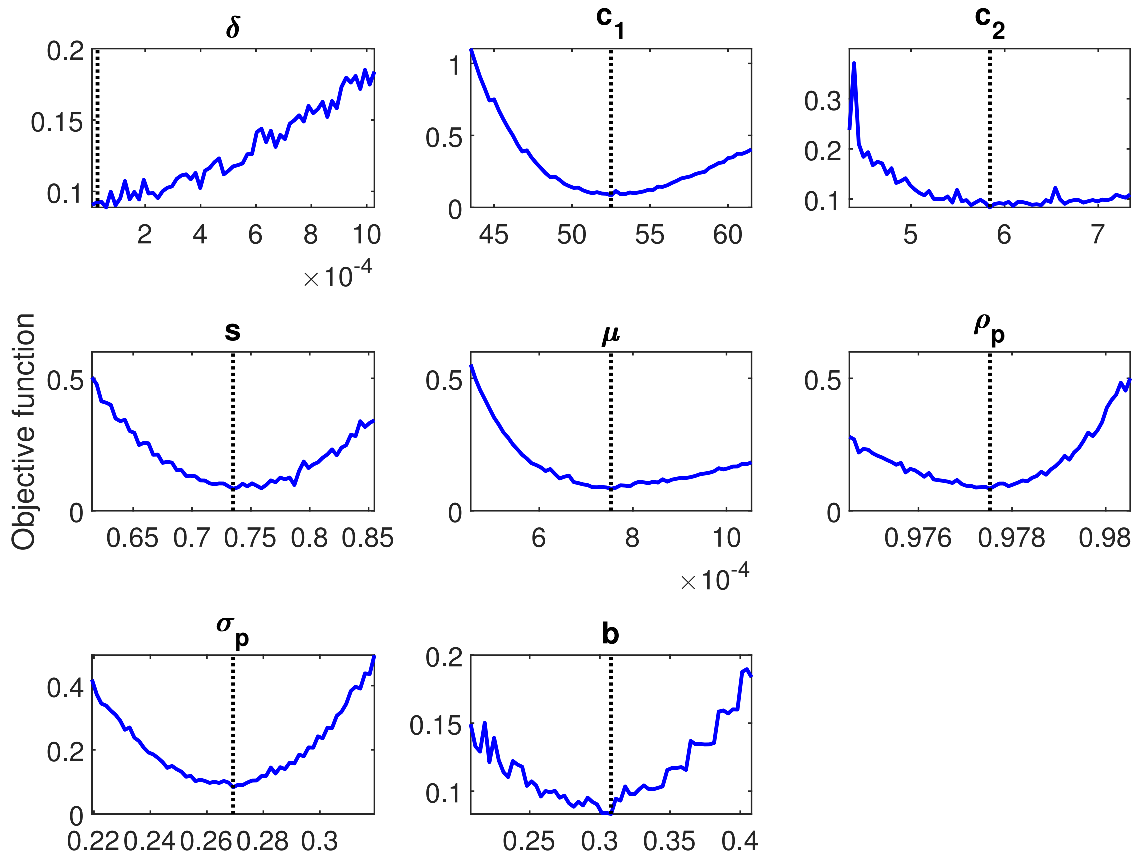


Figure 9: Slices of objective function for each parameter. The vertical dotted line is the corresponding parameter value in $\hat{\xi}$.

C.2 Calibration

The parameters are calibrated by targeting the moments listed in Table 2. They are obtained by solving the following minimization problem

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} \left(\mathbf{g}_{\text{data}} - \mathbf{g}_{\text{model}}(\xi)\right)^{\intercal} \mathbf{\Lambda} \left(\mathbf{g}_{\text{data}} - \mathbf{g}_{\text{model}}(\xi)\right)$$

where ξ denotes the parameter vector, $\mathbf{g}_{\text{data}}$ the vector of data moments, and $\mathbf{g}_{\text{model}}(\xi)$ the corresponding model generated vector of moments. Each moment is rescaled by the inverse of the square of its empirical value: $\mathbf{\Lambda} = \operatorname{diag}(1/\mathbf{g}^2_{\text{data}})$. Figure 9 further shows slices of the objective function around the estimated parameter values.

C.3 Aggregate shocks solution

As explained in the main text, the simulation algorithm in the presence of aggregate shocks relies on two approximations. First, the measure of employment at firms of different productivity is summarized by a set of $N_m + 1$ moments

$$
\begin{aligned}
m_t^0 &:= u_t = 1 - \int dL_t(p) \\
m_t^1 &:= \int p d\bar{L}_t(p) \\
&\dots \\
m_t^{N_m} &:= \int p^{N_m} d\bar{L}_t(p),
\end{aligned} \tag{23}
$$

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 choice of N_m in Appendix C.4.

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

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 (23) 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 values of $\{\theta_U, \theta_{p_1}, \dots, \theta_{p_{N_p}}\}$, 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 shock is discretized using Tauchen procedure with $N_\omega = 15$ 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 at times close to collinear during the procedure, I use ridge regression to regularize the problem. As an example, 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

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

I proceed 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. 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.

C.4 Number of moments in approximation

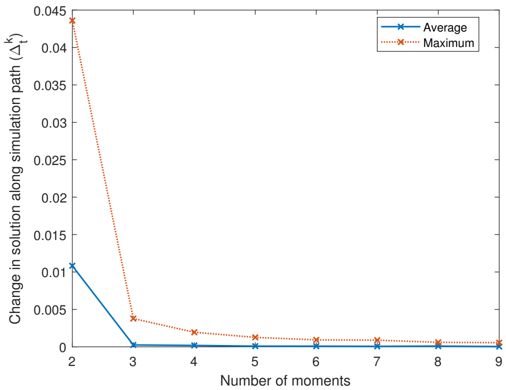
I assess the sensitivity of this solution method to the number of moments used in approximating L_t by means of the following test. I incrementally introduce up to $N_m = 9$ moments to summarize L_t, and solve the model using the same sequence of aggregate shocks each time. I can then compute a solution $\tilde{S}^k(p, \omega_t, \tilde{\boldsymbol{m}}_t; \theta_{p_i})$ 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 an additional moment k

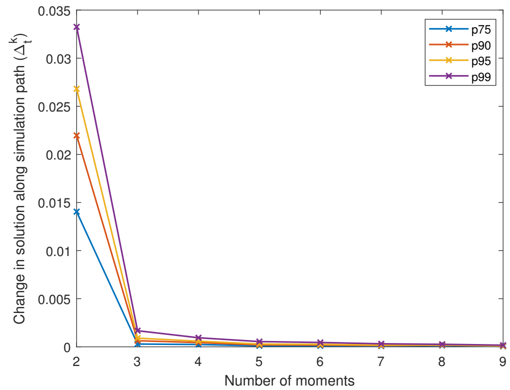
$$\Delta_t^k(p) := \left| \tilde{S}_t^k(p) - \tilde{S}_t^{k-1}(p) \right| = \left| \ln S_t^k(p) - \ln S_t^{k-1}(p) \right|$$

and similarly for $\tilde{U}_t^k$. Figure 10a reports the average and maximum $\Delta_t^k(p)$ along the simulated sequence of shocks as more moments are included. To check that this is not purely driven by outliers, Figure 10b further shows several percentiles of $\Delta_t^k(p)$.

These figures suggest that the largest changes in the approximation is obtained going from one to two moments ($k = 2$), with much smaller changes in the solution as additional moments are included. All results in the paper are for the solution $\{\theta_U, \theta_{p_1}, \ldots, \theta_{p_{N_p}}\}$ with $N_m = 2$.



(a) Average and Maximum



(b) Quantiles

Figure 10: Robustness to number of moments included in approximation.

	Absolute error (in %)				
Variable	Mean	p75	p90	p95	Max
Value Functions					
S_t	0.100	0.142	0.256	0.291	0.348
U_t	0.040	0.044	0.047	0.048	0.051
Moments L_t $(\boldsymbol{m}_t)$					
$m_t^0 (:= u_t)$	0.347	0.448	0.816	1.069	2.307
m_t^1	0.085	0.099	0.126	0.146	0.228
m_t^2	0.127	0.146	0.176	0.197	0.290

Table 8: Accuracy tests

C.5 Accuracy test

The accuracy of the procedure is assessed through the test 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 $\{\tilde{\omega}_\tau\}_{\tau=1}^T$, $S_t(p)$ and U_t can be obtained either using their respective approximation based on θ_{p_i} and θ_U, or computed directly solving the model backward in time and explicitly taking an expectation over $\tilde{\omega}_{t+1}$ in each period.

Table 8 reports these statistics for an alternative sequence of shocks, distinct from the one used to solve for the coefficients $\{\theta_U, \theta_{p_1}, \ldots, \theta_{p_{N_p}}\}$. I report the average and maximum absolute percent error between the approximation and explicit solutions, i.e. $100 \cdot (y_t^{\text{approx.}} - y_t^{\text{explicit}})$, taken at each point in time and each node, where $y_t^{\text{approx.}}$ denotes either $\tilde{S}(p, \tilde{\omega}_t, \hat{\boldsymbol{m}}_t; \theta_p)$ or $\tilde{U}(\tilde{\omega}_t, \hat{\boldsymbol{m}}_t; \theta_U)$. These tests suggest that the proposed algorithm is overall accurate, though even with a large number of grid points ($N_p = 401$), some small inaccuracy in the value functions can result in a slightly larger difference in the simulated unemployment rate (2.3% at most).