Firm Dynamics, On-the-Job Search and Labor Market Fluctuations

Michael W. L. Elsby Axel Gottfries*

August 2021

Abstract

We devise a tractable model of firm dynamics with on-the-job search. The model admits analytical solutions for equilibrium outcomes, including quit, layoff, hiring and vacancy-filling rates, as well as the distributions of job values, a fundamental challenge posed by the environment. Optimal labor demand takes a novel form whereby hiring firms allow their marginal product to diffuse over an interval. The evolution of the marginal product over this interval endogenously exhibits gradual mean reversion, evoking a notion of imperfect labor market competition. This in turn contributes to dispersion in marginal products, giving rise to endogenous misallocation. Quantitatively, the model provides a parsimonious reconciliation of leading estimates of rent sharing, the negative association between wages and quits, the link between job and worker flows, and the cyclicality of labor market quantities and prices.

JEL codes: E24, E32, J63.

Keywords: Labor turnover, job creation/destruction, unemployment, business cycles.

---

* University of Edinburgh (mike.elsby@ed.ac.uk, axel.gottfries@ed.ac.uk).

We are grateful to the Editor Dirk Krueger, and anonymous referees for their constructive suggestions. We also thank Alex Clymo, Karl Harmenberg, Rasmus Lentz, Guido Menzio, John Moore and Jonathan Thomas as well as seminar participants at numerous institutions for helpful comments. All errors are our own.

We gratefully acknowledge financial support from the UK Economic and Social Research Council (ESRC), Award reference ES/L009633/1.

The labor market is in a perpetual state of flux. Large flows of unemployed workers find new jobs, while large flows of employed workers lose them (Blanchard and Diamond 1990). Many firms grow through job creation, while many others shrink through job destruction (Davis and Haltiwanger 1992). And, in tandem, substantial numbers of employed workers move directly from one employer to another (Fallick and Fleischman 2004). These worker and job flows vary considerably over the business cycle, and exhibit clear cross-sectional correlations (Davis, Faberman and Haltiwanger 2012, 2013).

The purpose of this paper is to understand the economics underlying this rich array of empirical regularities. To do so we devise a model that integrates firm dynamics with on-the-job search. Firms subject to hiring costs face idiosyncratic shocks that drive changes in their desired employment, and thereby job creation and destruction. Workers search for jobs across firms while both unemployed and employed, driving worker flows. Direct employer-to-employer transitions emerge naturally from the heterogeneity across firms induced by idiosyncratic shocks. And we show how the model can be extended to accommodate aggregate shocks, and thereby business cycles. The result is a framework in which an understanding of the economics of the foregoing stylized facts is feasible.

Attaining this goal is easier said than done, however. The interplay of firm dynamics with on-the-job search poses a seemingly daunting analytical challenge. In general, the rate of worker turnover faced by a firm will depend on the firm's position in the hierarchy of job values in the economy. Firms further up in the hierarchy will face lower turnover. Steady-state labor market equilibrium thus involves finding a fixed point of a *distribution* of job values, one that both sustains firms' labor demand decisions and is implied by aggregation of those same decisions. Out of steady state, equilibrium further involves finding a fixed point of the dynamic path of the distribution.

This paper proposes two contributions. First, it develops a benchmark model that admits an analytical characterization of labor market equilibrium and, crucially, the distribution of job values, induced by firm dynamics and on-the-job search. Second, a quantitative assessment of the model reveals that it is able to provide a parsimonious account of a wide range of stylized facts of labor market outcomes, both in the cross section, and over the business cycle.

In section 1, we devise a baseline environment that greatly simplifies the analytical challenge noted above. This is aided by a model of *ex post* wage bargaining that synthesizes insights from credible bargaining (Binmore et al. 1986) and multilateral bargaining (Bruegemann et al. 2018) in the presence of on-the-job search (Gottfries 2019).[1] The environment gives rise to a normalization in which the value of jobs to workers and firms are monotone functions of a single idiosyncratic state variable, the marginal product of labor. The distribution of job values can thus be summarized by the distribution of marginal products. Furthermore, optimal labor demand can be decoupled into two regions for the marginal product. Mirroring canonical models of firm dynamics (Bentolila and Bertola 1990; Hopenhayn and Rogerson 1993; Abel and Eberly 1996), there is a *natural wastage region*. At its lower boundary, firms shed workers into unemployment. On its interior, firms neither hire nor fire, and turnover occurs at a maximal constant quit rate.

A novel implication of the presence of on-the-job search, however, is the addition of a nondegenerate *hiring region*. Importantly, this emerges even in the absence of heterogeneity in marginal hiring costs. The key intuition is that hiring firms face a novel trade off in the presence of on-the-job search. On the one hand, they value the additional output generated by new hires. On the other, they value reductions in turnover associated with a higher marginal product. We show that this tradeoff is resolved by a novel solution: Firms allow their marginal products to diffuse across an interval, a strategy that is supported by a quit rate that declines with the marginal product at an appropriate rate. We show that the latter force is captured by a simple differential equation that gives rise to a closed-form solution for the quit rate. Crucially, this in turn gives rise to a closed-form solution for the distribution of marginal products offered to new hires—a key result in light of the analytical challenge noted above.[2]

The implications of the preceding behavior for aggregate labor market equilibrium are not obvious: Optimal labor demand and turnover are heterogeneous across firms, and evolve in a nonlinear fashion with idiosyncratic shocks. Nonetheless, we show how it is possible to derive an analytical characterization of steady-state labor market equilibrium. We begin by aggregating microeconomic behavior, obtaining expressions for the separation rate into unemployment, as well as the hiring rate, the vacancy-filling rate, and the distribution of workers at each marginal product. These in turn imply two conditions for

[1] Interestingly, the model also provides a rationale for the absence of offer matching if job offers are private information, since it is not credible to elicit them through the use of layoff lotteries (Moore 1985).

[2] In a model of consumer search, Burdett and Menzio (2018) derive an equilibrium pricing policy with strikingly similar features. Striking because their result arises for different reasons. In their model, an analogue of our natural wastage region emerges from the presence of *lump-sum* menu costs which induce firms to adjust *discretely* at the region's boundary. In our model, linear hiring costs induce *infinitesimal* control; the natural wastage region is instead the outcome of the presence of idiosyncratic shocks.

aggregate steady-state equilibrium that mirror those in the canonical Mortensen and Pissarides (1994) model: a *Beveridge curve* implied by steady-state unemployment flows; and a *job creation curve* that summarizes aggregate labor demand.

A host of insights follow on the nature of labor market behavior induced by the model. A first insight emerges from the fact that hiring rates are increasing in the marginal product. Coupled with decreasing quit rates, this gives rise to endogenous *gradual mean reversion* in marginal products among hiring firms. Positive innovations raise a firm's hiring rate and reduce its quit rate. Firms thus accumulate more workers and the marginal product reverts back in expectation. An appealing interpretation is that the latter is a manifestation of *imperfect labor market competition*; perfect competition would imply infinite mean reversion to a law of one marginal product. We show that this is a distinctive implication of the interaction of firm dynamics and on-the-job search in the model: limiting economies without these ingredients do not exhibit this property.

Second, the model reveals a novel paradox in the interplay between on-the-job search and misallocation. As in canonical models of on-the-job search (Burdett and Mortensen 1998), equilibrium in our model involves dispersion in marginal products across workers, and thereby misallocation. In stark contrast to canonical models, however, on-the-job search *contributes to*, rather than resolves, such misallocation by inducing turnover costs on firms, and thereby the presence of a nondegenerate hiring region. The model thus captures a novel notion of *endogenous misallocation*, driven by the interaction of firm dynamics and on-the-job search.

We turn to a quantitative assessment of the model in section 2. We explore a calibration that targets standard estimates of the levels of labor market stocks and flows, hiring costs, wage gains to on-the-job search, and inaction in hiring across firms. Strikingly, the calibrated model is able to replicate a wide array of nontargeted cross-sectional stylized facts.

First, the model can accommodate quintessential symptoms of imperfect labor market competition noted by Manning (2011). Hiring costs generate employer rents. *Ex post* bargaining generates a rent-sharing link between labor productivity and wages. And a defining implication of the model is a quit rate that endogenously declines in productivity. Together, these give rise to a negative association between quits and wages. Quantitatively, the model performs well on all these dimensions. In addition to matching the size of employer rents through calibration of the hiring cost, the model aligns well with recent leading estimates from the empirical rent-sharing literature (Kline et al. 2019), and delivers a wage-elasticity of quits that lies in the range of estimates reported by Manning, and Kline et al. We are unaware of prior work that has been able to match these moments jointly.[3]

Second, the model naturally generates cross-sectional relationships between worker flows and firm growth that mirror those documented in the empirical work of Davis et al. (2012, 2013). Firm growth in the model is monotone in the marginal product. Faster-growing firms are thus less likely to lay off workers, more likely to hire and post vacancies and, crucially, will face lower quit rates and higher vacancy yields. Davis et al. highlight the latter as important channels missing from conventional models. Quantitatively, the calibrated model again performs well on these dimensions, as well as with a host of indicators of the incidence and persistence of desired hiring, and of hires without a prior vacancy, emphasized by Davis et al. A contribution of the model is that this large set of outcomes emerges naturally from the environment; recent work has instead provided potential explanations for a subset of these outcomes in isolation.[4]

Finally, in section 3, we explore the aggregate dynamics implied by the model out of steady state. Recall that, in general, this involves a fixed point in the dynamic path of the distribution of job offers. Note that this problem is distinctly more intractable than those that arise in standard heterogeneous agent models in which agents must forecast a market price. Here the analogue of the market price is a whole function, the offer distribution. Nonetheless, we are able to make progress by generalizing our earlier results. In particular, the same forces that give rise to a closed-form solution for the offer distribution of marginal products in steady state allow us to infer the functional form of the offer distribution out of steady state. Doing so reduces the problem to one of inferring the dynamic path of a single scalar, labor market tightness.

We use our approach to study the transition dynamics following an MIT shock in the model calibrated as in section 2. Since the latter is informed solely by steady-state moments, this allows a quantitative assessment of aggregate dynamics implied by the model. Much like many models in the search tradition, we find that the model generates

[3] To the contrary, Manning (2011) suggests standard models of imperfect labor market competition are unable to reconcile estimates of the wage-elasticity of quits with estimates of employer rents (hiring costs) without invoking an extremely convex hiring technology.

[4] Kaas and Kircher (2015) and Gavazza et al. (2018) explain the behavior of vacancy yields by allowing for imperfect substitutability between vacancies and other recruitment margins. Conversely, Schaal (2017) explains the relation between quit rates and job flows in a model of directed on-the-job search.

limited internal propagation. The implied amplitudes of labor market outcomes, however, are plausible. Based on an update of the methods of Shimer (2005), we find that the model accounts for around 60 percent of the empirical volatility of unemployment, vacancies and the job-finding rate; captures the spike in job loss at the onset of recessions; and essentially replicates the volatility of the job-to-job transition rate, a central feature of the model. Furthermore, it does so while simultaneously replicating influential estimates of the procyclicality of real wages (Solon et al. 1994).

Taken together with the cross-sectional results of section 2, the model thus provides a parsimonious quantitative account of key endogenous outcomes—from markers of imperfect competition, to the interaction of worker and job flows, to the cyclical behavior of labor market quantities and prices. A key contribution of the model is that it matches these moments jointly, with few degrees of freedom.

In the closing sections of the paper, we show how the baseline model can be extended in several directions. Most prominently, we show that the hiring region varies in an interesting way with the structure of wage determination. We extend the sequential auctions approach of Postel-Vinay and Robin (2002) to allow for multi-worker firms and partial offer matching. A revealing implication is that firms' turnover costs, and thereby the size of the hiring region, are declining in firms' ability to match offers. In the limit in which firms can tailor their responses perfectly to the idiosyncratic outside offers of their workers, firms become *indifferent* to turnover, and the hiring region collapses. Our results then give rise to a simple novel analytical characterization of the equilibrium in this limit. Absent an ability to match offers, a firm has one instrument—the marginal product—to respond to a *continuum* of outside offers. In the presence of constraints to its ability to match such a continuum of outside offers, the firm will face costs of turnover, and a nondegenerate hiring region emerges.

The model can also be adapted to include several extensions often invoked in the literature on firm dynamics. Training costs, convex hiring and vacancy costs, and firm entry, growth and exit all can be accommodated. The key insight is that the aggregation results we develop hold more generally, allowing these extended problems to be distilled into systems of differential equations that are amenable to solution.

We conclude by offering thoughts on the direction of future work. The tractability of our framework rests on strong assumptions on wage determination that imply that firms are unable to commit to future wages. Although our quantitative results suggest this approach to wage determination can nonetheless reconcile important stylized facts on wage outcomes, there is much more work to be done to understand the economic implications of (limited) commitment. This in turn would further refine a key theme of the present paper, by providing a more complete synthesis of the frictional forces raised by the presence of on-the-job search, and the neoclassical forces underlying firm dynamics.

Related literature. The model set out in this paper provides a new theory of firm dynamics with (random) job search, both off- and on-the-job. In addition to the work already cited, it relates to three further strands of literature.

First, our model builds on recent work that has developed so-called "large-firm" search models that fuse firm dynamics with off-the-job search. These have been used to study firm growth (Acemoglu and Hawkins 2014), worker flows over the business cycle (Elsby and Michaels 2013), the role of wage posting and directed search in recruitment (Kaas and Kircher 2015), and cyclical recruitment intensity (Gavazza, Mongey and Violante 2018). None of these papers incorporates on-the-job search, however.

Second, a further strand of related literature has incorporated a business cycle into models of on-the-job search. Menzio and Shi (2011) demonstrate that the presence of directed search imparts on equilibrium a "block-recursive" structure that allows characterization of aggregate labor market dynamics without having to solve for the dynamics of distributions of job values. Closer to our environment is a strand of random search models that must, and do, confront this challenge (Moscarini and Postel-Vinay 2013; Coles and Mortensen 2016; Lise and Robin 2017). More recently, Audoly (2019) and Gouin-Bonenfant (2018) study related models that incorporate entry, exit and firm lifecycles. In contrast to our model, however, all such work has maintained the assumption of linear production technologies.

Third, and most closely related to our work, a handful of recent papers has sought to integrate firm dynamics with on-the-job search. Lentz and Mortensen (2012) focus on firm lifecycles and steady-state wage and productivity dispersion in a model without idiosyncratic or aggregate shocks. In a related model with idiosyncratic shocks, Trapeznikova (2017) incorporates an intensive margin of hours adjustment. Fujita and Nakajima (2016) study the relation between worker and job flows over the business cycle, but assume for tractability that workers have no bargaining power. Elsby, Michaels and Ratner (2019) study an environment related to ours, but focus on the interaction between replacement hiring and on-the-job search across firms in generating vacancy chains.

In an important contribution, Schaal (2017) devises a tractable theory of firm dynamics and on-the-job search. His insight is that, if search is directed, and firms can commit to complete state-contingent contracts, equilibrium has a block-recursive structure such that the distribution of job values no longer shapes turnover decisions, as in Menzio and Shi (2011). This allows a complete characterization of aggregate dynamics, as well as extensions to consider the effects of time-varying idiosyncratic risk. By contrast, our model considers a case in which search is random, there is no commitment in wage contracts, and the distribution of job values affects turnover. Unlike Schaal (2017), an exact solution for the aggregate dynamics is not feasible. However, a contribution is to show that the analytical challenge that accompanies our case can largely be surmounted: We develop a simplified numerical scheme that can solve for the responses to MIT shocks.[5]

Bilal, Engbom, Mongey and Violante (2019) make two important contributions relative to our analysis. First, they provide sufficient conditions based on limited commitment and mutual consent that distil the firm's problem into one of surplus maximization. Second, they present a much fuller exploration of a model with a convex vacancy cost, and firm entry and exit, enabling a novel quantitative study of worker flows and employment dynamics over firms' lifecycles. Our analytical framework can accommodate some of these features—convex vacancy costs, and endogenous firm entry, for example. However, central to the tractability of our model is the availability of a normalization of the firm's problem that reduces its idiosyncratic state to a single variable, the marginal product. A prominent case considered by Bilal et al. in which this fails is one where firms may endogenously choose to exit in the presence of fixed operating costs. This yields richer implications for firm lifecycles in their model.

Importantly, relative to Schaal (2017) and Bilal et al. (2019)—and, by extension, the preceding literature—we offer two main contributions. First, we provide several novel analytical results: The identification and characterization of a nondegenerate hiring region, the associated equilibrium quit rate, the role of wage determination in shaping these, the analytics of aggregation, and the use of all of these in simplifying and solving for aggregate dynamics are new results of this paper. Second, our model yields predictions for wages—as opposed to just values—which admit a novel reconciliation of many of the salient features of the empirical relationship between wages, firm productivity, and turnover.

[5] Furthermore, given that we show that the market prices that agents need to forecast can be reduced to a scalar—labor market tightness—our numerical scheme naturally can be extended to provide approximate solutions to stochastic aggregate shocks, along the lines of Krusell and Smith (1998).

1. Model

In this section, we devise a new model of the interaction of firm dynamics and on-the-job search, and present an array of analytical properties of its solution. We begin by describing the economic environment, and a baseline protocol for wage determination, according to which firms and workers interact. Given these, we present the problems facing firms and workers. We show that these admit analytical solutions for optimal labor demand and equilibrium turnover. In turn, we show how aggregation of this behavior also can be inferred analytically, allowing a characterization of steady-state labor market equilibrium. The section closes by noting a range of novel properties of equilibrium behavior in the model.

1.1 Environment

Time is continuous and the horizon is infinite. The labor market is comprised by two sets of agents, firms and workers, that we now describe.

Firms. There is a unit measure of firms. Each firm employs a measure of workers, denoted n, to produce a flow of output, denoted y, using an isoelastic production technology $y = xn^{\alpha}$, where $\alpha \in (0,1)$. Idiosyncratic productivity x is the source of uncertainty to the firm, and of heterogeneity across firms. It evolves according to the geometric Brownian motion

$$\mathrm{d}x = \mu x \mathrm{d}t + \sigma x \mathrm{d}z, \tag{1}$$

where $\mathrm{d}z$ is the increment to a standard Brownian motion.

Firms hire workers subject to a per-worker hiring cost c. Denoting the cumulative sum of a firm's hires by H, and its increment over the time interval $\mathrm{d}t$ by $\mathrm{d}H$, the firm faces flow hiring costs of $c \cdot \mathrm{d}H$. Separations occur through two channels. First, the firm's employees quit at rate δ. Second, additional separations may be implemented at zero cost; we denote their cumulative sum by S, and its increment $\mathrm{d}S$.[6] It follows that the firm's employment evolves according to the law of motion

$$\mathrm{d}n = \mathrm{d}H - \mathrm{d}S - \delta n \mathrm{d}t. \tag{2}$$

[6] We use this notation to allow for the possibility that a firm may choose a continuous, but non-differentiable path for cumulative hires and separations.

Firms' hires are mediated through vacancies. The firm faces a vacancy-filling rate q. By a law of large numbers, the firm posts the requisite vacancies v to implement its desired hires $\mathrm{d}H = qv\mathrm{d}t$. This vacancy policy is strictly optimal if, in addition to the hiring cost c, vacancy posting further incurs an arbitrarily-small cost. We denote the aggregate measure of vacancies per firm by V.

Workers. There is a unit measure of households, each of which is comprised by a measure L of workers. Workers are risk-neutral, discount the future at rate r, and occupy one of two employment states: employment and unemployment. We denote the aggregate measures of unemployed and employed workers per household by U and $L - U$ respectively. While unemployed, workers receive a flow payoff b. While employed, they receive a flow wage w, determined according to a protocol described below.

Matching. Firms hire workers by posting vacancies. Workers search while unemployed, and while employed with exogenous relative search intensity s. Frictions are embodied in a standard increasing, continuous, constant-returns-to-scale meeting function $M(U + s(L - U), V)$ that regulates the total flow of contacts M arising from V vacancies, U unemployed searchers, and $s(L - U)$ employed searchers. The ratio of vacancies to searchers, $\theta \equiv V/[U + s(L - U)]$ is thus a sufficient statistic for contact rates: Vacancies contact a searcher at rate $\chi(\theta) = M/V = M(1/\theta, 1)$. Unemployed searchers receive job offers at rate $\lambda(\theta) = M/\big(U + s(L - U)\big) = M(1, \theta)$, and employed workers at rate $s\lambda(\theta)$. We assume that $M(1,0) = 0$, and $M(1, \theta) \to \infty$ as $\theta \to \infty$. To economize on notation, in what follows we suppress dependence on θ, except where necessary.

Analytical challenge. A defining consequence of on-the-job search is that not all offers are accepted: an offer will be accepted only if the worker's valuation of the prospective offer exceeds that of her current firm. This feature of the environment poses two related analytical challenges.

First, depending on the nature of wage determination, workers' valuations of offers can in principle depend on (arbitrarily) many state variables—for example, the wages of all other co-workers in the firm, in addition to the firm's employment n, and productivity x. Collect these into the vector $\mathbf{x}$. A consequence is that the turnover rates faced by a firm will inherit the state variables $\mathbf{x}$, and firms must keep track of them.

Consider the quit rate δ. Each of the firm's employees receives an offer from another firm at rate $s\lambda$. And each contacted employee will choose to quit if the worker values the outside offer $\tilde{\mathbf{x}}$ above that at the current firm $\mathbf{x}$; that is, if $\tilde{\mathbf{x}}$ is in the *acceptance set* $a(\mathbf{x})$. Note that the latter is an endogenous outcome of the model that depends on the contractual environment (that we specify shortly). Denoting the joint distribution of states among *job offers* by $\Phi(\cdot)$, we can write the quit rate faced by the firm as

$$\delta(\mathbf{x}) = s\lambda \int_{\tilde{\mathbf{x}} \in a(\mathbf{x})} \mathrm{d}\Phi(\tilde{\mathbf{x}}). \tag{3}$$

The vacancy-filling rate q faced by the firm likewise entails similar considerations. Each of the firm's vacancies contacts a searcher at rate χ. With probability $\psi = U/[U + s(L - U)]$ the searcher is unemployed, and hired with certainty (since no firm will post a vacancy unattractive to an unemployed searcher). With probability $1 - \psi$, the searcher is employed, and is hired only if she values the firm's offer $\mathbf{x}$ above that of her existing firm $\tilde{\mathbf{x}}$. Denoting the joint distribution of states among *employees* by $\Gamma(\cdot)$, the vacancy-filling rate faced by the firm can thus be written as

$$q(\mathbf{x}) = \chi \left[\psi + (1 - \psi) \int_{\tilde{\mathbf{x}} : \mathbf{x} \in a(\tilde{\mathbf{x}})} \mathrm{d}\Gamma(\tilde{\mathbf{x}}) \right]. \tag{4}$$

Thus, firms must keep track of the states $\mathbf{x}$ that determine workers' turnover decisions, as well as the endogenous ordering of workers' valuations over those states, summarized by the acceptance set $a(\mathbf{x})$.

The second analytical challenge posed by the environment is that turnover is determined by a firm's position in a joint distribution of potentially-many states—summarized by $\Phi(\cdot)$ and $\Gamma(\cdot)$ above. Firms must know this distribution in order to make labor demand decisions; and the distribution in turn is determined by aggregation of those same decisions. Steady-state equilibrium thus involves a fixed point in this distribution. And out of steady state equilibrium further involves a fixed point in the dynamic path of the distribution.

Note that this challenge is distinct from that posed in standard models of aggregate equilibrium in heterogeneous agent economies (as in, for example, Krusell and Smith 1998). In this frictional firm dynamics context, the latter involves firms having to forecast the path of the equilibrium market tightness—a *scalar* (see, for example, Elsby and Michaels 2013). The presence of on-the-job search overlays on top of this the higher-dimensional challenge of firms having to forecast the path of the *functions* $\delta(\cdot)$ and $q(\cdot)$.

Central to these challenges is the structure of wage determination, which determines workers' valuations of offers, the vector of states $\mathbf{x}$ that inform them and, thereby, turnover decisions. In what follows we show how progress can be made on these challenges by devising tractable protocols for wage determination.

1.2 Wage setting

Our baseline model uses a simple protocol in which wages are determined entirely *ex post*—that is, after all search decisions have been completed—according to a model of bargaining between a firm and its many workers. A corollary is that all workers in a given firm are paid a common wage with a simple structure. Firms do not engage in offer matching in response to their employees' outside offers. Later, in section 4, we study an alternative protocol that accommodates offer matching, generalizing the sequential auctions model of Postel-Vinay and Robin (2002) to a multi-worker firm context.

Bargaining in the absence of offer matching. For now, though, we begin by describing a simple model of *ex post* bargaining between a firm and its many workers in the absence of offer matching. To clarify our meaning of *ex post*, it is helpful first to return to the environment faced by firms and workers and consider the order of events within each $\mathrm{d}t$ period. At the beginning of the period, productivity is realized, and hiring and separation decisions are made. Upon completion, a bargaining stage then begins in which wages are negotiated between the firm and its many workers—it is in this sense that bargaining is *ex post*. Once bargaining is complete, production takes place, agreed wages are paid, and the period concludes.

The bargaining stage takes the following form. The firm and its workers bargain over the flow wage for the current period, $w\mathrm{d}t$, according to the bargaining game proposed by Bruegemann, Gautier and Menzio (2018). The firm engages in a sequence of bilateral bargaining sessions with each of its workers subject to breakdown risk. The sequence of play is devised such that the strategic position of each worker within the firm is symmetric. They characterize an equilibrium[7] of the game in which all workers within the same firm receive the same wage, and this wage coincides with that implied by a marginal surplus-sharing rule proposed by Stole and Zwiebel (1996).

[7] Specifically, the limit as the probability of breakdown goes to zero of their no-delay subgame perfect equilibrium. In their static setting, they show that there is a unique no-delay equilibrium. A sufficient condition for the latter to hold in our dynamic setting is the presence of non-history-dependent strategies.

The relevant marginal surplus that the firm and its workers share is determined by the threats that the firm and each of its workers can credibly issue in the event of a breakdown of negotiations. Binmore et al. (1986) and, more recently, Hall and Milgrom (2008) emphasize that threats of *permanent* suspension of negotiations are not plausibly credible in this setting: Regardless of a breakdown in the current period, the firm will wish to resume negotiations with the same workers in the subsequent period. Instead, breakdown is credibly associated only with a *temporary* disruption of production due to delayed agreement. Since wages are renegotiated every period, turnover and wages in subsequent periods will be independent of the currently-agreed wage, and the effective surplus that the firm and its workers share will be the marginal *flow* surplus.

This approach to wage bargaining has several appealing properties. First, wage outcomes take a particularly simple form. Following Hall and Milgrom, suppose that, in the event of breakdown, each employee receives a flow payoff ω_e, and a firm incurs a per-worker flow cost ω_f. Then, marginal flow surplus sharing implies

$$\beta\left(x\alpha n^{\alpha-1} - w - w_n n + \omega_f\right) = (1-\beta)(w-\omega_e), \tag{5}$$

where $\beta \in (0,1)$ indexes worker bargaining power. Defining $\omega_0 \equiv \beta\omega_f + (1-\beta)\omega_e$,[8] it is straightforward to verify that the wage solution takes the following simple form,

$$w = \frac{\beta}{1-\beta(1-\alpha)} x\alpha n^{\alpha-1} + \omega_0. \tag{6}$$

The wage equation captures some familiar forces: Wages are increasing in the marginal product $x\alpha n^{\alpha-1}$, and the flow payoffs from breakdown, summarized by ω_0. Due to decreasing returns in production, $\alpha \in (0,1)$, failure to agree with an individual worker will result in higher bargained wages for all remaining workers. Using these threats, workers are able to capture some of the inframarginal product, giving rise to the leading coefficient. Because breakdown of negotiations does not involve permanent severance of a match, the option values to search (both off- and on-the-job) do not play a role in wage outcomes. In this respect, the wage bargain resembles aspects of Hall and Milgrom (2008), extended to accommodate multi-worker firms and continual renegotiation.

A further virtue of this approach to wage bargaining is that it can be reconciled with the presence of on-the-job search, in two important senses. First, it is not subject to the

[8] Strictly, the wage equation holds in the event of agreement, which occurs provided the marginal flow surplus is positive, $\{x\alpha n^{\alpha-1}/[1-\beta(1-\alpha)]\} + \omega_f - \omega_e > 0$. We assume this holds in what follows. A sufficient condition is $\omega_f > \omega_e$.

concern noted in Shimer (2006) that the effects of bargained wages on turnover will render the bargaining set nonconvex. Since bargaining pertains only to the current flow wage, which in turn is re-bargained each period, current wages have no effect on future wages, and thereby turnover (see Nagypal 2007, and Gottfries 2019).[9] Second, this approach to wage bargaining also suggests a natural rationale for the absence of offer matching. Suppose job offers are privately observed by workers and unverifiable. A firm would be able to elicit the value of such offers if it were able to confront its (potential) workers with a set of appropriately-devised layoff lotteries (Moore 1985). But, echoing our earlier discussion of the bargaining stage, such layoff lotteries will not be credible *ex post*: the firm will wish to resume its relationship with a worker after any such layoff realization. Thus, inability to commit to permanent severance provides a simple reconciliation of wage bargaining, on-the-job search, and absence of offer matching. Mortensen (2003, section 5.1) discusses at further length these and other possible impediments to offer matching.

1.3 Firm and worker problems

A key implication of the wage solution in (6) is that all workers within a firm are paid a common wage w which depends only on the firm's marginal product, $x\alpha n^{\alpha-1}$. This implies that the potentially high-dimensional vector $\mathbf{x}$ of state variables for the worker and firm can be reduced to the firm's productivity x, and employment n.[10] This in turn allows a statement of the problems faced by firms and workers, to which we now turn.

Firm problem. Given the environment described above, and recalling the laws of motion for the firm's productivity and employment in (1) and (2), standard methods (Dixit 1993; Stokey 2009) imply that the Bellman equation for the value of the firm $\Pi(n,x)$ satisfies

$$r\Pi \mathrm{d}t = \max_{\mathrm{d}H\geq 0, \mathrm{d}S\geq 0} \left\{ \left[xn^\alpha - wn - \delta n \Pi_n + \mu x \Pi_x + \frac{1}{2}\sigma^2 x^2 \Pi_{xx} \right] \mathrm{d}t - (c - \Pi_n)\mathrm{d}H - \Pi_n \mathrm{d}S \right\}, \quad (7)$$

where the flow of hires $\mathrm{d}H = qv\mathrm{d}t$ is generated by posting the appropriate vacancies v.

[9] An increase in the flow wage will increase worker values in proportion to the duration of the contract, $\mathrm{d}t$. By contrast, the impact on turnover is determined by the latter increase in worker values multiplied by the contract duration, and is thus proportional to $(\mathrm{d}t)^2$. Turnover and surplus therefore become independent of the bargained flow wage as renegotiation becomes very frequent (Gottfries 2019).

[10] We do not consider possible equilibria in which the state includes variables that are not directly payoff relevant. These are formally ruled out later in our definition of an m-solution.

Here, the firm's value Π, wage w, quit rate δ, and vacancy-filling rate q are all functions of the reduced state (n, x). The firm chooses its hires $\mathrm{d}H$ and separations $\mathrm{d}S$ to maximize the expected present discounted value of its profit stream. Its flow profits are given by the flow revenue xn^{α}, less wage payments wn and hiring costs $c \cdot \mathrm{d}H$. The firm faces capital gains from two sources. First, the firm's employment n evolves according to the law of motion (2). Each incremental change $\mathrm{d}n$ is valued by the firm according to the marginal value Π_n. The second source of capital gains to the firm arises from the idiosyncratic shocks to firm productivity x, which evolve according to the stochastic law of motion (1). Application of Ito's lemma yields the form in (7).

Worker problem. A worker currently employed at a firm with employment n and productivity x must choose the set of outside offers $a(n, x)$ she will accept, if contacted. The environment implies a simplification of this decision. Workers share a common valuation of each firm, summarized by its *worker surplus*—the value each firm offers its workers in excess of unemployment—which we denote W. An employee thus accepts outside offers with a worker surplus higher than that of her current firm. It follows that the quit and vacancy-filling rates in (3) and (4) can be written more succinctly as

$$\delta(W) = s\lambda[1 - \Phi(W)], \text{ and, } q(W) = \chi[\psi + (1 - \psi)\Gamma(W)], \tag{8}$$

where $\Phi(\cdot)$ can now be interpreted as the offer distribution of worker surpluses, and $\Gamma(\cdot)$ the associated worker distribution.

It remains to determine the worker surplus $W(n, x)$. Consider first the value of employment $\Omega(n, x)$ to a worker currently employed in a firm offering worker surplus W. This satisfies

$$\begin{aligned} r\Omega \mathrm{d}t = \max\Bigg\{ & \left[w + s\lambda \int_W \big(\widetilde{W} - W\big) \mathrm{d}\Phi\big(\widetilde{W}\big) - \delta n \Omega_n + \mu x \Omega_x + \frac{1}{2}\sigma^2 x^2 \Omega_{xx} \right] \mathrm{d}t \\ & + \Omega_n(\mathrm{d}H^* - \mathrm{d}S^*) - W\frac{\mathrm{d}S^*}{n}, r\Upsilon \mathrm{d}t \Bigg\}, \end{aligned} \tag{9}$$

where the value of employment Ω, the wage w, the worker surplus W, and the firm's optimal hiring and firing flows $\mathrm{d}H^*$ and $\mathrm{d}S^*$ are all functions of the reduced state (n, x).

An employed worker receives a flow wage w given by (6), and faces capital gains from three sources. First, at rate $s\lambda$ she contacts an outside firm with worker surplus $\widetilde{W}$ drawn from the offer distribution of worker surpluses $\Phi(\cdot)$. She accepts the outside job only if it offers a larger worker surplus, $\widetilde{W} > W$. Second, employment at her current firm will evolve according to the law of motion (2). If the worker remains employed by the firm, she values each incremental change $\mathrm{d}n$ by Ω_n. If the firm implements layoffs, $\mathrm{d}S^* > 0$, the worker faces a uniform risk of being laid off and realizing a capital loss equal to the worker surplus W. Since the flows of hires and fires are chosen by the firm, they are evaluated at the equilibrium values that maximize the firm's problem in (7), $\mathrm{d}H^*$ and $\mathrm{d}S^*$. Third, her current firm's idiosyncratic productivity evolves according to the stochastic law of motion (1) and, by Ito's lemma, gives rise to the remaining capital gain terms.

Finally, note that the worker retains an option to quit, which she will exercise whenever Ω falls below the value of unemployment Υ to a worker. This in turn satisfies

$$r\Upsilon = b + \lambda \int \widetilde{W} d\Phi(\widetilde{W}). \tag{10}$$

While unemployed, a worker receives a flow payoff b. At rate λ she receives an offer with worker surplus $\widetilde{W}$ drawn from the offer distribution of worker surpluses $\Phi(\cdot)$. Since it is never optimal for a firm to make an offer that would not be accepted by an unemployed searcher, the worker accepts with certainty.

Recalling that the worker surplus is the additional value to a worker of employment over unemployment, $W(n,x) \equiv \Omega(n,x) - \Upsilon$, and noting that the value of unemployment Υ is independent of any firm's idiosyncratic state, the worker surplus satisfies

$$\begin{aligned} rW\mathrm{d}t = \max\Bigg\{\Bigg[& w - b - \lambda \int \widetilde{W} d\Phi(\widetilde{W}) + s\lambda \int_W (\widetilde{W} - W)\mathrm{d}\Phi(\widetilde{W}) - \delta n W_n + \mu x W_x \\ & + \frac{1}{2}\sigma^2 x^2 W_{xx}\Bigg]\mathrm{d}t + W_n(\mathrm{d}H^* - \mathrm{d}S^*) - W\frac{\mathrm{d}S^*}{n}, 0\Bigg\}. \end{aligned} \tag{11}$$

In what follows we assume that the worker's reservation wage is sufficiently low such that the firm (weakly) initiates all separations into unemployment, and optimality decisions over hires and fires can be inferred from solving the firm's problem. However, the alternative case can be accommodated by a similar analysis.

1.4 Labor demand and turnover

We can now proceed to consider optimal labor demand and turnover decisions. Recall that the latter is a key challenge that arises from the interaction of firm dynamics and on-the-job search, as labor demand decisions and turnover rates are intertwined in this environment. In this subsection, we provide a solution in which the joint determination of labor demand and turnover takes a surprisingly simple and tractable form.

We begin by returning to the firm's problem in (7). Optimal hires and separations satisfy (see, for example, Harrison and Taksar 1983)[11]

$$(-c + \Pi_n)\mathrm{d}H^* = 0, \text{ and, } \Pi_n \mathrm{d}S^* = 0. \tag{12}$$

The marginal value of labor Π_n is set equal to the marginal hiring cost c in the event of hiring, $\mathrm{d}H^* > 0$, and to zero in the event of firing, $\mathrm{d}S^* > 0$. It follows that the maximized value of the firm satisfies

$$r\Pi = xn^\alpha - wn - \delta n \Pi_n + \mu x \Pi_x + \frac{1}{2}\sigma^2 x^2 \Pi_{xx}. \tag{13}$$

The proximate effects of on-the-job search on the firm are thus distilled in the *turnover costs* $\delta n \Pi_n$. Intuitively, each of the firm's n employees quits at rate δ, and is valued on the margin by the firm at Π_n. The magnitude of these turnover costs, and the firm's response to them, will play a central role in the model.

Equation (12) provides conditions on the marginal value of labor to the firm Π_n. For brevity, in what follows we shall denote the latter by $J \equiv \Pi_n$. Differentiating the firm value in (13) implies that

$$rJ = x\alpha n^{\alpha-1} - \frac{\partial(wn)}{\partial n} - \frac{\partial(\delta n J)}{\partial n} + \mu x J_x + \frac{1}{2}\sigma^2 x^2 J_{xx}, \tag{14}$$

The marginal value of labor to the firm is determined by the flow marginal product $x\alpha n^{\alpha-1}$ net of the marginal cost of labor $\partial(wn)/\partial n$ and the marginal turnover costs $\partial(\delta n J)/\partial n$, together with the capital gains associated with shocks to the firm's idiosyncratic productivity.

Solution approach. As stated, the worker surplus in (11), the firm's value in (13) and the firm's marginal value in (14) require two idiosyncratic state variables: the firm's employment n, and productivity x. We now show how the structure of the problem admits a normalization that allows one to distill these into a single idiosyncratic state, namely the firm's flow marginal product, which we shall hereafter denote $m \equiv x\alpha n^{\alpha-1}$. Thus, we propose and verify a solution in which the marginal product m is a sufficient statistic for worker and firm behavior. We gather this together in the following definition.

[11] The reader may wonder whether the firm's optimality conditions also should include terms that capture potential effects of the firm's choice of hires $\mathrm{d}H$ and separations $\mathrm{d}S$ on turnover, via effects on the worker surplus in (11). Note, however, that the terms in $\mathrm{d}H^*$ and $\mathrm{d}S^*$ in (11) capture the present discounted value of the effects of the firm adhering to its *optimal* hiring and separation policy *in the future*.

Definition *An* m-solution *is a solution to the firm and worker problems such that, for any aggregate state, firms' optimal hiring and firing rates,* dH^*/n *and* dS^*/n*, and workers' optimal turnover decisions, are uniquely determined by the marginal product* m*.*

In what follows, we characterize equilibrium under an m-solution.

Optimal worker turnover. Consider first worker turnover. Confronted with an outside offer, the worker's optimal acceptance rule selects the firm that offers the higher worker surplus. The following result establishes that such decisions take a particularly simple form under the proposed m-solution.

Lemma 1 *Under an* m*-solution, the worker surplus* W *in (11) is uniquely determined by, and monotonically increasing in, the marginal product* m*. Workers' optimal acceptance set is therefore* $a^*(m) = \{\tilde{m}: \tilde{m} > m\}$*.*

This verifies the requirement of an m-solution that optimal worker turnover depends only on the marginal product m. The intuition for the monotonicity result in Lemma 1 comes from two channels. First, a direct benefit of being employed in a firm with a higher marginal product is a higher flow wage in (6). Second, under the proposed m-solution, a higher marginal product in the current period also implies a weakly higher path of future marginal products for any sequence of realizations of idiosyncratic shocks in (1).

The upshot of Lemma 1 for what follows is that optimal turnover decisions take a simple form, as orderings of worker surpluses coincide with orderings of marginal products. Thus, all job-to-job switches involve worker transitions from low-m firms to high-m firms. The marginal product thus becomes a sufficient statistic for worker turnover. Recall from (8) that the quit rate δ, and the vacancy-filling rate q, depend respectively on the distributions of worker surpluses among offers $\Phi(W)$, and workers $\Gamma(W)$. With a slight abuse of notation, it follows that we can rewrite these as

$$\delta(m) = s\lambda[1 - F(m)], \text{ and, } q(m) = \chi[\psi + (1-\psi)G(m)], \tag{15}$$

where $F(m) = \Phi[W(m)]$ is the offer distribution, and $G(m) = \Gamma[W(m)]$ the worker distribution, of marginal products.

Optimal labor demand. Now consider the determination of the firm's marginal value of labor. Applying the proposed m-solution, and the wage equation (6), the marginal value in (14) can be rewritten as a function solely of the marginal product m. With a slight abuse of notation, we will henceforth write this as $J(m)$, which satisfies the recursion

$$rJ(m) = (1-\omega_1)m - \omega_0 - [\delta(m) - (1-\alpha)m\delta'(m)]J(m) + [\mu + (1-\alpha)\delta(m)]mJ'(m) + \frac{1}{2}\sigma^2 m^2 J''(m), \tag{16}$$

where $1-\omega_1 \equiv (1-\beta)/[1-\beta(1-\alpha)]$ is the firm's share of the marginal product implied by the wage bargaining solution.

Optimality conditions for hires and separations provide boundary conditions for the firm's marginal value in (16). We will show that these are solved by a labor demand policy with three thresholds for the marginal product, $m_l < m_h < m_u$; respectively, the *layoff*, *hiring* and *upper* boundaries.

Optimal hires and separations are zero whenever the firm's marginal value J lies in the interval $(0, c)$. Because the presence of quits will induce employment to decline over time in this region, we shall refer to it as the *natural wastage region*. The firm will undertake non-zero separations $\mathrm{d}S^* > 0$ whenever the firm's marginal value J reaches the lower boundary 0, where the marginal product is m_l. Likewise, the firm will undertake non-zero hires $\mathrm{d}H^* > 0$ as soon as the firm's marginal value J reaches the boundary c, where the marginal product is m_h.

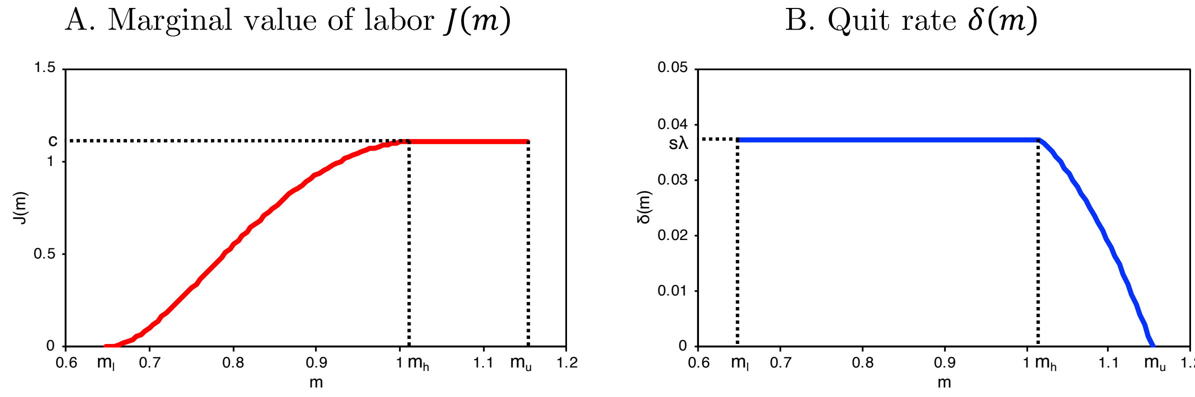
We shall see, however, that a distinctive implication of the interaction of on-the-job search with firm dynamics is the additional presence of a *hiring region* in which optimal hires $\mathrm{d}H^*$ are positive for all $m \in (m_h, m_u)$ such that the firm's marginal value J is equal to the marginal hiring cost c. That this interval may be nondegenerate is a novel and surprising feature of this environment. It also provides a key solution to the challenge of solving for the equilibrium distributions that, as we have discussed, are fundamental to models of on-the-job search.

We now characterize each of these two regions.

The natural wastage region. The natural wastage region is the more straightforward of the two. Under the proposed m-solution, the lowest-value hiring firm has marginal product m_h, which exceeds that for any firm in the natural wastage region where $m \in (m_l, m_h)$. Firms thus face the maximal quit rate $\delta(m) = s\lambda$, and thus $\delta'(m) = 0$, for all m on the interior of this region—hence *natural wastage*.

This considerably simplifies the recursion for the firm's marginal value (16),

Figure 1. Optimal labor demand and the equilibrium quit rate



Notes. Parameter values are based on the model calibrated as described in section 3.

$$(r + s\lambda)J(m) = (1 - \omega_1)m - \omega_0 + [\mu + (1 - \alpha)s\lambda]mJ'(m) + \frac{1}{2}\sigma^2 m^2 J''(m). \tag{17}$$

Two pairs of boundary conditions determine the marginal value $J(m)$ and the boundaries m_l and m_h. The first pair reiterates (12). Expressed in the vocabulary of Dumas (1991) and Stokey (2009), these are the *smooth-pasting* conditions,

$$J(m_l) = 0, \ \text{ and, } \ J(m_h) = c. \tag{18}$$

The second pair comprises the *super-contact* conditions,

$$J'(m_l) = 0, \ \text{ and, } \ J'(m_h) = 0. \tag{19}$$

Applying these yields the following solution.

Proposition 1 *In the natural wastage region, the quit rate is constant* $\delta(m) = s\lambda$, *and there is a unique solution for the firm's marginal value given by*

$$J(m) = \frac{(1 - \omega_1)m}{\rho(1)} - \frac{\omega_0}{\rho(0)} + J_1 m^{\gamma_1} + J_2 m^{\gamma_2}, \tag{20}$$

for all $m \in (m_l, m_h)$. *The coefficients* J_1 *and* J_2, *and the boundaries* m_l *and* m_h, *are known implicit functions (provided in the Appendix) of the parameters of the firm's problem, and are unique.* $\gamma_1 < 0$ *and* $\gamma_2 > 1$ *are roots of the fundamental quadratic,*

$$\rho(\gamma) = -\frac{1}{2}\sigma^2\gamma^2 - \left[\mu - \frac{1}{2}\sigma^2 + (1 - \alpha)s\lambda\right]\gamma + r + s\lambda = 0. \tag{21}$$

Constancy of the quit rate in the natural wastage region transforms the firm's labor demand decision into a canonical firm dynamics problem. An extension of the approach devised by Abel and Eberly (1996) yields the solution for the firm's marginal value in Proposition 1, and establishes its uniqueness.

The first two terms in (20) characterize the value to the firm of a marginal employee absent the option to hire and fire. The final two terms in (20) capture the marginal value of the options to separate from employees in adverse future states, and to hire employees in favorable future states. In combination, these yield a marginal value that is S-shaped in the natural wastage region, a shape that is characteristic of firm dynamics models with constant depreciation and infinitesimal control (Dixit 1993). Figure 1 illustrates.

Optimal labor demand in the natural wastage region thus corresponds closely to that in existing models of firm dynamics. We will see, however, that firm behavior differs importantly from this benchmark in the hiring region, to which we now turn.

The hiring region and the equilibrium quit rate. A distinctive feature of the interaction of firm dynamics and on-the-job search is that labor demand and turnover are jointly determined among hiring firms. Formally, we seek solution for the firm's marginal value $J(m)$ and the quit rate $\delta(m)$ that are mutually consistent in this case.

The model offers a considerable simplification, however. Proposition 2 first establishes that the quit rate $\delta(m)$ is continuous, differentiable, and strictly decreasing in any region in which there is strictly positive hiring. Equivalently, the offer distribution $F(m)$ has no mass points, and strictly positive density in any hiring region. Intuitively, an individual hiring firm can profitably deviate from any mass point by delaying hiring, allowing its marginal product to drift above the mass point, and realizing a discrete reduction in turnover costs, $\delta n \Pi_n$ in (13).[12] Equation (12) then stipulates that a hiring firm's marginal value be set equal to the marginal hiring cost, $J(m) = c$, and thus $J'(m) = J''(m) = 0$, for all m on the interior of any nondegenerate hiring region. This observation transforms the recursion for the marginal value in (16) into a differential equation for the quit rate $\delta(m)$,

$$rc = (1 - \omega_1)m - \omega_0 - [\delta(m) - (1 - \alpha)m\delta'(m)]c. \tag{22}$$

[12] Formally, note from (13), (18) and (19) that any discontinuity in the quit rate at the hiring threshold would contradict the optimality of hiring at m_h established in Proposition 1. Although, in principle, the implied discontinuity in turnover costs $\delta n \Pi_n$ in (13) could be offset by an appropriate discontinuity in the remaining terms in (13), the latter is ruled out by the smooth-pasting and super-contact conditions in (18) and (19) that underlie the optimality of labor demand in Proposition 1.

It follows that the quit rate must be differentiable in any hiring region. Finally, that the quit rate is strictly decreasing implies that the hiring region is a unique interval (m_h, m_u). Observing that $\delta(m_h) = s\lambda$ and $\delta(m_u) = 0$ in turn gives rise to a simple solution.[13]

Proposition 2 *There is a unique hiring region in which the firm's marginal value is constant,* $J(m) = c$*, and in which there is a unique solution for the quit rate given by*

$$\delta(m) = s\lambda + \frac{1}{c}\left\{\frac{(1-\omega_1)(m-m_h)}{\alpha} \right. \\ \left. - \left[\frac{(1-\omega_1)m_h}{\alpha} - \omega_0 - (r+s\lambda)c\right]\left[\left(\frac{m}{m_h}\right)^{\frac{1}{1-\alpha}} - 1\right]\right\}, \tag{23}$$

for all $m \in (m_h, m_u)$*, where the upper boundary solves* $\delta(m_u) = 0$ *and is unique. Furthermore,* $\delta(m)$ *is strictly decreasing and concave for all* $m \in (m_h, m_u)$*.*

Proposition 2 is an important result. By establishing the equilibrium quit rate $\delta(m)$, it in turn implies a solution for the equilibrium offer distribution of marginal products, $F(m)$ in (15). Proposition 2 thus provides a key part of the solution to the challenge of how to determine equilibrium turnover, and thereby the equilibrium distributions of marginal products. We will see that this provides a key building block to the determination of steady-state aggregate equilibrium, as well as out-of-steady-state aggregate dynamics.

Proposition 2 also has a surprising implication: Hiring firms that face a *homogeneous* per-worker hiring cost c nonetheless allow their marginal products to *vary* over an interval, giving rise to a *non-degenerate* distribution of worker values across hiring firms.

The intuition for why is as follows. Consider a firm at the middle boundary m_h. Following a positive innovation to its productivity x, and thereby its marginal product m, the firm faces a novel tradeoff in the presence of on-the-job search.

On one hand, the firm values the net additional output generated by new hires. If this were the firm's only consideration, it would simply hire until its marginal product returns to the middle boundary; formally, m_h would become a *reflecting barrier*. This is the force captured in canonical models of firm dynamics (such as Bentolila and Bertola 1990). On the other hand, in the presence of on-the-job search, this is not the firm's only

[13] The equilibrium quit rate in Proposition 2 renders hiring firms indifferent over their hiring rate. However, uniqueness of optimal hires for each firm can be restored by the introduction of an arbitrarily-small convexity in the hiring cost.

consideration: it also values reductions in turnover costs afforded by the lower quit rate that accompanies a higher marginal product.

The economy resolves this tradeoff in a novel way. Firms no longer hire until their marginal products are reflected back to m_h. Instead, they hire less aggressively, allowing their ms to diffuse across an interval (m_h, m_u). This policy is supported by a quit rate $\delta(m)$ in (23) that declines in m at an appropriate rate to maintain firms' desire to hire. Intuitively, the rate at which $\delta(m)$ declines in m is shaped by the density of offers $f(m)$. Suppose the density at some m were higher than implied by (23), any firms at that m would face excessive turnover costs. Their marginal value of labor would be below the hiring cost c, they would not find it optimal to hire, and the density of offers at that m would be zero—a contradiction. Conversely, suppose the density at some m were lower than implied by (23), any firms at that m would face excessively low turnover costs, and value labor on the margin strictly in excess of the hiring cost c. Such firms would seek to hire a mass of workers, inducing a mass point in the distribution of offers at that m—a contradiction. The density prescribed by Proposition 2 exactly balances these forces.

Finally, note that $\delta(m)$ is declining throughout the hiring region, and so the implied offer distribution $F(m)$ is rising in m. By Lemma 1, it follows that the quit rate in (23) is consistent with optimal worker turnover.

1.5 Aggregation and steady-state equilibrium

We now infer the implications of the preceding microeconomic structure for equilibrium labor market dynamics. An important first step toward this end is to aggregate individual firm and worker behavior for a given aggregate state. Given a solution to this aggregation problem, we then characterize conditions for steady-state labor market equilibrium.

To aid these steps, we restrict the drift of the stochastic process for idiosyncratic shocks in (1) to ensure that aggregate labor demand is stationary. This obtains when frictionless employment, which is proportional to $x^{1/(1-\alpha)}$, has no drift. Applying Ito's lemma, this requires that

$$\mu + \frac{1}{2}\frac{\alpha}{1-\alpha}\sigma^2 = 0. \tag{24}$$

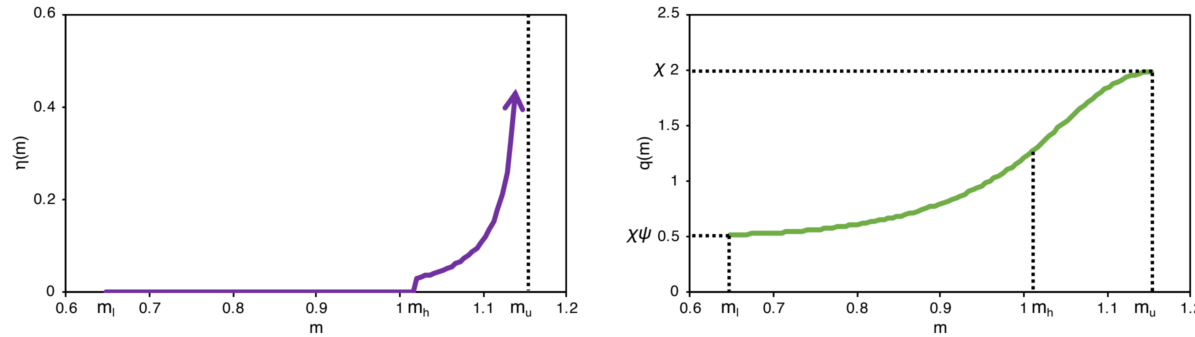
This assumption is made purely to simplify the analysis by abstracting from growth.[14]

---

[14] The condition in (24) ensures that the mean of the distribution of employment across firms is stationary. Analogous results hold for arbitrary μ with appropriate balanced-growth assumptions on c and ω_0.

Figure 2. Steady-state hiring and vacancy-filling rates

A. Hiring rate $\eta(m)$ B. Vacancy-filling rate $q(m)$



Notes. Parameter values are based on the model calibrated as described in section 3.

Aggregation. Steady-state aggregate labor market stocks and flows in the model are summarized by solutions for the separation rate into unemployment (denoted ς), the hiring rate (denoted η), and the density of employees g, at each marginal product m.

Proposition 3 *In steady state, (i) the separation rate into unemployment is given by*

$$\varsigma = \frac{\sigma^2/2}{1-\alpha} m_l g(m_l). \tag{25}$$

(ii) The hiring rate is given by

$$\eta(m) = -\frac{\sigma^2/2}{1-\alpha}\frac{m\delta'(m)}{\delta(m)}. \tag{26}$$

(iii) The vacancy-filling rate is given by

$$q(m) = \chi \exp\left[-\frac{1-\alpha}{\sigma^2/2}\int_m^{m_u}\frac{\delta(\widetilde{m})}{\widetilde{m}}\mathrm{d}\widetilde{m}\right], \tag{27}$$

with $\psi = exp\left[\frac{1-\alpha}{\sigma^2/2}\int_{m_l}^{m_u}\frac{\delta(\widetilde{m})}{\widetilde{m}}d\widetilde{m}\right]$. *Using (15), this yields the worker distribution* $G(m)$.

The most standard element of Proposition 3 is the solution for the separation rate into unemployment ς. All such separations arise at the lower boundary m_l. There, a density of $g(m_l)$ employees receives shocks to their log marginal product of variance σ^2.

Nonstationarity of the variance of employment can be remedied quite simply: Suppose that, for each m, firms exit at some rate and are replaced by an equal measure of firms with the mean employment of firms at m. Then all the results that follow will hold with a stationary distribution of employment across firms. In the Online Appendix, we also provide a model of firm exit that additionally accommodates firm growth.

Following negative shocks, employees are shed into unemployment until the marginal product is replenished, at a rate proportional to the elasticity of labor demand, $1/(1-\alpha)$.

The remaining results in Proposition 3 are novel features of this environment, however, and are illustrated in Figure 2. The economic content of these will be explained in more detail in the next section. For now, we simply note that the hiring rate $\eta(m)$, the worker distribution $G(m)$ and, thereby, the vacancy-filling rate $q(m)$ are inferred from the Fokker-Planck (Kolmogorov Forward) equation that describes the flow of workers across marginal products. Intuitively, $\delta(m)$ determines the flow of hires at each m (via $F(m)$), and the outflow of employees from each m. In this way it shapes the hiring rate $\eta(m)$, and the worker distribution $G(m)$. An important implication is that the solution for the equilibrium quit rate $\delta(m)$ in Proposition 2 is sufficient to solve for all of these aggregate outcomes, which in turn are functions solely of the marginal product m.

This completes the m-solution that we sought to derive. We now show how the elements of Proposition 3 can be used to characterize steady-state equilibrium.

Steady-state equilibrium. The matching structure implies that all outcomes of the model described thus far in Propositions 1, 2, and 3 depend on a single endogenous aggregate state, labor market tightness θ, via the contact rates $\lambda(\theta)$ and $\chi(\theta)$. Given this, we can characterize steady-state equilibrium in terms of two conditions reminiscent of the standard search model of Mortensen and Pissarides (1994).

Proposition 4 *There exists*[15] *a steady-state equilibrium in which unemployment U and labor market tightness θ satisfy (i) a Beveridge curve condition,*

$$U_{BC}(\theta) = \frac{\varsigma(\theta)}{\varsigma(\theta) + \lambda(\theta)} L; \tag{28}$$

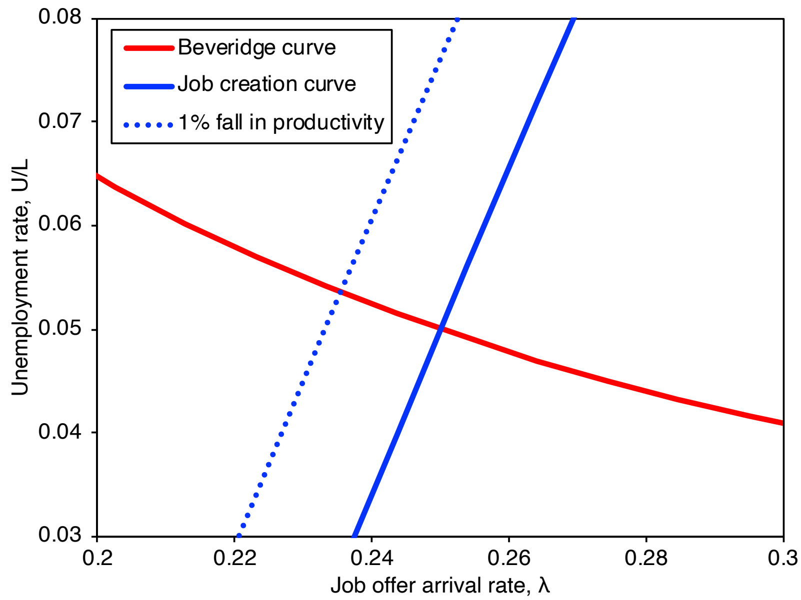
and (ii) a job creation condition,

$$U_{JC}(\theta) = L - \left[X \Big/ \int m^{\frac{1}{1-\alpha}} g(m;\theta) \mathrm{d}m \right], \tag{29}$$

where $X \equiv \mathbb{E}\left[(\alpha x)^{1/(1-\alpha)}\right]$.

[15] We have been unable to establish generally that (28) and (29) yield a unique steady-state equilibrium. The source of difficulty is that the comparative statics of the natural wastage region are analytically intractable. However, in numerical analyses we have confirmed that, for a wide range of parameters, the Beveridge curve condition in (28) slopes downward, while the job creation condition in (29) slopes upward.

Figure 3. Steady-state equilibrium and comparative statics



Notes. Based on the model calibrated as described in section 2. The figure illustrates the steady-state response to a one-percent decline in aggregate labor productivity.

The Beveridge curve condition emerges from the law of motion for unemployment. Making explicit the dependence of the separation rate on tightness, this reads

$$\frac{\mathrm{d}U}{\mathrm{d}t} = \varsigma(\theta)(L-U) - \lambda(\theta)U. \tag{30}$$

In steady state, unemployment is stationary, and we obtain the Beveridge curve (28).

The job creation condition is implied by aggregation of firms' labor demand. Aggregate employment is the mean of employment across firms, $N = \mathbb{E}\left[(\alpha x/m)^{1/(1-\alpha)}\right]$. Observing that the latter is equal to the ratio of the mean of $(\alpha x)^{1/(1-\alpha)}$ across firms and the *employment-weighted* mean of $m^{1/(1-\alpha)}$ gives rise to the job creation condition in (29).

This completes our constructive characterization of labor market equilibrium. Since it is founded on an m-solution for labor demand and turnover, we label it an m-*equilibrium*.

Definition *An* m-equilibrium *is a collection of optimal worker acceptance, and firm hiring and firing decisions* $\{a^*; dH^*, dS^*\}$*; worker and firm (marginal) values* $\{W, J\}$*; quit, layoff, hiring and vacancy-filling rates* $\{\delta, \varsigma, \eta, q\}$*; aggregate unemployment and job-finding rate* $\{U, \lambda\}$*; and offer and worker distributions* $\{F, G\}$ *such that (i) optimal worker and firm decisions, their associated (marginal) values, and the quit rate satisfy Lemma 1 and Propositions 1 and 2; (ii) layoff, hiring and vacancy-filling rates are given by Proposition*

3; (iii) aggregate unemployment and labor market tightness are given by Proposition 4; and (iv) the offer and worker distributions solve (15). Vacancies V, market tightness θ*, and the vacancy contact rate* χ *can then be inferred from the matching function and* λ*.*

Figure 3 illustrates the steady-state job creation and Beveridge curves, and depicts the upward shift of the job creation curve induced by a decline in aggregate labor productivity. Specifically, it plots the effect of modifying the production function to pxn^{α}, such that p falls by one percent. The positive slope of the job creation curve reflects the nature of search equilibrium in the model: Labor market tightness equilibrates the labor market via its effects on firms' turnover costs, in contrast to its effects on the cost of hiring, as in many conventional search models.

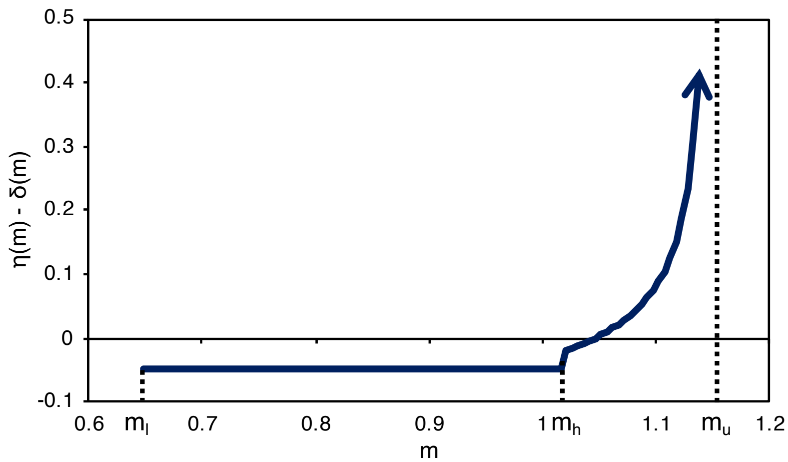
1.6 Equilibrium properties

The labor market equilibrium characterized in the preceding sections yields novel insights on labor market behavior. Here we highlight two distinctive features that emerge from the equilibrium: a novel form of imperfect labor market competition; and, relatedly, a novel form of misallocation. A set of instructive limiting economies underscores that the interaction of firm dynamics and on-the-job search is crucial to these new results.

Labor market competition. Proposition 3 reveals that, in steady state, the hiring rate $\eta(m)$ is proportional to minus the elasticity of the quit rate—or, equivalently, the hazard function of the offer distribution of log marginal products, $mf(m)/[1-F(m)]$. Intuitively, a firm's hiring rate is thus determined by the intensity of offers at m, $f(m)$, *relative to* the intensity of offers at *higher* ms, as captured by $1-F(m)$. Offers at *lower* ms are not directly relevant, since all such offers are dominated by those issued by firms at m.

In combination with Proposition 2, this result yields further intuitive insights. First, and most simply, since the quit rate is a constant (equal to $s\lambda$) in the natural wastage region, (26) confirms that the hiring rate is zero for $m < m_h$. Second, because $\delta(m)$ is strictly decreasing and concave in m in the hiring region, it follows that the hiring rate $\eta(m)$ is strictly positive and increasing in m for $m > m_h$. Third, since the quit rate equals zero at the upper boundary, the hiring rate asymptotes to infinity at m_u, as in Figure 2A. Consequently, firms' net employment growth, given by the difference between the hiring rate and the quit rate, $\eta(m) - \delta(m)$, takes the form depicted in Figure 4.

Figure 4. Steady-state net employment growth rate, $\eta(m)-\delta(m)$



Notes. Parameter values are based on the model calibrated as described in section 2.

These properties have an important bearing on labor market behavior. A key implication is that the marginal product m endogenously displays *gradual mean reversion* in the hiring region. The stochastic law of motion for m takes the form

$$\mathrm{d}m=\{\mu-(1-\alpha)[\eta(m)-\delta(m)]\}m\mathrm{d}t+\sigma m\mathrm{d}z. \tag{31}$$

In the hiring region, positive innovations to the marginal product m increase the hiring rate $\eta(m)$, and lower the quit rate $\delta(m)$, such that the firm accumulates more employees, as in Figure 4. The marginal product thus gradually reverts back down in expectation.

The presence of a region with gradual mean reversion in the marginal product is a novel manifestation of *imperfect labor market competition*. Its novelty lies in it being a distinctive consequence of the interaction of on-the-job search with firm dynamics. The following lemma formalizes this point by describing two limiting economies—those without on-the-job search ($s\to 0$), and without a notion of firm size ($\alpha\to 1$).

Lemma 2 *In the limits (a) as $s\to 0$, or (b) as $\alpha\to 1$ for fixed X and $\sigma^2/(1-\alpha)\equiv\tilde{\sigma}^2$, (i) the natural wastage region is bounded and nondegenerate, $0<m_l<m_h<\infty$; (ii) the hiring region is degenerate, $m_u=m_h$; and (iii) for all $m\in(m_l,m_h)$, the worker distribution simplifies to*

$$G(m) \to \begin{cases} \dfrac{\ln(m/m_l)}{\ln(m_h/m_l)} & \text{as } s \to 0, \\ \dfrac{(m/m_l)^{2s\lambda/\tilde{\sigma}^2} - 1}{(m_h/m_l)^{2s\lambda/\tilde{\sigma}^2} - 1} & \text{as } \alpha \to 1. \end{cases} \tag{32}$$

Eliminating either on-the-job search ($s \to 0$), or firm dynamics ($\alpha \to 1$), implies that the hiring region collapses to a point, $m_u \to m_h$. In these limits, deviations from competitive labor market outcomes take conventional forms. The $s \to 0$ limit mirrors models of firm dynamics in the tradition of Bentolila and Bertola (1990). There, the hiring boundary m_h becomes a reflecting barrier, and the presence of idiosyncratic shocks and hiring costs gives rise to dispersion in marginal products. The $\alpha \to 1$ case holds fixed X to ensure that aggregate job creation in (29) remains bounded in the limit, and $\sigma^2/(1-\alpha) \equiv \tilde{\sigma}^2$ to ensure that separations into unemployment (25) remain strictly positive and bounded in the limit. This case resembles a standard search and matching model, extended to accommodate on-the-job search and endogenous job destruction (e.g. Pissarides 2000, Chapter 4). The presence of idiosyncratic shocks, and *ex post* bargaining with *ex ante* investments, gives rise to productivity dispersion and, via rent sharing, wage dispersion. In both limits, gradual mean reversion in the marginal product vanishes, and the stochastic law of motion (31) becomes a geometric Brownian motion,

$$\mathrm{d}m = [\mu + (1-\alpha)s\lambda]m\mathrm{d}t + \sigma m\mathrm{d}z. \tag{33}$$

The reflecting barriers $m_l < m_h$ imply a stationary density that obeys a power law, (32).

Lemma 2 thus underscores that the novelty of the model's deviation from competitive outcomes lies in the presence of gradual mean reversion in marginal products in a nondegenerate hiring region, and the essential role of the interaction of on-the-job search and firm dynamics in generating these. Intuitively, the hiring region emerges as firms seek to manage their turnover costs by managing their position in the hierarchy of marginal products. Recalling the discussion of Proposition 2, firms shade the intensity of their hiring to allow their marginal products to rise and, thereby, reduce turnover. Absent on-the-job search ($s \to 0$), firms have no turnover costs to manage. Absent firm dynamics ($\alpha \to 1$), firms are unable to influence their marginal product by adjusting their hiring behavior.

This new manifestation of imperfect labor market competition has an intuitive appeal. Perfect competition would induce *infinite* mean reversion in marginal products such that the law of one wage (and marginal product) is maintained. Instead, as firms shade their hiring decisions to manage turnover, mean reversion in marginal products weakens, and additional dispersion in marginal products emerges among hiring firms. Gradual mean reversion and a nondegenerate hiring region are thus two sides of the same coin.

These forces in turn shape the steady-state distribution of employees $G(m)$. Lemma 2 is echoed in the natural wastage region, where constancy of the quit rate implies that the marginal product m evolves according to (33), so that the worker distribution obeys a power law for $m \in (m_l, m_h)$. But, in contrast to Lemma 2, mean reversion in the hiring region thins the tail of the steady-state worker distribution for $m \in (m_h, m_u)$. Formally, because the quit rate is strictly declining in m in the hiring region, the vacancy-filling rate $q(m)$ in (27), and thereby the worker distribution $G(m)$, rise ever more slowly in m relative to the power law in the natural wastage region. Because the quit rate is zero (and the hiring rate explodes) at m_u, mean reversion becomes so extreme that the stationary density of employees converges to zero at the upper boundary, as in Figure 2B.

On-the-job search and misallocation. A key feature of the aggregation results in Proposition 3 is the presence of dispersion in marginal products across workers, as summarized by $G(m)$, and thereby the presence of *misallocation.* A natural intuition suggests that on-the-job search might alleviate such misallocation, by allowing employees to transition faster to more productive jobs. Paradoxically, the preceding model cautions against this intuition. The following Lemma provides a stark example of this paradox.

Lemma 3 *Suppose there are no idiosyncratic shocks,* $\mu = \sigma = 0$*, separations into unemployment occur at exogenous rate* ς_0*, and workers quit with strictly positive probability when indifferent.*[16] *Then, (i) the hiring region and quit rate in Proposition 2 hold* mutatis mutandis *with* r *exchanged with* $r + \varsigma_0$*; (ii) the boundary* m_h *is such that* $(1 - \omega_1)m_h - \omega_0 = (r + \varsigma_0 + s\lambda)c$*; (iii) the natural wastage region is never entered; (iv) hires replace quits,* $\eta(m) = \varsigma_0 + \delta(m)$*; and (v) the worker distribution takes the form*

$$G(m) = \frac{\varsigma_0 F(m)}{\varsigma_0 + s\lambda[1 - F(m)]}. \tag{34}$$

Lemma 3 reveals that the hiring region induced by the interaction of firm dynamics and on-the-job search is present even in the absence of idiosyncratic shocks and endogenous job destruction. Importantly, this hiring region, and the accompanying

[16] This tie-breaking condition is needed only in the case of no idiosyncratic shocks to rule out equilibria with mass points in the offer distribution; see Shimer (2006), and Gottfries (2019).

dispersion in marginal products, would not emerge in the absence of on-the-job search ($s = 0$). A striking implication, then, is that on-the-job search in fact *gives rise to all* equilibrium misallocation in this case. We argue in what follows that Lemma 3 presents a stark point of contrast to existing canonical models of on-the-job search.[17]

On one hand, the hiring region shares interesting parallels with a large literature inspired by Burdett and Mortensen (1998). This emphasizes how *ex ante* wage posting and firms' turnover concerns generate "residual" *wage* dispersion among identical workers. By contrast, in our model *ex post* wage bargaining and firms' turnover concerns give rise instead to "residual" dispersion in *marginal products*, and thereby in wages. Both results can be traced to notions of imperfect labor market competition associated with on-the-job search and labor market frictions, as well as to the nature of wage setting. In Burdett and Mortensen (1998), wage dispersion arises from firms having to commit *ex ante* to wage payments that cannot respond to workers' current or future outside options. In the present model, marginal product dispersion arises from firms being unable to commit to wages, and managing turnover instead via hiring decisions. For the special case of no idiosyncratic shocks in Lemma 3, the worker distribution of *marginal products* in (34) becomes exactly analogous to the worker distribution of *wages* in Burdett and Mortensen (1998), yielding a stable job ladder whereby workers move towards higher-wage, more productive firms.

On the other hand, a crucial message of Lemma 3 is that these models have fundamentally different implications for misallocation. Wage posting models in the mold of Burdett and Mortensen (1998) invoke linear technologies. When extended to incorporate productive heterogeneity (Bontemps et al. 2000), an extreme implication is that allocative efficiency requires all workers to be employed in the most-productive firm. On-the-job search is thus a force toward *resolution* of misallocation in these models, since it accelerates worker transitions toward more productive firms.

The paradox of Lemma 3 is that this last implication is turned on its head. In Lemma 3, heterogeneity in marginal products emerges as an *equilibrium outcome*, rather than by assumption. And the presence of on-the-job search is the primitive force that *gives rise to* equilibrium misallocation, rather than solely being an equilibrium response to it.

---

[17] The reader may worry that the special case in Lemma 3 has the seemingly pathological implication that firms with higher marginal products have lower employment, even though they face lower quit rates. Item (iv) of Lemma 3 explains: higher-m firms also have lower *hiring* rates. Exogenous separations ς_0, and the absence of shocks, implies that firms hire only to replace quits in this case; thus, the hiring rate falls in m.

The key difference is the presence of diminishing returns. This provides an economic margin by which differences in firm marginal productivity can be resolved. Indeed, in the absence of on-the-job search, marginal products are equalized: $s \to 0$ implies $m_u \to m_h$. Instead, in the presence of on-the-job search, firms allow their marginal products to vary as a means to manage turnover, generating equilibrium misallocation. Thus, the interaction of frictions with neoclassical forces that militate toward equality of marginal products fundamentally alters the economic role of on-the-job search in misallocation.

Lemma 3 makes this point starkly, by abstracting from idiosyncratic shocks and endogenous job destruction. Returning to the general case, though, Proposition 2 and Lemma 2 imply that on-the-job search will at least give rise to greater misallocation *among hiring firms*. It follows that there must be configurations of the parameters of the model such that this effect dominates, and on-the-job search can raise misallocation *overall*.

2. Quantitative exploration

The model of the preceding sections has rich implications for both the cross sectional and time series behavior of labor markets. We now explore these implications quantitatively and confront them with relevant stylized facts that have emerged in recent literature.

An overview of our approach is provided in the tables and figures that follow. Table 1 summarizes our calibration strategy and its implied parameters, expressed at a monthly frequency. Table 2, together with Figures 5 and 6, then report the calibrated model's implications for a wide range of nontargeted moments, and contrast them with empirical analogues. These are split between cross-sectional implications for measures of imperfect labor market competition and establishment dynamics, and macroeconomic implications for aggregate labor market dynamics. We now describe our approach in detail.

2.1 Calibration

We begin with a normalization. Note that, in the limit in which the hiring cost c is zero, optimal labor demand implies that marginal products are equalized across firms at a level $m^* \equiv \omega_0/(1-\omega_1)$. It follows from (6) that there is a common wage in this case equal to $w^* \equiv \omega_0/(1-\beta)$. We normalize $w^* \equiv 1$ or, equivalently, $\omega_0 \equiv 1-\beta$. It follows that all flow parameters are expressed in terms of monthly frictionless wages.

The discount rate r, the curvature of the production function α, and the labor force L are calibrated externally to replicate, respectively, an annual real interest rate of 5 percent, the estimates of Cooper, Haltiwanger and Willis (2007, 2015), and an average firm size of 20 employees, consistent with data from the Small Business Administration.

The remaining parameters are then calibrated internally. Although, of course, all target moments inform all parameters, in what follows we provide an account of the empirical moments that intuitively are most relevant for the calibration of each parameter.

A central ingredient to the frictions in the model is the hiring cost c. In his original study, Oi (1962) reported two early empirical results. First, the majority of hiring costs pertain to training, rather than to recruiting. This provides some empirical justification for our choice to model hiring rather than vacancy costs. Second, hiring costs correspond to approximately one month's wages.[18] In his survey, Manning (2011) notes that, although evidence on the magnitude of hiring costs remains limited, Oi's initial estimates broadly are borne out in subsequent work. More recently, Gavazza et al. (2018) report estimates of hiring costs compiled by human resources professionals that reinforce this conclusion. Accordingly, we target a hiring cost equal to the average monthly wage.

Hiring costs interact with the presence of idiosyncratic shocks x to determine the size of the natural wastage region. For a given hiring cost c, a higher standard deviation of idiosyncratic shocks σ implies greater dispersion in innovations to firms' desired labor demand, and a smaller measure of employment in firms with zero hires. We thus discipline σ by targeting the share of employment at establishments with zero hires over a month. Using microdata from the Job Openings and Labor Turnover Survey (JOLTS), Davis et al. (2013) estimate this share at 34.8 percent.

Next, we target a set of moments relating to labor market stocks and flows. To do so, we adopt a conventional Cobb-Douglas matching function,

$$M = A[U + s(L - U)]^{\epsilon} V^{1-\epsilon}, \tag{35}$$

where A denotes match efficiency. In addition, we augment the model of the preceding sections to incorporate a portion of separations into unemployment that are exogenous. Specifically, we allow such exogenous separations to occur at rate ς_0.[19]

[18] Oi's Table 1 reports his cost-per-hire estimate for all employees of \$381.73, of which training costs comprise \$281.70. He further reports average hourly earnings equal to \$1.952. Assuming a workweek of 40 hours, and 52/12 weeks in a month, implies that total hiring costs correspond to 1.13 months' pay.

[19] The Appendix presents an extension of the analytical solutions in Proposition 3 for this case.

Table 1. Parameters and targeted moments of calibrated model (monthly frequency)

Parameter		Value	Reason / Moment	Model	Target
A. Externally calibrated					
ω_0	Flow breakdown payoff	0.488	Normalization	—	—
r	Discount rate	0.004	Annual real interest rate	0.05	0.05
α	Returns to scale	0.64	Cooper et al. (2007, 2015)	—	—
L	Labor force	21.05	Average firm size	20	20
B. Internally calibrated					
c	Per-worker hiring cost	1.107	Hiring costs / Monthly pay	1	1
σ	Std. dev. x shocks	0.066	Empl. at estabs. with no hires	0.348	0.348
X	Job creation curve shifter	19.68	U-to-E rate	0.25	0.25
ς_0	Exogenous separation rate	0.012	Unemployment rate	0.05	0.05
A	Matching efficiency	1.111	Vacancy rate	0.025	0.025
ϵ	Matching elasticity	0.285	Beveridge curve elasticity	-1	-1
s	Employed search intensity	0.148	E-to-E rate	0.032	0.032
β	Worker bargaining power	0.512	Avg. job-to-job wage gain	0.08	0.08

Notes. The rationale and source for each targeted moment are explained in detail in the main text.

Unemployment stocks and flows are then targeted as follows. Note that any steady-state level of labor market tightness θ^* and, thereby, of the job-finding rate of unemployed searchers $\lambda(\theta^*)$, can be supported by an appropriate choice of the job creation curve shifter X in (29). We choose X to replicate a monthly unemployment-to-employment transition rate of 0.25, consistent with Current Population Survey "gross flows" data.[20] Given this unemployment outflow rate, we choose the exogenous separation rate ς_0 to replicate a steady-state unemployment rate of 5 percent.

As is standard in search and matching models, matching efficiency A, and the matching elasticity ϵ, then determine respectively the steady-state level of vacancies, and the slope of the Beveridge curve relation between vacancies and unemployment. We choose A to target a steady-state vacancy rate of 2.5 percent (Davis et al. 2013), and ϵ to target a Beveridge curve elasticity of -1 (Shimer 2005).

[20] See https://www.bls.gov/cps/cps_flows.htm.

Finally, we use the remaining parameters of the model to target empirical moments relating to the role of on-the-job search in the labor market. The search intensity of employed searchers s naturally shapes the magnitude of direct transitions from one employer to another. We therefore choose s such that the model replicates a monthly job-to-job transition rate of 0.032, as in Moscarini and Thomsson (2007). It remains to determine worker bargaining power β, which we choose to replicate an average wage gain associated with job-to-job transitions of 8 percent, based on Barlevy (2008). Intuitively, job-to-job transitions in the model involve workers moving up a hierarchy of marginal products. Recalling the wage solution (6), greater bargaining power raises the gradient of wage increases as workers move up this hierarchy through on-the-job search.

2.2 Cross-sectional implications

Given this calibration, the sections that follow explore the model's implications for empirical moments that were not targeted. Recall that the model has implications both for the cross-sectional behavior of the labor market, as well as for its aggregate dynamics. We begin in the present subsection by exploring the former—in particular, the model's implications for conventional empirical diagnostics of labor market competition, and for modern empirical findings on establishment dynamics.

Imperfect labor market competition. In his survey, Manning (2011) highlights a set of quintessential symptoms of imperfect competition in labor markets, and reviews their empirical relevance. Here, we describe how the model is able to accommodate these, and further confront the implications of the above calibration with available estimates.

First, the wage solution in (6) displays a close resemblance to estimating equations used in an empirical rent-sharing literature that dates back to the early work of Abowd and Lemieux (1993). Most recently, Kline et al. (2019) refine and extend that literature by estimating the effects of shocks to labor productivity induced by plausibly-exogenous shocks to patent approval. They present two sets of estimates that are most straightforward to compare to model outcomes. First, "pass-through" measures of the change in wages induced by a unit rise in value-added per worker. Second, "elasticity" estimates that multiply pass-through by the ratio of average value-added per worker to average wages. Their Table VIII suggests ranges for pass-through in region of 0.2 to 0.4, and for elasticities in the region of 0.4 to 0.7.[21]

Panel A of Table 2 presents the results of performing analogous regressions in data generated by the calibrated model. Note that nothing in the calibration procedure summarized in Table 1 assures that the model can match estimated measures of rent sharing. The targeted moment closest in spirit is the average wage gain from job-to-job transitions. But note that, in the model, the latter measures the wage returns to productivity changes *across* jobs, as opposed to rent sharing *within* jobs.

Table 2A suggests, however, that the model does a good job of replicating recent rent-sharing estimates. Its pass-through measure of 0.40 is at the upper end of the range reported by Kline et al.; its elasticity of 0.56 closer to the middle of their range. Viewed through the lens of the model, plausible levels of wage gains from job-to-job transitions thus dovetail well with plausible degrees of rent sharing.[22]

A second symptom of imperfect labor market competition accommodated by the model is the notion that the rate of turnover faced by a firm may be negatively associated with the wage it pays. In standard models of dynamic monopsony, such considerations play a central role by shaping the elasticity of labor supply to the firm. The model of the preceding sections provides a novel perspective on this, however. There, higher wages are associated with higher marginal products which, in turn, are associated with lower quit rates, as in Proposition 2. It follows that the model can speak to empirical estimates of the relationship between separations and wages.

Surveying the literature up to 2010, Manning (2011) reports a wide range of estimates of the wage elasticity of separations, but notes "perhaps a suggestion that those studies which have higher quality information [...] find elasticities in the region 1.5-2." This tentative conclusion has since been reinforced by Kline et al. (2019), who further exploit the wage effects of shocks to patent approval to identify the wage elasticity of separations.

[21] Kline et al.'s (2019) estimates align with the early studies of Abowd and Lemieux (1993) and Van Reenen (1996), and with the structural estimates based on a related model in Bagger et al. (2014). However, they are larger than many of the estimates surveyed by Card et al. (2018), particularly those based on the association between changes in firms' productivity and the wage growth of their workers. Kline et al.'s estimates suggest that this may be due to a failure to instrument for firm productivity in the latter studies.

[22] The model is unable to replicate richer aspects of Kline et al.'s estimates, however—greater rent-sharing for employees in the upper part of the within-firm earnings distribution; little rent sharing among new hires. In part, this is because the model abstracts from worker heterogeneity. But it may also reflect a deeper mismatch with the contract structure—for example, firms' inability to commit to future wages in the model.

Table 2. Nontargeted moments

Moment			**Model**	**Data**
A. Imperfect labor market competition				
Rent-sharing measures	Pass-through		0.40	0.2 to 0.4
	Elasticity		0.56	0.4 to 0.7
Wage elasticity of quits			-1.55	-1.5 to -2.0
Frictional log wage dispersion	Std. dev.		0.06	—
	90-10 differential		0.15	—
Mean-min wage ratio			1.24	—
Var. of log wages in natural wastage region / Total var.			0.75	—
B. Establishment dynamics				
		Raw v	Adj. $\tilde{v}$	
Employment at estabs. with no vacancies		0.54	0.45	0.45
Vacancies at t at estabs. with no vacancies at $t-1$		0.17	0.18	0.18
Hires in $(t-1,t)$ at estabs. with no vacancies at $t-1$		0.09	0.37	0.42
Hires rate elasticity of daily vacancy-filling rate		0.25	0.74	0.82
Std. dev. employment growth	Monthly		0.08	0.08
	Quarterly		0.19	0.29
	Annual		0.49	0.39
C. Aggregate dynamics				
Response relative to output per worker			|Elasticity|	Relative sd.
Unemployment rate			7.6	14.0
Vacancy rate			7.6	12.6
U-to-E rate			6.7	11.6
E-to-U rate			1.3	3.6
E-to-E rate			5.5	5.7
Response relative to unemployment rate			Semi-elasticity	Semi-elasticity
Average wage			-1.4	≈ -1

Notes. Data ranges are from the following sources. Panel A: Manning (2011); Kline et al. (2019). Panel B: Davis et al. (2012, 2013); Haltiwanger et al. (2013). Panel C: an update and extension of Shimer (2005) for labor market stocks and flows; a summary of Solon et al. (1994) and Elsby et al. (2016) for wages. "Adjusted $\tilde{v}$" model outcomes in Panel B are based on a model of mismeasurement of vacancies reported in the text. Model outcomes in Panel C are steady state (semi-)elasticities. Further detail is provided in the main text.

Their Table IX reports a full-sample elasticity equal to -1.62, and estimates for subsamples broadly in the range suggested by Manning.

Table 2A reports the results of an analogous exercise using model-generated data.[23] Strikingly, the model yields a wage elasticity of quits equal to -1.55, very much in the neighborhood of empirical estimates. In the model, the magnitude of this elasticity is shaped by the association between the quit rate and productivity, and the pass-through from productivity to wages. As already discussed, the model does a good job of matching the latter. It follows that the model's ability to replicate a plausible wage elasticity of separations further suggests that the association between quits and productivity also is empirically reasonable. Since the latter is a defining implication of the model, this is an especially reassuring quantitative outcome.

A success of the model, then, is that it is quantitatively consistent with key indicators of imperfect labor market competition (Table 2A), and can reconcile these in a parsimonious framework with more-conventional stylized facts of establishment dynamics, labor market stocks and flows, and job-to-job flows (Table 1). Manning (2011) highlights the outcomes in Table 2A, together with the magnitude of the hiring cost, as central moments for the theory of imperfect competition in the labor market. We are not aware of any previous model that has been able to match these moments jointly, and so we see this as a useful contribution of the model.

The remaining rows of Table 2A document the model's implications for "residual" wage dispersion—differences in wages among identical workers. Reliable measures of the latter are notoriously hard to estimate empirically, but we report model-implied measures for reference. Overall, the model implies a degree of residual wage dispersion that is nontrivial, but also not enormous. The standard deviation of log wages across workers in the model is 0.06, with a 90-10 percentile differential of 15 log points. Likewise, the mean-min wage ratio statistic proposed by Hornstein et al. (2011) is equal to 1.24, similar to their calculations for conventional on-the-job search models without firm dynamics.

But recall that the addition of firm dynamics to the model offers a novel perspective: Frictional wage dispersion instead becomes a symptom of misallocation of workers across firms; and, relatedly, in the hiring region this is aggravated, not resolved, by presence of on-the-job search. To provide a sense of this, the final row of Table 2A reports the share of the variance of log wages that would emerge in the absence of a hiring region—as in a

[23] These results are based on a simulation of 2,000,000 firms over a year.

standard firm dynamics model with exogenous quits (at rate $\varsigma_0 + s\lambda$).[24] This is 75 percent in the calibrated model. One interpretation, then, is that a quarter of the overall variance of log wages is accounted for by the presence of a hiring region in the calibrated model.

Establishment dynamics. In a pair of influential papers, Davis, Faberman and Haltiwanger (2012, 2013) document a set of stylized facts on the relationships between gross worker flows and job flows at the establishment level. They highlight two stark empirical deviations from the "iron link" between employment growth and gross hires and layoffs predicted by standard firm dynamics models. First, quits vary negatively with establishment growth, driving a wedge between job flows and gross worker flows (Davis et al. 2012). Second, vacancy yields vary positively with establishment growth, driving a wedge between gross hires and vacancies (Davis et al. 2013).

A novel implication of the model of the preceding sections is that it is naturally able to accommodate Davis et al.'s stylized facts, and in particular those that deviate from an iron link between worker and job flows. The key observation is that the marginal product m is a sufficient statistic for a firm's net employment growth $\eta(m) - \delta(m)$: Higher marginal products are associated with faster firm growth, as in Figure 4. It is then immediate from Propositions 2 and 3 that expanding firms will face lower quit rates, and larger hiring and vacancy-filling rates, as observed by Davis et al.

To illustrate this, Figure 5 and Table 2B report the results of applying the methods of Davis et al. to data simulated from the model calibrated as in Table 1. Since, as Davis et al. note, the concept of a vacancy is inherently more subjective than hires and separations, we explore two interpretations of vacancies in the model: first, treating model vacancies as one-to-one with empirical vacancies ("raw vacancies"); second, allowing for a simple model of mismeasurement of vacancies ("adjusted vacancies") described below. Mirroring the data, both interpretations of vacancies in the model are measured at a point in time; hires, layoffs and quits are cumulated over the subsequent month.[25]

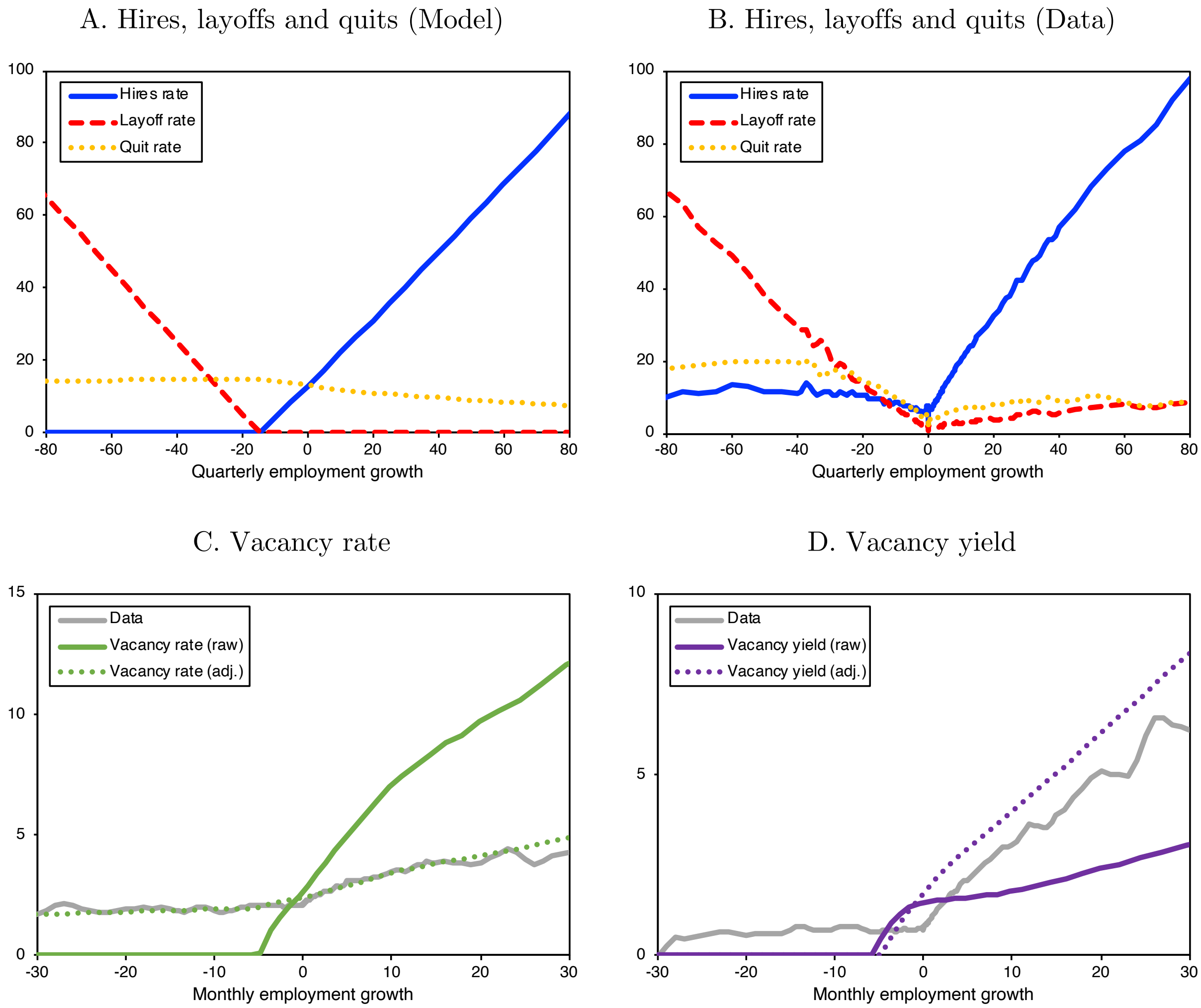
Figure 5 reveals that model outcomes qualitatively resemble those documented by Davis et al. A contribution of the model is that it provides a parsimonious account of *both*

---

[24] Interestingly, we will see later in section 4 that this also corresponds to model outcomes in an extension in which firms are perfectly able to match the outside offers of all of their employees. There we further show, however, that less-than-perfect offer matching will restore the presence of a hiring region.

[25] As is common in the literature, firms in the model are treated as analogous to establishments in the data.

Figure 5. Gross worker and job flows implied by the model



Notes. Application of Davis et al. and JOLTS methodologies to model-generated data simulated over one quarter in Panels A and B, and one month in Panels C and D. Data in Panel B are from Davis et al. (2012). Data in Panels C and D are estimates from Davis et al. (2013) with controls for establishment fixed effects. "Adjusted" model-generated data are based on a model of mismeasurement of vacancies reported in the text.

of Davis et al.'s documented deviations from an iron link between worker and job flows, generating a quit rate that declines, and a vacancy yield that rises, in firm growth.

By contrast, recent work has sought to explain subsets of the same data in isolation. Kaas and Kircher (2015) explain the behavior of vacancy yields by invoking convex vacancy costs and directed search. There, vacancies and wages are imperfect substitutes in recruiting, and growing firms use increased wage offers to attract workers. In a random search environment, Gavazza et al. (2018) explain the same pattern by invoking convex costs of recruiting effort, so that vacancies and recruiting effort are imperfect substitutes.

While these models can break an iron link between gross hires and vacancies, both abstract from on-the-job search, and so cannot address the decline in quits with firm growth. Conversely, the latter is addressed by Schaal's (2017) model of firm dynamics with on-the-job search. But, there, directed search and linear vacancy costs imply an indeterminate relationship between vacancy-filling rates and firm growth among hiring firms.

Table 2B then confronts the model with an array of additional nontargeted features of the link between worker and job flows emphasized by Davis et al. Consider first the results for raw vacancies in the model. Recall that the calibration targets the fraction of employment in firms with zero monthly hires. Table 2B reveals that the model also does a good job of matching "instantaneous" measures of the desire to hire across firms, and its persistence across time: Both the share of employment at firms with zero vacancies at a point in time (0.54), and the share of vacancies held at firms without any vacancies one month ago (0.17), are close to their empirical analogues (respectively, 0.45 and 0.18).

But the raw-vacancy outcomes in Figure 5 and Table 2B also differ from the empirical results of Davis et al. (2013) in a few, related dimensions. In Figure 5, the rise in the vacancy yield with firm growth is around half as steep as its empirical analogue; and the rise in the vacancy rate with firm growth is about twice as steep as in the data. In Table 2B, the model-implied share of hires at establishments with no vacancies at the end of the prior month is one-quarter of its empirical analogue. And, echoing the discrepancy in vacancy yields in Figure 5D, the model implies an elasticity of the daily vacancy-filling rate with respect to the hires rate that is one-third of its counterpart in the data.[26]

These discrepancies share a common source: In the model, all vacancies are posted by hiring firms that, in turn, are unlikely to shrink substantially. In the data, however, a nontrivial fraction of aggregate vacancies is accounted for by establishments that are shrinking, often at substantial rates (see Figure 5C). As we noted above, there is a compelling case to explore the role of measurement errors in vacancy data, as vacancies are inherently more subjective than hires and separations.

To illustrate this point, we briefly study the implications of a simple form of mismeasurement of vacancies in the model. Specifically, we allow the measured vacancy rate in firm i at time t, $\tilde{v}_{it}$, to be related to the actual vacancy rate, v_{it}, as follows

[26] We infer a model analogue to Davis et al.'s (2013) measure of daily vacancy-filling rates by applying their adjustment for time aggregation to data generated by the model. Although our model implies a different structure of time aggregation, we apply Davis et al.'s adjustment to be comparable with their results.

$$\tilde{v}_{it} = \max\{\kappa v_{it} + \varepsilon_{it}, 0\}\text{, where } \varepsilon_{it} = \rho_v \varepsilon_{it-1} + \iota_{it}\text{, and } \iota_{it} \sim \mathcal{N}(0, \sigma_v^2). \tag{36}$$

Firms in the model thus make two errors in their reporting of vacancies: First, errors in units, captured by the scaling parameter κ; and, second, errors in vacancy reports for a given understanding of units, ε_{it}, that are allowed to be persistent within firm over time. The measured vacancy rate $\tilde{v}_{it}$ is then reported subject to a nonnegativity constraint.

We set the scale parameter κ to replicate an aggregate vacancy rate of 2.5 percent (as before), and the persistence of individual firm errors ρ_v to match the empirical share of vacancies at establishments with no previous vacancy of 18 percent (almost as before). Crucially, the dispersion of firm errors σ_v is set to target the empirical vacancy rate of 1.7 percent among establishments with monthly employment growth of -30 percent.[27] The "adjusted" vacancy entries in Figure 5 and Table 2B then report the results of reapplying the methods of Davis et al. to data on $\tilde{v}$ from the model.

This simple adjustment aligns several nontargeted model outcomes even closer to their empirical counterparts. The behavior of the vacancy rate in model and data in Panel C of Figure 5 is essentially resolved. In turn, the model-implied gradient of the vacancy yield in firm growth is much closer to its empirical analogue. Likewise, in Table 2B, the incidence of hires without a prior vacancy, as well as the hires rate elasticity of the vacancy-filling rate, rise in line with the data. And the share of employment at firms with zero vacancies falls to replicate the data exactly. Of course, this exercise does not establish definitively that such measurement error in vacancies is present; only that it is one plausible and parsimonious resolution of model and data.

Finally, the remaining rows of Table 2B document the model's implications for a more conventional moment of establishment dynamics: the cross-sectional dispersion of employment growth. The latter provides a natural check on the plausibility of the calibration in Table 1. Despite not having been targeted, Table 2B suggests that the model implies a reasonable standard deviation of employment growth. It exactly replicates the monthly estimate based on JOLTS data (Davis et al. 2013), and is in the ballpark of quarterly estimates using Business Employment Dynamics data (Davis et al. 2012), and annual estimates using the Longitudinal Business Database (Haltiwanger et al. 2013). Furthermore, in the Appendix, we study a reinterpretation of the calibrated model that accommodates exogenous firm exit and entry. Under that interpretation, the model also

[27] This yields κ=0.28, ρ_v=0.604, and σ_v=0.041. As anticipated, significant measurement error is necessary to reconcile a vacancy rate in shrinking establishments of 1.7 percent with an aggregate rate of 2.5 percent.

replicates standard features of empirical firm dynamics—the Pareto shape of the firm-size distribution—and broadly matches rates of hiring and separation by firm size and age—most notably the elevated hiring rate of young firms.

3. Aggregate dynamics

The analysis thus far has addressed the first analytical challenge posed by the interaction of firm dynamics with on-the-job search—namely that of inferring the steady-state offer and worker distributions, $F(m)$ and $G(m)$. We begin this section by showing how these results also inform the solution to the second analytical challenge—inferring the out-of-steady-state dynamics of the distributions. We then use this approach to solve for the transition path of model outcomes, including the distributions $F(m)$ and $G(m)$, following an MIT shock to aggregate productivity. The section concludes with a comparison of the amplitudes of labor market stocks and flows, and wages, with empirical counterparts.

3.1 Solution method

The key insight of our approach is that the form of the quit rate in Proposition 2 also will hold out of steady state, subject to the modification that the middle boundary and the job offer arrival rate will vary over time, m_{ht} and λ_t. The intuition is simple. Out of steady state dynamics give rise to additional capital gains in the firm's marginal value relative to its steady-state form in (16),

$$\begin{aligned} rJ_t(m) &= (1-\omega_1)m - \omega_0 - [\delta_t(m) - (1-\alpha)m\delta_t'(m)]J_t(m) \\ &\quad + [\mu + (1-\alpha)\delta_t(m)]mJ_t'(m) + \frac{1}{2}\sigma^2 m^2 J_t''(m) + \frac{\partial J_t(m)}{\partial t}. \end{aligned} \tag{37}$$

Optimality in the hiring region, however, requires that the firm's marginal value of labor is a constant, equal to the marginal hiring cost, $J_t(m) = c$ for all $m \in (m_{ht}, m_{ut})$. As before, this implies that $J_t'(m) = 0 = J_t''(m)$ in the hiring region. But, crucially, it also implies that any such out-of-steady-state capital gains are zero in the hiring region, $\partial J_t(m)/\partial t = 0$. Thus, the quit rate shares the same functional form as in Proposition 2. This is a considerable simplification, as the solution for the dynamic path of the quit rate—or, equivalently, the offer distribution $F_t(m)$—is thus known up to the path of two scalars, m_{ht} and λ_t, a much simpler prospect.

This in turn aids the solution for the time path of the worker distribution. Just as the quit rate informs the steady-state vacancy-filling rate in (27), and thereby the steady-state worker distribution, its time path induces the dynamics of $G_t(m)$ via the out-of-steady-state Fokker-Planck (Kolmogorov Forward) Equation. Thus, the dynamic path of the worker distribution $G_t(m)$ is known up to the path of three scalars m_{lt}, m_{ht} and λ_t.

Finally, consider the natural wastage region. Here, the quit rate is maximal and equal to $s\lambda_t$. The job offer arrival rate λ_t is thus the sole aggregate state in this region. Given a time path for λ_t, the firm's marginal value $J_t(m)$, and the boundaries m_{lt} and m_{ht}, can then be inferred out of steady state. This implies a further simplification: the path of λ_t is also sufficient to determine the paths of m_{lt} and m_{ht}.

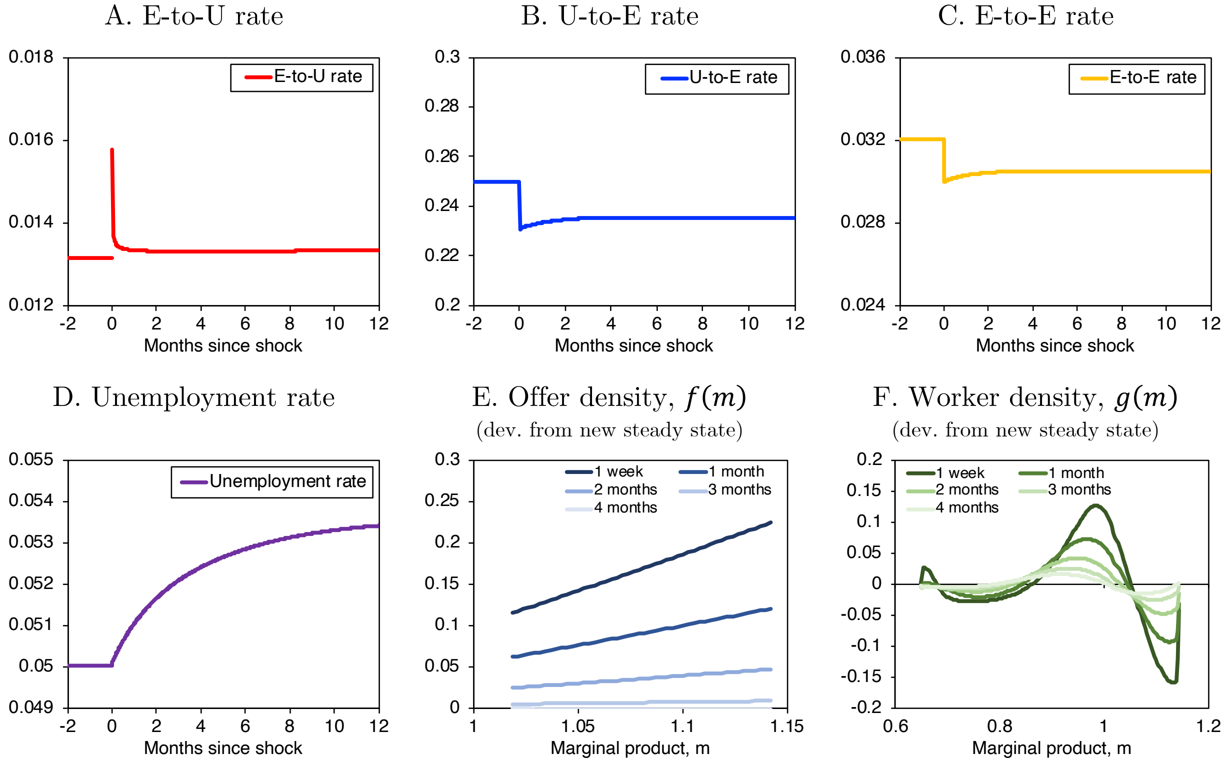
The upshot is that the dimensionality of the problem of inferring the model's transition dynamics is greatly reduced by the analytical results developed earlier in the paper. Absent these results, solving the model out of steady state would involve forecasts of the *unknown functions* $\delta_t(m)$ and $q_t(m)$. With these results, we can distil the problem to one which requires a forecast of the dynamic path of just one scalar, λ_t.

3.2 Transition dynamics and quantitative assessment

We now demonstrate the feasibility of the latter recipe. Given the calibration described in the previous section, we solve for the transition path of model outcomes in response to a permanent unanticipated decline in aggregate labor productivity p (as in Boppart, Krusell, and Mitman 2018), recalling that firm output in is given by $y = pxn^{\alpha}$. Our analytical results provide us with solutions for the initial and final steady states. Given a (conjectured) path for λ_t, we first solve backwards from the final steady state for the implied sequence of firm marginal value functions (in the natural wastage region). This implies sequences for the boundaries m_{lt} and m_{ht}, and thereby for the quit rate $\delta_t(m)$. Given these, we then use the Fokker Planck (Kolmogorov Forward) Equation to solve forward for the implied sequence of distributions of marginal products across employees $G_t(m)$ and thereby the vacancy-filling rate $q_t(m)$. We then iterate over the path of λ_t until a measure of excess labor demand at each point in time is reduced to zero (up to numerical error). The Appendix provides further detail.

Transition dynamics. Figure 6 depicts the results of this exercise. It plots the evolution of worker flows, the unemployment rate, and the offer and worker distributions following a permanent, unanticipated one-percent decrease in aggregate productivity.

Figure 6. Transition dynamics of calibrated model



Notes. Based on simulation of the model calibrated as described in Table 1. The figure illustrates the dynamic response to an unanticipated, permanent one-percent decline in aggregate labor productivity.

Figure 6 exhibits some familiar qualitative features. In panel A, the rate of job loss from employment to unemployment features a mass point of fires on impact of the shock, followed by a gradual descent to a new, higher steady-state level. Turning to rates of job finding, panels B and C confirm that both unemployed and employed workers transition to new jobs at a slower rate following the negative shock. Consequently, the unemployment rate in panel D rises gradually toward a new, higher steady state.

But Figure 6 also reveals some new features. First, an interesting aspect of the response of unemployment inflows is that it is magnified by the presence of on-the-job search: The shock not only renders a mass of marginal matches unprofitable, but also slows the rate of natural wastage due to on-the-job search. Since the latter is a force that replenishes firms' marginal products near the firing boundary, its relative absence after the shock induces still greater rates of job destruction.

Figure 6 also reveals a second novel effect: the dynamics of job finding rates exhibit (mild) *overshooting* relative to their steady-state responses. The intuition is as follows. On impact of the shock, marginal products $m = px\alpha n^{\alpha-1}$ jump down with p among non-firing firms. However, employment among these firms cannot move on impact; n is a state variable. Consequently, the job finding rate λ jumps down to alleviate turnover costs and induce firms to keep hiring. Subsequent to the shock, though, non-firing firms' marginal products gradually rise as employment adjusts downward along the transition path. As a result, the reduction in λ (relative to the initial steady state) necessary to equilibrate the labor market lessens; λ *rises* toward its new steady-state level, and there is overshooting.[28]

A further novel feature of Figure 6 is the solution for the evolution of the offer distribution $f(m)$, and the worker distribution $g(m)$. These are depicted in panels E and F, expressed as deviations from their counterparts in the new steady state. Recall that an important goal of the exercise is to demonstrate the feasibility of solving for the steady states and dynamics of these distributions, using the results of the preceding sections.

In addition, however, these endogenous distribution dynamics of the model paint a picture that complements the responses of the worker flows. Recall that, in the wake of the initial impact of the shock, it takes time for marginal products to rise toward their new steady-state distribution. Thus, relative to the new steady state, there are too few employees at high marginal products, and too many at lower marginal products along the transition in panel F. The worker density $g(m)$ "twists" toward its new steady state. The upward jump and subsequent decline in the excess mass of workers near the lower firing boundary is the counterpart of the response of the job loss rate in panel A. Similarly, the gradual accumulation of mass in the hiring region is the counterpart of the behavior of the job finding rates in panels B and C. This in turn dovetails with the response of the offer density $f(m)$ in panel E. Mirroring the overshooting response of job finding rates, the support of the offer distribution in panel E, (m_h, m_u), contracts in the immediate wake of the shock, and then widens again along the transition. Consequently, the density of offers on the support rises and then falls toward its new steady state.

The practical implication of Figure 6, however, is that labor market dynamics in the model are fast. We will see that this feature of the transition dynamics in turn facilitates a simple quantitative assessment of the model, to which we now turn.

---

[28] Overshooting need not emerge in a model with pure vacancy costs. In this case, a slackening of the labor market lowers effective hiring costs. Hiring firms' marginal products can then fall along the transition to the new steady state, and an initial jump down in λ will be instead be followed by a further *gradual decline*. We believe this is why prior work has found incremental transition dynamics (Elsby and Michaels 2013).

Quantitative assessment. We confront the quantitative predictions of the calibrated model with canonical empirical results on the aggregate dynamics of labor market quantities and prices. These are summarized in Table 2C.

For labor market stocks and flows, the empirical evidence in Shimer (2005) has become the standard by which models of this class have been assessed. For that reason, we adopt this benchmark, subject to two changes. First, given the central role of on-the-job search in the model, it is crucial to assess its predictions with respect to the aggregate dynamics of job-to-job flows. We therefore augment the series used by Shimer to include measures of the job-to-job transition rate estimated by Fallick and Fleischman (2004). These are available from 1994. Second, we update the vacancy series using Barnichon's (2010) HWI-JOLTS composite, which is available up to 2016. All other series remain as in Shimer (2005), and are publicly available from the Bureau of Labor Statistics. Table 2C then reports the results of reapplying Shimer's methods to these series for the period 1994 to 2016;[29] specifically, it reports the relative standard deviations of quarterly log-detrended outcomes with respect to output-per-worker.

For wages, the influential work of Solon et al. (1994) provides estimates of the cyclicality of real wages that take account of changes in worker composition over the cycle. They find that a percentage-point rise in the unemployment rate is associated with a 1.4 percent decline in real wages for U.S. men. Elsby et al. (2016) perform related analyses on more recent microdata, again finding that real wages are substantially procyclical. Taken together, a ballpark summary is that the semi-elasticity of real wages with respect to the unemployment rate is in the neighborhood of minus one, as reported in Table 2C. Roughly, real wages are about as procyclical as employment.

Returning to the model, recall that a key lesson of Figure 6 is that the model exhibits fast transition dynamics. Viewed at the quarterly or annual intervals applied to the data, the responses of the job-finding rates in Panels B and C are essentially jump. Together with the high level of the empirical job-finding rate, an implication is that the model displays limited internal propagation. Although the latter is a well-known property of canonical models in the search tradition (Shimer 2005), this outcome was not assured in our richer model of firm dynamics—hence the value-added of our ability to solve for the transition dynamics. A useful corollary is that much of the model's content with respect

[29] The difference in sample period relative to Shimer *raises* the volatility of the job-finding rate that he stresses. In this sense, our assessment of the model relative to these more volatile recent data is conservative.

to aggregate labor market dynamics is effectively conveyed by the steady-state elasticities of outcomes with respect to output-per-worker. Accordingly, Table 2C reports (the absolute value of) these elasticities as a counterpoint to the empirical volatilities.

The contrast between model and data in Table 2C is a reassuring one. On the quantity side, the model accounts for around 60 percent of the empirical volatility of unemployment, vacancies, and the job-finding rate from unemployment. The model thus goes a considerable way toward resolving Shimer's (2005) well-known puzzle that standard search models generate volatility an order of magnitude smaller than in the data. Further reassurance is provided by the response of the job-to-job transition rate, which almost exactly replicates its empirical analogue. The model thus has reasonable predictions for the cyclicality of on-the-job search, a central feature of the theory.

The response of the inflow rate into unemployment is more nuanced. Table 2C reveals that, steady state to steady state, the movement in the model's employment-to-unemployment rate accounts for only one-third of its empirical volatility. However, recall from Figure 6A that the response of the E-to-U rate in the model overshoots substantially relative to its steady-state response. This has empirical support: the majority of the cyclicality of unemployment inflows is concentrated in spikes at the onset of recessions (Elsby et al. 2009). The steady-state response of job losses reported in Table 2C should thus be viewed as a lower bound on the variation in job losses implied by the model.[30] A fair summary, then, is that the response of job loss in the model is qualitatively accurate, and quantitatively nontrivial, relative to the data.

Turning now to wages, Table 2C further reveals that the model generates an empirically-plausible degree of procyclicality in real wages. Strikingly, the model-implied semi-elasticity of real wages with respect to unemployment of -1.4 replicates exactly the estimate in Solon et al.'s (1994) classic analysis. The implication, then, is that the plausible volatility in labor market quantities noted above dovetails with plausible responses in prices through the lens of the model.

The overall message of Table 2C is thus an encouraging one, especially given that none of these outcomes was targeted as part of the calibration exercise. We explore the sources of this result in Appendix C. Procyclical turnover costs moderate aggregate volatility in the model. This is offset by three sources of amplitude. First, as Pissarides

[30] A more definitive quantitative conclusion would require a judgment on the variation in empirical job losses that is driven by such overshooting. Under the model, the latter depends on the precise timing and sequencing of aggregate shocks, which is difficult to infer from the data.

(2009) notes, a consequence of a fixed per-worker hiring cost, as opposed to a pure vacancy cost, is that recruitment costs no longer decline following adverse shocks, and so job creation falls more precipitously. Second, and quantitatively most important, the presence of credible bargaining limits the procyclicality of wages, amplifying quantity responses, as in Hall and Milgrom (2008). Finally, echoing Elsby and Michaels (2013), the presence of decreasing returns implies that marginal jobs generate smaller surpluses, amplifying responses still further. The magnitude of volatility, its pattern across search off- and on-the-job, and across quantities and prices, need not have turned out to be so quantitatively plausible. We see these as signs that the model is both sensible and useful.

4. Extensions

We now show that the model is amenable to several extensions, provided the homogeneity that underlies an m-solution is preserved. Here we provide two examples: a model of offer matching in wage determination; and richer structures of labor market frictions.

4.1 Offer matching

We study the role of offer matching via a generalization of the sequential auctions approach of Postel-Vinay and Robin (2002). As in their model, firms are assumed to have all the bargaining power. In a simple extension, we further allow for a variable propensity for offer matching among competing firms, indexed by a parameter ζ. Echoing our discussion of the impediments to offer matching, one interpretation of ζ is the probability that both firms are credibly informed over the presence of both job offers (with $1-\zeta$ the probability neither firm is informed). An alternative interpretation is that the firm and its workers will tolerate unequal treatment up to some limit, expressed for convenience as a fraction ζ of the firm's marginal value of labor. These interpretations are analytically equivalent. The special case of $\zeta = 1$ corresponds to the model of Postel-Vinay and Robin.

An implication of this setting is that the effective cost of hiring now includes both a base hiring cost as well as the recruitment bonuses that firms expect to pay. It is important in this case that the base hiring cost is incurred prior to meeting a searcher. Otherwise, a firm would have an incentive to hire unemployed workers to save on recruitment bonuses. We therefore proceed in this section by assuming that there is a vacancy cost which is sunk at the time of meeting. For comparability with the previous sections, though, we maintain the presence of a constant effective hiring cost equal to c for all hiring firms. Alternative functional forms for hiring costs are discussed in the following subsection.

To map worker and firm values to flow wages, firms are assumed to be able to commit to payments to workers only in the current $\mathrm{d}t$ period (as in Moscarini 2005). This aids comparability of this case with the preceding sections, and simplifies the contract structure as we will see that workers within a firm are almost always paid the same flow wage.

Equilibrium then takes a simple form. Consider a worker employed in a firm with marginal value Π_n. Upon realization of an outside offer from a firm with marginal value $\widetilde{\Pi}_n$, the worker chooses the firm with the higher marginal value. If she quits from her current firm (at rate δ), she receives (in expectation) a lump-sum recruitment bonus equal to $\zeta\Pi_n$. If she stays with her current firm (at rate $s\lambda - \delta$), she receives (in expectation) a lump-sum retention bonus equal to $\zeta\widetilde{\Pi}_n$. In the absence of an outside offer, the worker receives a flow wage payment such that she is indifferent to unemployment and the worker surplus is zero, $W = 0$. The option value to search while unemployed is thus also zero, and the firm's hiring and firing behavior has no effect on worker values.

Applying similar arguments to those underlying (11) and (13) above, we can write the firm and worker values implied by this environment as follows:

$$\begin{aligned} r\Pi &= xn^{\alpha} - wn - \delta n\Pi_n - (s\lambda - \delta)n\zeta\mathbb{E}\big[\widetilde{\Pi}_n \big| \widetilde{\Pi}_n < \Pi_n\big] + \mu x\Pi_x + \frac{1}{2}\sigma^2 x^2 \Pi_{xx}, \text{ and} \\ rW &= w - b + \delta\zeta\Pi_n + (s\lambda - \delta)\zeta\mathbb{E}\big[\widetilde{\Pi}_n \big| \widetilde{\Pi}_n < \Pi_n\big] - \delta n W_n + \mu x W_x + \frac{1}{2}\sigma^2 x^2 W_{xx}. \end{aligned} \tag{38}$$

The flow wage paid in the absence of outside offers solves $W = 0$, and takes the form

$$w = b - \delta\zeta\Pi_n - (s\lambda - \delta)\zeta\mathbb{E}\big[\widetilde{\Pi}_n \big| \widetilde{\Pi}_n < \Pi_n\big]. \tag{39}$$

The wage equals the unemployment payoff b, less expected capital gains from recruitment and retention bonuses from future outside offers. The firm's value is thus

$$r\Pi = xn^{\alpha} - bn - (1 - \zeta)\delta n\Pi_n + \mu x\Pi_x + \frac{1}{2}\sigma^2 x^2 \Pi_{xx}. \tag{40}$$

Equation (40) yields an important insight. Recall that the key channel through which on-the-job search interacts with firm decisions is through turnover costs; these are now given by $(1 - \zeta)\delta n\Pi_n$. The upshot of (40), then, is that the presence of offer matching implicitly reduces the turnover costs faced by the firm, and does so in proportion to the firm's propensity to match offers, ζ. The intuition stems from the wage equation (39). The prospect of future recruitment and retention bonuses leads the worker to accept lower flow wages. The firm implicitly recoups the entirety of the cost of its retention bonuses in this way. To the extent that firms' propensity to match offers is incomplete ($\zeta < 1$), the wage reductions implied by prospective recruitment bonuses only partially offset the firm's turnover costs. Thus, the degree of offer matching plays an important role in shaping the effective costs of turnover to the firm, and thereby the nature of labor market equilibrium.

In the limiting case of complete offer matching ($\zeta = 1$), the firm recoups all of its turnover costs, and thereby becomes *indifferent* to turnover. Interestingly, labor market equilibrium in this case takes a standard form summarized in the following lemma.

Lemma 4 *Suppose there is complete offer matching, $\zeta = 1$, and firms are subject only to a linear vacancy cost c_v. Then, (i) Proposition 1 holds* mutatis mutandis *with $\omega_0 = b$, $\omega_1 = 0$, $s\lambda = 0$, and $c = \left[c_v + \int_{m_l}^{m_h} J(\tilde{m}) dq(\tilde{m})\right]/\chi$; (ii) the hiring region is degenerate; and (iii) Propositions 3 and 4 hold with vacancy-filling rate $q(m) = \chi(m_h/m)^{-2(1-\alpha)s\lambda/\sigma^2}$.*

The complete offer matching limit thus provides a model of firm dynamics with efficient on-the-job search that can be solved analytically, a novel contribution to the literature. The effective hiring cost c has a direct correspondence in this case to a linear vacancy cost, plus an expected recruitment bonus. Furthermore, because effective turnover costs vanish, optimal labor demand is as if there is no turnover, and the hiring region collapses. Interestingly, labor market equilibrium resembles that in a model without on-the-job search, similar to Elsby and Michaels (2013).

Intuitively, when ζ equals one, the firm can tailor its offer matching to the idiosyncratic offers of all of its contacted employees. This has a nonlinear pricing interpretation. Absent an ability to match offers, a firm faces a quandary: it has one instrument—the marginal product m—to respond to a *continuum* of outside offers. In the presence of constraints to its ability to set a continuum of such prices, the firm will face costs of turnover, and the insights of the preceding sections will apply.

Indeed, the resemblance between the firm's problem with offer matching (40) and its counterpart with *ex post* wage bargaining and no offer matching (13) makes it clear that optimal labor demand and equilibrium turnover will have the same qualitative form when $\zeta \in [0,1)$. The following lemma confirms this for a case analogous to Proposition 2.

Lemma 5 *Suppose there is partial offer matching, $\zeta \in [0,1)$. Then, there exists a linear vacancy cost $c_v(m)$ such that (i) Proposition 1 holds* mutatis mutandis *with $\omega_0 = b$, $\omega_1 = 0$, and $s\lambda$ exchanged with $(1-\zeta)s\lambda$; (ii) the hiring region is nondegenerate with quit rate*

$$\delta(m) = s\lambda + \frac{1}{(1-\zeta)c}\left\{\frac{m-m_h}{\alpha} - \left(\frac{m_h}{\alpha} - b - [r + (1-\zeta)s\lambda]c\right)\left[\left(\frac{m}{m_h}\right)^{\frac{1}{1-\alpha}} - 1\right]\right\}; \quad (41)$$

and (iii) Propositions 3 and 4 hold.

Lemma 5 underscores the role of offer matching in shaping the presence of a hiring region and the competitiveness of the labor market in the model. Intuitively, the greater the propensity of offer matching (indexed by ζ), the smaller the hiring region, the greater the degree of mean reversion in marginal products, and the greater the degree of labor market competition among hiring firms. The nonlinear pricing interpretation of offer matching dovetails intuitively: To the extent that the firm can tailor wages to the idiosyncratic outside offers of its workers, competitive outcomes can be achieved.

Empirical evidence on the propensity for employers to match offers remains limited, but the evidence available suggests only a modest propensity. Based on questions appended to the Survey of Consumers, Brown and Medoff (1996) report that about a third of respondents thought their employers would match. Similarly, based on his interviews with employers, Bewley (1999, p.99) reports that most "made no counteroffers, or made them only rarely or to key people." These in turn dovetail with Bewley's classic finding of the importance of internal wage structure in constraining firms' ability to pay their employees different wages, a theme taken up in Snell and Thomas's (2010) model of equal treatment concerns. Most recently, Di Addario et al. (2020) devise a decomposition of variance for wages that accommodates worker, firm origin, and firm destination effects. Using Italian microdata, they find that only a small share of wage variation can be attributed to firm origin effects, contrary to the implications of pervasive offer matching. These threads of evidence support an intermediate value of ζ in the preceding model.

4.2 The structure of frictions

Our baseline analyses assume frictions take the form of a per-worker hiring cost c. Here we summarize a few examples of deviations from this baseline; formal details are provided in the Online Appendix. Key to tractability is the availability of an m-solution. The simplest such extension is thus a state-dependent linear hiring cost, $c(m)$; for example, a training cost whereby the cost of generating productive new hires depends on the marginal product of existing employees required to train them. All results generalize in this case.

But it is also possible to extend our results to cases in which there is convexity in the recruitment technology, which are widely implemented in models of firm dynamics. Here, an m-solution is preserved provided that the marginal hiring (vacancy) cost is a function solely of the hiring (vacancy) rate. The key insight is that the aggregation results in Proposition 3 continue to hold, in particular for the hiring rate $\eta(m)$, and the vacancy rate $\nu(m) \equiv \eta(m)/q(m)$. We show in the Online Appendix that this observation distils the problem into a system of differential equations—in the marginal value $J(m)$, the quit rate $\delta(m)$, the hiring rate $\eta(m)$, and the vacancy-filling rate $q(m)$—that is straightforward and efficient to solve numerically. The relative simplicity of this result further widens the scope of application of the current framework.

A final ingredient common to models of firm dynamics is firm entry, exit and growth. We show how it is also possible to embed these in the model, and its various extensions considered above. Again, the key is that, upon entry, firm entrants can be summarized by a distribution of marginal products. The Online Appendix shows how this, together with exogenous firm exit, can be accommodated straightforwardly into our framework.

5. Summary and discussion

This paper has presented a synthesis of firm dynamics and on-the-job search. The result is an environment in which some of the key empirical regularities of the labor market can be understood jointly. Firms with concave revenue functions face idiosyncratic shocks that drive job creation and destruction (Davis and Haltiwanger 1992). These in turn drive flows of workers in and out of unemployment (Blanchard and Diamond 1990) and, through on-the-job search, directly from one employer to another (Fallick and Fleischman 2004).

A set of novel contributions naturally emerges. First, the model admits an analytical characterization of equilibrium outcomes. This is a particular challenge posed by the presence of on-the-job search in the environment, since the rate of turnover faced by firms in general will depend on the firm's position in an endogenous distribution of job values. We devise an environment in which this distribution can be derived analytically and, as a consequence, quit, layoff, hiring and vacancy-filling rates can all be solved in closed form in steady-state equilibrium. We further show how these analytical results can be used to render feasible an analysis of out-of-steady-state transition dynamics.

Second, a host of new economic insights follow. Firms' desire to manage their turnover costs gives rise to a novel manifestation of imperfect labor market competition. In contrast to the competitive limit, differences in marginal products across firms are closed only incrementally. Formally, there is endogenous, gradual mean reversion in marginal products. A consequence is that there is additional dispersion in marginal products across firms in equilibrium—there is endogenous misallocation—that arises from the interaction of firm dynamics and on-the-job search.

Third, the endogenous hierarchy of firms that emerges from the environment naturally captures several stylized facts of imperfect labor markets and establishment dynamics. Firms higher up in the distribution of marginal products pay higher wages, consistent with recent estimates of rent sharing. These firms also face lower quit rates; the resulting negative association between turnover and wages again mirrors leading estimates in the empirical literature. Turning to establishment dynamics, the model captures the empirical correlation between job and worker flows noted by Davis, Faberman and Haltiwanger (2012, 2013). Firms higher up in the hierarchy hire more intensively, and are less likely to lay off employees. Crucially, they also face lower quit rates, and are able to fill their vacancies more quickly. We show that the model generates cyclical fluctuations in labor market stocks and flows that resemble standard measures of their empirical behavior (Shimer 2005). Strikingly, a quantitative assessment of the model reveals that, in all these dimensions, it generates moments in the region of their empirical counterparts.

Finally, the model is amenable to an array of extensions. Most importantly, we show how it is possible to accommodate a theory of partial offer matching into our multi-worker firm environment. An instructive implication is that the degree of misallocation among hiring firms due to on-the-job search is ameliorated by firms' ability to match offers. In the limit in which firms can respond perfectly to each of the idiosyncratic outside offers of its employees, the distribution of marginal products among hiring firms becomes degenerate. In further extensions, we show how the model can accommodate richer structures of labor market frictions, such as convex hiring and vacancy costs, as well as firm entry, exit, and growth, while preserving much of its analytical tractability.

There remain several avenues for future research not taken up in our framework. Central to the tractability of our model are a lack of commitment in wage setting and, relatedly, the availability of what we term an "m-solution." Although the model is able to account for many of the empirical features of wages—their procyclicality, and the presence of rent sharing, for example—a more satisfying understanding of the economics would also accommodate the implications of (limited) commitment in wage determination. Likewise, an m-solution is unlikely to be available in environments with more general technologies, shocks, frictions, and wage protocols. Our hope is that the present paper provides a first step toward a more complete synthesis of labor market frictions and firm dynamics.

Appendix

A. Proofs of Lemmas and Propositions

Proof of Lemma 1. We first verify that, under an m-solution, the worker surplus in (11) is a function solely of m. Denoting the firm's hiring and separation rates by $\mathrm{d}H^*/n = \eta(m;\mathrm{d}t)$ and $\mathrm{d}S^*/n = \varsigma(m;\mathrm{d}t)$, we can rewrite (11) as a function only of m,

$$
\begin{aligned}
rW(m)\mathrm{d}t = \bigg\{ & \frac{\beta m}{1-\beta(1-\alpha)} + \omega_0 - b - \lambda \int \widetilde{W} d\Phi(\widetilde{W}) + s\lambda \int_{W(m)} [\widetilde{W} - W(m)] \mathrm{d}\Phi(\widetilde{W}) \\
& + [\mu + (1-\alpha)\delta(W(m))] m W'(m) + \frac{1}{2}\sigma^2 m^2 W''(m) \bigg\} \mathrm{d}t \\
& - (1-\alpha)[\eta(m;\mathrm{d}t) - \varsigma(m;\mathrm{d}t)] m W'(m) - \varsigma(m;\mathrm{d}t) W(m).
\end{aligned}
\tag{42}
$$

Likewise, one can confirm that the firm's marginal value J in (14) is a function only of m.

We now establish monotonicity of W in m. First, we verify that all separations into unemployment occur at a lower reflecting boundary for m. Suppose, to the contrary, that the firm implements strictly positive fires such that the marginal product m diffuses over an interval $[m_1, m_2]$. By optimality, it must be that the firm's marginal value $J(m) = 0$ for all $m \in [m_1, m_2]$, and thus that $J'(m) = J''(m) = 0$ for all $m \in (m_1, m_2)$. Inserting the latter into (14) yields $J(m) = \{(1-\beta)/[1-\beta(1-\alpha)]\}m - \omega_0 = 0$ for all $m \in (m_1, m_2)$, a contradiction. Thus, all fires occur at a lower reflecting barrier; the firm's firing rate ς is thus weakly decreasing in m.

Now consider two firms with different initial marginal products $m' > m$. Fix, for both firms, a given sample path for changes in idiosyncratic productivity, arrivals of job offers, and layoff shocks. Furthermore, suppose that the worker employed in firm m' implements, for all future periods, the same job acceptance policy as the optimal policy for the worker employed in firm m. Denote by T the first time one of the following events occurs for the worker in firm m: the worker is fired; the worker accepts a job; the marginal product equals that in firm m'. Further denote by V_T the continuation value thereafter. Since we have fixed the same sample paths of shocks and job acceptance strategy, T and V_T are the same in firm m'. Since the worker surplus is based on expectations over sample paths, and since the worker in firm m' implements a weakly suboptimal job acceptance policy, $W(m') \geq \mathbb{E}\left[\int_0^T \mathrm{e}^{-rt} w(m_t') \mathrm{d}t + \mathrm{e}^{-rT} V_T | m_0 = m'\right] > \left[\int_0^T \mathrm{e}^{-rt} w(m_t) \mathrm{d}t + \mathrm{e}^{-rT} V_T | m_0 = m\right] = W(m)$, as required.

Proof of Proposition 1. The recursion for the firm's marginal value in the natural wastage region in (17) resembles canonical firm dynamics problems studied by Bentolila and Bertola (1990) and Abel and Eberly (1996). It can be verified that the stated solution for $J(m)$ in (20) satisfies (17) and the two pairs of boundary conditions in (18) and (19). Furthermore, the coefficients J_1 and J_2, and the boundaries m_l and m_h, that satisfy the boundary conditions can be inferred from the solution provided by Abel and Eberly (1996). Applying their result *mutatis mutandis* yields the coefficients

$$J_1 = -\frac{(1-\omega_1)\vartheta(\mathcal{G})m_l^{1-\gamma_1}}{\gamma_1\rho(1)}, \text{ and, } J_2 = -\frac{(1-\omega_1)[1-\vartheta(\mathcal{G})]m_l^{1-\gamma_2}}{\gamma_2\rho(1)}, \tag{43}$$

where

$$\mathcal{G} \equiv \frac{m_h}{m_l}, \text{ and, } \vartheta(\mathcal{G}) \equiv \frac{\mathcal{G}^{\gamma_2} - \mathcal{G}}{\mathcal{G}^{\gamma_2} - \mathcal{G}^{\gamma_1}}. \tag{44}$$

Continuity of the coefficients of the differential equation (17) implies that the latter constitutes a unique solution of (17), (18) and (19), by the Picard-Lindelöf theorem. In turn, the geometric gap between the middle and lower boundaries $\mathcal{G}$ is the solution to

$$\frac{\omega_0 + \rho(0)c}{\omega_0}\varphi(\mathcal{G}) - \mathcal{G}\varphi(\mathcal{G}^{-1}) = 0, \text{ where } \varphi(\mathcal{G}) \equiv \frac{1}{\rho(1)}\left\{1 - \frac{\vartheta(\mathcal{G})}{\gamma_1} - \frac{1-\vartheta(\mathcal{G})}{\gamma_2}\right\}. \tag{45}$$

Abel and Eberly (1995, 1996) prove that there exists a unique $\mathcal{G} \geq 1$ that solves the latter for any finite $c \geq 0$. Finally, the boundaries solve

$$(1-\omega_1)m_l = \frac{\omega_0}{\rho(0)\varphi(\mathcal{G})}, \text{ and, } (1-\omega_1)m_h = \frac{\omega_0 + \rho(0)c}{\rho(0)\varphi(\mathcal{G}^{-1})}. \tag{46}$$

Abel and Eberly further show that $0 < \varphi(0) < \varphi(1) < \varphi(\infty)$. Thus, $0 < m_l < m_h < \infty$.

Proof of Proposition 2. We begin by establishing three preliminary results. First, we show that the quit rate δ must be continuous—or, equivalently, that the offer distribution F has no mass points. Note that the maximized value of the firm $\Pi(n, x)$ must be continuous. Furthermore, under an m-solution we can write $\Pi(n, x) = \pi(m)n$. At any (n, x) at which there is strictly positive hiring, the smooth pasting and super contact conditions in (18) and (19) must hold (Dumas 1991; Stokey 2009). It follows that the firm's value Π is twice differentiable in n and x, and satisfies the Bellman equation (13), when there is strictly positive hiring. Together, these observations imply that the quit rate δ must be continuous in (n, x), and thereby in m.

Second, we demonstrate that, in any region over which there is strictly positive hiring, the quit rate is differentiable—that is, $\delta'(m)$ and, thereby, the offer density $f(m)$ exist. To see this, observe that in any such region the marginal value $J(m)$ must satisfy the recursion (16), with $J(m) = c$ and $J'(m) = 0 = J''(m)$; thus, $\delta(m)$ must be differentiable.

Third, we establish that the hiring region cannot have any "gaps" in which the offer density $f(m)$—or, the marginal quit rate $\delta'(m)$—is zero. Suppose, to the contrary, that there is such a gap for some interval $m \in (m_1, m_2)$. In any such gap, the quit rate would be a constant, $\delta(m) = \delta(m_1)$, and the marginal value $J(m)$ would satisfy (17) with $s\lambda$ replaced by $\delta(m_1)$, and boundary conditions $J(m_1) = c = J(m_2)$, and $J'(m_1) = 0 = J'(m_2)$. These can be satisfied only in the degenerate case $m_1 = m_2$, a contradiction.

Given these, it follows that the hiring region is a unique interval (m_h, m_u) over which $J(m) = c$ and $J'(m) = 0 = J''(m)$, yielding the recursion for the quit rate in (22). It is then straightforward to verify that the solution for the quit rate takes the form

$$\delta(m) = \frac{(1-\omega_1)m}{\alpha c} - \frac{\omega_0}{c} - r + \delta_1 m^{\frac{1}{1-\alpha}}, \tag{47}$$

for all $m \in (m_h, m_u)$. The coefficient δ_1, and the upper boundary for the marginal product in the hiring region, m_u, are determined by boundary conditions,

$$\delta(m_h) = s\lambda, \text{ and, } \delta(m_u) = 0. \tag{48}$$

It follows from the first boundary condition that

$$\delta_1 = \left(r + s\lambda + \frac{\omega_0}{c}\right) m_h^{-\frac{1}{1-\alpha}} - \frac{1-\omega_1}{\alpha c} m_h^{1-\frac{1}{1-\alpha}}. \tag{49}$$

Inserting the latter into (47) yields the stated solution for $\delta(m)$. Continuity of the coefficients of the differential equation (22) implies that the latter constitutes a unique solution of (22), and the boundary conditions in (48), by the Picard-Lindelöf theorem.

Turning now to the upper boundary m_u, the second condition in (48) implies

$$\left[m_h - \alpha\frac{\omega_0 + (r+s\lambda)c}{1-\omega_1}\right]\left[\left(\frac{m_u}{m_h}\right)^{\frac{1}{1-\alpha}} - 1\right] = \frac{\alpha c s\lambda}{1-\omega_1} + (m_u - m_h). \tag{50}$$

Using the solution for m_h in (46), we can write the leading coefficient in the latter as

$$m_h - \alpha\frac{\omega_0 + (r+s\lambda)c}{1-\omega_1} = \frac{\omega_0 + (r+s\lambda)c}{1-\omega_1}\left[\frac{1}{(r+s\lambda)\varphi(\mathcal{G}^{-1})} - \alpha\right]. \tag{51}$$

Abel and Eberly (1996) prove that $\mathcal{G} > 1$ implies that $\varphi(\mathcal{G}^{-1}) < \varphi(1) = 1/(r+s\lambda)$. Thus,

$$m_h - \alpha \frac{\omega_0 + (r + s\lambda)c}{1 - \omega_1} > \frac{\omega_0 + (r + s\lambda)c}{1 - \omega_1}(1 - \alpha) > 0. \tag{52}$$

This implies that there exists a unique $m_u > m_h$ that satisfies (50).

Now consider the slope of $\delta(m)$. Differentiating (23), applying the solution for m_h in (46), and once again noting that $\mathcal{G} > 1$ implies that $\varphi(\mathcal{G}^{-1}) < \varphi(1) = 1/(r + s\lambda)$ yields

$$\delta'(m) = \frac{1 - \omega_1}{\alpha c}\left\{1 - \frac{1}{1 - \alpha}[1 - \alpha(r + s\lambda)\varphi(\mathcal{G}^{-1})]\left(\frac{m}{m_h}\right)^{\frac{\alpha}{1-\alpha}}\right\} < \frac{1 - \omega_1}{\alpha c}\left[1 - \left(\frac{m}{m_h}\right)^{\frac{\alpha}{1-\alpha}}\right]. \tag{53}$$

It follows that $\delta'(m_h^+) < 0$ and that $\delta(m)$ is declining for all $m \in (m_h, m_u)$. Finally, differentiating (23) once more, and following the same steps,

$$\delta''(m) = -\frac{1 - \omega_1}{c(1 - \alpha)^2}\frac{1}{m_h}[1 - \alpha(r + s\lambda)\varphi(\mathcal{G}^{-1})]\left(\frac{m}{m_h}\right)^{\frac{2\alpha-1}{1-\alpha}} < 0. \tag{54}$$

Proof of Proposition 3. (i) Denote the logarithm of the marginal product $\mathscr{m} \equiv \ln m$. In the natural wastage region, this evolves according to the stochastic law of motion

$$\mathrm{d}\mathscr{m} = \mathrm{d}\ln x - (1 - \alpha)\mathrm{d}\ln n = \left[\mu - \frac{1}{2}\sigma^2 + (1 - \alpha)s\lambda\right]\mathrm{d}t + \sigma \mathrm{d}z \equiv \mu_{\mathscr{m}}\mathrm{d}t + \sigma\mathrm{d}z. \tag{55}$$

This can be approximated by a discrete-time, discrete-state process (Dixit 1993):

$$\mathscr{m}_{t+\mathrm{d}t} = \begin{cases} \mathscr{m}_t + \Delta & \text{with probability} \quad \mathscr{p}, \\ \mathscr{m}_t - \Delta & \text{with probability} \quad \mathscr{q}, \end{cases} \tag{56}$$

where $\Delta = \sigma\sqrt{\mathrm{d}t}$, $\mathscr{p} = \frac{1}{2}\left(1 + \frac{\mu_{\mathscr{m}}}{\sigma}\sqrt{\mathrm{d}t}\right)$, and $\mathscr{q} = \frac{1}{2}\left(1 - \frac{\mu_{\mathscr{m}}}{\sigma}\sqrt{\mathrm{d}t}\right)$.

Consider a worker at $\mathscr{m}_l$. With probability $\mathscr{q}$, her firm crosses the lower boundary and fires a fraction $\Delta/(1 - \alpha)$ of its employees such that it returns to $\mathscr{m}_l$. Denoting the stationary density of employees at $\mathscr{m}_l$ by $\mathscr{g}(\mathscr{m}_l)$, the fraction of total employment that separates into unemployment is given by

$$\varsigma\mathrm{d}t = \mathscr{q}\frac{\Delta}{1 - \alpha} \cdot [\mathscr{g}(\mathscr{m}_l) \cdot \Delta] = \frac{\sigma^2/2}{1 - \alpha}\mathscr{g}(\mathscr{m}_l)\mathrm{d}t + o(\mathrm{d}t). \tag{57}$$

Mapping back from logarithms to levels, $\mathscr{g}(\mathscr{m}_l) = m_l g(m_l)$, yields the stated result,

$$\varsigma = \frac{\sigma^2/2}{1 - \alpha}m_l g(m_l). \tag{58}$$

It will be useful in what follows to derive the flow-balance condition for the steady-state density at the lower boundary $\mathscr{g}(\mathscr{m}_l)$. Setting outflows equal to inflows,

$$\mathscr{p}\mathscr{g}(m_l) + \mathscr{q}\frac{\Delta}{1-\alpha}\mathscr{g}(m_l) + s\lambda \mathrm{d}t\mathscr{g}(m_l) = \mathscr{q}\mathscr{g}(m_l + \Delta). \tag{59}$$

Expanding $\mathscr{g}(m_l + \Delta)$, using the definitions of $\mathscr{p}$, $\mathscr{q}$ and Δ, collecting terms in orders of $\sqrt{\mathrm{d}t}$, and eliminating terms of order higher than $\mathrm{d}t$ yields

$$\begin{aligned}\left[\left(\mu_m + \frac{\sigma^2/2}{1-\alpha}\right)\mathscr{g}(m_l) - \frac{1}{2}\sigma^2\mathscr{g}'(m_l)\right]\sqrt{\mathrm{d}t}& \\ = \frac{\sigma}{2}\left[\left(\frac{1}{1-\alpha}\mu_m - 2s\lambda\right)\mathscr{g}(m_l) - \mu_m\mathscr{g}'(m_l) + \frac{1}{2}\sigma^2\mathscr{g}''(m_l)\right]\mathrm{d}t.&\end{aligned} \tag{60}$$

As $\mathrm{d}t \to 0$, the terms of order $\sqrt{\mathrm{d}t}$ dominate, and therefore must cancel,

$$\left(\mu_m + \frac{\sigma^2/2}{1-\alpha}\right)\mathscr{g}(m_l) - \frac{1}{2}\sigma^2\mathscr{g}'(m_l) = 0. \tag{61}$$

Noting that $\mathscr{g}(m_l) = m_l g(m_l)$ and $\mathscr{g}'(m_l) = m_l g(m_l) + m_l^2 g'(m_l)$, recalling the definition of μ_m, and imposing the aggregate stationarity condition $\mu + \frac{1}{2}\sigma^2\frac{\alpha}{1-\alpha} = 0$ yields

$$\left[(1-\alpha)s\lambda - \frac{1}{2}\sigma^2\right]g(m_l) = \frac{1}{2}\sigma^2 m_l g'(m_l). \tag{62}$$

(ii) and (iii). To infer the stationary distribution of marginal products across employees $g(m)$, and thereby the vacancy-filling rate $q(m) = \chi[\psi + (1-\psi)G(m)]$, we first infer the stochastic law of motion for the marginal product, $\mathrm{d}m/m = (\mathrm{d}x/x) - (1-\alpha)(\mathrm{d}n/n)$, on the interval $m \in (m_l, m_u)$. The evolution of productivity x is given by (1). The evolution of employment n is as follows: There are outflows of employment due to quits, $\delta(m)n\mathrm{d}t$. But there are also potential inflows due to hires: The hiring rate at m, denoted $\eta(m)$, can be written as the total measure of hires at m, $f(m)Vq(m)$, divided by the total measure of employment at m, $g(m)N$; or, using (15), and recalling that $\lambda = M/(U+sN)$, $\chi = M/V$, and $1-\psi = sN/(U+sN)$, we can write more succinctly as

$$\eta(m) = \frac{f(m)Vq(m)}{g(m)N} = -\frac{\chi(1-\psi)V}{s\lambda N}\frac{\delta'(m)q(m)}{q'(m)} = -\frac{\delta'(m)q(m)}{q'(m)}. \tag{63}$$

Thus, the stochastic law of motion for the marginal product is

$$\frac{\mathrm{d}m}{m} = \left\{\mu + (1-\alpha)\left[\frac{\delta'(m)q(m)}{q'(m)} + \delta(m)\right]\right\}\mathrm{d}t + \sigma\mathrm{d}z. \tag{64}$$

The latter describes the motion of the marginal product for an employee that remains in a given firm. However, additional flows of employees across marginal products arise due to the presence of search. Specifically, the net inflow of density into $g(m)$ from this channel is given by the measure of hires less quits,

$$[\eta(m)-\delta(m)]g(m) = -\frac{\partial}{\partial m}\frac{[\delta(m)q(m)]}{\chi(1-\psi)}. \tag{65}$$

The Fokker-Planck (Kolmogorov Forward) equation for the worker density $g(m)$ is thus

$$\begin{aligned}\frac{\partial g(m)}{\partial t} = &-\frac{\partial}{\partial m}\frac{[\delta(m)q(m)]}{\chi(1-\psi)} - \frac{\partial}{\partial m}\left[\left\{\mu + (1-\alpha)\left[\frac{\delta'(m)q(m)}{q'(m)} + \delta(m)\right]\right\} mg(m)\right] \\ &+ \frac{1}{2}\sigma^2\frac{\partial^2}{\partial m^2}[m^2 g(m)].\end{aligned} \tag{66}$$

Noting that $g(m) = q'(m)/\chi(1-\psi)$, and that $\partial g(m)/\partial t = 0$ in steady state, we can rewrite the latter as

$$\frac{\partial}{\partial m}[\delta(m)q(m)] + \frac{\partial}{\partial m}\left\{\mu m q'(m) + (1-\alpha)m\frac{\partial}{\partial m}[\delta(m)q(m)]\right\} = \frac{1}{2}\sigma^2\frac{\partial^2}{\partial m^2}[m^2 q'(m)]. \tag{67}$$

Integrating once,

$$\delta(m)q(m) + \mu m q'(m) + (1-\alpha)m\frac{\partial}{\partial m}[\delta(m)q(m)] = \frac{1}{2}\sigma^2\frac{\partial}{\partial m}[m^2 q'(m)] + C_1, \tag{68}$$

where C_1 is a constant of integration. Evaluating at $m = m_l$, imposing the boundary condition for $g(m_l) = q'(m_l)/\chi(1-\psi)$ in (62), noting that $\delta(m_l) = s\lambda$, $\delta'(m_l) = 0$, $q(m_l) = \chi\psi$, and recalling the aggregate stationarity condition, $\mu + \frac{1}{2}\sigma^2\frac{\alpha}{1-\alpha} = 0$, yields

$$C_1 = s\lambda\chi\psi - \frac{\sigma^2/2}{1-\alpha}m_l q'(m_l) = \chi(1-\psi)\varsigma\left(\frac{\lambda U}{\varsigma N} - 1\right) = 0, \tag{69}$$

where the second and third equalities follow from the solution for the separation rate into unemployment ς in (25), established above, the definition of $\psi = U/(U+sN)$, and the fact that unemployment inflows ςN must equal outflows λU in steady state.

Expanding and collecting terms in (68), we can now write

$$(1-\alpha)\frac{\partial}{\partial m}\left[m^{\frac{1}{1-\alpha}}\delta(m)q(m)\right] + (\mu - \sigma^2)m^{\frac{1}{1-\alpha}}q'(m) = \frac{1}{2}\sigma^2 m^{1+\frac{1}{1-\alpha}}q''(m). \tag{70}$$

Integrating again, applying integration by parts to the right-hand side, collecting terms, and imposing the aggregate stationarity condition $\mu + \frac{1}{2}\sigma^2\frac{\alpha}{1-\alpha} = 0$, yields a first-order differential equation in $q(m)$,

$$(1-\alpha)\delta(m)q(m) = \frac{1}{2}\sigma^2 m q'(m) + C_2 m^{-\frac{1}{1-\alpha}}, \tag{71}$$

where C_2 is a further constant of integration. Evaluating once again at $m = m_l$ implies

$$C_2 = (1-\alpha)\chi(1-\psi)\varsigma m_l^{\frac{1}{1-\alpha}}\left(\frac{\lambda U}{\varsigma N} - 1\right) = 0. \tag{72}$$

Thus we have

$$\delta(m)q(m) = \frac{\sigma^2/2}{1-\alpha} m q'(m). \tag{73}$$

Noting that $q(m_l) = \chi\psi$ and $q(m_u) = \chi$, it is straightforward to verify that the solution for $q(m)$, and the share of searchers that are unemployed ψ, take the form stated in the Proposition. Finally, it follows that the hiring rate $\eta(m)$ is as stated.

Proof of Proposition 4. The derivations of (28) and (29) are provided in the main text, so we focus on establishing existence here.

Consider first the limit as $\theta \to 0$, which implies $\lambda \to 0$. It can be verified that Proposition 1 and its proof apply, and that the boundaries satisfy $0 < m_l < m_h < \infty$ in this limit. Furthermore, from Proposition 2, the hiring region becomes degenerate, $m_u \to m_h$. Using Proposition 3, and applying L'Hôpital's rule, the worker distribution $G(m) \to \ln(m/m_l)/\ln(m_h/m_l)$. It follows from (28) and (29) that $\lim_{\theta\to 0} U_{JC}(\theta) < L = \lim_{\theta\to 0} U_{BC}(\theta)$.

Now consider the limit as $\theta \to \infty$, which implies $\lambda \to \infty$. It can be verified that the roots of the fundamental quadratic (21) satisfy $\gamma_1 \to -\infty$ and $\gamma_2 \to 1/(1-\alpha)$. Abel and Eberly (1996) prove that

$$0 < \varphi(0) = \frac{\gamma_1}{\gamma_1 - 1}\frac{1}{\rho(0)} < \varphi(1) = \frac{1}{\rho(0)} < \frac{\gamma_2}{\gamma_2 - 1}\frac{1}{\rho(0)} = \varphi(\infty). \tag{74}$$

It follows that the solution to (45) satisfies $\mathcal{G} \to \infty$, with $\varphi(\mathcal{G}) \to 1/\alpha\rho(0)$ and $\varphi(\mathcal{G}^{-1}) \to 1/\rho(0)$. Thus, the boundaries in (46) satisfy $(1-\omega_1)m_l \to \alpha\omega_0$, and $(1-\omega_1)m_h \to \infty$. Now, note from Proposition 3 that $G(m) \to 0$ for all $m \in (m_l, m_h)$. Thus, $g(m_l) \to 0$, and $\varsigma \to 0$. Furthermore, since $m_h \to \infty$, it must be that $\int m^{1/(1-\alpha)} g(m)\mathrm{d}m \to \infty$. It follows from (28) and (29) that $\lim_{\theta\to\infty} U_{JC}(\theta) = L > 0 = \lim_{\theta\to\infty} U_{BC}(\theta)$.

Since all objects in (28) and (29) are continuous in λ, and thereby θ, it follows that there must exist at least one $\theta \in (0,\infty)$ that satisfies (28) and (29).

Proof of Lemma 2. (a) $s \to 0$. To establish (i), simply note that Proposition 1 and its proof apply for all $s \geq 0$. Property (ii) follows directly from Proposition 2: since $\delta(m_h) \to 0$ as $s \to 0$, and since $\delta(m)$ is declining for $m > m_h$ for all $s > 0$, it follows that $m_u \to m_h$. Finally, (iii) emerges from Proposition 3 and application of L'Hôpital's rule,

$$G(m) \to \frac{(m/m_l)^{\frac{1-\alpha}{\sigma^2/2}s\lambda} - 1}{(m_h/m_l)^{\frac{1-\alpha}{\sigma^2/2}s\lambda} - 1} \to \frac{(m/m_l)^{\frac{1-\alpha}{\sigma^2/2}s\lambda}\ln(m/m_l)}{(m_h/m_l)^{\frac{1-\alpha}{\sigma^2/2}s\lambda}\ln(m_h/m_l)} \to \frac{\ln(m/m_l)}{\ln(m_h/m_l)}, \text{ as } s \to 0. \tag{75}$$

(b) $\alpha \to 1$, holding fixed X and $\tilde{\sigma}^2 \equiv \sigma^2/(1-\alpha)$. For (i), note that, combining the latter with the aggregate stationarity condition (24), the fundamental quadratic (21) is

$$\rho(\gamma) = -\frac{1}{2}\tilde{\sigma}^2(1-\alpha)\gamma^2 - \left[-\frac{1}{2}\tilde{\sigma}^2 + (1-\alpha)s\lambda\right]\gamma + r + s\lambda = 0. \tag{76}$$

It follows that $\gamma_1 \to -2\rho(0)/\tilde{\sigma}^2 \equiv \tilde{\gamma}_1$, and $\gamma_2 \to \infty$, as $\alpha \to 1$. Thus, (44) and (45) become

$$\vartheta(\mathcal{G}) \to \begin{cases} 1 & \text{for} \quad \mathcal{G} \geq 1 \\ \mathcal{G}^{1-\tilde{\gamma}_1} & \text{for} \quad \mathcal{G} < 1 \end{cases}, \quad \text{and} \quad \varphi(\mathcal{G}) \to \frac{1}{\rho(1)}\left[1 - \frac{\vartheta(\mathcal{G})}{\tilde{\gamma}_1}\right]. \tag{77}$$

The latter and (45) imply that the solution for $\mathcal{G}$ satisfies $\lim_{\alpha\to 1} \mathcal{G} > 1$, and that therefore

$$\lim_{\alpha\to 1} \varphi(\mathcal{G}) = \frac{1}{\rho(1)}\left(1 - \frac{1}{\tilde{\gamma}_1}\right) > \frac{1}{\rho(1)}\left(1 - \frac{\mathcal{G}^{\tilde{\gamma}_1 - 1}}{\tilde{\gamma}_1}\right) = \lim_{\alpha\to 1} \varphi(\mathcal{G}^{-1}). \tag{78}$$

It follows from (46) that $0 < m_l < m_h < \infty$. To verify (ii), note from (53) that $\delta'(m) \to -\infty$ as $\alpha \to 1$ for all $m > m_h$. (iii) follows from Proposition 3 and the definition of $\tilde{\sigma}$.

Proof of Lemma 3. (i) In the absence of idiosyncratic shocks, $\mu = \sigma = 0$, and with exogenous job destruction at rate ς_0, the firm's marginal value satisfies

$$\begin{aligned} rJ(m) = (1-\omega_1)m - \omega_0 - [\varsigma_0 + \delta(m) - (1-\alpha)m\delta'(m)]J(m) \\ + (1-\alpha)[\varsigma_0 + \delta(m)]mJ'(m). \end{aligned} \tag{79}$$

It follows that there is a hiring region such that $J(m) = c$ and $J'(m) = 0$ on its interior, and in which the quit rate is given as in Proposition 2, with r exchanged with $r + \varsigma_0$.

(ii) Evaluating (79) to the left and right of m_h implies

$$(\varsigma_0 + s\lambda)m_h J'(m_h^-) = m_h \delta'(m_h^+)c. \tag{80}$$

Noting that $J'(m_h^-) \geq 0$ and $\delta'(m_h^+) \leq 0$ implies that $J'(m_h) = \delta'(m_h) = 0$. This in turn implies that m_h solves $(r + \varsigma_0 + s\lambda)c = (1-\omega_1)m_h - \omega_0$, as claimed.

(iii) and (iv). Retracing the steps of the proof of Proposition 3, imposing $\mu = \sigma = 0$, and noting that the total separation rate from a firm is in this case given by $\varsigma_0 + \delta(m)$, gives rise to the following analogue to (67),

$$\frac{\partial}{\partial m}\{[\varsigma_0 + \delta(m)]q(m)\} + \frac{\partial}{\partial m}\left\{(1-\alpha)m\frac{\partial}{\partial m}\{[\varsigma_0 + \delta(m)]q(m)\}\right\} = 0. \tag{81}$$

Integrating once,

$$[\varsigma_0 + \delta(m)]q(m) + (1-\alpha)m\frac{\partial}{\partial m}\{[\varsigma_0 + \delta(m)]q(m)\} = C_1, \tag{82}$$

where C_1 is a constant of integration. This has solution

$$[\varsigma_0 + \delta(m)]q(m) = C_1 + C_2 m^{-\frac{1}{1-\alpha}}. \tag{83}$$

Evaluating at $m = m_h$, noting that $\delta(m_h) = s\lambda$, and $q(m_h) = \chi\psi$,

$$(\varsigma_0 + s\lambda)\chi\psi = C_1 + C_2 m_h^{-\frac{1}{1-\alpha}}. \tag{84}$$

Likewise, evaluating at $m = m_u$, noting that $\delta(m_u) = 0$, and $q(m_u) = \chi$,

$$\varsigma_0\chi = C_1 + C_2 m_u^{-\frac{1}{1-\alpha}}. \tag{85}$$

Solving for the constants yields $C_1 = (\varsigma_0 + s\lambda)\chi\psi = \chi\varsigma_0$, and

$$\left[\left(\frac{m_u}{m_h}\right)^{\frac{1}{1-\alpha}} - 1\right]C_2 = \chi[\psi s\lambda - (1-\psi)\varsigma_0]m_u^{\frac{1}{1-\alpha}} = \chi(1-\psi)\varsigma_0\left(\frac{\lambda U}{\varsigma_0 N} - 1\right)m_u^{\frac{1}{1-\alpha}} = 0, \tag{86}$$

where the latter uses the definition of $\psi = U/(U + sN)$, and the fact that inflows into unemployment $\varsigma_0 N$ must equal outflows from unemployment λU in steady state. We thus obtain the following solution for the vacancy-filling rate,

$$q(m) = \frac{\varsigma_0\chi}{\varsigma_0 + \delta(m)}. \tag{87}$$

The stated solution for the worker distribution $G(m)$ can be inferred from (15) and the fact that, in steady state, $U = \varsigma_0 L/(\varsigma_0 + \lambda)$, $N = \lambda L/(\varsigma_0 + \lambda)$, and $\psi = \varsigma_0/(\varsigma_0 + s\lambda)$. In turn, it follows that the hiring rate can be written as

$$\eta(m) = -\frac{\delta'(m)q(m)}{q'(m)} = \varsigma_0 + \delta(m). \tag{88}$$

The drift for each m is thus zero, firm marginal products are constant over time, and the natural wastage region is never entered.

Proof of Lemmas 4 and 5. (i) Applying the same methods as those underlying Propositions 1 and 2, the marginal value of labor to the firm $J = \Pi_n$ can be written

$$\begin{aligned} rJ(m) = m - b - (1-\zeta)[\delta(m) - (1-\alpha)m\delta'(m)]J(m) \\ + [\mu + (1-\zeta)(1-\alpha)\delta(m)]mJ'(m) + \frac{1}{2}\sigma^2 m^2 J''(m). \end{aligned} \tag{89}$$

In the natural wastage region, the latter simplifies to

$$[r+(1-\zeta)s\lambda]J(m)=m-b+[\mu+(1-\alpha)(1-\zeta)s\lambda]mJ'(m)+\frac{1}{2}\sigma^2m^2J''(m). \tag{90}$$

Thus, Proposition 1 holds *mutatis mutandis* with ω_0, ω_1 and $s\lambda$ exchanged respectively with b, 0, and $(1-\zeta)s\lambda$. We postpone the form of the effective hiring cost until after verification of (ii). For that, note simply that, in the hiring region, we can write

$$\{r+(1-\zeta)[\delta(m)-(1-\alpha)m\delta'(m)]\}c=m-b. \tag{91}$$

Combining with the boundary condition $\delta(m_h)=s\lambda$, one can verify that the solution for $\delta(m)$ takes the stated form, and that the hiring region is degenerate for $\zeta=1$.

Now return to the effective cost per hire. For general ζ, the latter is the sum of vacancy costs and expected recruitment bonuses as a ratio of the vacancy-filling rate,

$$c(m)=\left[c_v(m)+\zeta\int_{m_l}^{m}J(\widetilde{m})\mathrm{d}q(\widetilde{m})\right]/q(m). \tag{92}$$

For $\zeta=1$, the hiring region is degenerate on m_h, where the vacancy-filling rate is $q(m_h)=\chi$. It can then be verified that a linear vacancy cost, $c_v(m)=c_v$ for all m, implies $c'(m)<0$ for all $m<m_h$, and $c'(m_h)=0$. Thus, no firm with $m<m_h$ will wish to hire, and the effective hiring cost for hiring firms is as stated, $c=\left[c_v+\int_{m_l}^{m_h}J(\widetilde{m})\mathrm{d}q(\widetilde{m})\right]/\chi$. For $\zeta\in[0,1)$, there exists a vacancy cost that sets (92) equal to c for all $m\in[m_l,m_u)$, as stated.

Finally, to verify (iii), the proofs of Proposition 3 and 4 are unchanged. When $\zeta=1$, there is no hiring region, and the vacancy-filling rate simplifies to the stated expression.

B. Computational appendix

Steady-state outcomes. We compute steady-state outcomes using the analytical results stated in the main text, with two exceptions. First, we extend these analytical results to accommodate the case in which there are also exogenous separations into unemployment at rate ς_0. These are provided by the following lemma.[31]

Lemma 6 *Suppose additional separations into unemployment occur at exogenous rate ς_0. Then, (i) prior results for the steady-state marginal value $J(m)$ hold* mutatis mutandis *with $\delta(m)$ exchanged for $\varsigma_0+\delta(m)$; (ii) the separation rate into unemployment is*

[31] The proof is a straightforward extension of the proof of Proposition 3, and so is omitted.

$$\varsigma = \varsigma_0 + \frac{\sigma^2/2}{1-\alpha} m_l g(m_l), \tag{93}$$

and (iii) the vacancy-filling rate is given by $q(m) = q_0(m) + q_1(m)$ *where*

$$q_0(m) = \chi\psi \exp\left[\frac{1-\alpha}{\sigma^2/2}\int_{m_l}^{m} \frac{\varsigma_0 + \delta(\widetilde{m})}{\widetilde{m}} \mathrm{d}\widetilde{m}\right], \text{ and } q_1(m) = -\frac{1-\alpha}{\sigma^2/2}\chi\varsigma_0 \int_{m_l}^{m} \frac{q_0(m)}{z q_0(z)} \mathrm{d}z. \tag{94}$$

Notice that the presence of exogenous separations affects the worker distribution both by raising the effective quit rate to $\varsigma_0 + \delta(m)$ in the $q_0(m)$ term, and by eroding the distribution of workers in the $q_1(m)$ term.

The second exception is our computation of employment growth, and worker flows by employment growth underlying the "hockey sticks" in Figure 5, and the establishment dynamics moments in Table 2B. To compute these, we simulate the dynamics of the marginal product and employment (m, n) over a year. Vacancies are measured at the beginning and end of each month; layoffs, hires, and quits (which also includes exogenous separations at rate ς_0) are cumulated over the month. In practice, we simulate 2,000,000 firms using 200 time steps per day, and a maximum firing and hiring rate of 2000 percent.

Out-of-steady-state outcomes. We do not have analytical solutions for the distribution of workers $G(m)$, and the marginal value function $J(m)$ out of steady state. We solve for these objects using a finite difference method similar to, for example, that used in the recent work of Achdou, Han, Lasry, Lions and Moll (2017).

To calculate the marginal value function J, we use a grid for the log marginal product, $\ln m$, which is denser around m_l and m_m where the function is especially nonlinear. We use the half-implicit (Crank-Nicolson) scheme and impose the smooth-pasting conditions via a penalty method whereby deviations above or below the exercise option are penalized. Each time step can then be reduced to the solution of a system of nonlinear equations.

Similarly, for the worker distribution, we use a grid that is especially dense in the neighborhoods of m_l, m_h and m_u. To improve the accuracy of the algorithm, we integrate the Fokker-Planck (Kolmogorov Forward) equation once to infer the law of motion of the worker distribution function G, as opposed to the density function g, and use a fully implicit scheme. The central difference is used everywhere, except at the boundaries.

To solve for the response of model outcomes to an aggregate shock, a simple scheme is used whereby we iterate over the path for the job finding rate λ until excess demand is sufficiently small. In particular, we implement the following steps:

1. First, we solve for the two steady states. We use our analytical solutions to solve for the job offer arrival rate λ in each steady state. We then use our numerical scheme to solve the marginal value function J and worker distribution G on the grid.
2. We make an initial guess for the transition path for the job offer arrival rate $\lambda(t)$. (We begin with a constant job offer arrival rate equal to that in the new steady state.)
3. We solve the marginal value function (HJB) equation backwards within the natural wastage region in order to calculate $m_l(t)$, and $m_h(t)$.
4. We compute $m_u(t)$ using $m_h(t)$, $\lambda(t)$, and the known functional form for the quit rate, δ. With this information, we can then solve forward for the worker distribution using the integrated Fokker-Planck equation.
5. Lastly, we calculate excess demand. If excess demand is sufficiently small, we stop. Otherwise, we update the time path of $\lambda(t)$ based on each period's excess demand, and return to step 3. We find that a sluggish updating rule, with relatively more updating in earlier periods, helps with stability of the solution.

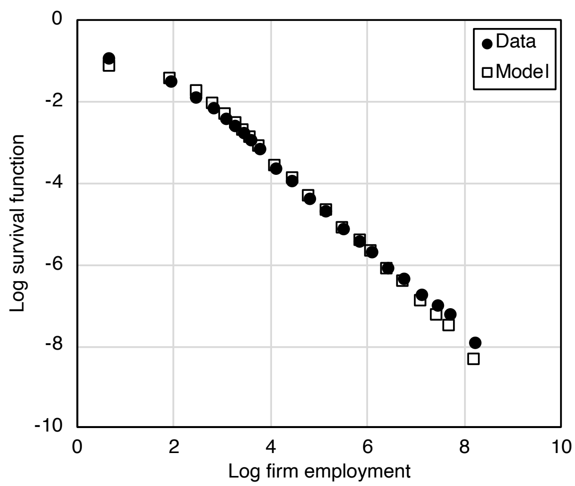
We examine the accuracy of our numerical scheme by comparing its steady-state outcomes with our steady-state analytical results for the marginal value J and worker distribution G. In all cases, errors induced by the numerical scheme are very small.

C. Additional quantitative results

Firm dynamics and worker flows. Here, we report the implications of an interpretation of the model calibrated as in Table 1 for the firm-size distribution, and worker flows by firm size and age. The interpretation we explore is one in which incumbent firms exit at exogenous rate ξ, and are replaced by an equal measure of entrant firms with initial productivity given by a constant x_0, and initial employment such that the measure of workers at each marginal product m among entrant firms is given by $H(m) = \xi G(m)$. The latter preserves all the results stated in the main text, but the presence of firm entry and exit gives rise to a stationary firm-size distribution. The calibration in Table 1 can then be applied subject to one change in interpretation: r now reflects the sum of the discount rate and the firm exit rate ξ.

Figure C1. Firm dynamics and worker flows

A. Firm-size distribution



B. Worker flows by size and age

	Hiring rate		Sep. rate	
A. Firm size	Model	Data	Model	Data
0-19 employees	0.119	0.122	0.144	0.110
20-49	0.159	0.106	0.134	0.102
50-249	0.140	0.102	0.134	0.097
250-499	0.140	0.101	0.134	0.096
500+	0.143	0.087	0.134	0.084
B. Firm age				
0-1 years	0.254	0.212	0.127	0.140
2-3	0.134	0.150	0.135	0.144
4-5	0.137	0.134	0.135	0.130
6-10	0.135	0.125	0.135	0.121
11+	0.144	0.086	0.133	0.084

Notes. Data on firm-size distribution are from the 2016 Statistics of U.S. Businesses. Data on worker flows by size and age taken from Bilal et al. (2019) who use the Job-to-Job Flows provided by the Census Bureau.

Figure C1 reports the results, and contrasts them with data on the firm-size distribution from the Statistics of U.S. Businesses, and worker flows by firm size and age reported by Bilal et al. (2019) based on Job-to-Job Flows data from the Census Bureau. Bilal et al. report an employment-weighted annual exit rate of 2 percent. Accordingly, in our monthly calibration, we set $\xi = 0.02/12$. All other aspects of the calibration are as in Table 1. A convenient implication of the latter is that it implies an annual discount rate of 3 percent, which remains within the reasonable range of data on the real interest rate. We simulate 2 million firms in the calibrated model. To ensure that the right tail of firm size is not driven by very old firms (Gabaix et al. 2016), and to aid numerical accuracy, we restrict attention to firms of age 50 years or less.

Figure C1 reveals that the model does a good job of capturing these dimensions of the data. First, the model-implied firm-size distribution almost exactly replicates its empirical analogue in Panel A.[32] Second, worker flows by firm size and age are broadly in the range

[32] It is well-known that exogenous firm exit gives rise to a stationary distribution of firm productivity x with a Pareto right tail (e.g. Gabaix 2009). Given isoelastic production, firm size $n = (\alpha x/m)^{1/(1-\alpha)}$ for a given marginal product m. In a frictionless labor market, the latter is constant across firms, and so the distribution of frictionless firm size directly inherits a Pareto right tail, mirroring the data. Frictions in the

of the data in Panel B. The most notable feature of the data—the strong decline in hiring rates with age among young firms—is captured well by the model as new firms hire upon entry to reach their optimal size. In addition, separation rates are mildly declining in size, and mildly hump-shaped in age in both model and data. Where the model deviates from the data is in failing to replicate the lower rate of worker reallocation among very large and old firms. Taken together with the fact that these outcomes were not targeted by our calibration, though, the model does a reasonable job of capturing these additional dimensions of the data.

Origins of aggregate volatility. Table 2 revealed that the calibrated model gives rise to a considerable degree of amplitude in aggregate labor market volatility. Here we explore the origins of this result, and contrast it with the standard linear Pissarides (1985) model. Relative to the latter, our model has five differences: idiosyncratic shocks and endogenous job destruction; decreasing returns to scale; credible bargaining; hiring costs (as opposed to vacancy costs); and on-the-job search, with the associated costs of turnover. Accordingly, starting with our model, we chart a course back to Pissarides (1985) by adjusting each of these in turn.

It remains to specify a calibration strategy. Our approach is to hold constant the ratio of the fixed component of wages as a fraction of output-per-worker. The latter summarizes the average rent (to the firm) from employment relationships, and has been highlighted as a key determinant of the aggregate volatility implied by standard linear search models (Shimer 2005; Mortensen and Nagypal 2007; Hagedorn and Manovskii 2008; Ljungqvist and Sargent 2017). In our model with credible bargaining, this involves holding constant $\omega_0/(Y/N)$ at 0.317. In bargaining models with unemployment as the outside option, this involves holding constant $(1-\beta)b/(Y/N)$ at 0.317—or, equivalently, holding constant $b/(Y/N)$ at 0.650—where b is the flow payoff from unemployment. We adjust the hiring (vacancy) cost to maintain these measures of average rent. Otherwise, we set the dispersion of productivity σ to hold constant the unemployment rate at 5 percent, and set the matching elasticity ϵ equal to 0.5 in each model variant. The latter approximately maintains a Beveridge elasticity of minus one, as in the baseline model of this paper. Otherwise, all relevant parameters are as reported in Table 2.

---

model induce a deviation from the latter, but that deviation is *bounded*, since marginal products are bounded, $m \in (m_l, m_u)$. As a result, the frictional firm-size distribution also exhibits a Pareto right tail.

Table C1. Origins of aggregate labor market volatility

Moment	Model variants				Data
	(1)	(2)	(3)	(4)	(5)
	This paper	(1), no OJS, vacancy cost	Elsby and Michaels (13)	Pissarides (85)	Rel. sd.
Response relative to Y/N					
Unemployment rate	7.6	6.3	3.1	1.5	14.0
Vacancy rate	7.6	6.1	3.2	1.6	12.6
U-to-E rate	6.7	6.2	3.2	1.5	11.6
E-to-U rate	1.3	0.4	0.1	—	3.6
E-to-E rate	5.5	—	—	—	5.7
Response relative to U/L					
Average wage	-1.4	-1.7	-6.2	-13.5	≈ -1

Notes. Model outcomes are the absolute value of steady-state elasticities for labor market stocks and flows, and steady-state semi-elasticities for wages.

Table C1 summarizes. For reference, columns (1) and (5) repeat the model outcomes and empirical analogues reported in Table 2. Recall that labor market tightness θ equilibrates the model of column (1) in a novel way by determining the costs of turnover faced by firms. Column (2) suspends this new channel by eliminating on-the-job search ($s = 0$), and replaces it with a conventional linear vacancy cost, denoted c_v. There is no hiring region, and the hiring boundary m_h becomes a standard reflecting barrier. This implies a per-worker hiring cost of $c_v/\chi(\theta)$, where $\chi(\theta)$ is now the vacancy-filling rate. Tightness θ thus equilibrates the model of column (2) in this conventional way. Table C1 reveals that exchanging these two sources of labor market equilibration implies a similar degree of aggregate volatility.

Column (3) then further alters the process of wage determination, as in Elsby and Michaels (2013). Specifically, it exchanges the model of credible intra-firm bargaining underlying the wage equation (6) for intra-firm bargaining with unemployment as the workers' outside option. In conjunction with the absence of on-the-job search, and a vacancy cost, this corresponds to a version of Elsby and Michaels (2013). They derive the analogous wage equation,

$$w = \frac{\beta}{1-\beta(1-\alpha)} m + \beta c\theta + (1-\beta)b, \tag{95}$$

where b is the flow payoff from unemployment. This model variant adds a further channel of labor market equilibration, since tightness θ gives rise to additional wage procyclicality. Table C1 confirms that the latter approximately halves the degree of labor market volatility implied by the model. And, consistent with the foregoing intuition, this is accompanied by excessively procyclical wages relative to the data. These results are quantitatively similar to those presented in Elsby and Michaels (2013).

Finally, column (4) reports analogous results for the standard Pissarides (1985) model. This iteration suspends both idiosyncratic shocks ($\sigma = 0$) and decreasing returns ($\alpha = 1$). Table C1 confirms that this implies an additional approximate halving of the implied labor market volatility, and a related doubling of real wage procyclicality, way in excess of its empirical analogue, a point noted by Pissarides (2009). Again, the quantitative results in column (4) are in line with those in the literature (such as Shimer 2005, and Mortensen and Nagypal 2007).

References

Abel, Andrew B. and Janice C. Eberly. 1995. "Optimal investment with costly reversibility." National Bureau of Economic Research, Working Paper 5091.

Abel, Andrew B. and Janice C. Eberly. 1996. "Optimal investment with costly reversibility." *Review of Economic Studies* 63(4): 581-593.

Abowd, John A., and Thomas Lemieux. 1993. "The Effects of Product Market Competition on Collective Bargaining Agreements: The Case of Foreign Competition in Canada." *Quarterly Journal of Economics* 108(4): 983-1014.

Acemoglu, Daron and William B. Hawkins. 2014. "Search with multi-worker firms." *Theoretical Economics* 9(3): 583-628.

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions and Benjamin Moll. 2017. "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach." NBER Working Paper No. 23732.

Audoly, Richard. 2019. "Firm Dynamics and Random Search over the Business Cycle." Mimeo, University College London.

Bagger, Jesper, Bent Jesper Christensen, and Dale T. Mortensen. 2014. "Productivity and Wage Dispersion: Heterogeneity or Misallocation?" Mimeo, Aarhus University.

Barlevy, Gadi. 2008. "Identification of Search Models using Record Statistics." *Review of Economic Studies* 75(1): 29-64.

Barnichon, Regis. 2010. "Building a composite Help-Wanted Index." *Economics Letters* 109: 175-178.

Bentolila, Samuel and Giuseppe Bertola. 1990. “Firing costs and labour demand: how bad is Eurosclerosis?” *Review of Economic Studies* 57(3): 381-402.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2019. “Labor Market Power.” National Bureau of Economic Research, Working Paper 25719.

Bewley, Truman F. 1999. *Why Wages Don’t Fall during a Recession.* Cambridge, MA: Harvard University Press.

Bilal, Adrien, Niklas Engbom, Simon Mongey and Giovanni L. Violante. 2019. “Firm and Worker Dynamics in a Frictional Labor Market.” Mimeo, Princeton University.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky. 1986. “The Nash bargaining solution in economic modelling.” *RAND Journal of Economics* 17: 176-188.

Blanchard, Olivier J. and Peter Diamond. 1990. “The cyclical behavior of the gross flows of US workers.” *Brookings Papers on Economic Activity* 1990(2): 85-155.

Bontemps, Christian, Jean-Marc Robin, and Gerard J. van den Berg. 2000. “Equilibrium Search with Continuous Productivity Dispersion: Theory and Non-Parametric Estimation.” *International Economic Review* 41(2): 1039-1074.

Boppart, Timo, Per Krusell, and Kurt Mitman. 2018. “Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative.” *Journal of Economic Dynamics and Control* 89: 68-92.

Brown, Charles, and James L. Medoff. 1996. “Employer Characteristics and Work Environment.” *Annales d’Économie et de Statistique* 41/42: 275-298

Bruegemann, Bjoern, Pieter Gautier and Guido Menzio. 2018. “Intra Firm Bargaining and Shapley Values.” Forthcoming, *Review of Economic Studies.*

Burdett, Kenneth, and Guido Menzio. 2018. “The (Q,S,s) Pricing Rule.” *Review of Economic Studies* 85: 892–928.

Burdett, Kenneth and Dale T. Mortensen. 1998. “Wage differentials, employer size, and unemployment.” *International Economic Review* 39: 257-273.

Bureau of Labor Statistics, U.S. Department of Labor. Civilian Labor Force Level, Employment Level, Unemployment Level, Unemployment Rate, Number Unemployed for Less than 5 Weeks, Nonfarm Business Employment, Nonfarm Business Output. Retrieved 25/04/2020 from https://data.bls.gov.

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

Coles, Melvyn G. and Dale T. Mortensen. 2016. “Equilibrium labor turnover, firm growth, and unemployment.” *Econometrica* 84(1): 347-363.

Cooper, Russell, John Haltiwanger, and Jonathan L. Willis. 2007. "Search frictions: Matching aggregate and establishment observations." *Journal of Monetary Economics* 54: 56-78.

Cooper, Russell, John Haltiwanger, and Jonathan L. Willis. 2015. "Dynamics of labor demand: Evidence from plant-level observations and aggregate implications." *Research in Economics* 69(1): 37-50.

Davis, Steven J., R. Jason Faberman, and John Haltiwanger. 2012. "Labor market flows in the cross section and over time." *Journal of Monetary Economics* 59(1): 1-18.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger. 2013. "The establishment-level behavior of vacancies and hiring." *Quarterly Journal of Economics* 128(2): 581-622.

Davis, Steven J. and John Haltiwanger. 1992. "Gross job creation, gross job destruction, and employment reallocation." *Quarterly Journal of Economics* 107(3): 819-863.

Di Addario, Sabrina, Patrick Kline, Raffaele Saggio, and Mikkel Solvsten. 2020. "It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages." IRLE Working Paper No. 104-20.

Dixit, Avinash K. 1993. *The Art of Smooth Pasting.* Vol. 55 of series Fundamentals of Pure and Applied Economics, eds. Jacques Lesourne and Hugo Sonnenschein. Reading, UK: Harwood Academic Publishers.

Dumas, Bernard. 1991. "Super contact and related optimality conditions." *Journal of Economic Dynamics and Control* 15: 675-685.

Elsby, Michael W. L. and Ryan Michaels. 2013. "Marginal Jobs, Heterogeneous Firms, and Unemployment Flows." *American Economic Journal: Macroeconomics* 5(1): 1-48.

Elsby, Michael W. L., Ryan Michaels and David Ratner. 2019. "Vacancy Chains." Mimeo, University of Edinburgh.

Elsby, Michael W. L. Ryan Michaels, and Gary Solon. 2009. "The Ins and Outs of Cyclical Unemployment." *American Economic Journal: Macroeconomics* 1(1): 84-110.

Elsby, Michael W. L., Donggyun Shin, and Gary Solon. 2016. "Wage adjustment in the Great Recession and other downturns: Evidence from the United States and Great Britain." *Journal of Labor Economics* 34(S1): S249-S291.

Fallick, Bruce, and Charles A. Fleischman. 2004. "Employer-to-employer flows in the US labor market: The complete picture of gross worker flows." Mimeo.

Fujita, Shigeru, and Makoto Nakajima. 2016. "Worker flows and job flows: A quantitative investigation." *Review of Economic Dynamics* 22: 1-20.

Gabaix, Xavier. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics* 1(1): 255-294.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The dynamics of inequality." *Econometrica* 84(6): 2071-2111.

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

Gottfries, Axel. 2019. "Bargaining with Renegotiation in Models with On-the-Job Search." Mimeo, University of Edinburgh.

Gouin-Bonenfant, Emilien. 2018. "Productivity Dispersion, Between-Firm Competition, and the Labor Share." Mimeo, University of California, San Diego.

Hagedorn, Marcus, and Iourii Manovskii. 2008. "The cyclical behavior of equilibrium unemployment and vacancies revisited." *American Economic Review* 98(4): 1692-1706.

Hall, Robert E. and Paul R. Milgrom. 2008. "The limited influence of unemployment on the wage bargain." *American Economic Review* 98(4): 1653-1674.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." Review of Economics and Statistics 95(2): 347-361.

Harrison, J. Michael, and Michael I. Taksar. 1983. "Instantaneous Control of Brownian Motion." *Mathematics of Operations Research* 8(3): 439-453.

Hopenhayn, Hugo and Richard Rogerson. 1993. "Job turnover and policy evaluation: A general equilibrium analysis." *Journal of Political Economy* 101(5): 915-938.

Hornstein, Andreas, Per Krusell, and Giovanni L. Violante. 2011. "Frictional Wage Dispersion in Search Models: A Quantitative Assessment." *American Economic Review* 101: 2873–2898.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin. 2020. "Granular Search, Market Structure, and Wages." National Bureau of Economic Research, Working Paper 26239.

Kaas, Leo and Philipp Kircher. 2015. "Efficient firm dynamics in a frictional labor market." *American Economic Review* 105(10): 3030-3060.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar. 2019. "Who Profits from Patents? Rent-Sharing at Innovative Firms." *Quarterly Journal of Economics* 134(3): 1343–1404.

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

Lentz, Rasmus, and Dale T. Mortensen. 2012. "Labor market friction, firm heterogeneity, and aggregate employment and productivity." Mimeo, University of Wisconsin.

Lise, Jeremy, and Jean-Marc Robin. 2017. "The macrodynamics of sorting between workers and firms." *American Economic Review* 107(4): 1104-35.

Ljungqvist, Lars, and Thomas J. Sargent. 2017. "The fundamental surplus." *American Economic Review* 107(9): 2630-65.

Manning, Alan. 2003. *Monopsony in Motion*. Princeton University Press.

Manning, Alan. 2011. "Imperfect competition in the labor market." In David Card and Orley Ashenfelter (eds.) *Handbook of Labor Economics* 4B: 973-1041 Elsevier.

Menzio, Guido, and Shouyong Shi. 2011. "Efficient Search on the Job and the Business Cycle." *Journal of Political Economy* 119 (3): 468-510.

Moore, John. 1985. "Optimal labour contracts when workers have a variety of privately observed reservation wages." *Review of Economic Studies* 52(1): 37-67.

Mortensen, Dale T. 2003. *Wage Dispersion: Why Are Similar Workers Paid Differently?* MIT Press.

Mortensen, Dale T., and Eva Nagypal. 2007. "More on unemployment and vacancy fluctuations." *Review of Economic dynamics* 10(3): 327-347.

Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job creation and job destruction in the theory of unemployment." *Review of Economic Studies* 61(3): 397-415.

Moscarini, Giuseppe. 2005. "Job matching and the wage distribution." *Econometrica* 73(2): 481-516.

Moscarini, Giuseppe and Fabien Postel-Vinay. 2013. "Stochastic search equilibrium." *Review of Economic Studies* 80(4): 1545-1581.

Moscarini, Giuseppe and Kaj Thomsson. 2007. "Occupational and Job Mobility in the US." *Scandinavian Journal of Economics* 109(4): 807-836.

Nagypal, Eva. 2007. "Labor Market Fluctuations and On-the-Job Search." Mimeo, Northwestern University.

Oi, Walter Y. 1962. "Labor as a Quasi-Fixed Factor." *Journal of Political Economy* 70(6): 538-555.

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory*. MIT Press.

Pissarides, Christopher A. 2009. "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?" *Econometrica* 77: 1339-1369.

Postel-Vinay, Fabien and Jean-Marc Robin. 2002. "Equilibrium wage dispersion with worker and employer heterogeneity." *Econometrica* 70(6): 2295-2350.

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

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

Shimer, Robert. 2006. "On-the-job search and strategic bargaining." *European Economic Review* 4(50): 811-830.

Snell, Andy, and Jonathan P. Thomas. 2010. "Labor Contracts, Equal Treatment, and Wage-Unemployment Dynamics." *American Economic Journal: Macroeconomics* 2(3): 98-127.

Solon, Gary, Robert Barsky, and Jonathan Parker. 1994. "Measuring the Cyclicality of Real Wages: How Important is Composition Bias?" *Quarterly Journal of Economics* 109(1): 1-25.

Stokey, Nancy L. 2009. *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton (NJ): Princeton University Press.

Stole, Lars A. and Jeffrey Zwiebel. 1996. "Intra-Firm Bargaining under Non-Binding Contracts." *Review of Economic Studies* 63(3): 375-410.

Trapeznikova, Ija. 2017. "Employment Adjustment and Labor Utilization." *International Economic Review* 58: 889-922.

Van Reenen, John. 1996. "The Creation and Capture of Rents: Wages and Innovation in a Panel of U.K. Companies." *Quarterly Journal of Economics* 111(1): 195-226.

Data availability statement

The data and code underlying this research is available on Zenodo at https://doi.org/10.5281/zenodo.5221647.