Equilibrium Job Turnover and the Business Cycle*

Carlos Carrillo-Tudela †
University of Essex,
CEPR, CESifo and IZA

Alex Clymo ‡
University of Essex

Melvyn Coles §
University of Essex

September 2022

Abstract

This paper develops and estimates a new equilibrium theory of unemployment, firm dynamics and on-the-job search over the business cycle. We investigate two seemingly unexplored facts. Firm job destruction is negatively correlated with cyclical unemployment and shares a similar cyclical correlation to job creation. We show that these dynamics explain why unemployment is highly persistent and so provide a new perspective on the behaviour of unemployment over the business cycle. Our model is rich enough to match a wide range of firm- and worker-level patterns in the cross-section, yet tractable enough to be estimated over the business cycle. A key success is that our framework jointly replicates the observed aggregate fluctuations in a both worker turnover and firm job flows. We show the importance of job destruction due to unreplaced quits in explaining why job destruction is procyclical and why unemployment is so persistent.

Keywords: Job search, Firm dynamics, Business cycle.
JEL: E24, E32, J62, J63.

*We would like to thank participants in seminars at the universities of Cambridge, Oxford, LMU Munich, FAU Nuremberg, VCU Virginia and the Cleveland FED for their comments and suggestions as well as participants in the 3rd Dale T. Mortensen Centre Conference "Labour Markets and Search Frictions", Essex SaM workshop, EEA-ESEM 2021, NBER Summer Institute Macro Perspectives 2021, and the MacCaLM conference, 2022. We would also like to thank Jason Faberman for sharing with us his data. The usual disclaimer applies.

†Correspondence: Department of Economics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: cocarr@essex.ac.uk.

‡Correspondence: Department of Economics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: a.clymo@essex.ac.uk.

§Corresponding author: Department of Economics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: mcole@essex.ac.uk.

1 Introduction

In their seminal work on unemployment dynamics, Mortensen and Pissarides (1994) focussed on the cyclical properties of aggregate job creation and job destruction, pointing out that job destruction flows are more volatile than job creation flows and that job destruction spikes in recessions (see Davis and Haltiwanger, 1992, and Figure 2a below). More recently Davis, Faberman and Haltiwanger (2012) use JOLTS/BED microdata to measure job creation and job destruction at the firm level as well as measuring hires, quits and layoffs. Using their data (updated to 2019) we highlight two seemingly unexplored facts.[1] Although layoffs steeply rise and fall early on in a recession, the overall job destruction rate is in fact procyclical; i.e. the estimated cyclical component of job destruction is negatively correlated with that of unemployment. The second fact is that job creation and job destruction share a similar degree of procyclicality – they typically increase together during economic recoveries. We show that these dynamics explain why unemployment is so persistent and so provide a new perspective on the cyclical behaviour of unemployment.

Based on the seminal contribution of Klette and Kortum (2004), we develop and estimate a new equilibrium theory of unemployment, firm dynamics and on-the-job search over the business cycle (see also Lentz and Mortensen, 2008, Coles and Mortensen, 2016, Audoly, 2021, and recent related works on equilibrium firm dynamics which include Schaal, 2017, Bilal, Engbom, Mongey and Violante, 2022, Elsby and Gottfries, 2022, and Elsby, Gottfries, Michaels and Ratner, 2021). Specifically we consider equilibrium hiring and firing by heterogenous firms, where some firms endogenously grow over time while others decline. There is optimal wage posting by firms where, following Burdett and Mortensen (1998), both employed and unemployed workers use optimal job search strategies. With also aggregate productivity shocks, the framework is consistent with previous work which finds the job ladder collapses in recession; i.e. recessions endogenously cause a steep decline in quit turnover (see Moscarini and Postel-Vinay, 2018). The model is rich enough to match key firm- and worker-level data in the cross-section, yet tractable enough to be estimated over the business cycle. A key success is that our framework jointly replicates the observed aggregate fluctuations in both worker turnover and firm job flows. Specifically the model reproduces (i) the cyclical properties of aggregate job creation and job destruction, (ii) the underlying distribution of employment growth rates across firms (by age and size), (iii) the resulting reallocation of workers across firms in the cross-section, (iv) a procyclical job ladder (hires and quits), and (v) the dynamics of unemployment.

An important new insight is the quantitative importance of an oft-overlooked job destruction channel: firms also destroy jobs when they do not replace workers who quit. To illustrate the importance of

[1] We thank Jason Faberman for kindly providing their updated time series. See Davis et al. (2012) for details of the construction of these series.

this job destruction channel consider Figure 1a, which plots the cyclical component of the job destruction and layoff rates. This figure confirms existing intuition that large spikes in job destruction coincide with large spikes in layoffs. Following these spikes, however, job destruction and layoffs diverge.[2] To reveal the underlying process, consider instead "job destruction net of layoffs"; obtained by subtracting layoffs from measured job destruction. Figure 1b shows the cyclical component of job destruction net of layoffs is not only strongly procyclical, it is directly correlated with the quit rate: when quits fall during recessions, so does job destruction net of layoffs and the opposite applies in booms. The fact that total job destruction is procyclical, while layoffs are countercyclical, then follows from the procyclical behaviour of job destruction net of layoffs.[3] We show this property of the data is consistent with the equilibrium job ladder dynamics in a Klette and Kortum (2004) framework where workers quit from [low paying] declining firms to [better paying] growing firms. Job destruction net of layoffs is then strongly procyclical because declining firms often destroy jobs by not replacing workers who quit, and worker quit rates are strongly procyclical.[4]

A second feature of the data is that the cyclical component of job creation is not much more procyclical than that of job destruction (see Table 2 in the text for actual estimates). To show this directly, Figure 2a plots the cyclical components of job creation and job destruction rates. As is well known, job creation rates are less volatile than job destruction rates over the cycle. However, *following* each job destruction spike, job creation and job destruction tend to increase together in the recovery. To demonstrate the economic impact of this interaction, we construct "net job creation" (henceforth njc) by subtracting job destruction flows from job creation flows. Figure 2b describes a scatter plot of the cyclical component of njc against unemployment to reveal the large, asymmetric hysteresis loops in US unemployment.

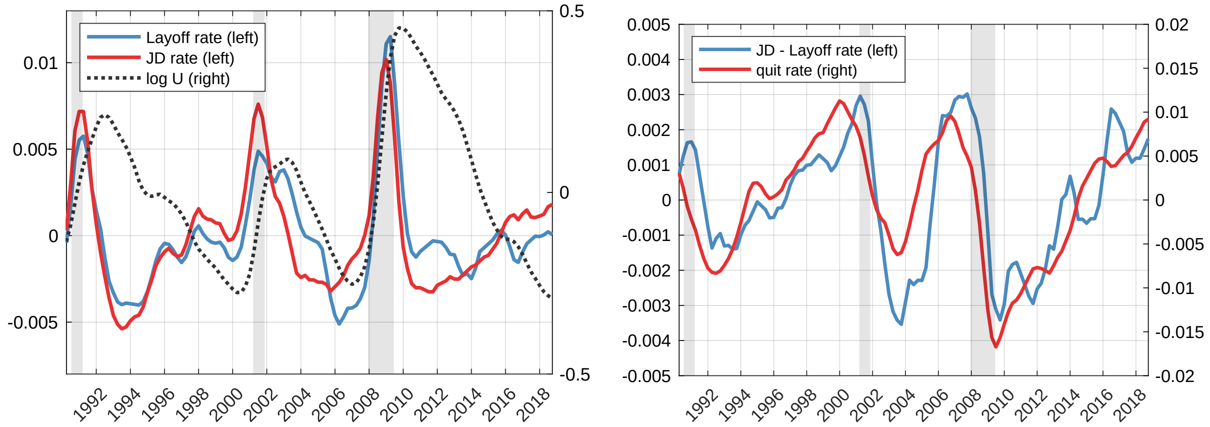
Figure 2a shows that each US recession coincides with a large spike in the job destruction rate and

[2] All flows are transformed into rates by dividing by aggregate employment. Layoffs measure the total number of employees laid off from jobs. Job destruction instead measures the number of jobs lost in the economy. This is computed as the sum of employment losses across all contracting establishments in the economy. See Davis, et al. (2012) for further details.

[3] In Appendix A.3 we provide a detailed discussion of the cyclical properties of job destruction along with extensive robustness. The procyclicality of job destruction is not a mechanical byproduct of the fact that it spikes at the beginning of recessions, as the same is true for layoffs, which are instead countercyclical. For the period 1990Q2 to 2018Q4, for example, we find that the correlation between cyclical job destruction and cyclical unemployment is -0.25, while for layoffs the same correlation is 0.17 (see Table 2 below). To obtain the cyclical components of the series we HP-filter the log quarterly time series with smoothing parameter of 10^5. We find the negative (positive) correlation between job destruction (layoffs) and unemployment remains robust to shortening the time period to 2010 (as in Davis et al., 2012) and alternative methods of filtering. We also discuss the distinction between measuring cyclicality by the correlation of flows with the *level* versus *change* in unemployment, as highlighted by Moscarini and Postel-Vinay (2012) and Haltiwanger et al. (2018).

[4] We estimate that, on average, firms replace 80% of workers who quit, and match this in our model (see Appendix B for details). This implies that only around 20% of total job destruction is due to the unreplaced quits channel in steady state. However, since quits are so volatile over the business cycle, this is sufficient for the unreplaced quits channel to play an important role in driving job destruction dynamics over the cycle.

Figure 1: Decomposing Job Destruction



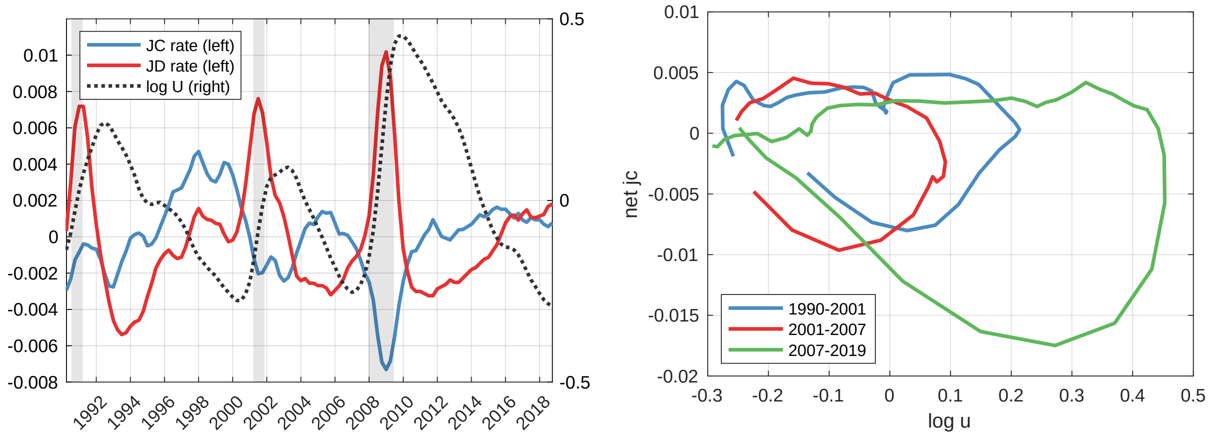
(a) Job Destruction and Layoffs (b) “JD net of layoffs” and quits

Note: Cyclical net job creation rate is constructed from the quarterly data used by Davis, Faberman and Haltiwanger (2012), updated by these authors, and HP filtered with parameter 10^5. Cyclical unemployment is constructed using quarterly data from the Current Population Survey and also HP filtered with parameter 10^5.

Figure 2b demonstrates the resulting steep drop in the njc rate (below trend) together with rapidly increasing unemployment. In the subsequent recovery (when njc is above trend), njc does not rise much above trend and unemployment recovers very slowly; i.e. high unemployment is persistent. The crucial observation, however, is that this slow recovery process occurs not because the job creation rate does not increase in the recovery, but rather the job destruction rate increase by a similar amount. The model reproduces these features of the data and shows that the slow recovery of unemployment is driven by job destruction due to unreplaced quits. Furthermore by directly crowding out the re-employment opportunities of the unemployed, the unreplaced quits process also explains why the job finding rate of the unemployed drops in the recession. Figure 2b thus makes clear that a data-relevant theory of the observed peristance and volatility of cyclical unemployment must be consistent with the underlying dynamic properties of job creation and job destruction.

Since job creation is not much more procyclical than job destruction, we also show that job creation flows are not well described by a free entry approach. Instead we adapt Coles and Moghaddasi (2018) who, in an otherwise standard Diamond-Mortensen-Pissarides (DMP) framework, estimate the elasticity of the vacancy creation rate with respect to firm value, where the free entry approach assumes this margin is infinitely elastic. Not surprisingly the latter case is strongly rejected by the data, but that does not imply an adjustment cost approach is appropriate. Rather Coles and Moghaddasi (2018) allow vacancies to evolve as a stock variable and find that (i) inventory stock dynamics play a central role in explaining short-run vacancy stock fluctuations, and (ii) an appropriately estimated vacancy creation elasticity resolves the well known persistence, volatility and Beveridge curve issues

Figure 2: Job Creation and Destruction dynamics



(a) Job Creation and Job Destruction

(b) Net job creation and unemployment

Note: Cyclical net job creation rate is constructed from the quarterly data used by Davis, Faberman and Haltiwanger (2012), updated by these authors, and HP filtered with parameter 10^5. Cyclical unemployment is constructed using quarterly data from the Current Population Survey and also HP filtered with parameter 10^5.

which are common to the standard DMP matching approach.

A technical difficulty of relaxing the free entry restriction in a stochastic equilibrium with heterogeneous firms is that it leads the distribution of unfilled vacancies to be a complex, high dimensional state variable. Both for clarity and tractability we abstract from vacancy stock dynamics by instead directly estimating job creation and job destruction elasticities relative to firm value. We formalise this approach by assuming search is still random and sequential but, reflecting that most vacancies are filled within a month,[5] we simplify by assuming vacancies are so very short-lived they effectively fill immediately. Essentially we assume trade is well approximated by stock-flow matching where the stock of unemployed workers [on the long side of a non-competitive labour market] chase the flow of new jobs created by firms, taking into account that employed workers also seek better paid employment. The firm's cost of filling a job is not the (frictional) time taken to find a worker, rather it is the substantial recruitment costs of screening job applicants and the training costs of new hires.[6] The crucial simplification when vacancies fill (almost) immediately is that the stock of (unfilled) vacancies across heterogeneous firms then plays no role. Using this simplification we find that the job creation rates of incumbent firms are slightly elastic to firm value, which in turn makes overall job creation respond relatively weakly to the cycle (as reflected by the data). Job creation by new entrants is more elastic, which is consistent with firm entry being more volatile in the data, but this has only a modest effect on

[5]See Coles and Smith (1998), Davis, et al. (2013), Carrillo-Tudela et al. (2022) and Mueller et al. (2022) for evidence on vacancy duration across several countries.

[6]This is consistent with the findings of Davis and Samaniego de la Parra (2020) who use data set on application flows and link it to online vacancy posting. They conclude that "the applicant gathering part of the search and matching process must be short compared to the parts devoted to screening, selection and post-offer recruiting".

aggregate job creation rates (at least in the short run) because firm entrants are small on average (both in the model and data).[7]

Related Literature The paper contributes to several important literatures. The first is to the equilibrium price dispersion literature where price posting firms set prices which are a best response not only to worker [or consumer] search strategies but also to prices set by competing firms; e.g. Diamond (1970), Burdett and Mortensen (1998) in what is a very large field. Because equilibrium is a fixed point argument in the space of wage distributions, only a handful of papers extend this price posting game to non-steady state [e.g. Coles (2001), Moscarini and Postel-Vinay (2012), Coles and Mortensen (2016), Audoly (2021)]. The central difficulty is that if wages depend on firm size, as is the case in Burdett and Mortensen (1998), then the distribution of employment across firms becomes a relevant state variable which is infinitely dimensional and evolves endogenously over the cycle. Using the Klette and Kortum (2004) approach we simplify this problem and identify stochastic equilibria where wage (and hiring) strategies are firm size independent and the distribution of employment across firms is no longer a relevant state variable. The aggregate state space reduces to the total measure of workers employed in firms by productivity and, with finite firm productivity types, the state space becomes finite and so highly tractable, even though wages and employment remain continuously distributed in equilibrium. The Gibrats' Law structure then implies the model generates a wide [Pareto] distribution of firm size consistent with the fact that most firms are small, yet most workers are employed in large firms. Furthermore the growth structure of firms implied by the estimated model generates the large-firm wage effect: that wages paid and firm size are positively correlated; e.g. Brown and Medoff (1989). Our approach also speaks to Bewley (2002): that cutting wages [relative to the market] triggers an excessive quit rate where this efficiency wage trade-off is only indirectly affected by the level of unemployment.

The paper also contributes to the rapidly growing literature on firm dynamics with on-the-job search; e.g. Coles and Mortensen (2016), Schaal (2017), Bilal et al. (2022), Elsby and Gottfries (2022), Audoly (2021) and Elsby et al. (2021). With the exception of Coles and Mortensen (2016) and Audoly (2021), this literature typically assumes there are decreasing returns to scale in production. This yields an easily understood steady state structure - that firms grow to the point where their marginal return to labour "equals" the wage. Because firm level wages are determined by bargaining, however, diminishing returns to labour implies wages depend on firm size and the extended aggregate state space in a stochastic equilibrium is then complex. Schaal (2017) solves this complication by using a directed search approach with a block recursive equilibrium structure. This allows him to

[7] This is consistent with the fact that young firms having the highest job creation rates in the economy, as shown by Haltiwanger et al. (2013). The direct job creation from firm entry (age 0 firms) is around 15% of total job creation in the Business Dynamics Statistics dataset, and firms aged 6 years and above account for 70% of job creation. These statistics are also reproduced by our model.

quantitatively analyse the full stochastic equilibrium. Bilal et al. (2022), Elsby and Gottfries (2022) and Elsby et al. (2021) follow instead a random search approach, but to investigate the non-steady state properties of their models they use a perturbation approach based on an unexpected aggregate shock (see Boppart et al., 2018). In contrast the properties of our model allow us to easily solve for the full stochastic equilibrium, while remaining consistent with a wide range of micro-level information on firm dynamics and worker turnover patterns. As in Bilal et al. (2022) our model is able to quantitatively reproduce the firm age and size distributions, the exit rates by firm age, the net poaching patterns by firm age (see Haltiwanger et al., 2017). As in Elsby and Gottfries, (2022), the model is also quantitatively consistent with the distribution of firm employment growth and the "hockey stick" relationships between firm employment growth and hires, quits and layoffs (see David et al., 2012), but our model also replicates their cyclical behaviour. In addition, the calibration of the full stochastic equilibrium reveals that our model remains consistent with the observed cyclical volatility and persistence of the aggregate worker and job flow series as well as of the unemployment rate (see also Schaal, 2017).

Our focus on job destruction due to unreplaced quits complements the recent literature which emphasises replacement hires; e.g. Faberman and Nagypál (2008), Mercan and Schoefer (2020) and Elsby et al. (2021). To the best of our knowledge showing that job destruction due to unreplaced quits is an important determinant behind unemployment fluctuations is novel.[8] Elsby et al. (2021) provide empirical evidence on the importance of replacement hiring and net inaction using US firm level data. An important contribution of their paper is to show that replacement hires increase the persistence of unemployment in response to an unexpected productivity shock.[9] Our analysis differs from Elsby et al. (2021) in that we incorporate endogenous firm entry and exit, which allows us to study how the job ladder is directed by age. This feature, along with our tractable out-of-steady-state model, also allows us to study how job creation, firm entry, layoffs, and endogenous job destruction through unreplaced quits co-move over the business cycle to shape unemployment fluctuations. We therefore provide a new perspective on the drivers of aggregate unemployment using firm-side job creation and job destruction flows, and show how they relate to and remain fully consistent with worker-side analysis of the "ins and outs of unemployment" (Shimer, 2012) which shows that unemployed job finding rates are the main driver of cyclical unemployment rate in the US data.

[8]Barlevy (2002) shows how the cyclical job ladder generates a sullying effect on aggregate productivity during recessions. While complementary, our focus is different as we allow for replacement hiring and show how unreplaced quits drive overall job destruction and unemployment dynamics.

[9]Faberman and Nagypál (2008) provides an early analysis of replacement hires using the standard search and matching model, but focus on a steady state analysis. Mercan and Schoefer (2020) investigate replacement hires using German data based on the Job Vacancy Survey and use a simple search and matching model with exogenous job-to-job transitions to explore its aggregate properties.

The paper is structured as follows. Section 2 describes the model, while Section 3 derives the stochastic equilibrium. Section 4 details the estimation of the model and provides the steady-state results. In Section 5 we discuss the business cycle implications of our model. Section 6 concludes. Proofs and estimation details are relegated to several online appendices.

2 The Model

Time t is continuous, has an infinite horizon, and we consider a stochastic equilibrium with aggregate shocks. There is a unit measure of equally productive workers who are risk neutral, infinitely lived and each has the same discount rate $r > 0$. At any point in time each worker is either employed (earning a wage w) or unemployed (with home production $b > 0$). Unemployed workers receive job offers according to a Poisson process with time varying parameter λ_{0t}, where a job offer is considered a random draw from the set of hiring firms. There is on-the-job search where employed workers receive job offers at rate $\lambda_{1t} = \phi\lambda_{0t}$ where $\phi \in (0, 1]$ is an exogenous parameter. In what follows $\lambda_{0t}, \lambda_{1t}$ are endogenous objects where a Markov equilibrium determines $\lambda_{0t} = \lambda_0(\Omega_t)$ and $\lambda_{1t} = \phi\lambda_0(\Omega_t)$ with Ω_t describing the aggregate state.

Firms are heterogeneous, risk neutral and have the same discount rate $r > 0$. For ease of exposition we initially assume a continuum of firm productivity states $i \in [0, 1]$ but in the application shall restrict to finite states (see Appendix B.3 for details). There are constant returns to production: given aggregate productivity $s \in \{1, 2, .., S\}$, firm $i \in [0, 1]$ with integer $n \in \mathbb{N}^+$ employees generates flow revenue $np^s(i)$ which is strictly increasing in i and s.

Worker job search is random and sequential but rather than adopt the standard matching function approach, we instead suppose stock-flow matching where the inflow of new jobs [on the short-side of the non-competitive labour market] matches immediately with workers on the long-side; e.g. Shapiro and Stiglitz (1984), Coles and Smith (1998) among many others. The formal assumption is that by paying hiring cost c_0 the firm immediately and randomly fills a job from the set of workers who prefer this job to their current position (taking into account on-the-job search effectiveness $\phi < 1$). The parameter c_0 here describes the direct recruitment and training costs of hiring a new employee. On the other side of the market, a positive stock of unemployed workers and a finite flow of new jobs implies it takes time for unemployed workers to get a job.

Similar to Burdett and Mortensen (1998) and as motivated in Bewley (2002), firms post wages as a best response to employee quit strategies but, in a stochastic equilibrium, we follow Coles and Mortensen (2016) where instead firm productivity is private information and subject to shocks. Assuming no job fees[10] and no precommitment on future wages and employment, each firm pays a sequence of spot wages to its employees where, in a Bayesian equilibrium, the firm's posted wage is a signal of its productivity i. Of course a signalling equilibrium implies the firm's wage strategy potentially depends on its entire wage posting history. Furthermore different employees are hired at different dates and so observe different parts of that wage history. Here we consider only (Markov Bayesian) equilibria which have the *job ladder property: firms with higher productivities i pay higher wages* $w(i,.)$. An additional important assumption is firm productivity follows a positively autocorrelated Markov process. Along the equilibrium path, the job ladder property guarantees workers only quit from low wage paying firms to higher wage paying firms because they believe the higher wage firm is the more productive and expect higher wages in the entire future [in the sense of first order stochastic dominance]. Should a firm cut its wage, its employees believe it has received an adverse productivity shock and, anticipating lower wages in the future, worker quit rates increase. Because replacing a worker who quits is costly, each firm trades-off paying lower wages against a higher quit rate which, in equilibrium, finds higher productivity firms do indeed pay higher wages. We assume there is no recall of rejected job offers and adopt the tie-breaking conventions that the worker quits when indifferent and that the firm invests when indifferent.

The framework considers such job ladder dynamics in a stochastic equilibrium with aggregate productivity shocks and in a rich Klette and Kortum (2004) framework with micro-firm growth dynamics and endogenous firm entry and exit. For ease of exposition, we assume each new start-up has initial employment level $n = 1$, though for the application we shall endogenise this (because the data find many start-ups instead have $n > 1$. See Appendix B.2). Conditional on surviving to date t, a firm with productivity $i_t = i \in [0, 1]$ and employment $n_t = n \in \mathbb{N}^+$, is subject to a wide variety of shocks:
(i) **Aggregate productivity shocks:** given current state $s_t = s \in \{1, 2, .., S\}$, an aggregate productivity shock occurs at exogenous rate $\alpha_a \geq 0$ where transition matrix $\Upsilon_{ss'}$ describes the probability the new state is s';
(ii) **Firm specific productivity shocks:** at exogenous rate $\alpha_\gamma \geq 0$ a firm with current productivity i has new productivity $i' \in [0, 1]$ considered a random draw from c.d.f. $\Gamma(i'|i)$. For the moment we shall assume no mass points in $\Gamma(.)$ but shall relax this for the application. The transition probabilities $\Gamma(.)$ satisfy first order stochastic dominance so that higher state i firms are more likely to be higher

[10]Although a standard assumption, this restriction is an important issue. For example the efficient contract in the Burdett and Mortensen (1998) framework is to charge a job fee and then pay wage equal to marginal product. But here as well, firm productivity is private information and so the efficient contract is not enforceable. Indeed the information asymmetry generates a lemons problem: firms which are more likely to close have a greater incentive to collect job fees (for they don't expect to repay the money). A simple (Bayesian) motivation for the "no job fees" assumption is that should any [deviating] firm ask for a job fee, the worker believes the job has sufficiently low value [short-lived and possibly fraudulent] that he/she will not earn back the downpayment and so rejects the offer.

productivity firms in the entire future;

(iii) **Firm level job creation** follows a Klette and Kortum (2004) type growth process. An expansion opportunity, say to open a new product line, occurs at firm (i, n) according to a Poisson process with parameter $\mu_1 n$ where $\mu_1 > 0$ is the same for all firms. Associated with the expansion opportunity is an idiosyncratic (sunk) capital investment cost $c^{JC} \geq 0$ considered a random draw from cdf $H^{JC}(.)$. If the firm invests, it pays c^{JC} and so creates an unfilled job. The post is then immediately filled by paying the recruitment cost c_0 whereupon the firm's size increases to $n + 1$; i.e. a new job is created.[11] If the firm rejects the expansion opportunity, say the investment is too costly, its firm size n remains unchanged. The important role played by this investment structure is that firms with greater value are more likely to create new jobs and $H^{JC}(.)$ then determines the elasticity of job creation with respect to firm value;

(iv) **Firm level downsizing:** capital is a one-hoss shay whereby each unit is randomly and independently destroyed at an exogenous rate δ_D. If a capital unit is destroyed, the firm can re-invest at cost $c^{JD} \geq 0$ considered a random draw from $H^{JD}(.)$. If the firm re-invests the firm's size n is unchanged. If the firm does not re-invest, the capital is lost, the corresponding employee is laid-off into unemployment and the firm downsizes to $n - 1$; i.e. one job is destroyed. This investment structure determines the elasticity of job destruction to firm value;

(v) **Job ladder quits:** employees may receive a preferred outside offer and so quit. Whenever a quit occurs, the firm has the option to pay recruitment cost c_0 and hire a replacement employee. If it does so, firm size n remains unchanged. If instead the firm chooses not to hire a replacement employee, the unreplaced quit implies firm size falls to $n - 1$; i.e. a job is destroyed via an unreplaced quit;

(vi) **Exogenous firm exit shocks:** at exogenous rate δ_F a firm experiences an exit shock and closes down with all jobs destroyed;

(vii) **Exogenous separations:** an employee separates into unemployment at exogenous rate λ_u and the firm then decides whether or not to hire a replacement at cost c_0. For ease of exposition, the theory section considers exogenous separations as quits. Because in the data it is ambiguous whether such a separation is a layoff or a quit, the quantitative section instead calibrates λ_u to match the average layoff rate of firms.

There is endogenous firm entry and exit. We assume a unit measure of entrepreneurs independently seek new business ventures. At rate μ_0 an entrepreneur identifies a possible business venture whose investment cost $c^E \geq 0$ is considered an independent random draw from cost distribution $H^E(.)$. If the entrepreneur chooses not to invest, the venture is lost with no recall. If the entrepreneur invests, a new start-up firm is created with a single employee drawn randomly and costlessly from the pool of

[11] We are assuming a Leontieff production function where capital is sunk.

unemployed workers; i.e. the investment cost c^E includes the cost of the first hire. Each new start up is described by $(i, n) = (i_0, 1)$ with initial productivity $i_0 \sim U[0, 1]$.

Firm exit occurs exogenously via firm destruction shocks δ_F and endogenously whenever a firm declines to size $n = 0$ [which is an absorbing state]. This firm turnover process is consistent with the fact that most firms which exit are indeed small. Although firms might also close down if their value becomes negative, such closures do not occur in the empirical application.

Some Preliminary Comments: Events (iii)-(iv) describe hold-up problems at the job creation and job destruction margins. Outside of a competitive equilibrium, an optimal contract might require employees to contribute to [firm specific] re-investment costs. We rule out direct worker contributions to firm investment and so avoid a complicated negotiating problem where such investments may not be observable/verifiable by workers. The framework instead adopts a standard hold-up structure: the firm either immediately invests or the opportunity is lost. That is not to say wages are unaffected by such costs, for the re-investment process generates a positive user cost of capital which reduces match surplus. The hold-up problem essentially implies wages paid reflect the [expected] user cost of capital rather than specific cost realisations.

Because the start-up process directly recruits one unemployed worker, it is convenient to consider that recruitment channel separately from the analysis that follows. Of course we take all recruitments into account when describing gross flows (see (21) below). Throughout we distinguish between rates and flows by using lower case to describe rates; e.g. $jc(i, \Omega)$ will denote the job creation rate per employee at firm (i, n, Ω), and upper case $JC(i, \Omega)$ will denote the aggregated job creation flow across all firms in state i.

3 Equilibrium

The equilibrium framework is rich but also complex because wages are determined by a dynamic signalling game with repeated trade, aggregate productivity shocks and (privately observed) firm specific productivity shocks, where the set of wage strategies must be a best response to aggregate dispersion in wages and worker quit strategies, while all agents are also forward looking over the cycle. However signalling equilibria which have **the job ladder property**, where higher productivity firms (in equilibrium) post higher wages, are surprisingly tractable. To simplify the exposition we first adopt the Coles and Mortensen (2016) insight that in a constant-returns-to-scale Klette and Kortum (2004) type growth framework, the equilibrium wage posting strategies of firms are firm size-independent. With that simplifying property in hand, we first fully describe the optimal job creation/destruction decisions of firms at the micro-level. Given those investment choices, we then solve the aggregation problem and provide *analytic* solutions for gross turnover flows; e.g. job creation, job destruction, quits and hires over the cycle. The *stochastic equilibrium* is then fully determined by characterising the set of optimal wage offer strategies across firms and over the cycle, and verifying those strategies are indeed consistent with the job ladder property and firm size invariance. Equation (19) below describes the equilibrium functional equation for the set of firm values by type $i \in [0, 1]$ in each aggregate state Ω_t. Along with a description for how Ω_t varies over the cycle, the solution to the functional equation not only determines firm-level growth rates, the aggregation problem yields closed form solutions for gross flows over the cycle. The application with finite productivity states further finds Ω_t reduces to a finite vector and the stochastic equilibrium is directly computed using standard recursive methods.[12]

Let U_t denote the measure of workers who are unemployed at date t and let $G_t(i)$ denote the fraction of employed workers at firms no greater than $i \in [0, 1]$. For ease of exposition we assume $G_t(.)$ has a connected support and that its density exists. Because optimal wages here are firm size independent, the aggregate state is $\Omega_t = (s_t, U_t, G_t(.))$ and, in a (Markov Bayesian) equilibrium, each firm is fully described by (i, n, Ω). It is important to note this approach is not inconsistent with the large firm wage effect because wages and firm growth are correlated processes; e.g. new start-ups begin small and those with low productivity not only pay low wages they remain small. Also note that without firm size independence in wages, the aggregate state must then include the distribution of employment across firms of type i and this joint distribution function $G = \widetilde{G}(i, n)$ then makes the stochastic equilibrium far more complex; e.g. Coles (2001), Moscarini and Postel-Vinay (2013), Audoly (2021), and this issue also applies to stochastic equilibria with instead decreasing returns to scale. Finally note the quit turnover patterns are very different to Burdett and Mortensen (1998): here employees quit from declining to growing firms rather than from small to large firms and it is this property which allows the estimated framework to match well firm-level job turnover.

The following describes the structure of the stochastic equilibrium and our notation:

Definition 1 (**Stochastic equilibrium**). *For $\Omega = (s, U, G(.))$, a stochastic [Markov Bayesian] equilibrium is the following set of functions. For each firm (i, n, Ω)*

1. $w(i, \Omega)$ *is the profit maximising wage strategy which is size independent and increasing in i;*
2. $jc(i, \Omega) \geq 0$ *is the profit maximising job creation rate per employee, and so $n[jc(i, \Omega)]$ describes the expected gross job creation flow;*
3. $jd(i, \Omega) \geq 0$ *is the profit maximising job destruction rate per employee, and so $n[jd(i, \Omega)]$ describes its expected gross job destruction flow;*

[12] This occurs even though the endogenous wage and employment distributions remain continuous in equilibrium as in the Burdett and Mortensen (1998) model.

4. *$h(i, \Omega)$ is the optimal hiring rate per employee, and so $n[h(i, \Omega)]$ describes its expected gross hire flow;*

New start up entry:

5. *$P^E(\Omega)$ is the probability an entrepreneur invests in a start-up in state Ω, and so $\mu_0 P^E(\Omega)$ describes the inflow of new start-ups with $n = 1$ and $i \sim U[0, 1]$;*

Worker search:

6. *$F(w, \Omega)$ is the distribution of wage offers across hiring firms;*
7. *$\lambda_0(\Omega)$ and $\lambda_1(\Omega) = \phi\lambda_0(\Omega)$ are the job offer arrival rates for unemployed and employed workers respectively;*
8. *$\widehat{q}(w', \Omega)$ is the quit rate of a worker employed at a firm paying wage w' and $q(i, \Omega) = \widehat{q}(w(i, \Omega), \Omega)$ is the firm's quit rate along its equilibrium path;*
9. *Employed and unemployed workers use optimal job search strategies to maximise expected lifetime value where, given current wage paid w', the worker's belief on the firm's underlying state i is consistent with the set of equilibrium wage strategies and Bayes' rule;*

Markov restriction:

10. *Ω follows a first order Markov process consistent with the equilibrium strategies of firms, workers and entrepreneurs.*

In any stochastic equilibrium, firm (i, n, Ω) has expected lifetime value $\Pi(i, n, \Omega)$ which satisfies the following Bellman equation:

$$
\begin{aligned}
r\Pi(i, n, \Omega) = \max_{w'} \quad & n[p^s(i) - w'] \\
& + n\widehat{q}(w', \Omega) \max[\Pi(i, n-1, \Omega) - \Pi(i, n, \Omega), -c_0] \\
& + \mu_1 n \mathrm{E} \max[\Pi(i, n+1, \Omega) - \Pi(i, n, \Omega) - [c_0 + c^{JC}], 0] \\
& + \delta_D n E \max[\Pi(i, n-1, \Omega) - \Pi(i, n, \Omega), -c^{JD}] \\
& + \delta_F[-\Pi(i, n, \Omega)] \\
& + \alpha_\gamma \int_0^1 [\max[\Pi(j, n, \Omega), 0] - \Pi(i, n, \Omega)] \, d\Gamma(j|i) \\
& + \alpha_a \sum_{s'} \Upsilon_{ss'} [\Pi(i, n, \Omega(s')) - \Pi(i, n, \Omega(s))] + \frac{\partial \Pi(i, n, \Omega)}{\partial t}.
\end{aligned} \tag{1}
$$

Given the firm's posted wage w', the firm's flow value equals its flow profit, plus the capital gains which arise when (i) a quit occurs (where the firm has the option of paying c_0 to hire a replacement), (ii) an expansion opportunity occurs with cost $c^{JC} \sim H^{JC}$, (iii) a downsizing shock occurs with cost $c^{JD} \sim H^{JD}$, (iv) a firm exit shock occurs, (v) a firm specific productivity shock occurs, (vi) an aggregate productivity shock occurs and the final term is shorthand for describing the change in $\Pi(.)$ as the state variables $(U_t, G_t(.))$ evolve endogenously over time. Equation (1) demonstrates the firm's optimal wage w' is a direct trade off between reduced profit flow and reduced quit flow $\widehat{q}(.)$.

Suppose for the moment we know the firm's optimal wage strategy $w(i, \Omega)$ (see Proposition 2 below). Substituting this optimal wage strategy $w' = w(i, \Omega)$ into (1) then describes the value of each firm (i, n, Ω) *along its equilibrium path.* The constant returns structure implies $\Pi(i, n, \Omega) \equiv nv(i, \Omega)$, where (1) implies $v(.)$ is defined recursively by:

$$\begin{aligned}(r + \alpha_\gamma + \alpha_a + \delta_F)v(i, \Omega) = {} & p^s(i) - w(i, \Omega) \\ & - q(i, \Omega)\min[v(i, \Omega), c_0] \\ & + \mu_1 E \max[v(i, \Omega) - [c_0 + c^{JC}], 0] \\ & - \delta_D E_c \min[v(i, \Omega), c^{JD}] \\ & + \alpha_\gamma \int_0^1 \max[v(j, \Omega), 0] d\Gamma(j|i) + \alpha_a \sum_{s'} \Upsilon_{ss'}[v(i, \Omega(s'))] + \frac{\partial v(i, \Omega)}{\partial t},\end{aligned} \tag{2}$$

where each line corresponds directly to its equivalent in (1). For example the second line finds if an employee quits, the firm either pays for a replacement worker or destroys the job depending on whichever option is more profitable. Firm value $v(i, \Omega)$ is the key object in what follows for it determines the optimal job creation and job destruction choices at every firm (i, n, Ω):

1. **job creation:** if an expansion opportunity arises, the firm creates a new job if and only if $v(i, \Omega) \geq c^{JC} + c_0$, and this occurs with probability $H^{JC}(v(i, \Omega) - c_0)$;

2. **job destruction through layoffs**: if a capital destruction shock occurs, the firm destroys the job and lays off the worker if and only if $v(i, \Omega) \leq c^{JD}$, which occurs with probability $1 - H^{JD}(v(i, \Omega))$;

3. **job destruction through unreplaced quits**: if an employee quits, the firm pays recruitment cost c_0 to hire a replacement worker if and only if $v(i, \Omega) \geq c_0$. Otherwise the worker is not replaced and the job is destroyed.

These three rules describe optimal firm level job creation and job destruction which not only depends on current firm productivity i, but also on the aggregate state Ω (booms and busts). $H^{JC}(v(i, \Omega) - c_0)$

describes how firm level job creation rates respond to variations in firm value $v(.)$ which, when aggregated, then determines the elasticity of gross job creation flows over the cycle. $H^{JD}(v(i,\Omega))$ plays the same role but at the job destruction margin. By relaxing the free entry restriction, estimation of the stochastic equilibrium identifies the elasticity of job creation and destruction across firms and over the cycle. Moreover, because Ω evolves endogenously over the cycle, estimation takes into account not only the cleansing effect of recessions [low productivity firms are most likely to layoff workers in the recession] but also the cleansing effect of recoveries [low productivity firms are unlikely to replace workers who quit in the recovery]. Or differently put, the job ladder property implies worker turnover is productivity enhancing (workers quit from low to high productivity firms), but the reallocation process, layoffs into unemployment or direct job-to-job quits, varies endogenously over the cycle (especially should the job ladder collapse in recessions as shown in Moscarini and Postel Vinay, 2018).

Because equilibrium firm values $v(i,.)$ are increasing in i [$p^s(i)$ is increasing in i in every state and $\Gamma(.|i)$ satisfies first order stochastic dominance], there are two important margins. First a firm closure threshold $i^c(\Omega)$ potentially occurs where $v(i^c,\Omega) = 0$ and so firms with $i < i^c$ might close down (rather than have negative value). An important possibility, however, would be to allow temporary layoffs wherein workers in firms $i < i^c$ are laid-off, but the firm's capital is not destroyed (or possibly maintained at some cost) and employees recalled should i or s subsequently improve e.g. Fujita and Moscarini (2017). Although temporary layoffs are important in the US, here we abstract from this process by ensuring $v(0,.) > 0$ in the application and so hereon consider the case in which $i^c = 0$.

The most important margin for what follows is the firm hiring margin $i^h(\Omega)$ defined where $v(i^h,\Omega) = c_0$. Firms with $0 \leq i < i^h$ have $0 < v(i,\Omega) < c_0$ and so neither replace workers who quit nor invest in expansion opportunities. Such firms are in decline and have a zero hiring rate. Together these results yield the following description of job creation and destruction at the firm level.

Proposition 1 (Job Creation and Job Destruction at the Firm Level). *A stochastic equilibrium implies:*

(i) hiring firms with $i \geq i^h(\Omega)$ have:

$$\begin{aligned} jc(i,\Omega) &= \mu_1 H^{JC}(v(i,\Omega) - c_0) \\ jd(i,\Omega) &= \delta_D[1 - H^{JD}(v(i,\Omega))] \end{aligned}$$

and because these firms replace workers who quit, their hiring rate is

$$h(i,\Omega) = jc(i,\Omega) + q(i,\Omega); \tag{3}$$

(ii) for non-hiring firms with $i \in [0, i^h)$,

$$\begin{aligned} jc(i,\Omega) &= 0 \\ jd(i,\Omega) &= \delta_D[1 - H^{JD}(v(i,\Omega))] + q(i,\Omega) \end{aligned}$$

because these firms do not replace workers who quit, and hires $h(i,\Omega) = 0$.

(iii) firm entry,

$$P^E(\Omega) = H^E(Ev(\Omega))$$

where $Ev(\Omega) = \int_0^1 v(i,\Omega))di$ is the expected value of a new start-up.

We now turn to the aggregation problem.

3.1 Aggregate Turnover, Job Offer Rates, and the Job Ladder

Aggregation is complicated by the job ladder process. To illustrate suppose the highest productivity firm $i = 1$ creates a new job and the job ladder property implies this firm posts the highest wage $\overline{w} = w(1,\Omega)$. This job offer not only attracts the unemployed, it attracts every employed worker at firms $i' \leq 1$. If an employed worker gets the job, then one of two things happens. If the worker quits from a firm $i' \geq i^h$ that firm immediately hires a replacement but at a lower wage $w(i',\Omega) < \overline{w}$. Such a quit does not crowd out the re-employment rates of the unemployment [because an unfilled job remains available] but it does crowd out re-employment wages [the replacement job is a lower wage $w(i',\Omega) < \overline{w}$]. Thus on-the-job search directly crowds out the re-employment wages of the unemployed. If instead $i' < i^h$ the firm does not hire a replacement worker which then directly crowds out the job finding rates of the unemployed for there is no longer an unfilled job post.

Proposition 1 implies the equilibrium flow of hires at each existing firm (i, n, Ω) is $nh(i,\Omega) = n[q(i,\Omega) + jc(i,\Omega)]$ if $i \geq i^h$, and 0 otherwise. Because $[1-U]G'(i)$ is the measure of workers employed at type i firms, aggregation over those firms implies the gross hire flow at type i firms is

$$H(i,\Omega) = \begin{cases} 0 & \text{if } i < i^h \\ [1-U]G'(i)[q(i,\Omega) + jc(i,\Omega)] & \text{if } i \geq i^h. \end{cases} \tag{4}$$

Although job offers are randomly made, $\phi < 1$ implies the employed receive relatively fewer offers. Let $\lambda = U\lambda_0 + (1-U)\lambda_1$ denote the total flow of job offers made and so fraction $\alpha \equiv [U\lambda_0]/\lambda = U/(U + \phi(1-U))$ of job offers go to the unemployed, the remaining $1-\alpha$ go to employed workers. Note that $\alpha(U) = U/(U + \phi(1-U))$ is increasing in U; i.e. employed workers are less likely to receive job offers when unemployment U is high.

To determine λ, note the job ladder property implies an equilibrium job offer by firm $i \geq i^h$ is only accepted by unemployed workers and those employed at firms $i' \leq i$. Hence random contacts implies a job offer by hiring firm i is accepted only with probability $\alpha + (1-\alpha)G(i)$. Given state i firms have gross hiring flow $H(i,\Omega)$, their corresponding flow of job offers must then be $H(i,\Omega)/[\alpha+(1-\alpha)G(i)]$ where the denominator takes into account that not all job offers are accepted. Hence aggregating across the flow of job offers across all firms yields total flow of job offers:

$$\lambda(\Omega) = \int_0^1 \frac{H(i,\Omega)}{\alpha + (1-\alpha)G(i)} di. \tag{5}$$

To determine equilibrium quit rates, note now that conditional on receiving a job offer, $\frac{1}{\lambda(\Omega)} \int_i^1 \frac{H(j,\Omega)}{\alpha+(1-\alpha)G(j)}$ describes the probability the offer is from a $j \geq i$ firm. Hence in any equilibrium with the job ladder property, the equilibrium quit rate at a type i firm is

$$\begin{aligned} q(i,\Omega) &= \lambda_u + \frac{\lambda_1}{\lambda} \int_i^1 \frac{H(j,\Omega)}{\alpha + (1-\alpha)G(j)} dj & (6) \\ &= \lambda_u + \phi \frac{\alpha(U)}{U} \int_i^1 \frac{H(j,\Omega)}{\alpha(U) + (1-\alpha(U))G(j)} dj. & (7) \end{aligned}$$

because an employee at firm i only quits to outside offers $j \geq i$. Furthermore $\alpha(U)/U$ decreasing in U now implies worker quit rates are directly crowded out by higher unemployment U.

Armed with these expressions we can now solve the aggregation problem: given the set of firm level job creation rates $\{jc(i,.)\}_{i\in[0,1]}$ and Ω, equations (4), (5) and (6) jointly determine firm level quit rates $q(.)$, gross hire flows $H(.)$ and the aggregate flow of job offers $\lambda(.)$. Lemma 1 now describes their closed form solution where $\lambda_1(\Omega)$ is given by (11) below.

Lemma 1. *A stochastic equilibrium implies*

$$q(i,\Omega) = \begin{cases} \lambda_u + \lambda_1(\Omega) & \text{if } i \in [i^c, i^h) \\ \lambda_u + \frac{\phi[1-U]}{U+\phi[1-U]G(i)} \int_i^1 \{jc(j,.) + \lambda_u\} G'(j) dj & \text{if } i \geq i^h, \end{cases} \tag{8}$$

$$h(i,\Omega) = \begin{cases} 0 & \text{if } i \in [i^c, i^h) \\ jc(i,\Omega) + \lambda_u + \frac{\phi[1-U]}{U+\phi[1-U]G(i)} \int_i^1 [jc(j,.) + \lambda_u] G'(j) dj & \text{if } i \geq i^h. \end{cases} \tag{9}$$

Proof of Lemma 1 is in Appendix B.

This solution reflects an important and useful property of a stochastic equilibrium. Suppose an existing firm $j > i^h$ creates a new job, either through investment in job creation or because an existing employee exogenously separates into unemployment and this firm then hires a replacement. With on-the-job search each new job created causes a hiring chain as described above. A stochastic equilibrium, however, implies

$$\lambda_0 U + \lambda_1[1-U]G(i^h) = \int_{i^h}^{1} \{jc(j,.) + \lambda_u\}[1-U]G'(j)dj \tag{10}$$

because the left hand side describes the flow death of hiring chains [ended when the last job hires an unemployed worker or an unreplaced quitter] which must equal flow birth of hiring chains [as described on the right hand side] as, with stock-flow matching, there is no stock of unfilled jobs. Using $\lambda_1(\Omega) = \phi\lambda_0(\Omega)$ in (10) now determines the job offer arrival rates,

$$\lambda_0(\Omega) = \frac{\int_{i^h}^{1} \{jc(i,.) + \lambda_u][1-U]G'(i)\}\, di}{U + \phi[1-U]G(i^h)} \tag{11}$$

$$\lambda_1(\Omega) = \phi\lambda_0(\Omega), \tag{12}$$

as functions only of firm level job creation rates $\{jc(i,.)\}_{i\in[0,1]}$ and Ω. Equation (8) has exactly the same intuition but reflects the hiring chain process from the perspective of a worker employed at a firm $i \geq i^h$. The integral in equation (8) describes the rate at which new jobs are created at higher wage firms $j > i$ which a worker employed in firm i would accept. For this employee, however, the relevant chain is destroyed once the job is either filled by an unemployed worker or by an employed worker at firm $i' < i$, for the worker employed in firm i is not interested in any replacement job $i' < i$. In this case $U + \phi[1-U]G(i)$ describes the death rate of an employee in firm i's (relevant) hiring chain while the integral in equation (8) describes the birth flow. Equation (8) then describes firm i's equilibrium quit rate in a stochastic equilibrium.

Equation (9) shows that hire flows depend not only on new job creation flows but also on quit replacement. This generates a **multiplier effect on hires**, where for every job created (and filled) at a firm $j > i^h$, hire flows additionally increase at every firm $i \in [i^h, j]$ for such firms replace any employees who quit to firm j. In this way on-the-job search magnifies gross job creation flows into larger gross hire flows. The critical economic insight here, however, is that increased job creation rates also increase job destruction via unreplaced quits. Indeed the flow of job destruction due to unreplaced job-to-job quits is

$$JD^Q(\Omega) = \frac{\phi[1-U]G(i^h)}{U + \phi[1-U]G(i^h)} \int_{i^h}^{1} \{jc(j,.) + \lambda_u\}[1-U]G'(j)dj \tag{13}$$

because the fraction describes the proportion of chain births which end with the hiring of a worker $i < i^h$. JD^Q thus directly increases with firm job creation rates $jc(.)$ and is crowded out by higher unemployment [the new job created is more likely to be filled with an unemployed worker]. Aggregate job creation, $JC(\Omega)$, and job destruction, $JD(\Omega)$, are calculated by integrating over the decisions of individual firms, with formal definitions deferred to Appendix C.1. We now complete the description of the stochastic equilibrium by determining the set of equilibrium wages $w(.)$.

4 Equilibrium Wages and Firm Values

Similar to Burdett and Mortensen (1998), a wage posting equilibrium requires simultaneously determining the equilibrium quit function $\widehat{q}(w', \Omega)$ should a firm post any wage w', while the equilibrium wage strategies must in turn maximise firm profits given this quit function and the wages offered by all other firms. This is a standard fixed point problem in the equilibrium price dispersion literature, but here we are in a stochastic equilibrium where firms cannot precommit to future wages and so the wage strategies must also be dynamically consistent. An additional complication is the reservation wage of workers $R_t = R(\Omega_t)$ is stochastic. Although determining $R(\Omega_t)$ remains feasible, the added complexity is not interesting and so we simplify by assuming a binding minimum wage policy: the government imposes a minimum wage $w_{\min}$ below which firms cannot pay.

Without restrictions on out-of-equilibrium beliefs, Coles and Mortensen (2016) show it is possible to support a plethora of signalling equilibria and so we adopt the following restriction.

Assumption 1. *For any Ω, worker beliefs on the firm's state i is first order stochastically increasing in the posted wage.*

This restriction rules out punishment beliefs: should a firm post a higher wage, workers might instead believe it has lower productivity and punish the firm with an increased quit rate [because they anticipate lower wages in the future] and no firm will then post such a wage in equilibrium. Coles and Mortensen (2016) establish the wage signalling equilibrium with monotone beliefs is unique. Rather than repeat their analysis, we refer the reader to that paper and here focus on constructing the resulting wage equilibrium.

As previously described, firms with $i \in [0, i^h)$ survive but do not recruit, while firms $i \geq i^h$ have a strictly positive hire flow for they (at least) replace workers who quit. The following characterises a [Markov Bayesian] wage signalling equilibrium where, consistent with the job ladder property, wage strategies have the following property:

1. $w(i, \Omega) = w_{\min}$ for $i \in [0, i^h]$;
2. $w(i, \Omega)$ is continuous and strictly increasing in$i$for$i \geq i^h$. (14)

Although the distribution of wage offers across hiring firms contains no mass points, that does not imply there is no mass point in the distribution of wages paid. Low productivity firms with $i \in [0, i^h)$ are in decline and equilibrium finds all such firms pay the binding minimum wage. Hiring firms $i \geq i^h$ however post higher wages $w(i, \Omega) \geq w_{\min}$ which are fully revealing. Bayes rule, equation (14) and the restriction to monotone beliefs now imply workers have the following beliefs:

Belief 1: if a firm posts wage $w' \in (w_{\min}, \overline{w}]$ where $\overline{w} = w(1, \Omega)$, the worker believes the firm's productivity $i = \widehat{i}(w', \Omega)$ where $\widehat{i}$ is the unique solution to $w(\widehat{i}, \Omega) = w'$; i.e. beliefs $\widehat{i}$ are the inverse of the equilibrium wage function;

Belief 2: if a firm posts wage $w' = w_{\min}$ and it is an outside job offer the worker believes the firm's productivity $\widehat{i} = i^h$ for it is a hiring firm. At a non-hiring firm an employee instead believes $\widehat{i} \in [0, i^h]$ where the specific choice plays no important role;

Belief 3: if a firm posts wage $w' > \overline{w} = w(1, \Omega)$, monotone beliefs require the worker believes firm productivity $\widehat{i} = 1$.

Because this wage structure has the job ladder property, a worker quits if and only if the wage offered by the outside firm is (weakly) higher than the worker's current wage w'. This follows because a higher outside offer and Beliefs 1-3 imply the outside firm is believed to have greater productivity, and first order stochastic dominance in $\Gamma(.)$ and equation (14) then imply the outside firm is more likely to post higher wages in the entire future.

Now given this optimal quit strategy, recall $F(., \Omega)$ is the distribution of wage offers across hiring firms. The optimal quit rate of a worker employed at a firm which posts wage w' is thus

$$\widehat{q}(w', \Omega) = \lambda_1(\Omega)[1 - F(w', \Omega)] + \lambda_u. \tag{15}$$

To solve for $F(.)$ consider any posted wage $w' \in [w_{\min}, \overline{w}]$ and note $\widehat{i}(w', \Omega)$ is the unique firm type $\widehat{i} \in [i^h, 1]$ solving $w(\widehat{i}, \Omega) = w'$ [Belief 1]. The fraction of job offers paying more than wage w' is thus

$$1 - F(w', \Omega) = \frac{\int_{\widehat{i}}^1 \frac{H(j,\Omega)}{\alpha+(1-\alpha)G(j)} dj}{\int_0^1 \frac{H(j,\Omega)}{\alpha+(1-\alpha)G(j)} dj} = \frac{1}{\lambda} \int_{\widehat{i}}^1 \frac{H(j, \Omega)}{\alpha + (1-\alpha)G(j)} dj \tag{16}$$

because $\frac{H(j,\Omega)}{\alpha+(1-\alpha)G(j)}$ describes total flow offers made by type j firms (who post wage $w(j, .) \geq w'$). Equations (2), (15) and (16) now imply the firm's optimal wage solves the following problem:

Lemma 2. *The optimal wage strategy* $w(i, \Omega)$ *solves:*

$$w(i, .) = \text{argmin}_{w' \geq w_{\min}} \left[w' + \left[\frac{\lambda_1}{\lambda} \int_{\widehat{i}(w',\Omega)}^1 \frac{H(j, .)}{\alpha + (1-\alpha)G(j)} dj \right] \min[v(i, .), c_0] \right]. \tag{17}$$

Each firm's optimal wage strategy involves a trade off between paying lower wages and having a higher quit rate, where a lower wage signals lower productivity and given the resulting beliefs regarding future wages, quit rates to better paying firms increase. Determining equilibrium wages is now trivial. Consider any hiring firm $i \in (i^h, 1)$ and suppose it posts optimal wage $w' \in (w_{\min}, \overline{w})$. The necessary condition for optimal w' implies

$$1 - \frac{\lambda_1}{\lambda} \frac{H(\widehat{i}, \Omega)}{\alpha + (1-\alpha)G(\widehat{i})} c_0 \frac{d\widehat{i}}{dw'} = 0,$$

where paying a marginally higher wage signals a marginally higher productivity $\widehat{i}$ and so yields a corresponding marginal fall in the quit rate. But equilibrium requires the optimal wage strategy $w(i,\Omega)$ must be the solution to this first order condition. Furthermore because $\widehat{i}$ is the inverse wage function, the necessary condition for optimality implies the differential equation

$$\frac{dw}{di} = \frac{\lambda_1}{\lambda}\frac{H(i,\Omega)}{\alpha+(1-\alpha)G(i)}c_0 \text{ for } i \in (i^h, 1).$$

Integration now yields Proposition 2.[13]

Proposition 2 (Optimal wages). *A stochastic equilibrium is characterised by equilibrium wage strategies*

$$w(i,\Omega) = \begin{cases} w_{\min} & \text{if } i \in [0, i^h) \\ w_{\min} + \frac{c_0\lambda_1}{\lambda}\int_{i^h}^{i}\frac{H(j,\Omega)}{\alpha+(1-\alpha)G(j)}dj & \text{if } i \geq i^h. \end{cases} \tag{18}$$

where beliefs on the firms type are given by Beliefs 1-3, optimal worker quit strategies are given by equation (15 with $F(.)$ given by (16).

Proof of Proposition 2 is in Appendix B.

5 The Stochastic Equilibrium

Because all firms $i \in [0, i^h]$ strictly prefer to post wage $w = w_{\min}$, while all firms $i \geq i^h$ are indifferent to doing so, the equilibrium wage strategies (18) and (2) now yield the following equilibrium functional equation for the set of firm values $\{v(i,\Omega)\}_{i\in[0,1]}$:

$$\begin{aligned}(r+\alpha_\gamma+\alpha_a+\delta_F)v(i,\Omega) = \;& p^s(i) - w_{\min} \\ & - [\lambda_1(\Omega)+\lambda_u]\min[v(i,\Omega), c_0] \\ & + \mu_1 E\max[v(i,\Omega)-[c_0+c^{JC}], 0] - \delta_D E_c \min[v(i,\Omega), c^{JD}] \\ & + \alpha_\gamma\int_0^1 [v(j,\Omega)]d\Gamma(j|i) + \alpha_a\sum_{s'}\Upsilon_{ss'}[v(i,\Omega(s'))] + \frac{\partial v(i,\Omega)}{\partial t}.\end{aligned} \tag{19}$$

The equilibrium functional equation is remarkably simple: aside from the exogenous productivity shock processes, it is only the endogenous job offer arrival rate $\lambda_1(\Omega)$ which drives firm values over the cycle. This reflects the underlying efficiency wage structure, that heterogeneous firms trade off higher wages against lower quit rates, where quit rates are driven by outside offer rates $\lambda_1(\Omega)$. Equilibrium $v(.)$ is a standard fixed point problem but rather than cleared by an equilibrium market tightness parameter as done in the free entry approach, market clearing here is instead a job offer arrival rate λ_1 described by equation (11).

[13]Proposition 2 implies equilibrium wage dispersion depends directly on c_0. If hiring costs $c_0 \to 0$, so that it is near costless to hiring replacement workers, then $\overline{w} \to w_{\min}$ and equilibrium converges to the Diamond paradox.

The stochastic equilibrium remains complex, however, because the distribution of employment $G(.)$ evolves endogenously over time, and $G(.)$ is infinitely dimensional. The application instead specialises to finite firm productivities $p = p^{is}$ with $i \in \{1, 2, ..., I\}$. In Appendix B.3 we show that in this case the state space reduces to a finite vector $\Omega = (s, \underline{N})$ where $\underline{N} = \{N_1, ..., N_I\}$ and N_i is the measure of workers employed in firms i, whilst with discrete productivities the wage distribution and the distribution of employment remain continuous.[14] The vector of firm values $\underline{v}(\Omega) = \{v_i(\Omega)\}_{i=1}^{I}$ continues to solve equation (19) but with job offer arrival rate given by:

$$\lambda_1(\Omega) = \frac{\phi \sum_{i \geq i^h} [jc(i, \Omega) + \lambda_u] N_i}{U + \phi \sum_{i < i^h} N_i} \tag{20}$$

and unemployment $U = 1 - \sum_{i=1}^{I} N_I$. Employment N_i is Markov and evolves according to:

$$\dot{N}_i(\Omega) = N_i \left[jc(i, \Omega) - jd(i, \Omega) - \delta_F\right] + \gamma_{0i}\mu_0 H^E(v^E(\Omega)) + \alpha_\gamma \sum_{j \neq i} N_j \gamma_{ji} - \alpha_\gamma \sum_{i \neq j} N_i \gamma_{ij} \tag{21}$$

where γ_{0i} describes the probability a new start-up is type i, $v^E(\Omega) = Ev(i, \Omega) = \sum_i \gamma_{0i} v(i, \Omega)$ is the expected value of a startup, γ_{ij} is the transition probability between states i, j should a firm specific productivity shock occur, and equation (21) takes into account jobs created via new start-ups. Standard recursive methods now apply and numerically computing $\{v_i(\Omega)\}_{i=1}^{I}$ in a stochastic equilibrium is straightforward.

6 Quantitative Analysis

To estimate the model we target worker and job flows as well as firm dynamics moments for the US economy. We use data from the Business Dynamics Statistics (BDS), Job Opening and Labor Turnover Survey (JOLTS), Compustat, and the Current Population Survey (CPS) for the period 1990Q2 to 2018Q4. In Sections 6.1 and 6.2 we describe the estimation. Steady state results are presented in Section 6.3 and business cycle results in Section 7. In Appendices A and C we provide a detailed discussion of the estimation procedure and construction of the data moments, including how we make model-consistent statistics from the different data sources.

[14]Because wages must be disperse across firms of the same type the cleanest approach is to assume firms select wage strategies as follows: i) On start-up, a firm is allocated a wage rank $\chi \sim U[0, 1)$. In the stochastic equilibrium, firm (i, χ, n, Ω) posts wage with rank χ in the firm i wage distribution. ii) On receiving a firm specific productivity shock with updated productivity i' the firm also updates to a new wage rank $\chi' \sim U[0, 1)$. Because all χ-wage strategies yield equal value, such wage selection is consistent with equilibrium. This wage selection process additionally preserves the job ladder property, that a worker will always quit to a higher wage offer, both across firm types and within firm types.

6.1 Parameterization

We set a period to be equal to a month and the time preference parameter to $r = 0.0043$ to match a yearly discount rate of 5%. We assume $S = 3$ aggregate productivity states indexed by $s = 1, 2, 3$. Let a_s denote the aggregate productivity shifter in state s such that it follows a discretised AR(1) process, where a new value is drawn at rate $\alpha_a = 1/3$ from the transition matrix $\Upsilon_{ss'}$. The latter is obtained using a modified Rouwenhurst approximation for a given autocorrelation parameter ρ_a and variance parameter σ_a. These parameters are set to match the simulated persistence (0.8482) and standard deviation (0.01) of quarterly-averaged log HP-filtered a_s to that of aggregate output per worker in the data.[15] We further suppose $I = 5$ firm productivity states which we show is enough to match well the firm dynamics data patterns, indexed by $i = 1, 2, ..., I$. A firm in state (i, s) has productivity $p^{is} = a_s p_i$ and equilibrium implies the aggregate state is $\Omega = (s, \underline{N})$ where $\underline{N} = \{N_1, N_2, .., N_5\}$ is the vector of employment across states i.

To estimate firms' productivity dynamics we define mature states $I^m = \{2, 3, 4\}$. While mature firms may transition across these states, we assume any firm in a mature state $i \in I^m$ cannot transit to states $i = 1, 5$; i.e. only entrant firms may have the extreme productivities $i = 5$ and $i = 1$. Allowing more extreme productivity states for entrant firms is important to account for the difference in growth outcomes between new start ups and existing (mature) firms. The reason for three mature states $i \in I^m$ is to capture disperse job creation and job destruction outcomes across existing firms. Conditional on survival, all firms receive a firm specific productivity shock at rate $\alpha_f = 1/3$ [i.e. roughly once a quarter], which is within the range of estimates from the data (see Appendix C). We assume firm $i \in I^m$ transits to state $j \in I^m$ with probability γ_j and so is independent of i. For parsimony we simply set $\gamma_3 = \gamma_4 = \frac{1}{2}[1 - \gamma_2]$. For firms $i = 5$ then γ_{55} describes the probability that a high productivity entrant remains in this state and so determines its persistency. With probability $1 - \gamma_{55}$, the firm otherwise becomes a mature firm $j \in I^m$ in proportions γ_j. Similarly if instead in state $i = 1$, γ_{11} is the probability the firm remains as a low productivity entrant. With probability $1 - \gamma_{11}$, the firm becomes mature $j \in I^m$ in the proportions γ_j. In this way the transition matrix $\{\gamma_{ij}\}$ is fully described by the choice of just three parameters $(\gamma_2, \gamma_{55}, \gamma_{11})$. Consistent with this structure, we assume an entrant firm is highly productive with probability γ_{05}, a low productivity entrant with probability γ_{01} and with complementary probability $(1 - \gamma_{05} - \gamma_{01})$ is a mature firm in states $j \in I^m$ with proportions γ_j. This productivity process, coupled with constant returns to scale in production, aims to capture in reduced form the growth processes (for example customer acquisition) and adjustment costs firms face

[15] Output per worker in the data is constructed using the ratio of quarterly real GDP over total employment, HP-filtered. In the model, output per worker and the productivity shifter a_s are not identical, due to endogenous composition effects. However, the differences are small so we calibrate the process of a_s directly in order to remove ρ_a and σ_a from the estimation routine and save on computation.

over their life-cycle. It not only allows for a simple solution, but as shown below is able to reproduce very well the observed firm age and size dynamics.

The specification of distributions H^{JC}, H^{JD} is central to determining the response of job creation and job destruction gross flows to aggregate shocks. Following Coles and Moghaddasi (2018), we specify distributions $H^{JC}(.), 1 - H^{JD}(.)$ which are isoelastic with respective elasticities ξ_{JC}, ξ_{JD}. In particular, for firms with $i \geq i^h$ the firm level job creation and job destruction rates are considered as $jc_i(\Omega) = \mu_1[v_i(\Omega) - c_0]^{\xi_{JC}}$ and $jd_i(\Omega) = \delta_D[v_i(\Omega))]^{-\xi_{JD}}$. A useful motivation is that the typical free entry approach assumes the job creation margin is infinitely elastic; i.e. entry is infinite when $c_0 < v(\Omega)$ and so free entry implies $v(\Omega) = c_0$. Our estimation instead recovers the job creation elasticity ξ_{JC} so that implied gross job creation flows are indeed consistent with the data, and similarly with ξ_{JD} and job destruction flows. Heterogeneity in firm types also allows potential cleansing or sullying effects of recessions: a negative aggregate productivity shock will trigger job destruction through layoffs at the lowest surplus states, but also protects them from quits as the job ladder slows down.

An important challenge is to make the notions of job creation and destruction in our model equivalent to how they are measured in the data. To do this, we make assumptions so that each productivity bin only performs job creation *or* job destruction, but not both. This is achieved by assuming a lower support for H^{JC} of $\underline{c}^{JC}$ and an upper support for H^{JD} of $\bar{c}^{JD}$. We set these support parameters to ensure that, in steady state, only firms in states 1, 2 and 3 destroy jobs after a δ_D job destruction shock, and only firms in states 4 and 5 create new jobs after a μ_1 job creation shock.

To estimate the entry process we parameterise $H^E(Ev(\Omega)) = (Ev(\Omega))^{\xi_E}$ to also generate a constant entry elasticity with respect to the expected value of entry. To match the average firm size at entry, we modify the model and allow firms to hire immediately after they have entered. In particular, firms enter with two employees drawn from unemployment, and the new firm also has N_0 immediate potential expansion jobs where $N_0 \in \mathbb{N}^+$ is exogenous. The job creation process is the same as for existing firms, such that for each N_0 positions the start-up draws a cost from H^{JC}, makes a decision whether to fill it or not, and may hire both from unemployment and by poaching from existing firms. This implies that each new firm begins life with initial employment $n_0 = 2 + \widetilde{n}_i$ where hires $\widetilde{n}_i$ are a binomially distributed random variable with N_0 independent trials and an [endogenous] probability of investment which depends on (i, Ω). In Appendix B.2 we present in more detail this extension.

If a firm ever reaches zero employees, the constant returns to scale structure implies that this is an absorbing state, where the firm never produces or has positive employment again. We thus treat these firms as having exited, and so we measure exit in our model as both the bankruptcy shock and firms who drop from one to zero employees. The bankruptcy rate is set to $\delta_F = 0.0004$ to give a

0.1% yearly exit rate from this shock, to roughly match the very low exit rate among firms with more than 500 employees in the BDS data. The remainder of exit in the model is driven by endogenous job destruction.

6.2 Estimation Strategy

To recover the remainder parameters we follow a two-step procedure in which we split the parameter set between an inner and outer loop. Given values for the outer loop parameters, we can directly calibrate those in the inner loop such that their values solve a non-linear system of equations, matching *exactly* the targeted moments in the non-stochastic steady state. We then iterate on the values of the outer loop parameters using a simulated minimum distance procedure until convergence, adjusting the inner loop parameters at each iteration. This estimation approach not only greatly simplifies the calibration but makes it very easy to implement (see Appendix C for details). Table 1 present all the model's parameters and their corresponding targeted moments. In Appendix C we provide a graphical representation of the identification of the outer loop parameters by evaluating the loss function at the optimal and showing its change in value as we perturbate each parameter in turn.

The inner loop contains the worker turnover parameters, where we target the average worker transition rates from Davis et al. (2012). In particular, we target an average quarterly layoff rate of 7.78%, which we equate with the EU rate in our model. We match this rate using layoffs due to job destruction from firms in states $1, 2, 3$ and add exogenous separations, λ_u, to capture that in growing firms in states $4, 5$ some workers transit into unemployment. The relative search intensity of employed workers, ϕ, is chosen to match an average quarterly quit rate of 6.81%. The exogenous minimum wage, $w_{\min}$, is used to target a labour share of 2/3.

The inner loop also contains the shifter and rates of the cost structure of JC and JD. The arrival rates μ_1 and δ_D control the average JC and JD rates, respectively. We target a monthly JD rate of 2.43%, the average documented in Davis et al. (2012) based on JOLTS data. We also target a 5.73% unemployment rate, the average from the CPS in our sample, which requires a 2.19% monthly JC rate (excluding JC from firm entry) to balance employment flows in steady state. We are careful to account for other sources of job creation (e.g. firm entry) and job destruction (e.g. unreplaced quits) when computing these series in the model. The firm entry flow shifter, μ_0, is chosen to control the number of firms in equilibrium. For a given targeted total employment N, μ_0 controls the equilibrium number of firms, and hence the average employment per firm. We choose μ_0 so that average firm size (total employment / total number of firms) is equal to 22.4, as in the BDS data in 2005.

In addition, we use p_4 to normalise aggregate labour productivity in the model to one: $Y/N = 1$, and p_3, particularly its value relative to p_4, to target the standard deviation of within-firm labour productivity obtained from Compustat, which we estimate to be 30% (see Appendix A.1 for details). Since in the model firms in states 2-4 typically are large and older and Compustat data is heavily biased towards larger firms, these data provides a consistent source to calibrate p_3.

Using data from Davis et al. (2012), we estimate that firms replace around 80% of workers who quit (see Appendix A.2) and target this fraction. We additionally validate this estimate by comparing other measures of replacement hiring in our model to the estimates of Elsby et al. (2021). This high level of replacement hiring identifies bounds on the hiring cost c_0 such that it must satisfy $v(2) < c_0 < v(3)$ in steady state. The latter arises as the data requires that firms in state $i = 3$ (which contains nearly 40% of employment in equilibrium) and above must replace workers who quit, while firms in states 1 and 2 destroy jobs after a worker quits. As firms in state 2 do not replace workers who quit, we choose the transition probability γ_2 to match that 80% of quits are replaced in steady state as this parameter controls the equilibrium mass of firms in state 2. In our estimated model, c_0 remains between $v(2)$ and $v(3)$ even during business cycle simulations.[16]

This procedure leaves twelve parameters that we jointly estimate in the outer loop of the calibration. To recover the entrant's productivity process and their maximum number of unfilled positions, N_0, we target the age and employment-age distributions of firms in 2005 obtained from the BDS (see Figure 3). In particular, we target the fraction of firms at age 0, 1, 2, 3-5, 6-10 and 11-15, and the fraction of aggregate employment at firms in the same age groups. The productivities of the entrant-specific productivity states p_1 and p_5, control how large is JC and JD among new entrants relative to older firms. N_0 controls the average initial size of entrants, particularly measured size at age 0. The transition probabilities γ_{11} and γ_{55} control for how long entrant JC and JD rates remain elevated. The entrant probabilities γ_{01} and γ_{05} control whether this is experienced by a small or large share of entrants. In the data, the shares of employment by firm age imply the net job creation rates of firms of different ages; while the shares of firms by age imply the different exit rates across age groups, thus informing these parameters. We also use the firm age distribution to recover the productivity of mature firms in state 2, p_2. In particular, since this productivity controls the overall JD and exit rates of state 2 firms, we target the average exit rate of firms consistent with on the overall distribution of firms by age. We find that the high JD rate in this state means that it drives 36% of total exit, despite only containing 14% of firms.[17]

[16]The values of c_0, $\underline{c}^{JC}$, and $\bar{c}^{JD}$ lead to a simple structure for job creation and destruction: Firms in states $i = 1, 2$ do not replace quits, and lay off workers due to the δ_D job destruction shock. Firms in state $i = 3$ still perform δ_D job destruction, but have high enough value to find it optimal to replace workers who quit. Finally, firms in states $i = 4, 5$ replace workers, do not perform δ_D job destruction, and create jobs in response to the μ_1 job creation shock. See Table 8 in Appendix B for a summary of the steady state values, policy functions, and distributions of firms in each state.

[17]The estimation also finds that the productivity grid is non-monotone, as $p_1 > p_2$ and $p_5 < p_4$. Nonetheless, firm *values* remain monotone, with $v_i < v_{i+1}$ for all i, which is sufficient for the job ladder to be directed monotonically by i, and hence for our notion of equilibrium to remain well defined. The disconnect between the ordering of productivities and

Table 1: Parameter values and target moments

	Interpretation	Value	Source
	Pre-set parameters		
r	Discount rate	0.0043	5% annual interest rate
α_a	Arrival rate of agg shocks	0.3333	Normalisation
α_γ	Arrival rate of firm prod shocks	0.3333	Autocorr. of idiosyncratic prod. (see text)
ρ_a	Persistence of aggregate productivity shock	0.7800	Autocorr. of aggregate labour prod.
σ_a	Std. of aggregate productivity shock	0.0120	Std. of aggregate labour prod.
δ_F	Arrival rate of bankruptcy shock	$8.3E-05$	Exit rate of firms with size > 500
	Firm productivity process		
p_1	Prod. in state 1	0.9342	Firm age distribution (SMD)
p_2	Prod. in state 2	0.6732	Firm age distribution (SMD)
p_3	Prod. in state 3	0.7127	Std. of idiosyncratic labour prod. = 30%
p_4	Prod. in state 4	1.3085	Normalise $Y/N = 1$
p_5	Prod. in state 5	1.0735	Firm age distribution (SMD)
γ_{11}	Prob. remain in state 1	0.9269	Firm age distribution (SMD)
γ_{55}	Prob. remain in state 5	0.6674	Firm age distribution (SMD)
γ_{01}	Share born with prod. 1	0.5063	Firm age distribution (SMD)
γ_{05}	Share born with prod. 5	0.1006	Firm age distribution (SMD)
γ_2	Prob. mature draw state 2	0.158	80% of quits replaced
N_0	Potential unfilled positions of entrants	75.105	Firm age distribution (SMD)
	Cost structure of JC and JD		
ξ_e	Elasticity of entry with respect to firm value	3.1723	Std deviation of cyclical firm entry (SMD)
ξ_{JC}	Elasticity of JC with respect to firm value	2.1673	Std deviation of cyclical JC (SMD)
ξ_{JD}	Elasticity of JD with respect to firm value	3.8197	Std deviation of cyclical JD (SMD)
$\underline{c}^{JC}$	Lower bound of $H^{JC(.)}$	0.9542	JC only in states 4 and 5
$\bar{c}^{JD}$	Upper bound of $H^{JD(.)}$	1.7277	JD only in states 1, 2 and 3
μ_0	Firm entry flow	0.0004	Average firm size 22.4 employees
μ_1	Arrival rate of capital investment shock	0.0116	Average JC rate = 2.17% per month
δ_D	Arrival rate of capital destruction shock	0.0140	Average JD rate = 2.43% per month
c_0	Worker replacement cost	0.7735	Autocorr. of JC and JD (SMD)
	Worker turnover		
λ_u	Quit rate to unemployment	0.0087	Worker EU rate 7.78% per quarter
ϕ	Employed fixed search intensity	0.1031	EE rate of 6.81% per quarter
$w_{\min}$	Minimum wage	0.6515	Labour share = 2/3

Note: Calibrated parameter values and source moments. Parameters marked "(SMD)" are chosen in the Outer Loop to jointly minimize the distance to a set of 18 moments, and the remaining parameters are chosen in the Inner Loop to exactly match their assigned moment. See text and Online Appendix for further details.

We also include in the outer loop the remainder parameters of the cost structure of JC and JD. To inform the JC and JD elasticity parameters ξ_{JC} and ξ_{JD} we simulate our stochastic model and target the volatility of the JC and JD rates series obtained from Davis et al. (2012) using JOLTS data. Both in the model and data, we use an HP-filter with parameter 10^5 to obtain the cyclical component of each of these series and compute their standard deviation. To inform the firm entry elasticity ξ_e we target the volatility of the firm entry series obtained from the BDS. To further inform the worker hiring cost, c_0, which the level of replacement hiring implies must satisfy $v(2) < c_0 < v(3)$, we target the autocorrelation of aggregate JC and JD rates in the data. Intuitively, the larger is c_0 the larger is the general equilibrium effect that rising unemployment in a recession raises firm value, since

values occurs simply because the entrant states $i = 1, 5$ are more persistent than the mature states.

firms must pay the replacement cost c_0 less often in recessions. The larger is this offsetting effect, the less persistent are JC and JD rates, as they recover faster in recessions. Note that by targeting the volatility of JC and JD we are not indirectly targeting the *cyclical correlation* between these series (or of their components) and that of unemployment, or between the latter and hires and quits. This is important as a key contribution of our model is to reproduce and explain these cyclical correlations.

6.3 Fit of the model

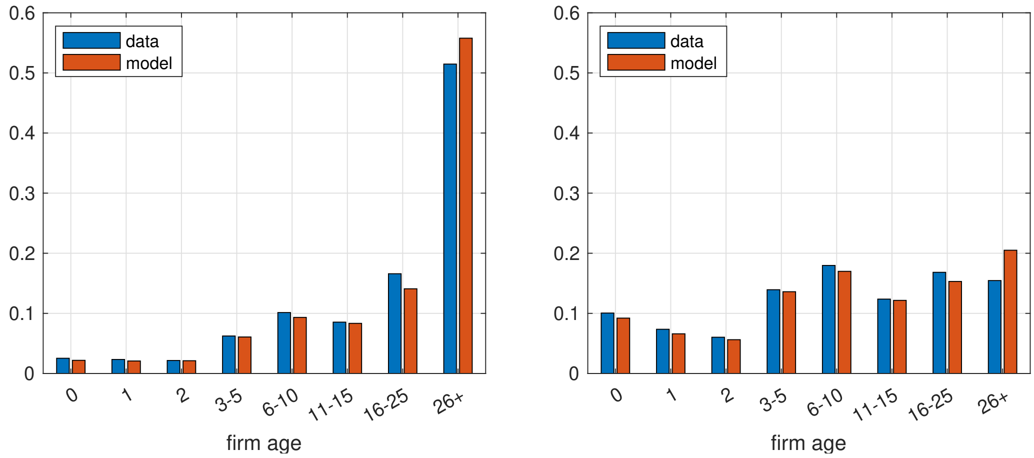
We now turn to discuss the fit of the model on the targeted cross-sectional data patterns. Since the inner loop moments are matched exactly we focus on the outer loop moments. In addition, we discuss the fit in several non-targeted dimensions and show the model matches the data very well.

6.3.1 Steady state results: Micro firm growth rates

Figure 3 shows that the model fits very well the targeted firm age distribution. Figure 3(a) describes the fraction of workers employed in firms within a particular age range, while Figure 3(b) describes the fraction of firms in that age range. In the model 9.6% of existing firms close per year which is similar to the 8.3% in the data. In steady state, these are replaced by an equivalent inflow of new start-up firms. Taking into account the range of each age bin, Figure 3(b) implies the average exit rate of firms falls steeply with age, as shown in Figure 4(a). Allowing ex-ante start-up heterogeneity is central to capturing this age structure. Matching the firm survival data, the calibration shows that 51% of new start-ups are born in the lower productivity state [$i = 1$] with a high associated firm death rate. In contrast, around 10% of new start-ups are born in the higher productivity state [$i = 5$] but, conditional on survival, their expected duration in this state is short, being only 2 months ($\gamma_{55} = 0.67$). In this way few start-ups remain highly productive for long and relatively few grow into very large firms. Conversely the struggling entrant state $i = 1$ is estimated to be much more persistent with an expected duration of 10 months ($\gamma_{11} = 0.93$). Their high exit rate then implies these firms are more likely to close than reach a mature state. Entrant firms that survive to age 25 are on average very large: more than 50% of all workers are employed in firms over 25 years old, yet there are relatively few such firms.

Given this firm heterogeneity structure, the estimated parameters imply that most employment N_i is in the mature states $i \in I^m$ and so job creation flows by mature firms (specifically, jc_4) are responsible for the larger part of gross job creation. But mature firms are also responsible for the larger part of gross job destruction flows through $jd_2 + jd_3$. Combining these two effects, Figure 4(b) shows that the model implied *net* job creation flows of mature firms are negative as in the data. It is only young firms who have positive net job creation, due to the extra job creation of entrants, and in particular firms in

Figure 3: Firm age distributions in the model and data

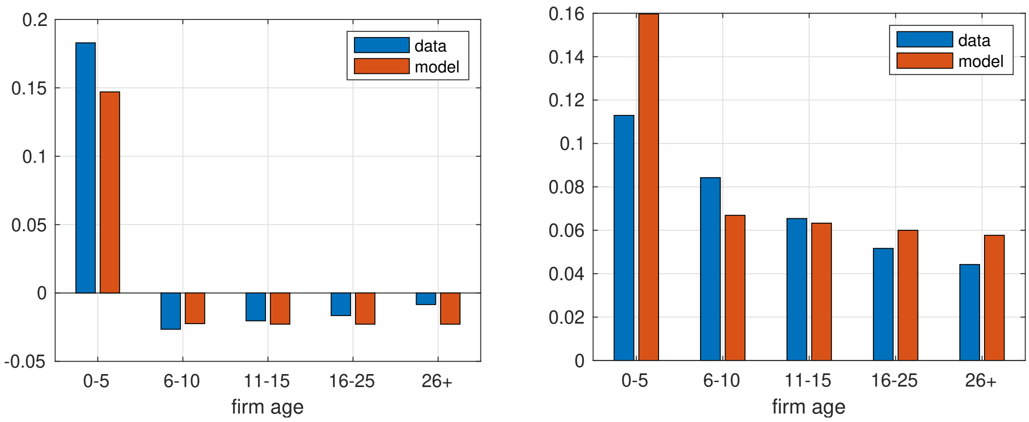


(a) Employment (b) Number of firms

Note: Left (right) panel plots the fraction of total employment (firms) contained in each firm age bin. Data corresponds to the BDS data for firms in 2005. Model corresponds to the steady state of the model.

state $i = 5$. At the same time, the model remains consistent with the higher exit rates of young firms, as shown in Figure 4(a), due to firms in state $i = 1$.

Figure 4: Net job creation and exit in the model and data



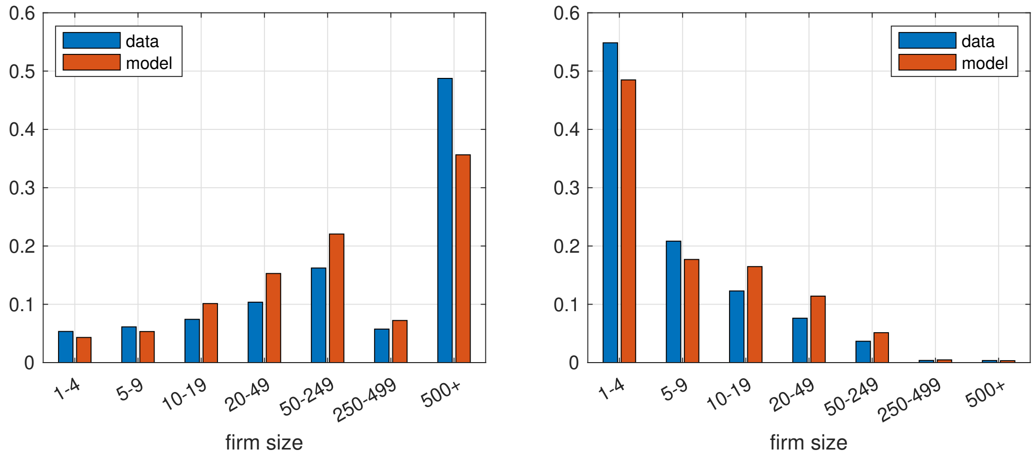
(a) Exit rate (b) Net JC rate

Note: Left panel plots the fraction of firms who exit per year. Right panel plots the yearly net job creation rate in firm age bin, computed as the net JC flow divided by the employment denominator. Data corresponds to the BDS data for firms in 2005. Model corresponds to the steady state of the model, where yearly rates are computed as $1 - e^{-12r}$ such that r are the theoretical monthly rates.

An important insight is that despite firm exit rates being high, the amount of job destruction due to such exits in the data is surprisingly small because most firm closures are small firms.[18] The model is consistent with this pattern, due to the high exit rate of small, young firms in state $i = 1$ and the gradual downsizing of employment by larger, mature firms in states $i = 2, 3$.

[18]For example, according to the 2005 BDS data, firm exit rates in the 1-4 employee size category is 12.3% per annum, while the exit rate for all larger size categories is much smaller, for example it is only 2.9% in the next size bin of firms with 5-9 employees. In the data, the overall annual exit rate thus reflects that 88% of firm exits involve the very smallest of firms.

Figure 5: Firm size distributions in the model and data



(a) Employment (b) Number of firms

Note: Left (right) panel plots the fraction of total employment (firms) contained in each firm size bin. Data corresponds to the BDS data for firms in 2005. Model corresponds to the steady state of the model.

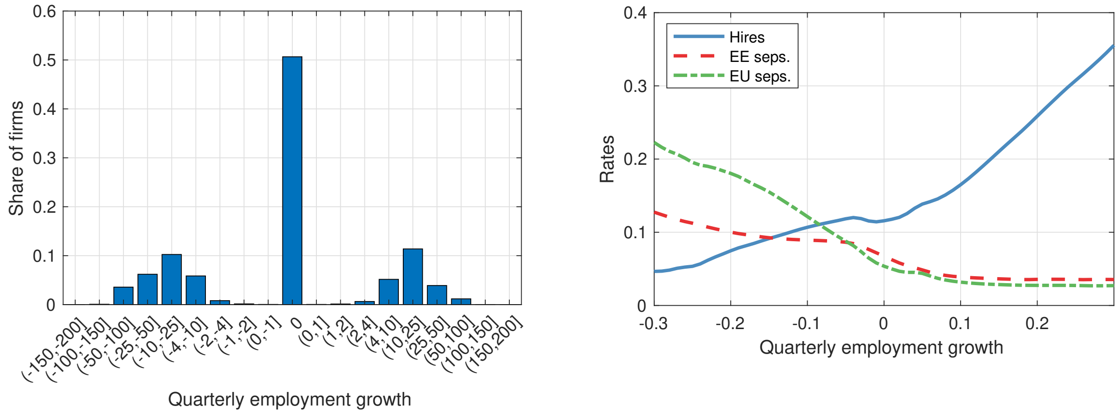
Although the model targets employment by firm age and survival rates, we do not target the distribution of firm size nor employment by firm size. The Gibrat's Law structure automatically implies substantial size dispersion across firms of the same age, because realised firm size depends on the firm's history of productivity outcomes and growth rates. If the model implied firm growth processes were a poor descriptor of actual firm size evolution, the simulated firm size distributions would not likely match the actual distributions of employment across firms. The fit turns out to be very good indeed, as shown in Figure 5. The figure reveals that in the data and model around half of firms are very small, in the 1-4 employee bin and constitute only around 5% of employment. In contrast, nearly 50% of total employment in the data is in firms which employ more than 500 workers while the number of such firms is very small. This property of the labour market is well known. The important point, however, is that the growth structure here also replicates well the (un-targeted) distributions of firm size and of employment by firm size, and provides (at least indirect) support for the Gibrat's Law approach taken.

Finally, Figure 6(a) shows the firm growth structure is also consistent with the (un-targeted) empirical employment growth distribution documented in Davis et al. (2012) and Elsby et al. (2021).[19] A key feature of the data is that around 55% of firms report zero growth in a given quarter (see Elsby et al., 2021), while a somewhat equal share of the remaining firms report either positive growth or negative growth. The model yields precisely this outcome: employment at most firms does not change over the quarter with an approximately even break of firms showing positive and negative growth. Of course for firms where employment does not change, hiring is not zero for those firms actively re-

[19] To save space we refer the reader to these papers to view the empirical employment growth distributions based on JOLTS and the Quarterly Census of Employment and Wages.

place workers who quit, as shown in Figure 6(b). Nevertheless the important point, as also argued in Bertola and Caballero (1994) and Cooper and Halitwanger (2006), is that many firms do not change employment and the smooth evolution of aggregate unemployment arises because of the aggregation of disperse employment decisions made by heterogeneous firms. A representative firm approach with smooth, convex adjustment costs is inconsistent with this microeconomic behaviour.

Figure 6: Employment growth distribution and relation with worker flows



(a) Employment growth

(b) Flows

Note: Left panel plots the distribution of quarterly employment growth rates across firms, excluding entry and exit, as in Elsby et al. (2021). Right panel plots the average hiring, EE and EU separation rates of firms with each employment rate (computed for bins of width 0.01 and smoothed with a 10 bin moving average). Both figures calculated from a simulation of a panel of firms in the steady state of the model.

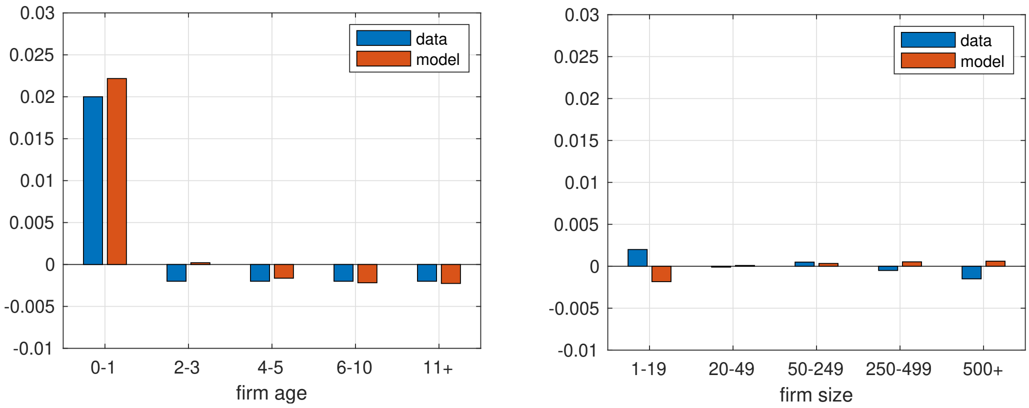
We now turn to describing the job ladder structure of the model and how we match data to information on quit and hire outcomes at the firm level.

6.3.2 Steady state results: Quit Turnover and Wage Competition

Quits are costly: a firm must either pay a cost c_0 to hire a replacement worker, or choose not to replace the worker and downsize. Although job creation rates jc_i and job destruction rates jd_i are the same for all firms with a given i, wage competition with on-the-job search implies firms in the same state i post different wages $w \in [w_i, w_{i+1})$, where paying a marginally higher wage marginally reduces worker quit rates. An important difference to the Burdett and Mortensen (1998) framework, however, is that wage offers here are not fundamentally related to firm size. Instead higher productivity firms post higher wages which fundamentally changes the structure of quit turnover dynamics over the cycle. Specifically it allows that small but fast growing entrants poach employees from lower productivity but possibly larger firms. Haltiwanger et al. (2018) show this is an important property of the data: there is clear evidence that workers typically quit to better wages, but there is no evidence of a systematic drift of workers from small to large firms.

Our estimation targets the average turnover rates in the economy, but it does not target the turnover properties across the firm growth rate distribution. Davis et al. (2012) document "hockey sticks" relationships between hires, layoffs and quits and firms' growth rates. Figure 6(b) shows that the model also generates such hockey sticks. Job-to-job quit rates in the model decline as we move from the [low wage] negative growth firms to the [high wage] positive growth firms; i.e. the highest growth firms pay the highest wages, have the lowest quit rates and expand with a high hiring rate. Figure 6(b) also shows that in the model the firms with the largest negative growth rates shrink more by layoffs than quits. Thus although un-replaced quits describe a significant channel for job destruction, high separation rates at the faster declining firms instead depend more on high layoff rates, consistent with the hockey stick relationships in the data.

Figure 7: Poaching flows by age and size



(a) Age (b) Size

Note: Left (right) panel plots the quarterly net poaching rate by firm age (size) bin. This is computed as total hires from poaching less total separations from poaching as a fraction of employment in each bin. Data are for 2005, from the Census job-to-job database. Model corresponds to the steady state of the model, with quarterly rates computed $1 - e^{-3r}$, where r are the theoretical monthly rates.

Figure 7(a) shows that the calibrated model also exhibits a job ladder by firm age. Following Haltiwanger et al. (2018), in this case we measure the job ladder through a firm's net poaching rate, defined as the difference between the rate at which it hires employed workers h_p less the number of workers it loses to other firms s_p and so describes net quit drift. These data were not targeted. By firm age, the model generates the observed poaching structure: young firms are large net poachers [from older firms] while older firms are net losers to young firms. This poaching structure reflects that in the model high productivity entrants $i = 5$ [10% of entrants] will hire many new workers while low productivity, struggling entrants $i = 1$ [51% of entrants] have few workers who can be poached.

Because firm specific productivity, and so firm growth, is positively autocorrelated, the model implies a positive, but weak, correlation between firm size and productivity. The wage setting process, in turn, then generates a positive correlation between firm size and wages, where a one standard deviation rise in firm size leads to a 14% standard deviation rise in wages (see also Brown and Medoff, 1989).[20]

[20]We regress wage on firm size (measured as number of employees) in the ergodic distribution. Specifically, we con-

But job-to-job quit turnover is confounded by high productivity entrants who are currently small but responsible for most of the net poaching in the data. As described in Haltiwanger et al. (2018), there is no simple relationship between the job ladder and firm size, as depicted in Figure 7(b).

7 Business Cycle Facts and Insights

This section considers the business cycle properties of the calibrated model, its match to the data and corresponding new insights into business cycle frequencies [further details are provided in Appendix C]. Table 2 describes the main business cycle statistics, where moments marked with † symbols are directly targeted. The top panel records the estimated volatility and persistence of the job and worker flow rates. The bottom panel presents the cyclical correlation between these aggregates and unemployment, employment and net job creation.

Table 2: Logged and HP-filtered Business Cycle Statistics. Data and Model

	jc	jd	quits	hires	layoffs	UE	U	N	net jc
Volatility and Persistence									
Data									
σ	0.042†	0.055†	0.116	0.058	0.048	0.169	0.204	0.018	0.006
ρ_{t-1}	0.433†	0.755†	0.945	0.904	0.761	0.959	0.977	0.973	0.722
Model									
σ	0.042	0.053	0.182	0.079	0.046	0.226	0.229	0.014	0.005
ρ_{t-1}	0.672	0.601	0.935	0.899	0.542	0.937	0.934	0.937	0.467
Cyclical Correlation									
Data									
U	-0.305	-0.254	-0.923	-0.760	0.174	-0.972	1.000	-0.922	0.012
N	0.061	0.386	0.780	0.539	-0.015	0.867	-0.922	1.000	-0.230
net jc	0.751	-0.870	0.245	0.511	-0.760	0.120	0.012	-0.230	1.000
Model									
U	-0.612	-0.586	-0.990	-0.939	0.295	-0.994	1.000	-0.987	0.070
N	0.597	0.573	0.983	0.923	-0.284	0.984	-0.987	1.000	-0.069
net JC	0.744	-0.847	0.068	0.275	-0.931	0.036	0.070	-0.069	1.000

Note: Time series in the model are obtained by simulating the model for 1,000 years and then aggregating to quarterly frequency as in the data. The cyclical components of the (log) of these time series are obtained using an HP filter with parameter 10^5. Net jc ($jc - jd$) is not logged as it takes negative values. jc, jd, quits, hires, and layoffs are rates relative to employment, and UE gives the job finding rate of unemployed, H^{UE}/U. At any time t the flows refer to flows between t and $t+1$, and the stocks (U and N) are measured at t. Moments targeted in the estimation are marked with a † symbol.

In the model aggregate productivity [the sole driving variable] evolves exogenously with a persistence ($\rho = 0.85$) and volatility ($\sigma = 0.01$) matched to that of measured aggregate output per worker Y/N. These productivity shocks are small and less persistent than unemployment, both as measured in the data and the calibrated model. The targeted moments in Table 2 are the volatility and persis-

struct an average wage for every firm size on our firm size grid, by integrating over the within-size productivity and rank distribution. We then regress this average wage on firm size using weighted OLS, with one observation per size node, with the weight for each size node given by its density in the ergodic distribution.

tence of job creation (jc) and job destruction (jd) (top panel). The calibrated model fits very well the cyclical volatility of these rates. While the calibrated persistence parameters of job creation and job destruction are of similar magnitudes, in the data the persistence of job creation is about half of that of job destruction. The main reason why the model does not produce a better fit in this dimension is because we only use the parameter c_0 to capture this property for both job creation and job destruction (see Section 4.2). The calibration procedure tries to resolve this tension by choosing a c_0 that places the values of ρ_{t-1} for jc and jd somewhere in the middle of their empirical targets.

The model also generates a volatile unemployment rate. This arises due to the endogenous response of firms and workers to productivity shocks, but also a result of targeting the volatility of job creation and job destruction. Similarly, as job destruction in the model is mostly made up of workers laid-off after a δ_D shock (70% of all job destruction), the volatility of the layoff rate follows closely that of job destruction.

Although not shown in Table 2, the calibration does generate relatively rigid aggregate wage dynamics. The volatility of average wages in the calibration is 0.0063 which is about 60% of the volatility of aggregate productivity. By regressing average wages on the unemployment rate in simulated data, we obtain a coefficient of -0.462, which is close to the elasticity of continuing worker wages of -0.426 estimated by Gertler et al. (2020, Table 2) using SIPP data.

7.1 Quits and hires

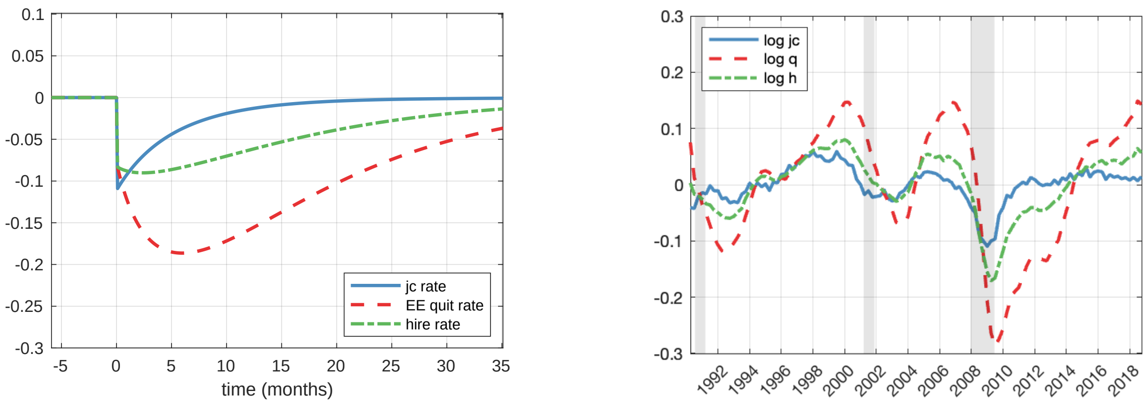
An important success of the model relative to the un-targeted moments described in the top panel of Table 2 arises from its ability to generate volatile and persistence quits and hires as in the data. The model is able to generate such properties due to replacement hiring, where only mature (typically large) firms in states 3 and 4 and new entrants in state 5 replace workers who quit. To illustrate the importance of replacement hiring over the cycle, Figure 8(a) presents the model's impulse response dynamics to a recessionary shock.[21] Specifically suppose that aggregate productivity is at the intermediate value $s = 2$ for a long period, so that the model converges to a "pseudo" steady state. At date 0 there is a recessionary shock to low productivity state $s = 1$. Productivity returns to $s = 2$ with probability 0.2079 per month, as per our estimated productivity process.[22] Given that our model is stochastic, and in order to capture the average dynamics across different possible lengths of recession, we simulate all possible lengths of recession and display an averaged impulse response path, weighting each possible path by its respective probability of occurrence. Note that the recessionary shock is fully anticipated

[21] In Appendix D we provide the impulse responses of the main aggregate variables, describing their behaviour under the same productivity shock.

[22] Our discretised productivity process is a Rouwenhurst approximation of an AR(1) process modified to additionally rule out transitions directly between states 1 and 3. Thus the economy must return to state $s = 2$ from state $s = 1$.

and agent expectations are always consistent with the full model. In this way, the impulse response dynamics summarise an average recession consistent with the true simulated moments from Table 2. As a comparison, Figure 8(b) plots the observed cyclical behaviour of the (log) job creation, gross hire h and gross quit q rates for the US (1990Q2-2018Q4).

Figure 8: JC, hires, quits in the data and model



(a) Model (b) Data

Note: Right panel plots Davis, Faberman and Haltiwanger (2012) data on the job creation, hire, and quit rates in the data, which have been logged and HP-filtered with parameter 10^5. Right panel plots the path for these variables in the model, during a "typical recession" experiment (see text for details), expressed as deviations from their initial values.

Figure 8 shows that in the model and data job creation, hire and quit rates are highly positively autocorrelated, where quits is the most volatile and job creation is the least (see also Table 2). To understand why our model reproduces these features, note first that in the estimation $\phi = 0.103 < 1$ and so changes in unemployment imply changes in aggregate search intensity. This generates important cyclical crowding out effects where higher unemployment crowds out on-the-job search. In particular, when the recessionary shock hits the economy the proportion of hires from unemployment rises from 51% to 60% at the trough of the recession in detriment of hires from employment. Because unemployment is more volatile and persistent than job creation (see Table 2), Equation (8) in Lemma 1 shows this crowding out effect implies the quit rate is also more volatile and more persistent than the job creation rate. Figure 8(a) shows that after its initial collapse the quit rate remains suppressed and recovers slowly. In contrast, the job creation rate is the first to start recovering and returns to its long-run value even while quits remain severely depressed.

An important feature of the data observed in Figure 8(b) is that when the quit rate is below trend [i.e. when $q < 0$] then the [detrended] gross hires rate typically is below the job creation rate, and vice versa. In the model this reflects the underlying replacement hiring process: that replacement hiring decrease (increases) as quits decrease (increase), and gross hires equal job creation plus replacement hiring (see equation (4)). As discussed in Section 3, replacement hiring causes hiring chains, where if a new job created is filled by an already employed worker then a "new job" continues to exist should the previous employer choose to hire a replacement worker. This effect magnifies the gap between gross hires and job creation. The volatile and persistent quit dynamic described in Figure 8(a) and Table 2 thus imply, through replacement hiring, that the gross hires rate are more volatile and more persistent than job creation.

Underlying the above cyclical dynamics, Davis et al. (2012) show that the "hockey sticks" characterising the relationship between firm-level employment growth and quits and hires (which we described in Figure 6(b) in steady state) shift downwards during recessions.[23] This shift implies that in the data the quit and hires rates not only fall during recessions among growing firms, but also among declining firms. Figures 9(a) and 9(b) show that the same un-targeted dynamics occur in the model. As the simulated economy enters into a recession the drop in quits and hires occurs throughout the employment growth rate distribution, but the decrease is more pronounced among the rapidly declining firms, typically those in the lower productivity states. This unequal cyclical shift in the hockey stick relationships is also consistent with the one presented by Davis et al. (2012).

Figure 9(c) further shows the cyclical shift in the hockey stick relationship between firm employment growth and the layoff rate. Also consistent with Davis et al. (2012), we find that the layoff rate is essentially invariant to the cycle across the growth rate distribution, except for rapidly declining firms. These firms experience noticeable increases in their layoff rates during recessions. We now show that these declining, low productivity firms play an important role in shaping unemployment dynamics. They do so by increasing layoffs early on during recessions and by not replacing quits as the economy rebounds.

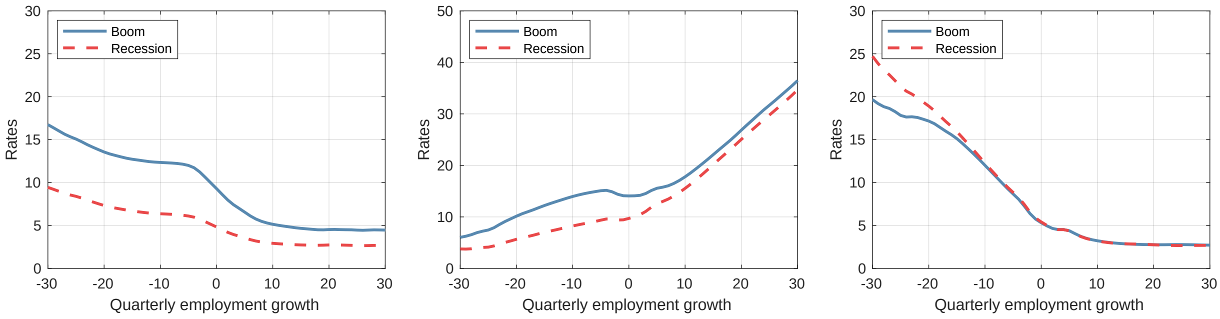
7.2 Decomposing job creation and job destruction

Unreplaced quits Recalling that JD_t denotes the aggregate job destruction flow, note that this can be decomposed into the sum of layoffs, EU_t, and destruction due to unreplaced job-to-job quits, JD_t^Q, giving $JD_t = EU_t + JD_t^Q$.[24] The lack of replacement hiring in firms by productivity state $i = 1, 2$ imply they are the source of job destruction due to unreplaced quits. Once the recessionary shock hits the economy and job creation and quits fall, these firms face less poaching leading to a collapse in JD_t^Q. As hires from employment begin to slowly recover, firms in states 1 and 2 start destroying

[23] To save space we refer the reader to Figure 8 in Davis el at. (2012) for the derived "hockey stick" relationships using JOLTS establishment data.

[24] Since some firms do not replace exogenous layoffs into unemployment due to the λ_u shock, the exact formula is $JD_t = EU_t + JD_t^Q - \lambda_u N_t^r$, where N_t^r is the mass of employment at firms who do replace quits (i.e. with $i \geq 3$). This last term is not cyclically important when looking at the aggregate JD rate, and so we exclude it from the discussion for simplicity.

Figure 9: Cyclical Relationship Between Firm Employment Growth and Quits, Hires and Layoffs



(a) Quits (b) Hires (c) Layoffs

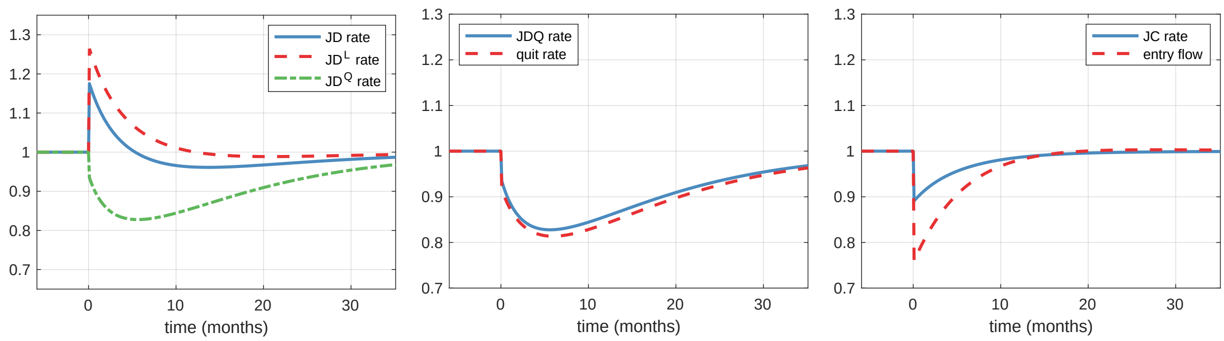
Note: Hockey stick relationships between firm employment growth and quits, hires and layoffs separately for expansion periods and recessions periods in the calibrated model. See Figure 6 for details of the hockey stick plot construction. The figures here are constructed in the same way, with the expansion line corresponding to periods with high unemployment (the 10% of periods with the highest unemployment rates) in our business cycle simulation, and contraction for the bottom 10%.

more jobs due to rising poaching rates. As discussed in Section 3.1, the crucial feature of this type of job destruction is that it prevents new jobs created higher up in the firm productivity ladder $(i = 4, 5)$ from lifting workers out of unemployment, because the new job is instead filled by stealing a worker from an unproductive firm. This is in contrast with job destruction due to layoffs which instead pushes employed workers into the unemployment pool.

Figures 10(a) and 10(b) show these dynamics by presenting the impulse responses characterising job destruction, layoffs, job destruction due to unreplaced quits, and quits. Here note that job destruction due to unreplaced quits solely drives job destruction net of layoffs. This figures show that the model is fully consistent with the motivating evidence presented in Figure 1 in the Introduction. As in the data, layoffs and job destruction net of layoffs exhibit very different cyclical dynamics. Unemployment steeply rises early in the recession (see Figure 11(a)) mainly due to the large spike in layoffs. While the layoff rate quickly returns to its long-run level, JD_t^Q falls and recovers slowly in tandem with quits. Crucially, the negative cyclical correlation between JD_t^Q and unemployment is the reason why the model generates procyclical job destruction (a negative correlation between the job destruction and unemployment rates), even though layoffs remain countercyclical. Combining the procyclicality of unreplaced quits with the countercyclicality of layoffs then leads job destruction to have a similar level of procyclicality as job creation. The bottom panel of Table 2 documents both of these model properties and show they are consistent with the data.

Firm entry The cyclical volatility, persistence and procyclicality of job creation in the calibration arises from two sources: job creation by incumbent firms JC_t^I and job creation by new entrants JC_t^E, giving $JC_t = JC_t^I + JC_t^E$. In line with the data, Figure 10(c) shows that firm entry is more volatile

Figure 10: Decomposing Job Destruction and Job Creation



(a) JD and components (b) JD net of layoffs vs. quits (c) Job creation and entry

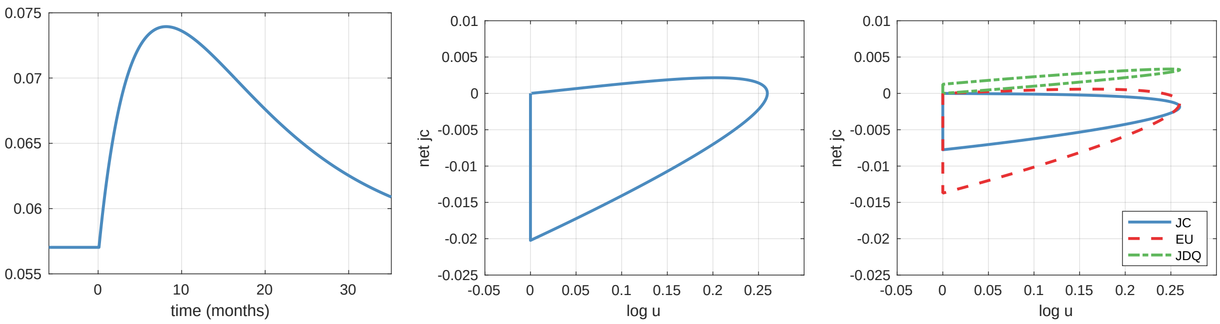
Note: All panels show simulated rates and flows from our typical business cycle experiment, expressed as deviations from their original value. The left panel plots the aggregate JD rate as well as its decomposition into JD from layoffs and unreplaced quits. The centre panel plots JD from unreplaced quits and the quit rate. The right panel plots the aggregate JC rate and the firm entry flow.

than gross job creation. Reflecting this feature, the estimated elasticities of firm entry and job creation to firm value satisfy $\xi_e = 3.1723 > \xi_{JC} = 2.1673$. Although their values are more than one, and so relatively elastic, both are a long way from infinity. This is in stark contrast to the free entry approach which assumes a penny increase in value yields an infinite number of entrants and this creation process is far too elastic for the data. The cyclical responses of firm entry and gross job creation also show that job creation by incumbents firms does not respond strongly to the cycle. Although firm entry is the more cyclically sensitive, it only has a modest effect on total job creation because new entrants begin typically small (see Section 6) and are unable to absorb the large amount of laid-off workers. Consequently the feedback of higher unemployment into greater job creation through firm entry is relatively weak and helps the model generate the observe cyclical variations in the unemployment rate. However, the model remains consistent with the fact that net job creation among young firms (0-5 years) is more cyclically sensitive to the business cycle than net job creation among older firms (6+ years) as documented by Fort et al. (2013) (see Figure D.2 in Appendix D).

7.3 Unemployment dynamics

To investigate how the job creation and job destruction dynamics described in Figure 10 interact to generate persistence in aggregate unemployment, first note that since $U_t = 1 - N_t$ in the model, unemployment dynamics are given by $\dot{U}_t = JD_t - JC_t$. That is, unemployment rises whenever net job creation, $JC_t - JD_t$, is negative, and falls whenever it is positive. In the initial pseudo steady state with $s = 2$ it holds that $JC = JD$ such that unemployment remains constant. As a recessionary shock hits the economy, Figure 10 implies the flow of jobs destroyed increases and the flow of jobs created decreases. $\dot{U}_t = JD_t - JC_t$ makes it clear that unemployment starts converging back to its steady state only once net job creation turns positive. Figure 11(a) shows that this recovery of unemployment is very sluggish.

Figure 11: Unemployment and Net Job Creation



(a) Unemployment (b) Net JC Loops (c) Net JC Loops: Components

Note: The left panel depicts the impulse response dynamics of unemployment in the calibrated model in the typical business cycle experiment. The middle panel present the cyclical relation between the unemployment rate and net job creation derived from the impulse response dynamics. The right panel depicts separate unemployment / net job creation anti-clockwise loops each varying only EU_t, JC_t, or JD_t^Q.

Figure 11(b) shows that the relationship between net job creation and unemployment in the model is characterised by anti-clockwise loops. At the beginning of the recession unemployment increases as net job creation falls. Subsequently the slow recovery of unemployment is due to a barely positive (and decreasing) net job creation. These dynamics are consistent with Figure 2(b) in the Introduction which shows the same pattern occurs in the data. Hence, to understand why unemployment is so persistent it is crucial to understand the dynamics of net job creation.

The role of unreplaced quits First note that (ignoring exogenous layoffs λ_u) equation (10) implies that the hire flow from unemployment, UE_t, satisfies $JC_t = UE_t + JD_t^Q$, where recall that JC_t describes the flow of new jobs created, re-interpreted as new hiring chains created.[25] The right hand side of this equation describes the corresponding destruction of hiring chains which occurs when either the job is taken by unemployed workers UE_t, or by workers who quit and the previous employer declines to hire a replacement, JD_t^Q. Using this relationship and $\dot{U}_t = EU_t - UE_t$, changes in the stock of unemployment can be expressed as

$$\dot{U}_t = EU_t - \underbrace{(JC_t - JD_t^Q)}_{UE_t}, \tag{22}$$

[25] Accounting for the λ_u layoff shock gives the full formula as $JC_t + \lambda_u N_t^r = UE_t + JD_t^Q$, where the $\lambda_u N_t^r$ term additionally accounts for hiring chains started by firms to replace exogenous layoffs. N_t^r is the mass of employment at firms who do replace quits and λ_u layoffs (i.e. with $i \geq 3$).

where EU_t captures the inflows from employment into unemployment, while $UE_t = JC_t - JD_t^Q$ captures the outflows.This equation highlights the impact of unreplaced quits in shaping the dynamics of net job creation and unemployment: Newly created jobs generate hiring chains that end either with a hire out of unemployment or by poaching from firms which do not replace their workers. As shown by equation (22), a hiring chain that ends with a hire out of unemployment helps reduce the unemployment pool, while one that ends with an unreplaced quit does not.

Since $NetJC_t \equiv EU_t - (JC_t - JD_t^Q)$ we can analyse each of its three drivers in turn. In Figure 11(c) we plot the contribution of EU_t, JC_t, and JD_t^Q in driving the loop depicted in Figure 11(b) by plotting net job creation if we allow only each component to vary, holding the others at their pre-shock values.[26] A surprising finding of this exercise is that job creation and layoffs do not contribute to the recovery of unemployment, because they drive loops which do not bring net job creation to be positive during the recovery (or do just barely, in the case of layoffs). The paths for the underlying values of EU_t, JC_t, and JD_t^Q can be found in Figure 10, making the reason for this clear: Job creation and layoffs do not significantly *overshoot* their initial value during the recovery. Job creation falls in the recession and then recovers back to its original value, which subtracts from the employment stock but then never adds any back. For layoffs a similar pattern occurs, but instead layoffs increase in the recession. This behaviour is precisely why job creation is negatively and layoffs positively correlated with the level of unemployment in the data, as shown in Table 2.

Figure 11(c) shows that job destruction from unreplaced quits instead drives clockwise loops which immediately push net job creation positive. As described in equation (13), when unemployment rises JD_t^Q decreases, *dampening* the increase in unemployment. Intuitively, at the trough of the recession newly created hiring chains have their highest chance of ending with a hire out of unemployment. As the economy recovers and unemployment starts to fall, job destruction from unreplaced quits starts increasing, which reduces net job creation and hence *slows down* the recovery of unemployment. This occurs as the crowding out of on-the-job search by unemployed workers decreases during the recovery, and the probability that newly created hiring chains end with an unreplaced quit rises. In summary, the cyclical behaviour of unreplaced quits dampens the peak rise in unemployment but makes unemployment more persistent during its recovery.[27]

[26] For example, for job creation we plot $\bar{EU} - (JC_t - \bar{JD}^Q)$, where $\bar{EU}$ is the value just before the shock hits.

[27] An alternative way to investigate the importance of JD_t^Q for aggregate dynamics is to ask what would need to happen in its absence. To see this, consider a simple accounting counterfactual in which unreplaced quits are not allowed to vary over the business cycle and are held at a fixed level JD^Q. For any simulated path of unemployment, U_t, and layoffs, EU_t, we can ask what the counterfactual paths for JC_t and JD_t would need to be to match this unemployment path if JD^Q is held constant. Counterfactual JC can be computed directly as $JC_t^{cf} = EU_t + JD^Q - \dot{U}_t$, where EU_t and $\dot{U}_t$ are a given simulated path and JD^Q is the fixed value of JD from unreplaced quits. Counterfactual JD is $JD_t^{cf} = EU_t + JD^Q$. Without cyclical changes in JD^Q, overall JD is positively correlated with unemployment because it inherits only the cyclical properties of layoffs. Matching our model's business cycle simulation paths (Table 2) this

The importance of c_0 Why do job creation and layoffs not respond strongly enough to high unemployment to adjust and help make net job creation positive? This is explained by the low estimated value of the cost of hiring a worker, c_0. This parameter affects how the value of a firm $v(.)$, as described in equation (19), responds to the aggregate quit rate through the term $-[\lambda_1(\Omega) + \lambda_u] \min[v(i, \Omega), c_0]$. Since firms have to pay the hiring cost in order to replace quits more often when unemployment is low and quits are high (i.e. λ_1 is high) a larger value of c_0 implies more responsive firm values to changes in aggregate conditions. After the initial outburst of layoffs as the recessionary shock hits, a large c_0 would imply a sharp increase in firm values that would increase job creation, and reduce layoffs, sufficiently to make unemployment quickly recover. In the data, however, we observe a high proportion of replacement hiring, which implies that c_0 cannot be very high. Indeed, Table 1 shows that the estimated value of c_0 is equivalent to 78% of one month's average output per worker. Given the relatively small value of c_0, firm values do not respond strongly to changes in aggregate conditions and imply quits and unreplaced quits recover slowly, consistent with a persistent unemployment rate.

To further demonstrate the role of c_0, we perform a counterfactual calibration of the model where we raise c_0 by 50%. In order to isolate the role of this channel without changing the amount of replacement hiring, we give firms a flow subsidy in order to raise values and hold $v_2 < c_0 < v_3$ despite the higher value of c_0. We recalibrate the elasticities of firm entry, job creation, and layoffs to hold their standard deviations at the same value as our estimated model (see Appendix C for details). The higher value of c_0 leads to a larger feedback from unemployment to job creation and layoffs which reduces the persistence of unemployment to 0.899. Given that the persistence of the exogenous aggregate shock process is 0.848, this represents a 40% reduction in the excess persistence of unemployment over the one of aggregate productivity. Moreover, the increased feedback from unemployment to firm value now leads layoffs to strongly decrease during the recovery phase of recessions, leading to a negative correlation between layoffs and unemployment (-0.1237) which is at odds with the positive correlation in the data (and our estimated model) shown in Table 2.

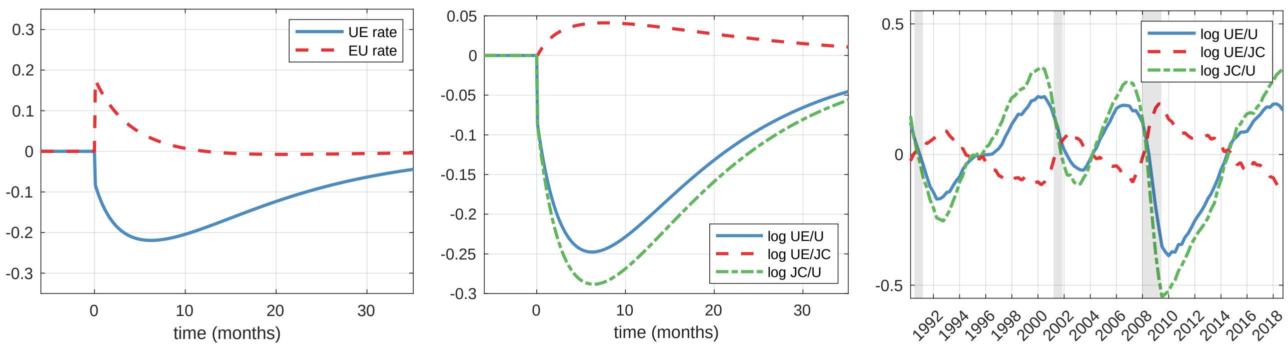
7.4 The role of job finding

Our emphasis on the importance of the cyclical dynamics of job destruction due to unreplaced quits and job creation in determining unemployment dynamics is fully consistent with the view of the labour market that argues unemployment fluctuations are primarily driven by unemployed workers' job find-

would imply a counterfactual JD-unemployment correlation coefficient of +0.20 instead of -0.25 as in the data and -0.57 in the estimated model. JC must also become positively correlated with unemployment, implying a counterfactual JC-unemployment correlation coefficient of +0.50 instead of -0.31 as in the data and -0.61 in the estimated model. Similar inconsistencies arise if we force the EU_t path to adjust instead of the JC_t path, since EU must now become incorrectly negatively correlated with unemployment.

ing rate (see Shimer, 2012, Elsby et al., 2008, among others). Figure 12(a) depicts the impulse response dynamics of the job finding rate and the layoff rate in the calibrated model. These show that the job finding rate is indeed the more important force behind the "in-and-out" dynamics of unemployment. Not only it exhibits a more pronounced fall following the recessionary shock, but it also takes much longer relative to the layoff rate to revert to its long-run level. Table 2 reconfirms this conclusion, showing that the cyclical component of the UE rate is more volatile and persistent than the cyclical component of the layoff rate, both in the model and in the data.

Figure 12: Decomposing Job Destruction and Job Creation - Model



(a) UE and EU rates (b) UE decomp: model (c) UE decomp: data

The left and centre panels plot variables during our typical recession experiment. The right panel plots our UE rate decomposition on real world data, which is logged and HP filtered with parameter 10^5. Cyclical job creation rate is constructed from the quarterly data used by Davis, Faberman and Haltiwanger (2012), updated by these authors. Cyclical unemployment is constructed using quarterly data from the Current Population Survey, and the UE flow is constructed following Shimer (2005).

Taking on-the-job search into account, we can gain further insights into the dynamics of the UE rate by using the decomposition: $\log \frac{UE_t}{U_t} = \log \frac{UE_t}{JC_t} + \log \frac{JC_t}{U_t}$. The first term UE_t/JC_t, describes *job creation yield*, the fraction of new jobs created which result in a worker being hired out of unemployment. It is clear that this term is not fixed at one in our model, as it would be in a model without a job ladder. The second term JC_t/U_t describes the gross flow of new jobs created per unemployed worker. Figure 12(b) plots the impulse responses of these two items. Although unemployed workers job finding rate is procyclical, the crowding out of on-the-job search by unemployed job search implies the job creation yield increases in the recession: a new job created is more likely to result in the hiring of an unemployed worker. The crucial insight is that despite the increase in the job creation yield, the steep fall in the UE rate is due to the even steeper fall in jobs created per unemployed worker JC_t/U_t during the recession. This occurs not because job creation flows dramatically fall during the recession but because they respond weakly to increasing unemployment, as explained in Section 7.3. Unemployment therefore remains high and recovers slowly due to inelastic job creation and its interaction with unreplaced quits. Figure 12(c) presents the same decomposition as above on US data and shows that the dynamics of the UE rate and job creation yield implied by the model are fully consistent with the data.

8 Conclusion

This paper has developed a new equilibrium business cycle model of the US labor market which is consistent both with the underlying distribution of firm growth rates across firms [by age and size] and macro-evidence regarding gross job creation and job destruction flows over the cycle. The framework not only successfully generates the (targeted) average firm size distribution by age but also the (untargeted) distributions of firms and of employment by firm size. The approach also provides an important new insight - that net job creation is uncorrelated with unemployment. We have shown that it is this property of the data which is central to explaining the business cycle frequencies of unemployment.

The approach has used aggregate productivity shocks, rather than discounting shocks, as the driver of the economy. It would seem unlikely that changing to discounting shocks would much affect our insights. For example any negative aggregate shock will always lead to a spike in layoffs because a key component of the job destruction process is low productivity firms have small surplus. Furthermore as described above, the efficiency wage distortion will always imply it is the quit process, rather than wages, which move the most over the cycle. That is, high unemployment always causes a steep fall in quit rates, the consequent collapse of the job ladder and a slow recovery because net job creation rates respond (at most) weakly to high unemployment.

References

[1] Audoly, R. 2020. "Firm Dynamics and Random Search over the Business Cycle," Mimeo.

[2] Barlevy, G. 2002. "The Sullying Effect of Recessions," *Review of Economic Studies*, 69(1): 65-96.

[3] Bertolla, G. and R. Caballero. 1994. "Cross-Sectional Efficiency and Labour Hoarding in a Matching Model of Unemployment," *Review of Economic Studies*, 61: 435-456.

[4] Bilal, A, N. Engbom, S. Mongey and G. Violante. 2021. "Firm and worker dynamics in a frictional labor market," *Econometrica*, 90(4): 1425-1462.

[5] Brown, C. and J. Medoff. 1989. "The Employer Size-wage Effect," *Journal of Political Economy*, 97(5): 1027-1059.

[6] Burdett, K. and D. Mortensen. 1998. "Wage Differentials, Employer size and Unemployment," *International Economic Review*, 39: 257-273.

[7] Carrillo-Tudela, C., H. Gartner and L. Kaas. 2022. “Recruitment Policies, Job-Filling Rates and Matching Efficiency”. Mimeo.

[8] Coles, M. G. and E. Smith. 1998. “Marketplaces and Matching,” *International Economic Review*, 39(1): 239-254.

[9] Coles, M. G. and A. Moghaddasi. 2018. “Do Job Destruction Shocks Matter in the Theory of Unemployment?,” *American Economic Journal: Macroeconomics*, 10(3): 118-136.

[10] Coles, M. G. and D. Mortensen. 2016. “Equilibrium Labor Turnover, Firm Growth, and Unemployment,” *Econometrica*, 84: 347-363.

[11] Coles, M. G. 2001. “Equilibrium Wage Dispersion, Firm Size and Growth,” *Review of Economic Dynamics*, 4(1): 159-187.

[12] Cooper, R and J. Haltiwanger. 2006. “On the Nature of Capital Adjustment Costs,” *Review of Economic Studies*, 73(3): 611-633.

[13] Davis, S. J., R. J. Faberman and J. Haltiwanger. 2012. “Labor market flows in the cross section and over time,” *Journal of Monetary Economics*, 59(1): 1-18.

[14] Davis, S. J., R. J. Faberman and J. Haltiwanger. 2013. “The Establishment-Level Behavior of Vacancies and Hiring,” *The Quarterly Journal of Economics*, 128(2): 581-622.

[15] Davis, S. J. and J. Haltiwanger. 1992. “Gross Job Creation, Gross Job Destruction, and Employment Reallocation,” *The Quarterly Journal of Economics*, 107(3): 819-863.

[16] Davis, S. J. and B. Samaniego de la Parra. 2020. “Application Flows”. Mimeo.

[17] Diamond, P. A. 1982. “Aggregate Demand Management in Search Equilibrium,” *Journal of Political Economy*, 90(5): 881-894.

[18] Elsby, M. and A. Gottfries. 2022. “Firm Dynamics, On-the-Job Search, and Labor Market Fluctuations”, *Review of Economic Studies*, 89(3): 1370-1419.

[19] Elsby, M., A. Gottfries, R. Michaels and D. Ratner. 2021. “Vacancy Chains”. University of Edinburgh, Department of Economics, mimeo.

[20] Fort, Teresa, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2008. “How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size”, IMF Economic Review 61, no. 3 (2013): 520–59.

[21] Faberman, J. and E. Nagypal. 2008. “Quits, Worker Recruitment, and Firm Growth: Theory and Evidence”, Working Paper No. 08-13, Federal Reserve Bank of Philadelphia, USA.

[22] 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.

[23] Haltiwanger, J. C., R. S. Jarmin, R. Kulick and J. Miranda. 2017. "High Growth Young Firms: Contribution to Job, Output, and Productivity Growth,". In J. Haltiwanger, E. Hurst, J. Miranda, and A. Schoar, (Eds). *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, University of Chicago Press, Chicago, USA.

[24] Hosios, A. 1991. "On the Efficiency of Matching and Related Models of Search and Unemployment". *Review of Economic Studies*, 57 (2): 279-298.

[25] Klette, T. J. and S. Kortum. 2004. "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 112(5): 986-1018.

[26] Mercan, Y. and B. Schoefer. 2020. "Jobs and Matches: Quits, Replacement Hiring, and Vacancy Chains", *American Economic Review: Insights*, 2(1): 101-124.

[27] Mortensen, D. and C. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment". *Review of Economic Studies*, 61(3): 397-415.

[28] Moscarini, G. and F. Postel-Vinay. 2018. "The Cyclical Job Ladder", *Annual Review of Economics*, 10: 165-88.

[29] Moscarini, G. and F. Postel-Vinay. 2013. "Stochastic Search Equilibrium," *Review of Economic Studies*, 80(4): 1545-1581.

[30] Moscarini, G. and F. Postel-Vinay. 2012. "The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment", *American Economic Review*, 102(6): 2509-2539.

[31] Mueller, A. I., D. Osterwalder, J. Zweimueller and A. Ketterman. 2022. "Vacancy Durations and Entry Wages: Evidence from Linked Vacancy-Employer-Employee Data". Mimeo.

[32] Lentz, R. and D. T. Mortensen. 2008. "An Empirical Model of Growth Through Product Innovation," *Econometrica* 76(6): 1317-1373.

[33] Schaal, E. 2017. "Uncertainty and Unemployment," *Econometrica*, 85(6): 1675-1721.

[34] Shapiro, C. and J. E. Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review*, 74(3): 433-444.

[35] Shimer, R. 2012. "Reassessing the Ins and Outs of Unemployment", *Review of Economic Dynamics*, 15(2): 127-148.

[36] Shimer, R. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95(1): 25-49.

ONLINE APPENDIX

A Data Appendix

A.1 Data sources and treatment

We use the following data throughout the paper. Our main sample period used for estimation purposes is the one for which we can get data on all variables simultaneously, which is 1990Q2 to 2018Q4.

Business Dynamics Statistics (BDS): We use the BDS to construct data stratified by firm age. We use the 2018 release, which is available yearly from 1978 to 2018, and take the national data, split by firm age. We use data on the number of firms (and total employment in firms) of each age bin to calibrate our model. We also measure firm entry using this dataset, as the number of firms aged 0 in each year, and firm exit, as given in the dataset. We also use the data on Job Creation and Destruction by age, but instead calibrate our model to the quarterly JC and JD data from Davis, et al. (2012).

Davis, Faberman, and Haltiwanger (2012, DFH): We extensively use the data provided by Davis et al. (2012) which underlies the analysis in their paper. We are grateful to the authors for sharing the (updated) data which underlies the plots in their paper. This data set consists of quarterly data from 1990Q2 to 2018Q4. We use their estimates of aggregate job creation and job destruction, as well as layoffs and quits. We add their measure of "other separations" to layoffs. Their data are given as rates of total employment. We calibrate our model to match the average of these data over our sample, as well as the HP-filtered time series. In addition, we use data from this paper to form an estimate of the fraction of worker quits that firms replace by undertaking a replacement hire. We discuss this further below.

Bureau of Labor Statistics (BLS): We use monthly aggregate data from the Current Population Survey. We aggregate these data up to a quarterly frequency by taking the simple average. We use data on total employment (CE16OV) and unemployment (UNEMPLOY), in levels and seasonally adjusted. We additionally use data on the total number of people unemployed for less than five weeks (UEMPLT5) to construct the unemployed worker UE rate, following the approach of Shimer (2005).

Bureau of Economic Analysis (BEA): To construct our measure of labour productivity (output per worker) we use data on quarterly real GDP (GDPC1) from the BEA. Labour productivity is calculated as real GDP divided by total employment from the BLS data.

Compustat: Since our model assumes constant returns to scale, we do not allow for permanent productivity differences across firms, as these would lead to permanent differences in employment growth rates (rather than levels, as would be true in a model with decreasing returns to scale). Therefore, we calibrate our firm-level productivity process to the within-firm standard deviation of productivity shocks, rather than the across firm standard deviation. To compute this measure, we use data from Compustat. We use data on all US based firms in their sample, and use data only on sales and total employment. We deflate sales using the GDP deflator (GDPDEF, from the BEA) to create a measure of real sales, and then define firm-level labour productivity each year as real sales over employment. We drop firm-year observations with missing or negative sales or employment, and winsorize the data by dropping the top and bottom 1% of data by both yearly sales growth and employment growth. We take the log of labour productivity, and regress it on firm and year fixed effects, and take the residual as our measure of firm-level productivity, corrected for firm-level averages and aggregate changes. We take the standard deviation of this measure, which yields a value of 28.38%, computed from 291,703 firm-year observations.

A.2 Estimating the fraction of quits which are replaced

To estimate the amount of replacement hiring in the model, we draw information both from gross flows and from underlying firm-level data. Firstly, we note that the amount of replacement hiring is not simple to observe from aggregate flows, due to the fact that firms may do replacement hiring for two reasons: either to replace workers who quit, or to replace workers they lay off for being a bad match but where the firm wants to keep the job open. Through the lens of the model, the total hiring rate is equal to

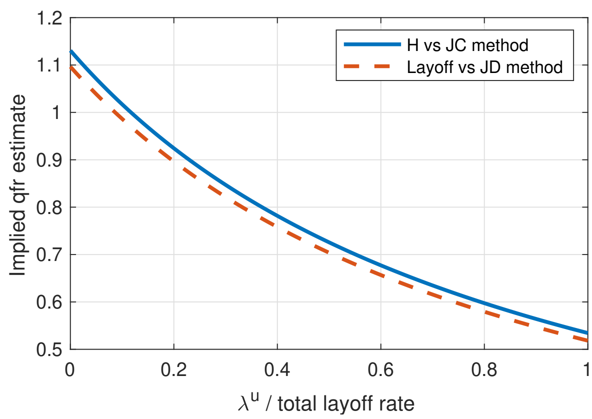
$$h_t = jc_t + qfr_t\left(q_t + \lambda^u\right) \tag{23}$$

where we define qfr_t as the fraction of worker quits (and layoffs due to bad worker match) which are replaced. Recall that λ^u is the rate at which workers are fired for being a bad match, but where the firm's capital remains intact so the firm has the option to hire to replace them. qfr serves as our calibration target for the amount of replacement hiring in the model. Notice that three objects in this equation are observable in the DFH dataset: hiring (h_t), job creation (jc_t), and quits (q_t). If we assumed that all layoffs were due to job destruction (and firms never fired workers with the aim of replacing them with another worker) then $\lambda^u = 0$, and estimating the degree of replacement hiring would be simple using this aggregate data alone. In this case, simply rearrange (23) to yield $qfr_t = (h_t - jc_t)/q_t$.

However, the data in DFH suggest that firms do indeed replace some of their workers who leave due to layoff, so this approach is likely not valid. In particular, in their well-known "hockeystick" plot (their Figure 7(b)) we observe that firms who have positive employment growth, and hence are expanding, still extensively use worker layoffs. In fact, our calculations below suggest that the average layoff rate for non-contracting firms appears to be around 2.73% per quarter. Given that these firms are expanding their employment on net, it is likely that they are replacing some of the workers who they have laid off, meaning that $\lambda^u > 0$. Indeed, through the lens of our model, firms which perform job creation necessarily replace worker quits.

Given that $\lambda^u > 0$ therefore seems like a more reasonable assumption, we can then return to (23) to understand the impact this has on estimated quit replacement. Expressing the relationship in steady state gives $h = jc + qfr\,(q + \lambda^u)$, where we take h, jc, and q directly from the average values in DFH's data. This gives the implied value of qfr for any assumed λ^u as $qfr = (h - jc)/(q + \lambda^u)$. Without any further information, λ^u is constrained to lie within the range 0 (in which case all layoffs are due to job destruction) and the total layoff rate in the data, l (in which case all layoffs are replaced, and not due to job destruction) [see equation (25) below]. We plot the implied value of qfr in this range in the blue line in Figure A.1 below.

Figure A.1: Estimated fraction of quits replaced vs. assumed λ^u



This procedure bounds the fraction of quits and layoffs which are replaced to be between around 50%, if all layoffs are assumed to be replaced, and 100%, if only 10% of layoffs are assumed to be replaced. Notice that for values below 10% the data implies that more than 100% of quits are replaced. Before going further, we therefore note that our chosen value for the estimation, $qfr = 80\%$, happens to lie approximately in the middle of the upper and lower bounds implied by the aggregate flow data.

One could potentially use other aggregate flow relationships, such as those between job destruction and layoffs, might help estimate the fraction of layoffs which are replaced, and hence pin down qfr. However, we found this challenging as the aggregate relationships are by definition collinear, and more information is needed. To see this, consider that job destruction is given by

$$jd_t = jd_t^l + jd_t^q = jd_t^l + (1 - qfr_t)(q_t + \lambda^u) \tag{24}$$

and layoffs by

$$l_t = jd_t^l + \lambda^u \tag{25}$$

where l_t is layoffs in the data, jd_t^l is job destruction shocks which induce layoffs, and jd_t^q is job destruction due to unreplaced quits and layoffs. Taking jc, jd, q, and l as data, (23), (24), and (25) appear to provide three equations which can solve for the three unknowns qfr_t, λ^u, and jd_t^l. However, the equations actually contain the same economic content and are collinear. Combining (24) and (25) to yield

$$qfr_t(q_t + \lambda^u) = l_t - jd_t + q_t \tag{26}$$

and rearrange (23) to yield

$$qfr_t(q_t + \lambda^u) = h_t - jc_t. \tag{27}$$

As the left hand sides of these equations are identical, the three equations together cannot be solved for a unique solution for qfr_t, λ^u, and jd_t^l. Instead, combining these two equations implies an adding-up condition which should hold in the data in theory: $jc_t - jd_t = h_t - l_t - q_t$. In practice, the adding up condition is very slightly violated, meaning that (26) and (27) provide very slightly different estimates of qfr_t for a given assumed λ^u. The estimate from (26) is given in Figure A.1 as the dashed red line, which is very similar the the previous estimate.

The discussion above shows that additional data must be included to estimate the fraction of quits which are replaced, and we investigate two approaches.

As a first approach, note that with knowledge of λ^u, the value of qfr can be calculated using the accounting relationship above. λ^u is the rate at which workers are fired for being a bad match, with the firm having the option to replace them if desired. Through the lens of the model, this can be identified as the layoff rate at expanding firms, who perform no job destruction and so any layoffs must be due to the λ^u shock. To estimate this in the data, we use the hockeystick and growth rate distribution plots in DFH.[28] We have access only to the growth rate distributions from 2006 and 2008-9 (as plotted in DFH) and so use data for the distribution and hockeysticks from 2006 to construct our estimate. Accordingly, this is data from a single year, which corresponds to an estimate for a typical non-recession year.[29] For each growth rate bin $i = -200, 199, ..., 200$, we have data on the

[28] The authors very kindly provided us with the data behind these plots, which consists of the hires, layoff, total separation, and quit rate at each growth rate bin (their Figures 6 and 8), and the (employment weighted) kernal density function of firms in each bin (their Figure 5). The data are provided on slightly different grids for each plot, and we interpolate the data onto an integer grid from -200 to 200.

[29] The results are robust to using the hockeysticks from all years (their Figure 6) integrated using the average of the 2006 and 2008-9 growth rate distributions to roughly attempt to form an estimate for all years. However, since the match of hockeysticks and growth rate distribution sample is not exact in this case, we prefer to use the data from 2006 only.

mass of employment at establishments with that growth rate (d_i) and the construct a layoff rate at that bin (l_i). We calculate the average layoff rate in all bins with non-declining employment growth as $\sum_{i=0}^{200} l_i d_i / \sum_{i=0}^{200} d_i = 2.73\%$. Under the identifying assumption that λ^u is constant across firms (as it is in the model), this implies an estimate $\lambda^u = 2.73\%$, which is 39% of the average layoff rate of 7.0% in the DFH data in 2006. Referring back to Figure A.1, we see that 39% of layoffs being potentially replaceable implies a value of qfr of approximately 80%, which is the value used in our calibration.[30] As an alternative, we also directly calculated the fraction of quits replaced from the 2006 hockeystick data for quits and hires, and found that 79.6% of quits were replaced. Specifically, we assume that expanding firm bins replace all quits and layoffs ($qfr_i = 1$). For contracting bins, we calculated the fraction of quits replaced as $qfr_i = h_i/(q_i + \lambda^u)$. Taking the $(q_i + \lambda^u)$-weighted average of qfr_i across the whole distribution yields 79.6%.

As a second approach, we consider the notion of replacement hiring in Elsby et al. (2021). They define a broad notion of replacement hiring using JOLTS data as follows. For each establishment, they consider replacement hires as the minimum of gross hires and quits in a given quarter. They then sum across establishments, and find that, by this definition, around 45% of all hires are replacement hires. Doing the same exercise on simulated data from our model finds that 40% of hires are replacement hires. As mentioned in the text, our model also generates that around 50% of firms have zero net employment change over 3 months, and since these firms also lose workers to quits, this serves as another measure of replacement hiring. Elsby et al. (2021) find this number to be around 55% and 65% in the QCEW and JOLTS data respectively. Finally, their strictest measure of replacement hiring is the total hiring at firms with zero net employment change as a fraction of total hiring. This number is 7.5% in their data, and 7.1% in our model. By all these measures our model generates a substantial amount of replacement hiring, close to the measures in the data. This provides an alternative justification for our calibrated value of $qfr = 0.8$, which delivers a sensible amount of replacement hiring by these alternative measures, and suggests that our results would be robust to instead using these measures as targets in our estimation.

A.3 Procyclicality of JD: robustness to alternative assumptions and data sources

In this section we discuss the correlations between JD, layoffs, and other variables with unemployment in more detail. In the Introduction, we argued that the overall JD rate is procyclical (positive correlation with unemployment) while the layoff rate is countercyclical. Highlighting this difference is, to our knowledge, is novel. For this reason, we provide extensive robustness showing that this

[30] If roughly 40% of layoffs are replaceable and the replacement rate is 80%, this implies that 32% of layoffs are actually replaced.

difference in cyclicality between JD and layoffs is robust. We first discuss the robustness of the result in relation to many different data treatments, still using our main DFH (Davis et al. 2012) dataset. We then show that the procyclicality of JD also holds on a different dataset, the yearly BDS dataset.

Robustness on main DFH dataset We provide an exhaustive battery of checks to show that no particular data treatment assumptions are driving our results. These are presented in the series of tables below, which we discuss in turn.

Firstly, our findings are robust to various detrending assumptions. In Table 3 we plot the correlation of job destruction (jd), layoffs (lo), job destruction net of layoffs ($jd - lo$), and job creation (jc) with the level of unemployment. We compute the correlation of unemployment at time t with the labour market flows between t and $t+1$ (as we also do in the model) so these correlations show the correlation between unemployment and how labour market flows then evolve in the next three months. Our baseline results detrend all variables using the HP-filter with parameter 10^5, following Shimer (2005), as shown in the first column. All of the flows are procyclical, apart from layoffs. The remaining columns apply different detrending methods: HP-filter with parameter 1600, Baxter-King approximate bandpass filter (frequencies 6 to 32 quarters), linear detrending, and finally the raw data without detrending. The correlations all maintain the same signs across all different detrending assumptions. The only exception is layoffs when not detrended, which now also become procyclical. Inspecting the raw data reveals that this is because layoffs have been trending downwards over time which creates a spurious correlation with unemployment, which highlights the importance of detrending data to deal with the well know slowing down of many labour market flows in recent decades.

Table 3: Cyclical correlations of JD, layoffs, and JC with unemployment

	HP 10^5	HP 1600	BK	Linear	Raw data
jd	-0.226	-0.170	-0.135	-0.228	-0.387
lo	0.143	0.156	0.211	0.164	-0.227
$jd - lo$	-0.607	-0.521	-0.700	-0.596	-0.457
jc	-0.270	-0.073	-0.068	-0.401	-0.446

Each row gives the correlation of the variable with unemployment. Each column detrends the data with a different method before computing the correlations, with HP indicating the HP-filter with a given parameter, BK the Baxter-King approximate bandpass filter (frequencies 6 to 32 quarters), Linear indicating linear detrending, and Raw data the data without detrending. JC, JD, and layoff rates are from the quarterly data used by Davis, Faberman and Haltiwanger (2012), updated by these authors. Unemployment is the number of unemployed individuals using quarterly data from the Current Population Survey. The data are from 1990Q2 to 2018Q4, and we compute the correlation between unemployment at date t with the flows (e.g. for JD) between t and $t + 1$.

Secondly, we stress the important distinction between the correlation of flows with the *level* of unemployment and the *change* in unemployment. The two approaches give different results, as shown and discussed by Moscarini and Postel-Vinay (2012) and Haltiwanger et al. (2018). Our empirical and model findings are robust to this distinction, because these correlations measure different concepts, which our model helps clarify. In Table 4 we give the correlations with the change in unemployment. This shows that job destruction is positively correlated with the change in unemployment, despite being negatively correlated with the level of unemployment. This does not cast doubt on the results presented in levels, but actually highlights the different cyclical features that the level and change in unemployment capture.

To understand this further, consider that in our model the change in unemployment is by construction driven by JC and JD as

$$\dot{U}_t = JD_t - JC_t \tag{28}$$

Thus, it is very natural that JD must be positively correlated with the change in unemployment, because (28) shows that an increase in JD is literally the mechanical driver of an increase in $\dot{U}$. In our model, it is also true that JD and layoffs (JC) are positively (negatively) correlated with the change in unemployment, and hence our model is simultaneously consistent with the correlations of JD, layoffs, and JC with *both* the level and change in unemployment. Specifically, we compute log HP-filtered data as we did for our main results in Table 2. The correlation of each variable with the first difference of unemployment in the data (model) is -0.5247 (-0.6789) for JC; 0.7302 (0.4598) for JD; and 0.6611 (0.7536) for layoffs.

Table 4: Cyclical correlations of JD, layoffs, and JC with Δu

	HP 10^5	HP 1600	BK	Linear	Raw data
jd	0.759	0.726	0.817	0.742	0.488
lo	0.721	0.699	0.746	0.734	0.459
$jd - lo$	0.147	0.073	0.045	0.038	-0.005
jc	-0.494	-0.417	-0.747	-0.536	-0.056

Each row gives the correlation of the variable with the change in unemployment. Each column detrends the data with a different method before computing the correlations, with HP indicating the HP-filter with a given parameter, BK the Baxter-King approximate bandpass filter (frequencies 6 to 32 quarters), Linear indicating linear detrending, and Raw data the data without detrending. JC, JD, and layoff rates are from the quarterly data used by Davis, Faberman and Haltiwanger (2012), updated by these authors. Unemployment is the number of unemployed individuals using quarterly data from the Current Population Survey. The data are from 1990Q2 to 2018Q4, and we compute the correlation between unemployment change ($u_{t+1} - u_t$) with the flows (e.g. for JD) between t and $t+1$.

At the same time, how can JD be negatively correlated with U while being positively correlated with ΔU? This is because the two correlations capture different frequencies of the data. The $(JD, \Delta U)$ correlation captures the very short term correlation between high JD and rising U which is driven by *layoffs*. The (JD, U) correlation instead captures the longer term correlation between JD and the level of unemployment, which is driven by *unreplaced quits*. Essentially, JD is made up by a fast moving component (layoffs) and by a slow moving component (unreplaced quits) with different correlations with unemployment, and the $(JD, \Delta U)$ and (JD, U) correlations have opposite signs because each picks up a different component respectively.

Thirdly, the correlations are robust to whether data are expressed in levels or logs. In Table 5 we repeat the correlations this time taking the logs of both the flows and the unemployment level. Since $jd - lo$ sometimes takes negative values it cannot be logged, and so is excluded from this exercise. The correlations retain the same signs and properties as the data in levels in Table 3.

Table 5: Cyclical correlations of JD, layoffs, and JC with unemployment (all logged)

	HP 10^5	HP 1600	BK	Linear	Raw data
$\log jd$	-0.254	-0.113	-0.058	-0.262	-0.389
$\log lo$	0.174	0.230	0.301	0.175	-0.211
$-$	0	0	0	0	0
$\log jc$	-0.305	-0.123	-0.159	-0.433	-0.447

Each row gives the correlation of the log of variable with log unemployment. Each column detrends the data with a different method before computing the correlations, with HP indicating the HP-filter with a given parameter, BK the Baxter-King approximate bandpass filter (frequencies 6 to 32 quarters), Linear indicating linear detrending, and Raw data the data without detrending. JC, JD, and layoff rates are from the quarterly data used by Davis, Faberman and Haltiwanger (2012), updated by these authors. Unemployment is the number of unemployed individuals using quarterly data from the Current Population Survey. The data are from 1990Q2 to 2018Q4, and we compute the correlation between unemployment at date t with the flows (e.g. for JD) between t and $t+1$.

Fourthly, an important data issue is how to treat "Other Separations" in the DFH dataset. Separations in their dataset are split into layoffs, quits, and other separations. Since other separations are classified neither as layoffs or quits, which make up the only two types of separation in our model, we must deal with this extra category somehow. In practice, this does not affect the results because other separations is a relatively small category. In our main dataset, we add other separations to layoffs and hence treat them as layoffs rather than quits. In Table 6 we instead add other separations to quits, and show that the correlations all retain the same signs as in the main specification in Table 3.

Table 6: Cyclical correlations with unemployment (alternative treatment of other separations)

	HP 10^5	HP 1600	BK	Linear	Raw data
jd	-0.226	-0.170	-0.135	-0.228	-0.387
lo	0.205	0.186	0.229	0.240	-0.184
$jd - lo$	-0.637	-0.527	-0.651	-0.652	-0.604
jc	-0.270	-0.073	-0.068	-0.401	-0.446

Each row gives the correlation of the variable with unemployment. Each column detrends the data with a different method before computing the correlations, with HP indicating the HP-filter with a given parameter, BK the Baxter-King approximate bandpass filter (frequencies 6 to 32 quarters), Linear indicating linear detrending, and Raw data the data without detrending. JC, JD, and layoff rates are from the quarterly data used by Davis, Faberman and Haltiwanger (2012), updated by these authors. Unemployment is the number of unemployed individuals using quarterly data from the Current Population Survey. The data are from 1990Q2 to 2018Q4, and we compute the correlation between unemployment at date t with the flows (e.g. for JD) between t and $t+1$.

Finally, we consider the robustness our our results to an alternative subsample of the data. Inspecting the time series for JD in Figure 2, we see that the period following the Great Recession exhibits the clearest negative correlation between JD and unemployment, and so one might worry that the post-Great Recession era is driving our results. It could be that the negative correlation is a feature novel to the Great Recession, which featured a very pronounced collapse in job-to-job quits. To investigate this, in Table 7 we repeat our correlations but excluding the post-Great Recession period, computing the correlations only up to 2010Q1. As can be seen in the table, all correlations maintain their same signs on this smaller subsample.

Overall, these various exercises indicate that the correlations we document are a robust feature of the data, and are not specific to any specific detrending and sampling assumptions.

Table 7: Cyclical correlations of JD, layoffs, and JC with unemployment (1990Q2 to 2010Q1)

	HP 10^5	HP 1600	BK	Linear	Raw data
jd	-0.075	-0.198	-0.082	-0.005	-0.167
lo	0.278	0.189	0.421	0.290	0.035
$jd - lo$	-0.613	-0.619	-0.750	-0.514	-0.518
jc	-0.346	-0.076	-0.320	-0.524	-0.434

Each row gives the correlation of the variable with unemployment. Each column detrends the data with a different method before computing the correlations, with HP indicating the HP-filter with a given parameter, BK the Baxter-King approximate bandpass filter (frequencies 6 to 32 quarters), Linear indicating linear detrending, and Raw data the data without detrending. JC, JD, and layoff rates are from the quarterly data used by Davis, Faberman and Haltiwanger (2012), updated by these authors. Unemployment is the number of unemployed individuals using quarterly data from the Current Population Survey. The data are from 1990Q2 to 2010Q1, and we compute the correlation between unemployment at date t with the flows (e.g. for JD) between t and $t + 1$.

Robustness to different dataset: BDS To further check the robustness of the correlation between JD and unemployment, we show that it holds on a different dataset. We turn to the BDS dataset, which is the yearly dataset we used for our firm age distribution calculations. This dataset features three differences from the DFH data. Firstly, it has a longer sample period, from and we use data from 1978 to 2018. Secondly, since it is yearly it measures the yearly JD and JC flows, which may in principle be different from the quarterly flows in DFH. Finally, it does not measure layoffs, and so we can only check the correlations of JD and JC with unemployment. Nonetheless, we identify the same patterns on this dataset.

We compute the correlations of the aggregate JD and JC rates with unemployment, where we calculate unemployment as the average unemployment within each year from the BLS data. We log and HP-filter all data with a HP-filter parameter of 100. Given that the dataset is yearly, it is very important to consider timing, and we correlate unemployment in year t with the JC and JD flows between year t and $t + 1$, to capture the correlation of unemployment with the rates going forward, as we do in the model. We find that both JC and JD are procyclical, just as in the quarterly DFH data: their correlations with unemployment in the yearly BDS data are -0.1035 and -0.1952 respectively. Hence, the procyclicality of JD appears to be a robust feature of both the quarterly and yearly data.

B Model Appendix

B.1 Proofs

Proof of Lemma 1: For convenience we suppress reference to Ω. It is immediate that $q(i) = \lambda_u + \lambda_1$ for $i \leq i^h$. Consider now $i > i^h$ where (16) implies equilibrium quit rate

$$q(i) = \lambda_u + \frac{\lambda_1}{\lambda} \int_i^1 \frac{h(j)[1-U]G'(j)}{\alpha + (1-\alpha)G(j)} dj.$$

Now $\lambda_1 = \phi\lambda_0$ and $\lambda = \lambda_0 U + \lambda_1(1-U)$ implies $\lambda_1/\lambda = \phi/[U + \phi(1-U)]$ while $\alpha \equiv \lambda_0 U/\lambda = U/[U + \phi(1-U)]$. Substituting out λ_1/λ and α in the above yields

$$q(i) = \lambda_u + \int_i^1 \frac{h(j)G'(j)}{\frac{U}{\phi(1-U)} + G(j)} dj.$$

Because $h(j) = JC(j) + q(j)$ for $j \geq i > i^h$ we also have

$$h'(j) = JC'(j) - \frac{h(j)G'(j)}{\frac{U}{\phi(1-U)} + G(j)}.$$

Now define $Z(j) = \frac{U}{\phi(1-U)} + G(j)$ and so $Z'(j) = G'(j)$. Integration by parts establishes:

$$\int_i^1 Z'(j)h(j)dj = [Z(j)h(j)]_i^1 - \int_i^1 Z(j)\left[JC'(j) - \frac{h(j)G'(j)}{\frac{U}{\phi(1-U)} + G(j)}\right] dj$$

and simplifying yields:

$$Z(1)h(1) - Z(i)h(i) = \int_i^1 Z(j)JC'(j)dj.$$

Integrating by parts then yields:

$$Z(1)h(1) - Z(i)h(i) = Z(1)JC(1) - Z(i)JC(i) - \int_i^1 Z'(j)JC(j)dj.$$

Substituting out $h(1) = \lambda_u + JC(1)$, $h(i) = JC(i) + q(i)$ implies:

$$\begin{aligned} Z(i)q(i) &= Z(1)\lambda_u + \int_i^1 Z'(j)JC(j)dj \\ &= Z(i)\lambda_u + \int_i^1 Z'(j)[\lambda_u + JC(j)]dj. \end{aligned}$$

Using $Z(i) = \frac{U}{\phi(1-U)} + G(i)$ and $Z'(j) = G'(j)$ then establishes the Lemma.

Proof of Proposition 2: Note these wage strategies are firm size invariant and have the job ladder property. By construction, beliefs B1-B3 are consistent with these equilibrium wage strategies, Bayes Rule and the restriction to monotone beliefs. Because these wage strategies have the job ladder property, the worker quit strategies (15) are optimal with the distribution of wage offers $F(.)$ given by (16). All that remains is to show is these wage strategies are indeed optimal. It is easy to verify that firms $i < i^h$ (who have $v < c_0$) strictly prefer to post wage $w' = w_{\min}$ [paying a higher wage reduces the employee quit rate but the firm's gain by doing so $v < c_0$ and this wage deviation is profit reducing]. Conversely each hiring firm is actually indifferent to posting any wage $w' \in [w_{\min}, \overline{w}]$ while posting wage $w' > \overline{w}$ is strictly profit decreasing [because quit rates cannot fall further]. Hence each firm's wage strategy is indeed privately optimal which completes the proof of Proposition 2.

B.2 Entry with more than one worker

In this section we show how to extend the model in the main text to allow a new firm to start with more than one employee, but $n_0 \in \{1, 2, .., N_0 + 1\}$ employees. This version of the model is the one used in the calibration. The reason for this extension is that in the data we observe that the majority of new firms have more than one employee in their first year. This extension allows the model to be closer to the data in this dimension as well for allowing start-up size to be potentially sensitive to the aggregate state Ω_t, while keeping the rest of its formulation as in the main text.

In particular, assume there is a unit measure of entrepreneurs who independently seek business ventures. At rate μ_0 an entrepreneur identifies a possible business venture whose investment cost $c^E \geq 0$ is considered an independent random draw from cost distribution $H^E(.)$. If the entrepreneur chooses not to invest, the venture is lost with no recall. If the entrepreneur invests, a start-up is created with $n_u \in \mathbb{N}^+$ employees drawn randomly from the pool of unemployed workers. Its productivity $i \sim U[0, 1]$ is then revealed at which point we refer to the start-up as a new firm. The new firm thus has n_u employees, is in state i, and subsequently pays wages and expands/contracts like all other existing firms. Each new firm also has N_0 immediate potential expansion job opportunities where $N_0 \in \mathbb{N}^+$ is exogenous. The job creation process is the same as for existing firms: associated with each potential expansion is an independent cost draw $c^{JC} \sim H^{JC}$ and recruitment cost c_0 to hire a worker. If the new firm invests its initial size n_0 increases by 1. If the new firm does not invest, the expansion opportunity is lost with no recall. Hence each new firm begins life with initial employment $n_0 = n_u + \widetilde{n}_i$ where hires $\widetilde{n}_i$ are a binomially distributed random variable with N_0 independent trials and an [endogenous] probability of investment which depends on (i, Ω).

Note that by allowing new firms to start with $n_0 > 1$ we need to amend the definition of equilibrium by adding: $P^E(\Omega)$ is the probability an entrepreneur invests in a start-up in state Ω and, given realised i, its starting size $n_0 = n_u + \widetilde{n}(i, \Omega)$ maximises expected profit. We now turn to derive $P^E(\Omega)$.

Firm Optimality [New Firms] Suppose in state Ω, an entrepreneur creates a new firm with revealed productivity $i \sim U[0, 1]$. If $i < i^c$ the entrepreneur closes the firm [because $v(i, \Omega) < 0$]. If

$i \in [i^c, i^h)$ the firm survives but the entrepreneur does not invest in new jobs [because $v(i,\Omega) < c_0$] and so initial firm size $n_0 = n_u$. For $i \geq i^h$, the entrepreneur invests in each expansion opportunity if and only if realised $c^{JC} \leq v(i,\Omega) - c_0$. Thus start-up employment $n_0 = n_u + \widetilde{n}(i,\Omega)$ where $\widetilde{n}(i,\Omega)$ is a binomially distributed random variable with expected value $N_0 H^{JC}(v(i,\Omega) - c_0)$. The expected value of a start-up is therefore

$$\Pi^{SU}(\Omega) = \int_{i^c}^{1} \left\{ n_u v(i,\Omega) + N_0 \int_{0}^{v(i,\Omega)-c_0} [v(i,\Omega) - c_0 - c'] dH^{JC}(c') \right\} di.$$

Hence given the investment opportunity, the entrepreneur proceeds with a new start-up when $c^E \leq \Pi^{SU}(\Omega)$ and so $P^E(\Omega) = H^E(\Pi^{SU}(\Omega))$. Initial firm size [in expectation] is then $En_0 = n_u + N_0 H^{JC}(v(i,\Omega) - c_0)$, noting that $H^{JC}(v(i,\Omega) - c_0) = 0$ for all $i < i^h$. Further, since for $i \geq i^h$ it holds that $H^{JC}(v(i,\Omega) - c_0) = jc(i,\Omega)/\mu_1$, then for these firms $En_0 = n_u + \frac{N_0}{\mu_1} jc(i,\Omega)$.

Job Creation and aggregate hires This extension also requires us to amend the expression for total hires in the main text. In particular, to calculate the total job creation flow in this case, by the definition of G, $[1-U]G'(i)jc(i,\Omega)$ describes the total job creation flow from all existing $i \geq i^h$ firms. Similarly the uniform distribution implies gross job creation flows (excluding the initial unemployed workers) at new $i \geq i^h$ firms is $\mu_0 P^E(\Omega) N_0 H^{JC}(v(i,\Omega) - c_0)$. Adding both flows together yields the total job creation flow (excluding the initial unemployed workers) by firm productivity

$$JC(i,\Omega) = \left\{ [1-U]G'(i) + \frac{\mu_0}{\mu_1} N_0 P^E(\Omega) \right\} jc(i,\Omega). \tag{29}$$

Finally, the modification to job creation flows implies that the expression for total hiring flows in the text is modified, since startups now hire more than one worker. In particular, (4) is replaced with

$$H(i,\Omega) = \begin{cases} 0 & \text{if } i < i^h \\ [1-U]G'(i)q(i,\Omega) + JC(i,\Omega) & \text{if } i \geq i^h, \end{cases} \tag{30}$$

where $JC(i,\Omega)$ now includes the hiring flow from incumbent firms and the hiring flow (excluding the first "free" hires from unemployment) of entrant firms. This modification has an important economic implication, since productive startup firms may now hire by poaching workers from other firms.

The remainder of the description of the model and its expressions remain unchanged.

B.3 Finite state space

Suppose finite productivity states so that firm productivity $p = p^{is}$ with $i \in \{1, 2, ..., I\}$. A very important property of the model is that the aggregate state now reduces to a finite vector, a result which holds even in the extended case with endogenous worker reservation wages $R = R(\Omega)$ and

$\phi < 1$. This property does not arise in the standard Burdett and Mortensen (1998) framework, e.g. Coles (2001), Moscarini and Postel Vinay (2013) and Audoly (2020), and only holds in Coles and Mortensen (2016) for the special case $\phi = 1$.

The reason for the finite state space result is simple but subtle and reflects the replacement hiring process. Define N_i as the measure of workers employed in firms in state i, the employment vector $\underline{N} = (N_1, ..., N_I)$ where adding up implies unemployment $U = 1 - \sum_{i=1}^{I} N_i$. We now show the aggregate state reduces to vector $\Omega = (s, \underline{N})$ with corresponding vector of firm values $\underline{v}(\Omega) = \{v_i(\Omega)\}_{i=1}^{I}$, job creation rates $\{jc_i(\Omega)\}_{i=1}^{I}$ and so on as previously determined in the main text.

The first step is to extend the notation because firms in the same state i post different wages; i.e. there is equilibrium wage dispersion within each state i. The cleanest approach is to assume firms select wage strategies as follows: i) On start-up, a firm is allocated a wage rank $\chi \sim U[0, 1)$. In the stationary equilibrium, firm (i, χ, n, Ω) posts wage with rank χ in the firm i wage distribution. ii) On receiving a firm specific productivity shock with updated productivity i' the firm also updates to a new wage rank $\chi' \sim U[0, 1)$. Because all χ-wage strategies yield equal value, such wage selection is consistent with equilibrium. We choose this wage selection process because it guarantees first order stochastic dominance in wages, and so a worker will always quit to a higher wage offer.

To match to the notation in the main text, consider the following partition of line $[0, 1]$ into a grid $\{x_0, x_1, .., x_I\}$ where $x_0 = 0$, $x_i = x_{i-1} + \gamma_{0i}$ and $x_I = 1$. A firm in state $i \in \{1, 2, ..., I\}$ with wage rank $\chi \in [0, 1)$ is correspondingly defined as being in state $x \in [0, 1]$ where $x = x_{i-1} + \chi[x_i - x_{i-1}]$. Each start-up is then equivalently defined as having initial state $x \sim U[0, 1]$, where $p^s(x) = p^{is}$ for $x \in [x_{i-1}, x_i) \subset [0, 1]$. The only material difference to the continue productivity case is that firm productivity $p^s(x)$ is increasing in $x \in [0, 1]$ but not strictly increasing. The underlying wage structure equation (14), however, continues to apply and equation (19) describes the equilibrium values $v_i(\Omega)$.

So why is the state space finite? The critical property is that despite there being wage and quit rate dispersion across firms $x \in [x_{i-1}, x_i)$ within a productivity level, all such firms have identical expected employment dynamics. Why? Because all firms with $i \geq i^h$ immediately replace any worker who quits and so their expected employment dynamics are independent of χ. Additionally, all firms with $i < i^h$ post the same wage, $w_{\min}$ and so their expected employment dynamics are also independent of χ. Now recall that $G(x)$ describes the distribution of employment across firms $x \in [0, 1]$. By definition of the partition above, firm $x = x_i$ has productivity $i + 1$ and rank 0 and so $G(x_i) = \sum_{j=1}^{i} N_i/(1 - U)$. Because firm size is orthogonal to rank we then have

$$G(x) = \frac{\sum_{j=1}^{i-1} N_j + \frac{x - x_{i-1}}{x_i - x_{i-1}} N_i}{1 - U} \text{ for all } x \in [x_{i-1}, x_i),$$

is continuous but $\underline{N}$ is a sufficient statistic for $G(.)$. Thus with finite productivity states the previous analysis all goes through with the added simplification that the aggregate space is a finite vector $\Omega = (s, \underline{N})$.

B.4 Finite productivity model summary

We now briefly summarise the equations of the finite productivity model as we use it in our quantitative work which include minor additions to the model made relative to the continuous productivity model. Our calibrated model features $i^c(\Omega) = 1$ at all times, and we present the equations for this case of the model. Recall from the previous section that we specialise to a finite number of productivities $i = 1, ..., I$, where within each productivity level firms additionally separate into different wage ranks $\chi \in [0, 1]$. We then define the overall wage rank across all firms as $x \in [0, 1]$.

Firm HJB and policy functions: All firms with the same productivity p_i achieve the same value v_i, regardless of their wage rank. With aggregate shocks the HJB includes the aggregate state $\Omega = (s, N_1, ..., N_I)$. If $i^c(\Omega) = 1$ at all times, the N_i evolve continuously over time. In this case, the HJB can be written

$$(r + \delta_F)v_i(\Omega) = a_s p_i - c_f - w_{\min} - (\lambda_1(\Omega) + \lambda_u)\min[v_i(\Omega), c_0] + \mu_1 E_c \max[v_i(\Omega) - [c_0 + c], 0]$$
$$- \delta_D E_c \min[v_i(\Omega), c] + \alpha_\gamma \sum_j \gamma_{ij}(v_j(\Omega) - v_i(\Omega))$$
$$+ \alpha_a \sum_{s'} \gamma_{s,s'}(v_i(s', \underline{N}) - v_i(\Omega)) + \sum_{j=1}^{I} \frac{\partial v_i(\Omega)}{\partial N_j}\dot{N}_j(\Omega). \quad (31)$$

Notice that we extend the model relative to the main text by introducing a flow cost of capital maintenance, c_f. This is a cost which must be paid each period to maintain each existing unit of capital. The introduction of c_f does not change the economics of the model, but it is useful as it allows us to more easily partition the productivity bins into those which do and do not replace quits (see below for more details), and is also used as the subsidy in our counterfactual calibration with a higher value of c_0 . The only aggregate "price" which affects firm value is the scalar $\lambda_1(\Omega)$. The expectations over JC and JD cost draws have closed form solutions under the assumed distributions. The hiring threshold $i^h(\Omega)$ is defined as the first i for which $v_i(\Omega) > c_0$.

In the finite productivity model, most policy functions – and in particular those which relate to net employment dynamics – depend only on i and not the wage rank. Specifically, the job creation rate per employee is $jc_i(\Omega) = \mu_1 H^{JC}(v_i(\Omega) - c_0)$. The job destruction rate is $jd_i(\Omega) = \delta_D[1 - H^{JD}(v_i(\Omega))]$ for firms with $i > i^h(\Omega)$ and $jd_i(\Omega) = \delta_D[1 - H^{JD}(v_i(\Omega))] + \lambda_1(\Omega) + \lambda_u$ otherwise. Entrants who draw productivity i have average initial employment $\bar{n}_{0,i}(\Omega) = n_u + \frac{N_0}{\mu_1} jc_i(\Omega)$. Here, $n_u = 2$ is the number of initial workers a firm can draw from unemployment for free upon startup. We found it helpful to set $n_u = 2$ rather than $n_u = 1$ to ensure that even unproductive ($i = 1$) entrants start with more than one employee. This slows down the firm exit process for unproductive entrants, who otherwise exit as soon as their one and only initial employee is poached or fired. We define $\hat{n}_{0,i}(\Omega) = \frac{N_0}{\mu_1} jc_i(\Omega)$ as average entrant size excluding the initial n_u free hires from unemployment.

Evolution of employment distribution: The total mass of employment at each productivity bin evolves according to:

$$\dot{N}_i(\Omega) = \mu_0 P^E(\Omega)\gamma_{0i}\bar{n}_{0,i}(\Omega) + N_i \left[jc_i(\Omega) - jd_i(\Omega) - \delta_F - \alpha_\gamma \sum_{j \neq i} \gamma_{ij} \right] + \alpha_\gamma \sum_{j \neq i} \gamma_{ji} N_j. \tag{32}$$

The first term on the right hand side is the inflow of employment from firm entry. The term in square brackets gives net job creation accounting for job creation and destruction including the firm exit shock. The terms proceeded by α_γ gives the transition of firms across productivity bins. Total unemployment is $U = 1 - \sum_i N_i$. The distribution across firm wage ranks can then be calculated from our closed form solution:

$$G(x, \Omega) = \frac{\sum_{j=1}^{i-1} N_j + \frac{x - x_{i-1}}{x_i - x_{i-1}} N_i}{1 - U} \text{ for all } x \in [x_{i-1}, x_i). \tag{33}$$

Recall that we define the boundaries $x_0 = 0$, $x_i = x_{i-1} + \gamma_{0i}$ and $x_I = 1$. A firm in state $i \in \{1, 2, ..., I\}$ with wage rank $\chi \in [0, 1)$ is correspondingly defined as being in state $x \in [0, 1]$ where $x = x_{i-1} + \chi[x_i - x_{i-1}]$.

Quits and hires across the x distribution: To close the model, we need to calculate the offer arrival rate $\lambda_1(\Omega)$. To do this, we must solve the quit rates across the wage rank distribution. As in the continuous productivity model, the quit rate for incumbent firms at any wage rank x is given by (8), which we rewrite in our x notation as

$$q(x, \Omega) = \begin{cases} \lambda_u + \lambda_1(\Omega) & \textit{if } x \in [0, x^h(\Omega)) \\ \lambda_u + \frac{\phi \int_x^1 \{JC(y,.) + \lambda_u [1-U] G'(y)\} dy}{U + \phi[1-U]G(y)} & \textit{if } x \geq x^h(\Omega) \end{cases} \tag{34}$$

where $x^h(\Omega) = x_{i^h(\Omega)-1}$ corresponds to the lowest ranked hiring firm, who has $i = i^h(\Omega)$ and $\chi = 0$. The total job creation flow at each x is $JC(x, \Omega) = \left\{ [1 - U]G'(x) + \frac{\mu_0}{\mu_1} N_0 P^E(\Omega) \right\} jc_i(\Omega)$. The closed form solution for $G(x)$ similarly defines a closed form solution for $G'(x)$, which is well defined except at the x_i boundaries where $G(x)$ is non-differentiable. Performing the integration in (34) yields a closed form solution for $q(x, \Omega)$ for any $x \geq x^h(\Omega)$:

$$q(x, \Omega) = \lambda_u + \frac{(1 - \chi) \left[(jc_i + \lambda_u) N_i + \mu_0 P^E(\Omega) \gamma_{0,i} \hat{n}_{0,i} \right] + \sum_{j=i+1}^{I} \left[(jc_j + \lambda_u) N_j + \mu_0 P^E(\Omega) \gamma_{0,j} \hat{n}_{0,j} \right]}{U/\phi + \sum_{j=1}^{i-1} N_j + \chi N_i} \tag{35}$$

This equation uses the reverse mapping $\chi(x)$ to find the χ associated with the current x from $\chi = \frac{x - x_{i-1}}{\gamma_{0i}}$, and similarly the productivity level $i(x)$ associated with the current x. The hiring rate for incumbent firms can then be simply computed as $h(x, \Omega) = jc_i(\Omega) + \mathbf{1}(i \geq i^h(\Omega))q(x, \Omega)$. Finally, $\lambda_1(\Omega)$ is just $q(x, \Omega)$ evaluated at $x = x^h(\Omega)$ and then subtracting λ_u, which gives

$$\lambda_1(\Omega) = \frac{\sum_{j=i^h(\Omega)}^{I} \left[(jc_j(\Omega) + \lambda_u)N_j + \mu_0 P^E(\Omega)\gamma_{0,j}\hat{n}_{0,j}(\Omega) \right]}{U/\phi + \sum_{j=1}^{i^h(\Omega)-1} N_j} \tag{36}$$

This closes the model, and provides sufficient information to simulate the model keeping track only of the finite employment vector $\underline{N}$. In particular, the model can be solved and simulated using only the HJB (31), the evolution of total employment by productivity bin (32), and the closed-form solution for the job offer arrival rate (36).

C Numerical Methods Appendix

Steady state: For given parameter values, solving the core equations of the model in steady state reduces to solving for a vector of I values v_i and employment stocks N_i, as well as the arrival rate λ_1. This is a simple problem to solve using the steady state versions of (31), (32), and (36). Intuitively, one can guess a value of λ_1, solve the HJB for the values v_i, use the implied policy functions to calculate the employment stocks N_i, and use these to update your guess for λ_1. In practice, we solve the model in steady state at the same time as calibrating our parameters, which involves calculating other statistics, which we detail further below.

Calculating average quit and hiring rates involves integrating over the wage rank distribution, x. To do this, we build an uniform grid over $\chi \in [0, 1]$ with 1,000 nodes. This is then combined with the x_i to build a grid over x with $I \times 1,000$ nodes. We calculate all integrals on this grid using trapezoid integration.

To calculate the firm age and size distribution we solve for the densities of firms on grids for age and size. This is thus reminiscent of the non-stochastic simulation approach of Young (2010), or the methods of Achdou et al. (2021). To calculate the size distribution, we build a grid over firm sizes. Recalling that the number of employees in a given firm is an integer, we define a size grid as integer values from 0 to 20,000 employees. We solve for the mass of firms at each size s and productivity i. We use the firm dynamics processes (job creation, destruction, entry, exit, and so on) to construct flow rates across these joint size-productivity bins, which we use to build a matrix of transitions. We can then solve for the steady-state density of the number of firms at each size-productivity bin by inverting this matrix.

To calculate the age distribution we follow a similar process. However, since age is a continuous number, we discretise the age grid on a uniform grid from age 0 to age 26 years old (since this is the maximum age bin recorded in the BDS data) with 1,000 nodes. We solve for the mass of firms and employment at each age-productivity bin. Finally, to compute the mass of active firms at each age, we actually need to compute the joint age-size distribution, since we define firms with 0 employees as having exited. To do this, we must solve for the joint age-size-productivity distribution, which we do using the same methods. Given the high dimension of this object, we solve for this distribution using a reduced firm size grid from 0 to 1,000 employees, and confirm that raising this maximum has no impact on the moments for which this distribution is used.

To compute the growth rate density and hockeystick plots (Figure 6 in the main text) we simulate a panel of firms for one quarter, with their initial states drawn from the steady-state size-productivity distribution. We simulate a panel of one million firms and calculate net employment growth and gross flows from the beginning to the end of the quarter.

Business cycle: For given parameter values, solving the core equations of the model over the business cycle reduces to solving for the functions $v_i(\Omega)$, $\dot{N}_i(\Omega)$, and $\lambda_1(\Omega)$ over the state space $\Omega = (s, N_1, ..., N_I)$, using the equations (31), (32), and (36). In terms of approximation, it actually suffices to approximate only the function $v_i(\Omega)$, as the values of $\dot{N}_i(\Omega)$, and $\lambda_1(\Omega)$ can then be calculated exactly using (32) and (36) at any grid point or point in a simulation.

We approximate $v_i(\Omega)$ using second order polynomials in $N_1, ..., N_I$, with different coefficients for each of the discrete productivity level pairs i, s. In particular, first adjust the value function notation to $v_i(\Omega) = v_{i,s}(N_1, ..., N_I)$ to acknowledge that aggregate productivity s is also a discrete state. Then for each i, s we approximate $v_{i,s}(N_1, ..., N_I)$ as

$$v_{i,s}(N_1, ..., N_I) \simeq h^0_{i,s} + \sum_{j=1}^{I} \left(h^1_{i,s,j} N_j + h^2_{i,s,j} N_j^2 \right) \tag{37}$$

where $h^0_{i,s}$ $h^1_{i,s,j}$ and $h^2_{i,s,j}$ are scalar coefficients to be estimated. $h^0_{i,s}$ is the intercept, and $h^1_{i,s,j}$ and $h^2_{i,s,j}$ capture the first and second order effect on the value of firms with state i of changing total employment of firms with productivity j . Notice that we exclude cross terms in the second-order approximation, since these are known to typically be unstable given that the $N_1, ..., N_I$ tend to be highly correlated. For each i, s this approximation uses $1 + 2 \times I = 1 + 2 \times 5 = 11$ coefficients. Since we use $I = 5$ idiosyncratic and $S = 3$ aggregate productivity nodes, this gives $5 \times 3 \times 11 = 165$ coefficients to estimate.

To solve the HJB, we need to know both the level of value and its derivative with respect to the aggregate employment bins. The derivatives are easy to compute given our approximation, as

$$\frac{\partial v_{i,s}(N_1, ..., N_I)}{\partial N_j} = h^1_{i,s,j} + 2h^2_{i,s,j}N_j \tag{38}$$

In order to solve the full business-cycle version of the model, we use the following procedure:

1. We need a grid of values for $(N_1, ..., N_I)$ to approximate our second-order polynomial on. To generate this, we use a Sobol set, which generates values of the $(N_1, ..., N_I)$ vector which are roughly equally spaced between a minimum and maximum value for each N_i. We generate 50 of such vectors, denoting the values of $(N_1, ..., N_I)$ at each candidate z as $\underline{N}_z = (N_{1,z}, ..., N_{I,z})$ for $z = 1, ..., 50$. Note that the aggregate state at any grid point is now denoted s, z, where s corresponds to aggregate productivity and z to the vector of $\bar{N}$ values.

2. Generate initial guesses for the parameters $h^0_{i,s}$ $h^1_{i,s,j}$ and $h^2_{i,s,j}$. Generate an initial guess for $\lambda_1(\Omega) = \lambda_{1,s,z}$ at each aggregate state. Generate an initial guess for $\dot{N}_i(\Omega) = \dot{N}_{i,s,z}$ at each aggregate state.

3. Given these guesses, solve the value function (31) for values $v_{i,s}(\Omega) = v_{i,s,z}$ at each idiosyncratic productivity node and aggregate state node. In solving (31), replace $\lambda_1(\Omega)$ with the current guess $\lambda_{1,s,z}$, and the drift term, $\sum_{j=1}^{I} \frac{\partial v_i(\Omega)}{\partial N_j} \dot{N}_j(\Omega)$, using i) the current guesses for the value function derivative implied by $h^1_{i,s,j}$ and $h^2_{i,s,j}$ and ii) the current guess for the drifts $\dot{N}_{i,s,z}$.

4. Using the new values of $v_{i,s,z}$, perform OLS regressions on (37) to update the parameters $h^0_{i,s}$ $h^1_{i,s,j}$ and $h^2_{i,s,j}$ with dampening.

5. Using the new values of $v_{i,s,z}$, calculate the new policy functions for job creation and destruction. Use these to update the drifts $\dot{N}_{i,s,z}$ using (32) and the offer arrival rates $\lambda_{1,s,z}$ using (36), both with dampening.

6. Return to step 3 and iterate to convergence.

As a measure of the accuracy of our second-order approximation, the R^2 of the regressions used to fit the polynomial is 99% on average across the $I \times S = 15$ regressions. This R^2 is a measure of the error between the predicted value from the second-order polynomial and the exactly computed value from the HJB of the $v_{i,s,z}$ on the nodes where the HJB is evaluated.

With our approximated policy function parameters $h^0_{i,s}$, $h^1_{i,s,j}$, and $h^2_{i,s,j}$ in hand, we can simulate the aggregate model, calculating all other objects exactly using the true nonlinear equations of the model. When estimating the model, we simulate using a one month aggregate time step $\Delta t = 1$, but finer grids do not affect the results. For most data comparisons, we aggregate up to quarterly data via averaging and HP-filter the model data as in the data.

Post estimation, we also simulate our impulse response to a typical recession using the same procedure. We additionally simulate the age and size distributions over time, by first solving for the aggregate dynamics, and then extending our age and size distribution codes to allow for aggregate dynamics.

Overview of estimation procedure: We pre-set six parameters, and our estimation procedure then chooses 23 parameters to minimize the distance to a large number of moments. To speed up the estimation, we split the estimation into two layers: an inner loop and an outer loop. Conditional on outer loop parameter values, in the inner loop we solve for the values of 11 parameters to exactly hit 11 moments. Intuitively, each of these 11 parameters has a tight link to a particular moment, which we are able to exploit to quickly solve for the value of these parameters. In the outer loop, we use the remaining 12 parameters to minimize the average distance to a set of 18 moments using a global minimization routine, every time repeating the inner loop procedure.

A key step in speeding up our estimation is that we calibrate many parameters in the non-stochastic steady state of the model (i.e. a version of the model without aggregate shocks) as is standard in heterogeneous agent modelling. In brief, the procedure operates as follows:

1. Guess values for the 12 outer loop parameters.
2. Given the current guess for the outer loop parameters, use the inner loop to exactly solve for the 11 inner loop parameters, to exactly hit the inner loop moments. These moments are calculated in the non-stochastic steady state of the model.
3. Given the values of the inner and outer loop parameters, now solve the full model (out of steady state), and simulate to construct aggregate time series.
4. Calculate the moments used in the outer loop. The moments related to the firm age distribution are calculated from the non-stochastic steady state, and the moments related to the business cycle are calculated from the business cycle simulation.
5. Calculate the distance measure of the outer loop moments to the moments in the data. Update the outer loop parameters using the global minimization routine, and return to step 2. Repeat until the global minimization routine completes.

For our global minimization routine in the outer loop, we program a simplification of the "TikTak" algorithm of Arnoud et al. (2019). Specifically, we draw 12,000 initial guesses of the outer loop parameters from a Sobol set, and calculate the outer loop moments at each guess. We then choose the five best performing guesses and run a local optimizer (pattern search) at each to find the local minima, and choose the lowest error among these as our final estimate.

C.1 Further details of estimation and parameterization

As we impose a relatively small number of productivity states, we use parameter choices to impose the key behaviour (perform JC or not, perform JD or not, replacing quits or not) of each state, rather than letting the estimation decide. The estimation is allowed to affect behaviour within each node (for example the level of JC in the node, if it is positive) and the probabilities that nodes are drawn.

Additional flow cost of capital maintenance: In the estimation we impose that $i^h = 3$ in steady state, which requires that $v_2 < c_0 < v_3$. This could be imposed using a penalty function approach, which penalises the SMD error whenever either of these inequalities is violated. We take a simpler approach, which speeds up the estimation (by allowing more parameters to be chosen in the inner loop) at the cost of introducing one new parameter, the flow cost of capital maintenance. Specifically, we firstly impose $p_2 < p_3$ in the estimation, which ensures that $v_2 < v_3$. Secondly, we then choose c_f in the inner loop to ensure that $c_0 = 0.5(v_2 + v_3)$, which guarantees that $v_2 < c_0 < v_3$. Intuitively, we thus introduce the flow cost of capital to shift the average value to ensure that c_0 lies exactly in the middle of v_2 and v_3 in steady state. We force this parameter to be small by imposing penalties if it falls below 0 or above 0.1. The estimated value of this cost is small, at 0.078, which is a small fraction of average productivity (which is one) and small compared to the hiring cost c_0. This flow cost c_f is also used as the flow subsidy (allowing $c_f < 0$) in our counterfactual calibration with a higher value of c_0.

We also impose that i^h should not move over the business cycle, so that $i^h(\Omega) = 3$ almost always during a simulation. We do this by computing the fraction of periods where $i^h(\Omega) \neq 3$ during our business cycle simulation and adding this as a penalty in the SMD function if it exceeds 1% of the simulation. At the estimated parameters, $i^h(\Omega) \neq 3$ only 0.77% of the time during our long business cycle simulation, and $i^h(\Omega) = 3$ at all times in our typical recession impulse response function plots.

Placement of cost function support parameters: We place the lower support of H^{JC} and upper support of H^{JD} so that (in steady state) i) firms with $i = 1, 2, 3$ perform JD in response to the H^{JD} shock, but never JC in response to the H^{JC} shock, and ii) firms with $i = 4, 5$ perform JC but not JD. To do this requires that i) $v_3 < \bar{c}^{JD} < v_4$, so that firms with $i = 4$ have high enough value to survive any draw of c^{JD}, and ii) $v_3 < \underline{c}^{JC} + c_0 < v_4$, so that firms with $i = 3$ have low enough value so that the minimum cost of performing JC is too high. In order to not introduce additional degrees of freedom into the model, we simply set $\bar{c}^{JD} = 0.5(v_3 + v_4)$ and $\underline{c}^{JC} = 0.5(v_3 + v_4) - c_0$, which we can impose simply at every iteration of the inner loop.

Definitions of aggregate flows: The formulas below give the aggregate flows (per unit of time) used in Table 2. These can be expressed as rates by dividing by an appropriate denominator, and are constructed to be comparable with their data counterparts.

Note that we treat the λ_u shock as layoffs into unemployment, not quits into unemployment. This means that in the model, the EU flow and layoffs are identical, as all EU moves are involuntary. Similarly, this means that all quits are job-to-job quits, so the EE flow and quit flow are identical.

Total JC flow: $JC_t = \sum_{i=1}^{I} \left[\mu_0 P^E(\Omega)\gamma_{0i}\bar{n}_{0,i}(\Omega) + N_{i,t} jc_i(\Omega)\right]$
Total JD flow: $JD_t = \sum_{i=1}^{I} \left[jd_i(\Omega) + \delta_F\right] N_{i,t}$
Total EU flow: $EU_t = \sum_{i=1}^{I} \left[\delta_D[1 - H^{JD}(v_i(\Omega))] + \delta_F + \lambda_u\right] N_{i,t}$.
Total UE flow: $UE_t = \lambda_0(\Omega)U_t + n_u\mu_0 P^E(\Omega)$
Total EE flow: $EE_t = \int_0^1 (q(x,\Omega) - \lambda_u)dG(x,\Omega) \times N_t$
Total hiring flow: $H_t = UE_t + EE_t$
Firm entry flow: $EN_t = \mu_0 P^E(\Omega)$
Firm exit flow: $EX_t = \delta_F M_t + \sum_{i=1}^{I} jd_i(\Omega) m_{i,t}^1$, where M_t is the total mass of firms with at least one employee, and $m_{i,t}^1$ is the mass of firms with productivity i and exactly one employee.

JC and JD definition in the model vs. the data: The definitions of JC and JD in the model are built to correspond as closely as possible to their notions in the data. However, practical computational limitations mean that their definitions are not identical. In particular, in the data JC is defined as the sum of employment increases across firms which saw an increase in employment between two dates, and JD is defined as the sum of employment falls at firms which saw a decrease in employment (see, e.g., DFH). Computing these measures exactly therefore would require simulating a panel of firms over the business cycle, which slows down the estimation and simulation of the model.

Instead, we are careful to segment our model so that firms in states $i = 1, 2, 3$ have declining employment at any instant of time, and firms in states $i = 4, 5$ have increasing employment at any instant in time. Abstracting for the moment from entry and exit, JC can therefore be measured as the employment change at state $i = 4, 5$ firms, which corresponds exactly to the job creation flow $\sum_{i=1}^{5} jc_i(\Omega)N_{i,t}$ in the model, since $jc_i(\Omega) = 0$ for $i = 1, 2, 3$. Similarly, JD can be measured as the employment change at state $i = 1, 2, 3$ firms, which corresponds to $\sum_{i=1}^{5} jd_i(\Omega)N_{i,t}$ since $jd_i(\Omega) = 0$ for $i = 4, 5$. This is complicated slightly by two factors. Firstly, firms may switch from being in state $i = 1, 2, 3$ to $i = 4, 5$ within the same quarter, which is the period over which JC and JD are measured in the DFH data. This means that measured and theoretical measures may differ, as a firm might receive a JC and JD shock in the same quarter. However, this is a rare occurrence, as our productivity shocks occur on average only once per quarter. Secondly, we must account for entry and exit. Entry is simple, as all entrants create jobs and therefore can simply be added to the job creation flow. Exit complicates the analysis somewhat, as even firms with $i = 4, 5$ might receive the δ_F exit shock. However, this shock is calibrated to be very rare so this does not matter much in practice.

In order to check the applicability of our JC and JD measures, we simulate a large panel of firms after estimating the model. We focus on the steady state, and simulate the firms for one quarter of data, with the firms drawn from the ergodic productivity and size distribution. We compute the job destruction rate exactly as is done on the data, and find a quarterly rate of 6.2%, while the theoretical rate as calculated by $jd = \sum_{i=1}^{I} [jd_i + \delta_F] N_i/N$ exactly equals the targeted value of 7.0%. While not identical, this difference of roughly 10% is in line with the average error of the other moments in the outer loop of our SMD routine.

Comparison of firm-level autocorrelation to data: Our parameter values generate an autocorrelation of 0.21 for yearly productivity in a year-averaged simulated firm-level productivity series, or 0.53 for quarterly productivity (within mature firms). Elsby et al. (2017) discuss empirical estimates of the persistence of idiosyncratic productivity, and find a wide range of values. Our estimate lies within this range. Specifically, Cooper, Haltiwanger, and Willis (2015) imply a quarterly autocorrelation of 0.4, which is below our value, while Abraham and White (2008) imply 0.68 which is above our value. While our productivity process has relatively low persistence, our constant returns to scale structure means that current productivity controls the growth rate of employment, not the level. Hence, even temporary productivity shocks will generate permanent effects on a firm's employment.

Further details of the Inner loop: We present below a list of the 11 parameters (plus the additional parameter c_f) chosen in the inner loop, and how the moment used to choose the parameter is calculated. The inner loop is terminated when the error in all moments is below 10^{-8}. All moments are computed in the non-stochastic steady state of the model.

The parameter p_4 is chosen to normalise aggregate labor productivity to one. The parameter p_3 is chosen to generate a standard deviation of idiosyncratic productivity of 30%. The parameter γ_2 is chosen to match that 80% of quits (and replaceable layoffs) are replaced, calculated as $qfr = \int_{x^h}^{1} q(x)dG(x)/\int_0^1 q(x)dG(x)$.

Parameters $\underline{c}_{JC}$, $\bar{c}_{JD}$, and c_f are set as discussed above. The firm entry flow μ_0 is set to hit the average size of firms, measured as N/M, where M is the mass of firms with at least one employee. Parameters μ_1, δ_D, λ_u, and ϕ are set to match the theoretically computed JC, JD, layoff, and EE quit rates. Finally, $w_{\min}$ is set to match the labor share, defined as $LS = Ew \times N/Y$, where $Ew = \int_0^1 w(x)dG(x)$.

Further details of the Outer loop: The 12 parameters chosen in the outer loop fall into two broad categories: those relating to the firm age distribution $(p_1, p_2, p_5, \gamma_{11}, \gamma_{55}, \gamma_{01}, \gamma_{05}, N_0)$ and those relating to business cycle moments $(\xi_e, \xi_{JC}, \xi_{JD}, c_0)$. The 12 age distribution moments are computed using the non-stochastic simulation of the age distribution in steady state. The six business cycle moments are computed by simulating the model for 1,000 years. The simulated data are then aggregated to a quarterly frequency and HP-filtered, as in the data. We apply a simple diagonal weighting to the moments, with the total weight given to business cycle moments slightly overweighted to ensure the model performs well on both business cycle and steady state dimensions.

We perform several parameter swaps in the outer loop estimation, searching over some hyper-parameters instead as these ensure that the estimation searches over sensible regions of the parameter space. Specifically, instead of searching over values of p_1, p_2, and p_5, we search over steady state values of JD_1, JD_2, and JC_5. Instead of searching over values of N_0, we search over values of the average number of employees of firms at the moment of entry.

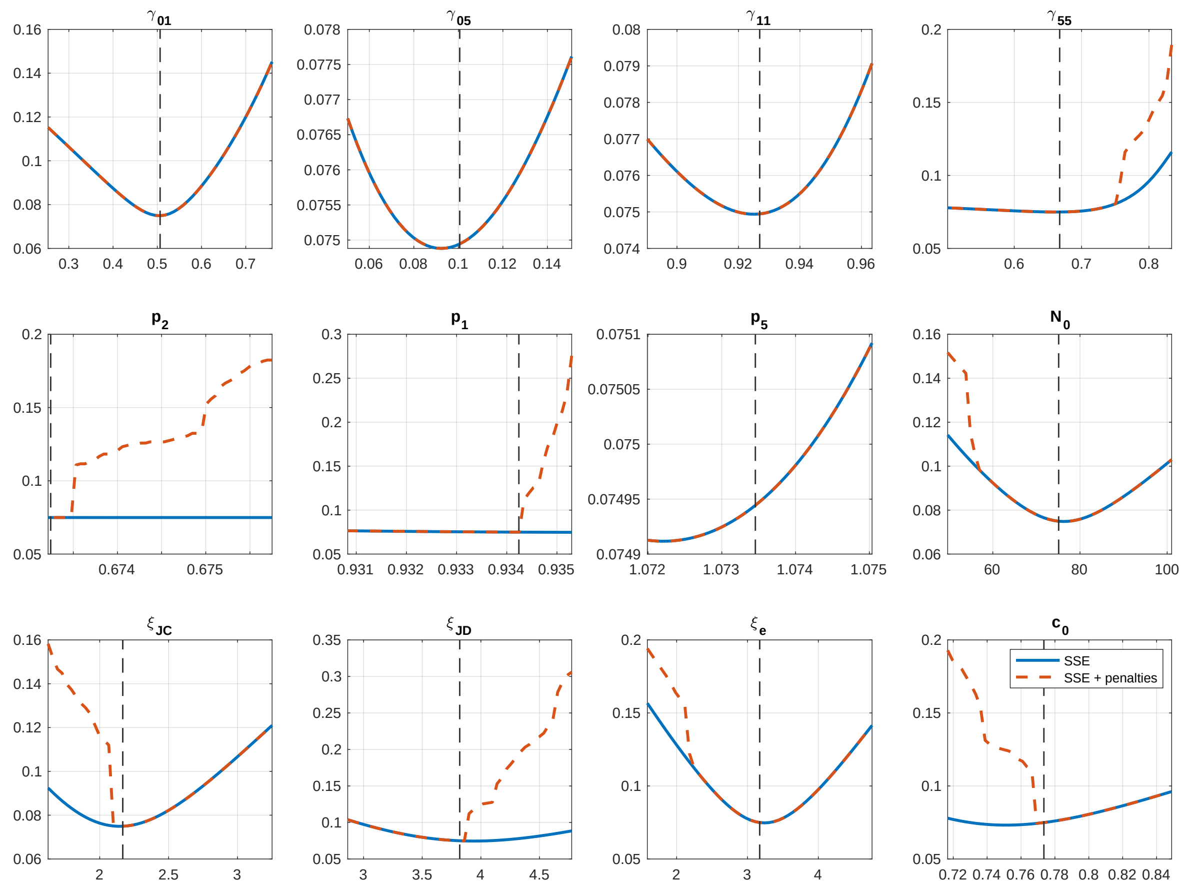
The estimation finds that the productivity grid is non-monotone, as $p_1 > p_2$ and $p_5 < p_4$. Nonetheless, firm *values* remain monotone, with $v_i < v_{i+1}$ for all i, which is sufficient for the job ladder to be directed monotonically by i, and hence for our notion of equilibrium to remain well defined. The disconnect between the ordering of productivities and values occurs simply because the entrant states $i = 1, 5$ are more persistent than the mature states.

Simulated Minimum Distance (SMD) results The weighted average error of model moments to data moments in the outer loop SMD estimation is 7.5% across 18 moments. The average error for the six business cycle moments is 8.0%, where the quarterly data and model moments can be seen in Table 2, and the yearly standard deviation of firm entry is 0.0688 in the model and 0.0701 in the data. The average error for the 12 firm age moments is 6.4%, where the data and model moments can be seen in Figure 3, and the firm exit rate is 9.6% in the model and 8.3% in the data. The errors for the 11 moments in the inner loop are zero by construction.

Identification All parameters in the inner loop are adjusted to match one assigned moment to the data, and hence we know these parameters are identified by the moments because individually adjusting their values successfully sets the errors in those moments to zero.

To check the identification of parameters in the outer loops we perform three experiments. First, we vary the values of our outer loop parameters one by one in a grid of values around their estimated values. The other parameters are held at their estimated values. We compute the sum of squared errors from the SMD at the new values, and plot them in Figure C.1. Each panel gives the effect of varying one parameter on the (square root of) the sum of squared errors from the outer loop. The blue line gives the sum of squared errors, and the red led the sum of squared errors plus the penalties imposed in the estimation for parameter values which lead to the model violating certain required conditions. As can be seen in the figure, the parameters are well identified with most sitting near the bottom of "U" shapes in the sum of squared error. The penalty function penalises the estimation if either: (i) the condition $v_2(\Omega) < c_0 < v_3(\Omega)$ is violated more than 1% of the time during a simulation. This would imply that the degree of replacement hiring experienced extreme counterfactual changes as entire groups of firms started or stopped replacement hiring, or (ii) matching $v_2 < c_0 < v_3$ in steady state required a value of c_f outside of the allowed small range. Some parameters are also partly identified because changing them would violate these penalties. This stresses that matching the degree of replacement hiring in the data is not something that our model can do "for free" and that it imposes constraints on the estimation. We highlight that the fact that the model can match the moments that it does while matching the degree of replacement hiring over the business cycle is not automatic, and is a success of the model.

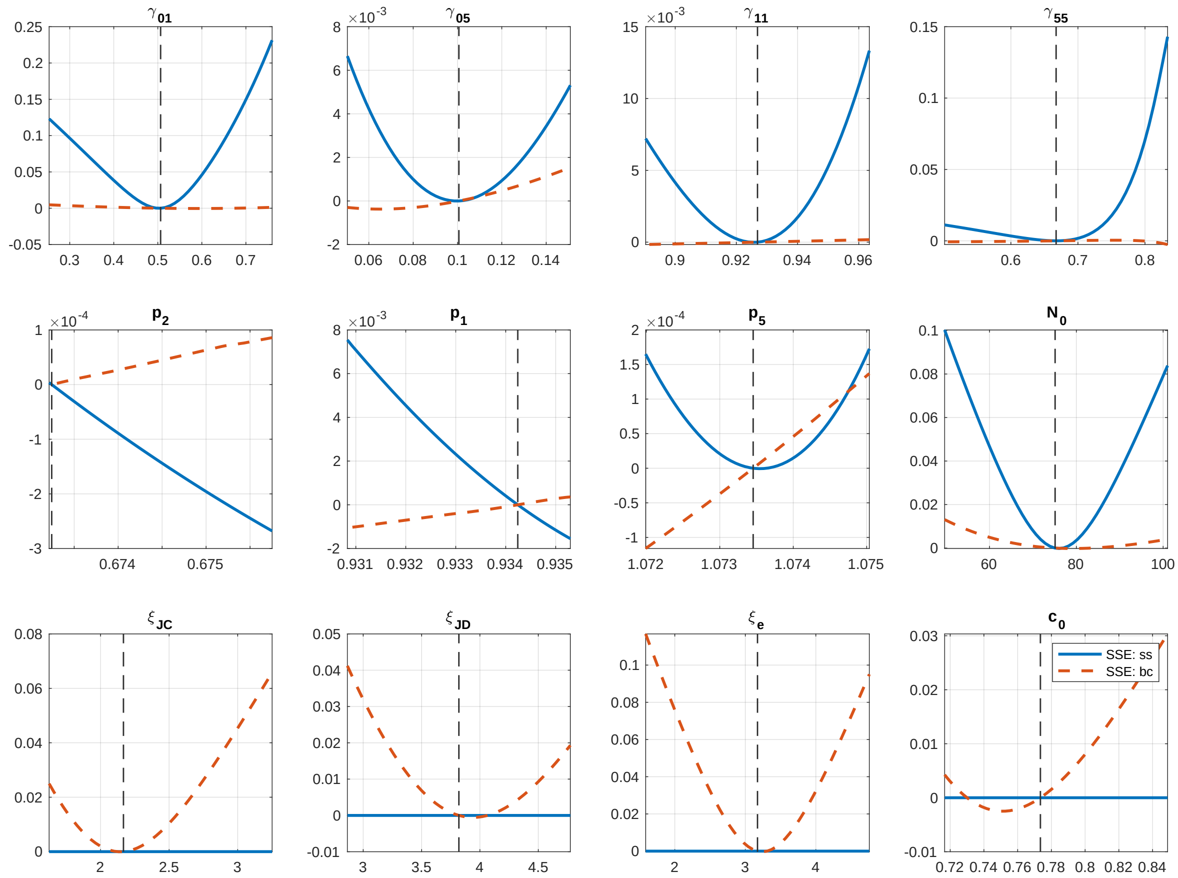
Figure C.1: Identification test: effect on sum of squared errors of varying each parameter



Effect on sum of squared errors of varying one parameter at a time around their estimated value. Blue line gives the sum of squared errors and red line also include the penalty functions from the estimation. The vertical dashed line indicates the estimated parameter value.

Secondly, we verify that the parameters that we informally associate with affecting the steady state moments (firm age distribution) versus the business cycle moments actually affect these moments more than the others. In Figure C.2 we split the sum of squared errors into the part coming from firm age moment errors and the part from business cycle errors. We plot these, subtracting the true estimated errors so that the values of each line is zero at the estimated parameter values. In blue we plot the steady state moment errors, and in red the business cycle errors. The parameters associated more with the steady state are in the top two rows, and the business cycle in the bottom row. We can see that the parameters in the top two rows affect the steady state moments (blue line) much more than the business cycle moments (red line), while the opposite is true for the parameters in the bottom row. Hence, our intuitions about the role of these parameters appears to be justified.[31]

Figure C.2: Identification test: steady state versus business cycle errors



Effect on sum of squared errors of varying one parameter at a time around their estimated value. Blue line gives the sum of squared errors for the firm age distribution moments, and red line for the business cycle moments. Both lines subtract the value of the moments at the estimated parameters, so are normalised to zero at the estimated parameter value. The vertical dashed line indicates the estimated parameter value.

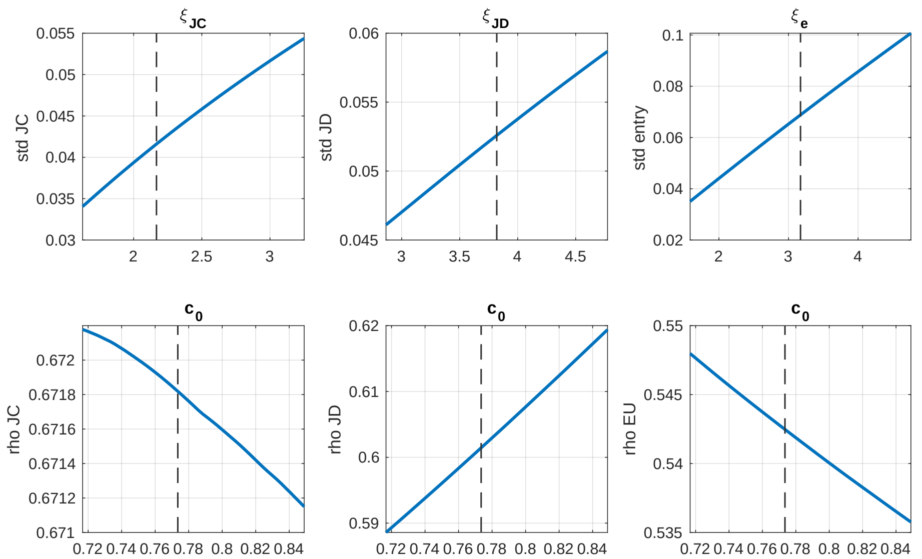
Finally, in our estimation we informally associated certain parameters with certain business cycle moments. We argued that ξ_{JC}, ξ_{JD}, and ξ_e controlled the standard deviations of job creation, job destruction, and entry respectively. In the top row of Figure C.3 we show that varying these parameters monotonically affects the associated moment in the expected way, with higher values increasing the associated elasticity and hence standard deviation. We argued that raising c_0 lowered the autocorrelation, by increasing the feedback from unemployment into firm value and hence job creation and layoff

[31] We actually perform a few parameter swaps in the estimation which helps keep the parameter choices in sensible ranges. For example, rather than directly searching over p_1, we search over values of jd_1, which are used to back out the required value of p_1. This explains why adjusting the parameter values in the bottom row has exactly zero effect on the steady state moments, since these parameter swaps hold the realised firm age distribution constant by construction.

policies. In the bottom row we show how varying c_0 affects the autocorrelation of JC, JD, and layoffs respectively. This shows that, as expected, raising c_0 lowers the autocorrelation of JC and layoffs. In isolation, raising c_0 actually raises the autocorrelation of JD, contrary to expectations. This is because total JD includes both layoffs and JD from unreplaced quits, and raising c_0 also affects the dynamics of unreplaced quits by making the overall volatility of unemployment lower. However, when recalibrating ξ_{JC}, ξ_{JD}, and ξ_e to maintain the same volatilities as in the baseline estimation, raising c_0 also lowers the autocorrelation of total JD (see the counterfactual experiment in the "The importance of c_0" section of the main text).

For the parameters we informally associated with the moments of the firm age distribution, we did not informally associate each individual parameter with a specific moment of the firm age distribution. Instead, we envisaged choosing the parameters as a whole to match the whole distribution. Hence we do not repeat the exercises of Figure C.3 for the firm age distribution, but note that these parameters do indeed affect the firm age distribution and are well identified, as we showed in our discussion of Figure C.2.

Figure C.3: Identification test: business cycle moments



Effect on the associated moment(s) of changing each parameter. This exercise is performed for each parameter we informally associated with business cycle moments. The vertical dashed line indicates the estimated parameter value.

Counterfactual calibration with higher c_0 In Section 7.3 we discussed a counterfactual calibration where we raised c_0 by 50%. This counterfactual is constructed as follows. Firstly, we fix c_0 at 50% higher than its estimated value. Secondly, we adjust the outer-loop parameters ξ_{JC}, ξ_{JD}, and ξ_e to hold the simulated standard deviations of JC, JD, and entry at the same values from our estimated model. We do not adjust the remaining outer-loop parameters and hyper-parameters from their original values, which holds the firm age distribution exactly at the original estimated distribution. Finally, given these outer-loop parameters, we re-run the inner loop, which adjusts the inner loop parameters to the values needed to continue exactly hitting the original inner loop moments. With this new calibration, we repeat our original business cycle simulation in order to compute the new moments discussed in the text.

D Additional Tables and Figures

Table 8: Equilibrium policies and values in steady state

	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
v_i	0.7242	0.7244	0.8226	2.6328	2.6561
p_i	0.9342	0.6732	0.7127	1.3085	1.0735
jc_i	0	0	0	0.0446	0.0458
jd_i	0.0995	0.0995	0.0294	0	0
njc_i	-0.0996	-0.0996	-0.0295	0.0445	0.0457
N_i/N	0.0031	0.1208	0.3841	0.4848	0.0072
M_i/M	0.0498	0.1423	0.3968	0.4035	0.0076

Table summarizes the value and policy functions in steady state across productivity levels $i = 1, ..., I$. v_i is firm value, and p_i productivity. Value is monotonically increasing across states. jc_i and jd_i are job creation and destruction rates per employee for incumbent firms, excluding the δ_F exit shock. njc_i is net job creation: $njc_i = jc_i - jd_i - \delta_F$. The final two rows give the fraction of employment and active firms (with at least one employee) repectively at each i in the ergodic distribution.

Figure D.1: Impulse Response Function - Cyclical behaviour of key aggregates

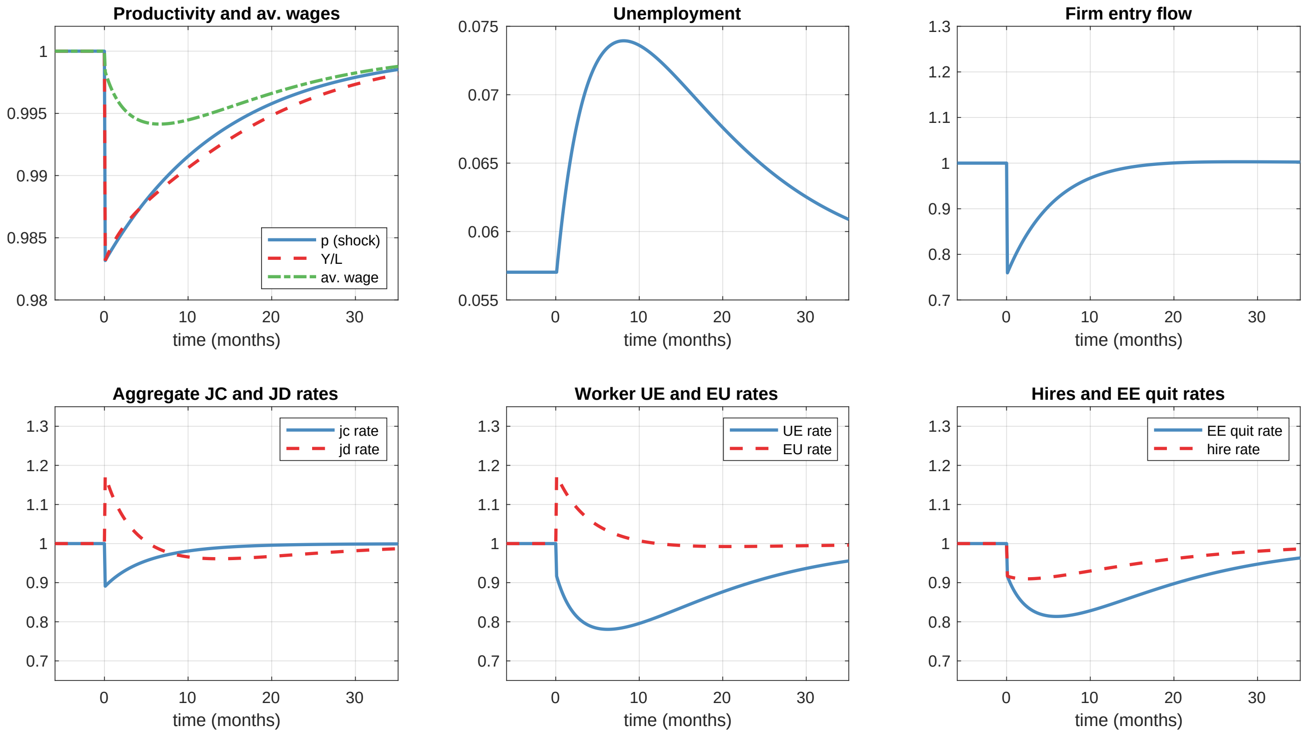


Figure plots additional aggregates from our typical recession impulse response function. See Section 7.1 for further details of the experiment.

Figure D.2: Model Impulse Response: Net JC by age

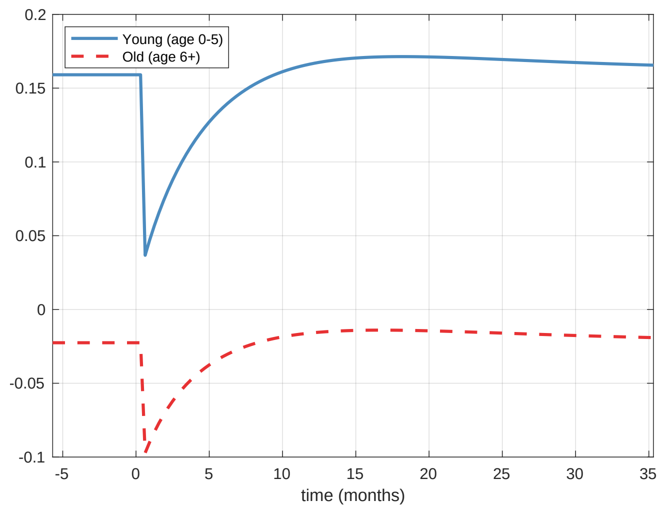


Figure plots Net JC rates by firm age for our typical recession experiment. See Section 7.1 for further details of the experiment. JC and JD flows are yearly, and computed as $1 - e^{-12r}$, where r are the average theoretical monthly rates within each bin.

References

[1] Achdou, Y, J. Han, J.-M. Lasry, P.-L. Lions, B. Moll. 2021. "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," *The Review of Economic Studies*, 2021.

[2] Arnoud, A, F. Guvenen, T. Kleineberg. 2019. “Benchmarking Global Optimizers,” *mimeo*.

[3] Audoly, R. 2020. “Firm Dynamics and Random Search over the Business Cycle,” Mimeo.

[4] Burdett, K. and D. Mortensen. 1998. “Wage Differentials, Employer size and Unemployment,” *International Economic Review*, 39: 257-273.

[5] Coles, M. G. and D. Mortensen. 2016. “Equilibrium Labor Turnover, Firm Growth, and Unemployment,” *Econometrica*, 84: 347-363.

[6] Coles, M. G. 2001. “Equilibrium Wage Dispersion, Firm Size and Growth,” *Review of Economic Dynamics*, 4(1): 159-187.

[7] Davis, S. J., R. J. Faberman and J. Haltiwanger. 2012. “Labor market flows in the cross section and over time,” *Journal of Monetary Economics*, 59(1): 1-18.

[8] Elsby, M., R. Michaels and D. Ratner. 2017. “Vacancy Chains”. University of Edinburgh, Department of Economics, mimeo.

[9] Moscarini, G. and F. Postel-Vinay. 2013. “Stochastic Search Equilibrium,” *Review of Economic Studies*, 80(4): 1545-1581.

[10] Shimer, R. 2005. “The Cyclical Behavior of Equilibrium Unemployment and Vacancies,” *American Economic Review*, 95(1): 25-49.

[11] Young, E. R. 2010. “Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm and non-stochastic simulations,” *Journal of Economic Dynamics and Control*, 34(1): 36-41.