A Theory of Firm Wage Dynamics *

Marc de la Barrera† Masao Fukui‡
IESE Business School *Boston University*

July 2026

Abstract

We develop a theory of firm wage dynamics that integrates the canonical wage-posting model à la Burdett and Mortensen (1998) with firm dynamics. Firms offer dynamic wage contracts under an equal-treatment constraint in the presence of search frictions. We provide an analytical characterization of equilibrium wage contracts and firm growth as functions only of the distribution of marginal surplus. Consistent with recent empirical evidence, the model implies that (i) firm wages are strongly linked to firm growth but not to firm size; (ii) firm wages decline over the firm life cycle; and (iii) the pass-through of permanent productivity shocks to wages is higher in the short run than in the long run. Firms at the top of the job ladder face excessive labor market competition, so the optimal policy subsidizes their hiring. At the macro level, the optimal policy elevates business dynamism in the steady state, and even more so along the transition.

*We are especially grateful to Philipp Grübener for encouraging us to write this paper. We thank Sebastian Di Tella, Santiago Franco, Gregor Jarosch, Guido Menzio, and Martin Souchier for helpful conversations.

†IESE Business School (e-mail: mdelabarrera@iese.edu).
‡Boston University (e-mail: mfukui@bu.edu).

1 Introduction

Firms play a pivotal role in wage determination. A long-standing literature since Abowd et al. (1999) documents that systematic differences in firm wage premia accounts for an important share of overall wage dispersion, contrary to the predictions of the competitive model of the labor market, as surveyed by Kline (2024). A workhorse model to interpret this empirical regularity is the Burdett and Mortensen (1998) (henceforth, BM) model, in which firms post wages to attract and retain workers in a labor market with search frictions. The BM model has been a highly influential framework for understanding the determinants of wage inequality and the consequences of various labor market policies (e.g., Engbom and Moser 2022, Morchio and Moser 2026).

BM makes two central predictions. The first prediction is that wages are fully endogenous to the job-ladder distribution, without introducing bargaining or preference parameters. Formally, the equilibrium wage of a firm with productivity z, which also governs the ranking of the job ladder, is given by

$$w(z) = \mathbb{E}_{\hat{G}^2}\left[\tilde{z} \mid \tilde{z} \leq z\right], \tag{1}$$

where both wages and productivities are expressed net of the reservation wage.[1] Expression (1) says that firms pay the weighted average productivity among the firms below them on the job ladder. The cumulative weight, $\hat{G}^2(z)$, is governed by the quadratic of the job-ladder distribution ($\hat{G}(z)$), which we will make precise later. This implies that the speed at which workers climb up the job ladder is the critical determinant of firm wages, and thereby of firms' monopsony power. The second prediction of BM is a tight link between firm wages and firm size. In BM, firm size is pinned down by productivity alone, $n(z)$. Since more productive firms post higher wages, they poach and retain more workers, ending up larger in the steady state. Because both $w(z)$ and $n(z)$ are strictly increasing, larger firms pay higher wages.

While BM provides an elegant departure from the competitive labor market, it faces empirical challenges. Recent evidence shows that firm size is not associated with firm wages, at odds with the central prediction of BM (Bloom et al. 2018, Diegmann et al. 2026, Haltiwanger et al. 2018).[2] Instead, firm growth appears to be a robust predictor of firm wages (Haltiwanger et al. 2018, Tanaka et al. 2023). Two additional empirical regularities support the view that firm growth, rather than firm size, is a key determinant of firm wages. First, firm wages decline over the firm life cycle once adjusted for worker composition (Schmieder 2023, Babina et al. 2019), reflecting

[1] Although (1) is not the expression BM originally derived, it is equivalent and, in our view, more interpretable. See Appendix F for the derivation and for the precise definition of $\hat{G}^2(z)$.

[2] Bloom et al. (2018) document that the firm-size-wage gradient has flattened over time. Our paper is silent on the changes over time. See Fukui et al. (2026) for a potential explanation of this trend.

the rapid employment growth of young firms. Second, following a near-permanent shock to firm productivity, the pass-through to wages is stronger in the short run than in the long run (Seegmiller 2025, Bloesch et al. 2026, Talesara and Wickard 2026), reflecting rapid firm growth immediately after the shock. Due to the lack of explicit firm dynamics in BM and subsequent work that builds upon it, there has been no framework to speak to these empirical regularities.[3]

In this paper, we fill this gap by developing a tractable framework that integrates firm dynamics into BM and show that the model naturally reproduces all aforementioned empirical regularities. Firms operate a decreasing-returns-to-scale production technology, face idiosyncratic productivity shocks, and enter and exit endogenously, as in Hopenhayn (1992). Firms pay the same wage to employees and hire workers under search frictions with on-the-job search, as in BM. Unlike BM, firms offer dynamic wage contracts that specify the wages and the firing probabilities contingent on the history of productivity shocks. The present discounted value of the contract governs the worker's job mobility decisions and thereby the ranking of the job ladder. Firms set contracts by trading off higher hiring and retention against higher wage costs.

Our first contribution is to provide a tight analytical characterization of the equilibrium wage contracts and firm dynamics. The first result shows that the equilibrium wage contracts in this rich environment take a surprisingly simple form that draws a parallel to BM:

$$\hat{V}(S_n) = \mathbb{E}_{\hat{G}^2}\left[\tilde{S}_n \mid \tilde{S}_n \leq S_n\right], \tag{2}$$

where $\hat{V}$ is the value of the contract in excess of the unemployment value, and S_n is what we call the marginal surplus. The marginal surplus is a dynamic counterpart of the static marginal product of labor and is defined as the marginal contribution of a worker to the joint surplus of the firm and workers. As $\hat{V}(S_n)$ is increasing in S_n, the ranking of the job ladder is governed by the marginal surplus. Mirroring BM, firms pay the weighted average of the marginal surplus among the firms below them on the job ladder, and the weight is given by the quadratic of the job-ladder distribution. The second result shows that the firm growth rate, dn/n, is solely a function of the marginal surplus. As the firms with higher marginal surplus have a stronger incentive to hire, the firm growth increases with the marginal surplus.

These two results naturally resolve the tensions that BM faces. First, since both the contract value and the firm growth rate are increasing in the marginal surplus, the model predicts a tight link between firm wages and firm growth. Second, when the model satisfies Gibrat's law—firm growth is independent of firm size—as it does in our calibration, wages are unrelated to firm size.

[3] As we will discuss in the literature review, existing work that integrates firm dynamics into the frictional labor market either assumes wage bargaining or allows for complete wage discrimination among employees. Our framework assumes wage posting under an equal-treatment constraint within a firm, in line with BM and the long-standing literature on monopsony models surveyed in Kline (2025).

Third, since younger firms have higher marginal surplus, the model predicts that firm wages decline over the firm life cycle. Fourth, in response to a permanent idiosyncratic productivity shock, the model predicts strong pass-through to wages in the short run, when the firm has a high marginal surplus and grows fast, which fades out in the long run, once firms have accumulated enough workers that the marginal surplus diminishes.

Beyond explaining these empirical regularities, our framework also offers a new perspective on the determinants of the firm's monopsony power and on how researchers can infer it from the data. In a standard monopsony framework, monopsony power, which is the difference between the marginal product of labor and the wage, is governed by the elasticity of firm size with respect to the wage. Since the marginal product of labor is inherently unobservable, researchers often use the pass-through from productivity shocks to wages and firm size to estimate the elasticity and thereby the monopsony power.

In our dynamic model, the firm's monopsony power is captured by the difference between the marginal surplus (S_n) and the contract value that firms offer ($\hat{V}(S_n)$). This difference is governed by the semi-elasticity of firm growth with respect to the contract value, holding the vacancy rate fixed, reiterating the point that firm growth, rather than firm size, is a key determinant of monopsony power. In principle, this semi-elasticity can be estimated from the responses of firm growth and of the present discounted value of wage payments to a productivity shock. We show, however, that the estimate must strip out the part of the firm growth response driven by vacancy creation. Failing to do so substantially understates monopsony power. This bias is particularly large for firms at the top of the job ladder, which grow primarily by posting more vacancies.

Our second contribution is to study the efficiency properties of the above equilibrium. As articulated by Menzio (2024a), the presence of wage dispersion or monopsony power under search frictions is not necessarily a sign of inefficiency. We show that, once firm dynamics are taken into account, optimal policy calls for elevating business dynamism. We reach this conclusion by identifying a novel force for excessive labor market competition at the top of the job ladder. When a firm hires workers, it destroys the surplus of other firms through poaching—a business-stealing effect. This effect is captured by the average loss in marginal surplus among firms ranked below on the job ladder:

$$\mathbb{E}_{\hat{G}}\left[\tilde{S}_n \mid \tilde{S}_n \leq S_n\right]. \tag{3}$$

By contrast, the private cost of hiring is the present discounted value of wage payments (2), which weights the same distribution quadratically. We show that the gap between the private cost (2) and the social cost (3) widens with the marginal surplus. This implies that high marginal surplus firms offer excessively high wages. Consequently, firms at the top of the job ladder under-hire, so the optimal policy subsidizes their hiring.[4]

We provide a complete characterization of the optimal policy implementation and then quantitatively study the macro consequences of the optimal policy. As the policy subsidizes the hiring of firms at the top of the job ladder, it increases overall business dynamism in the steady state. The reallocation of workers across firms increases substantially, and firm entry rises. We also find that this comes at the cost of increasing wage dispersion. Our final exercise studies the transition dynamics following the implementation of the optimal policy. We show that the policy elevates business dynamism even more along the transition than the steady-state comparison would suggest.

Related Literature. Our work builds on BM and the subsequent literature on wage-posting models with search frictions. We integrate two strands of this literature: (i) BM with idiosyncratic productivity shocks; and (ii) BM with decreasing returns to scale in production.[5] In the first strand, Coles and Mortensen (2016), Gouin-Bonenfant (2022), and Audoly (2026) develop such models while maintaining constant returns to scale in both production and vacancy (or hiring) technologies. These models therefore do not offer a satisfactory theory of firm size. In the second strand, Burdett and Vishwanath (1988), Menzio (2024b), Bilal and Lhuillier (2021), and Gottfries and Jarosch (2023) study a version of BM with decreasing returns to scale in production in the absence of firm dynamics. In this class of models, the distribution of firm growth rates is degenerate. Our contribution here is to integrate both idiosyncratic productivity shocks and decreasing returns to scale in production into a single tractable framework, thereby offering a strict generalization of Hopenhayn (1992). In so doing, our model is uniquely positioned to study how firm wages are linked to the firm life cycle and the dynamics of pass-through.

Beyond wage-posting models, another strand of the literature integrates firm dynamics into frictional labor markets with on-the-job search under alternative wage-setting protocols. Schaal (2017) assumes complete wage contracts. Bilal et al. (2022) studies a model with outside offer matching à la Postel-Vinay and Robin (2002), while McCrary (2022) considers collective bargaining without commitment. Elsby and Gottfries (2022) imposes static wage rules. With the exception of McCrary (2022), these papers do not focus on wages. Several pieces of evidence support the BM-style wage-setting protocol. Hall and Krueger (2012) documents that wages are not bargained in many low-wage labor markets. Di Addario et al. (2023) show that wages depend little

[4]There are other sources of inefficiency identified by Hosios (1990) and Fukui and Mukoyama (2025).

[5]We also build on the literature that introduces explicit dynamics into BM. Moscarini and Postel-Vinay (2013) studies aggregate productivity shocks. Fukui (2020) studies the first-order response to aggregate productivity shocks in BM with risk-averse workers, while Talesara and Wickard (2026) studies the first-order response to an idiosyncratic productivity shock. Bloesch et al. (2026) study the first-order response to an idiosyncratic productivity shock in a version of the BM model with a general recruiting technology. None of these papers features steady-state firm dynamics.

on past employers, which is hard to reconcile with models featuring outside offer matching or complete wage contracts. Finally, Hazell et al. (2022) provide evidence supporting equal treatment within firms, a central assumption in BM. We therefore view our BM-style wage-setting protocol as well suited to studying many, though not all, labor markets.

Finally, we contribute to the literature on the efficiency properties of the frictional labor market (Hosios 1990, Acemoglu 2001, Gautier et al. 2010, Cai 2020, Fukui and Mukoyama 2025). Our contribution here is to provide a complete characterization of the efficient allocation as well as the optimal policy implementation in the presence of firm dynamics. We furthermore identify a novel source of inefficiency in the equilibrium of our model: firms at the top of the job ladder face excessive labor market competition and thus under-hire.

2 Model

We start by describing the model in discrete time with a time interval $dt > 0$ in order to clarify the timing of the model. We will shortly take the continuous time limit $dt \to 0$. The time horizon is infinite. The economy is populated by a unit mass of workers and an endogenous mass of firms. For most of the paper, we focus on the steady state of the economy and drop time subscripts whenever there is no ambiguity. We study the transition dynamics in Section 7.

Production Technology. There is a single homogeneous good. Workers can be either unemployed or employed by a firm. When workers are unemployed, they produce b units of goods at home. A firm with productivity z employing n workers produces $y(z, n)$ units of goods. We assume $y(z, n)$ is strictly increasing in n and z, and weakly concave in n. Note that $y(z, n)$ can potentially include a fixed cost of operation, in which case $y(z, 0) < 0$.

Firm Demographics. Firm's productivity z evolves stochastically with the following stochastic process that is a discrete-time analogue of a diffusion process in continuous time:

$$z_{t+dt} = z_t + \mu(z_t)\, dt + \sigma(z_t)\, \sqrt{dt}\, \varepsilon_{t+dt}, \tag{4}$$

where $\mu(z)$ is the drift, $\sigma(z) \geq 0$ governs the volatility, and ε_{t+dt} follows a standard normal distribution, $\varepsilon_{t+dt} \sim \mathcal{N}(0, 1)$. Productivity shocks are idiosyncratic and independent across firms.

Firm's employment n evolves for four reasons. First, workers exogenously separate into unemployment at rate $s > 0$. Second, employed workers receive outside offers at rate $\lambda^E \geq 0$ and may quit to move to the other firms. Third, firms can fire workers at no cost. We let $\varsigma \in [0, 1]$ denote the fraction of workers fired. Finally, firms can hire workers by posting vacancies subject to vacancy costs. Vacancy costs of a firm with employment n take the form $c(v)n$, where v denotes the vacancy rate (number of vacancies divided by current employment). We assume the vacancy cost takes the isoelastic form:

$$c(v) = \bar{c}^v \frac{v^{1+\gamma}}{1+\gamma}. \tag{5}$$

We furthermore assume that the elasticity parameter satisfies $\gamma \geq 1$ throughout the paper, for the reason that will become clear later. Firms choose the vacancy rate subject to $v \geq \underline{v}$, where $\underline{v} > 0$ is an arbitrarily small lower bound that captures passive recruitment channels: a non-trivial share of hires occurs at establishments reporting no vacancies (Davis et al. 2013).[6]

The assumption on the vacancy cost function deserves some discussion. Note that our specification implies that the vacancy cost function exhibits constant returns to scale in the number of vacancies and the firm size. That is, doubling the number of vacancies and the firm size doubles the vacancy cost. We make this choice for two reasons. First, this assumption is necessary for the model to be consistent with Gibrat's law—an empirical regularity that firm growth is roughly independent of firm size. If the vacancy cost function exhibits decreasing returns to scale as in BM (i.e., doubling the vacancies and the firm size more than doubles the vacancy cost), then larger firms systematically grow more slowly.[7] Second, this assumption is precisely what makes our model tractable, as we will discuss in detail. We also note that this assumption has been commonly employed in the literature (e.g., Kaas and Kircher 2015, Gouin-Bonenfant 2022).

There is free entry. A potential firm can pay fixed cost c^e to become an operating firm. After paying the fixed cost, firms draw initial employment size, n_0, and initial productivity, z_0, from a joint distribution function $\Psi(z_0, n_0)$ with density $\psi(z_0, n_0)$. We assume entrants directly hire workers from unemployment.[8] Firms exit at an exogenous rate $\varkappa > 0$, and also can exit endogenously. We denote the total mass of firms in the economy as m and the mass of entrants per unit of time as m^0.

Preferences and Worker Demographics. All agents discount the future with the discount rate $r > 0$ and are risk-neutral. We normalize the total mass of workers to be one. Let $u \in [0, 1]$ denote the fraction of unemployed workers.

[6]The only role of the passive-hiring floor $\underline{v} > 0$ is to eliminate knife-edge equilibrium configurations in the proof of Proposition 2. All our results hold for every $\underline{v} > 0$, however small, and we set $\underline{v}$ effectively to zero in the quantitative analysis.

[7]This constant-returns-to-scale vacancy-cost assumption is the only reason our model does not nest BM. Appendix F shows that BM is recovered under the alternative cost specification $c(vn)$, which exhibits decreasing returns to scale. See also Bloesch et al. (2026) for a detailed discussion on how the returns to scale of recruiting technology affects the steady state wage markdown.

[8]We describe an alternative version of the model where entrants post vacancies as much as needed to hire n_0 workers in Appendix C.

Matching. An unemployed worker has a search efficiency of 1 and meets a firm at rate $\lambda^U > 0$. An employed worker has a search efficiency $\zeta \in (0, 1]$ and meets another firm at rate $\lambda^E = \zeta\lambda^U > 0$. A vacancy meets a worker at rate $\lambda^F > 0$. All these meeting rates are endogenously determined by a constant-returns-to-scale matching function $M(\tilde{u}, \tilde{v})$, where

$$\tilde{u} \equiv u + \zeta(1 - u) \tag{6}$$

is the total efficiency of search by workers and $\tilde{v}$ denotes the total number of vacancy postings. The equilibrium meeting rates are given by

$$\lambda^U = M(\tilde{u}, \tilde{v})/\tilde{u}, \quad \lambda^E = \zeta M(\tilde{u}, \tilde{v})/\tilde{u}, \quad \lambda^F = M(\tilde{u}, \tilde{v})/\tilde{v}. \tag{7}$$

Note that they can be expressed as functions of market tightness alone, $\theta = \tilde{v}/\tilde{u}$.

Timing. The timing within a period is as follows. First, the idiosyncratic productivity shocks are realized, and firms produce with the workers and pay wages. Second, firms exit with arrival rate $\varkappa$. Third, the exogenous separation takes place. Fourth, firms hire workers, and entrants enter and hire workers to achieve the initial firm size. At the same time, unemployed and employed workers contact their potential employers. Finally, firms fire workers and endogenously exit if they decide to do so.

Contracts. Firms offer wage contracts to workers. As in the tradition of BM, we impose equal-treatment constraint within a firm. That is, firms have to offer the same wage w to all the workers within a firm, and whenever firms fire, firms have to randomize which workers to fire. This assumption implies that all workers within a firm enjoy the same expected utility, and it rules out outside offer matching.

Firms have full commitment to the contracts. Workers cannot commit. The contract specifies wages and firing rates for all possible histories of productivity shocks. As in Spear and Srivastava (1987), firm's past commitment can be summarized by the promised utility to the workers. Let $W \geq 0$ denote the promised utility to the workers *in excess of unemployment value*, so that $W = 0$ corresponds to offering the value of unemployment.

We denote $F(V)$ as the vacancy-weighted distribution of continuation value across firms. This is the offer distribution of continuation values that the workers face upon meeting a vacancy. Likewise, we denote $G(V)$ as the search-intensity-weighted distribution of continuation value across firms. This is the distribution of continuation value that the vacancies face upon meeting a worker. Since G includes unemployed workers, it has a mass point at 0. We will formally define $F(V)$ and $G(V)$ as equilibrium objects below.

At time t, given the past commitment W_t as well as other state variables (z, n), firms offer a combination of real wage w_t, and continuation values (in excess of unemployment value) for each realization of productivity shocks z_{t+dt}. The contracts have to satisfy the promise-keeping constraints.

$$W_t = w_t dt + e^{-rdt}\left[e^{-\lambda^E dt}V_{t+dt} + (1 - e^{-\lambda^E dt})\int \max\{V_{t+dt}, \tilde{V}\}dF(\tilde{V})\right] - (1 - e^{-rdt})U, \quad (8)$$

where

$$V_{t+dt} \equiv e^{-(s+\varkappa)dt}(1 - \varsigma_t)(1 - \chi_t)\mathbb{E}[W_{t+dt}(z_{t+dt}; n, W_t)]$$

is the expected continuation value at the current firm, where χ is an indicator function that takes one if the firm exits and ς is the fraction of workers fired. The promise-keeping constraint states that the firms have to deliver the past commitment W_t either through the current wage payments w_t or through the continuation value. The continuation value derives either from staying in the current firm or moving to the new firm.

Given the contracts, the evolution of firm-level employment, conditional on no exit, is given by

$$n_{t+dt} = e^{-sdt}(1 - \varsigma_t)\left[\left(1 - (1 - e^{-\lambda^E dt})(1 - F(V_{t+dt}))\right)n_t + (1 - e^{-\lambda^F dt})G(V_{t+dt})v_t n_t\right]. \quad (9)$$

The term $e^{-sdt}(1 - \varsigma_t)$ captures the exogenous and endogenous separation. The next term $(1 - e^{-\lambda^E dt})(1 - F(V_{t+dt}))$ is the probability that a worker is poached, as workers receiving outside offers above V_{t+dt} will quit. Finally $(1 - e^{-\lambda^F dt})G(V_{t+dt})$ corresponds to the probability that a vacancy successfully hires a worker, as vacancies meeting workers with continuation value less than V_{t+dt} will hire them.

2.1 Firm's Problem

Given the state variables (z, n, W), the firm's problem is to choose the vacancy rate $v \geq 0$, the wage w, how many workers to fire $\varsigma \in [0, 1]$, whether to exit $\chi \in \{0, 1\}$, and how much utility to promise $W_{t+dt}(z_{t+dt}) \geq 0$, in order to maximize present discounted value of profits. Let $J(z, n, W)$ denote the firm's value function. The Bellman equation is

$$J(z_t, n_t, W_t) = \max_{w,v,\varsigma,\chi,n_{t+dt},W_{t+dt}} y(z_t, n_t)dt - wn_t dt - c(v)n_t dt \quad (10)$$

$$+ e^{-(r+\varkappa)dt}(1 - \chi)\mathbb{E}J(z_{t+dt}, n_{t+dt}, W_{t+dt}(z_{t+dt})) \quad (11)$$

$$\text{subject to} \quad (8) \text{ and } (9). \quad (12)$$

This original problem is quite complex. The Bellman equation features three state variables, z, n, W, and the control variables are high-dimensional, promised utility in each possible realization of productivity shocks, $W_{t+dt}(z_{t+dt})$.

We overcome this problem by showing that the firm's problem can be represented as a joint surplus of firms and workers. To this end, we guess and verify that the firm's value function takes the following form:

$$J(z, n, W) = S(z, n) - Wn. \tag{13}$$

Since the term $S(z, n)$ is the sum of firm's value and the workers' value (in excess of unemployment value), it can be interpreted as the total surplus of the firm and workers inside the firm. Importantly, the surplus function $S(z, n)$ is independent from W, which derives from the fact that utilities are transferable.

Proposition 1. *The firm's value function takes the form of $J(z, n, W) = S(z, n) - Wn$, where $S(z, n)$ is the joint surplus of the firm and workers. The surplus $S(z, n)$ and the policy functions solve the following Bellman equation:*

$$\begin{aligned} S(z_t, n_t) = \max_{v, \varsigma, \chi, n_{t+dt}, V_{t+dt}} & y(z_t, n_t)dt - c(v)n_t dt - (1 - e^{-rdt})Un_t \\ & - e^{-rdt}V_{t+dt}(1 - e^{-\lambda^F dt})G(V_{t+dt})vn_t + e^{-rdt}(1 - e^{-\lambda^E dt}) \int_{V_{t+dt}} \tilde{V} dF(\tilde{V})n_t \\ & + e^{-(r+\varkappa)dt}(1 - \chi)\mathbb{E}S(z_{t+dt}, n_{t+dt}) \end{aligned} \tag{14}$$

subject to

$$n_{t+dt} = e^{-sdt}(1 - \varsigma) \left[\left(e^{-\lambda^E dt} + (1 - e^{-\lambda^E dt})F(V_{t+dt}) \right) + (1 - e^{-\lambda^F dt})G(V_{t+dt})v \right] n_t. \tag{15}$$

Consequently, the policy functions only depend on a subset of state variables, (z, n).

The proof of this proposition as well as the rest of the proofs are collected in Appendix A. The first line in (14) is the net flow surplus of the firms and workers, which is the production minus the vacancy cost and the flow unemployment value. The first term in the second line captures the flow value that accrues to the new hires. From the viewpoint of the firm, raising the promised utility to workers has a cost of giving more resources to the new hires. The second term in the second line captures the additional value created whenever workers are poached by the other firms offering higher continuation value. We let $v(z, n)$, $\varsigma(z, n)$, $n_{t+dt}(z, n)$, $V_{t+dt}(z, n)$, $\chi(z, n)$ be the policy functions associated with the above Bellman equation, where $\chi(z, n)$ is an indicator function that takes the value of one if the firm exits.

There are several pieces to unpack in the above problem. First, the promised utility W drops from the state variable for policy functions. This is because, with transferable utility, the promised utility is irrelevant for the firm's decisions. Regardless of the past commitment, the firm can adjust current wage to achieve any desired level of next period's continuation value. This substantially simplifies the problem without the need to keep track of the promised utility as a part of the state variables.

Second, although a similar surplus representation has been established in the earlier literature on firm dynamics with on-the-job search (Schaal 2017, Bilal et al. 2022), the result here is distinct from them. In the literature, the surplus representation obtains under sufficiently rich contracts that they vary across the workers within a firm. Here, as in BM, the contracts are extremely simple and restrictive. Firms have to pay the same wage to all their workers, which rules out outside offer matching or any form of dependence on the worker specific history. Our contribution here is to show that the surplus representation still obtains under such contracts.[9]

Finally, the Bellman equation only determines the expected continuation value for the workers, V_{t+dt}, but how the continuation value is split across the realization of the states, $W_{t+dt}(z_{t+dt})$, is indeterminate. This comes from the risk-neutrality of the workers. Workers are indifferent between being offered V_{t+dt} with certainty and being offered $W_{t+dt}(z_{t+dt})$ that varies across the states, as long as the mean is the same. This indeterminacy naturally translates into the indeterminacy of the wage, which is determined residually from the promise-keeping constraint (8). It is important to note that there is no indeterminacy in terms of the allocations, such as hiring, firing, and job ladder.

However, we show in Appendix A.3 that wages are uniquely determined as long as workers have arbitrarily small risk aversion. Specifically, the split of $W_t(z_t)$ across z_t is uniquely determined so as to satisfy the following condition:

$$\tilde{w}\left(W_t(z_t^1), V_{t+dt}(z_t^1, n_t)\right) = \tilde{w}\left(W_t(z_t^2), V_{t+dt}(z_t^2, n_t)\right), \tag{16}$$

for any z_t^1 and z_t^2, where $\tilde{w}(W, V)$ is the wage implied by the promise-keeping constraint (8) given the promised utility W and the continuation value V:

$$\tilde{w}(W, V)dt \equiv W - e^{-rdt}\left[e^{-\lambda^E dt}V + (1 - e^{-\lambda^E dt})\int \max\{V, \tilde{V}\}dF(\tilde{V})\right] + (1 - e^{-rdt})U, \tag{17}$$

Therefore, wages are uniquely determined so as to minimize the fluctuations across the realizations of productivity shocks. This result is intuitive. While workers do not care about the fluctuations in wages whenever they are exactly risk-neutral, they would like to smooth the fluc-

[9]The surplus representation holds in BM as well. See Appendix F for details.

tuations the moment they are arbitrarily risk-averse. We view this as a natural refinement, and in what follows, we select the path of wages such that (16) holds.

2.2 Continuous Time Limit

Taking the continuous time limit will provide a particularly transparent and clean characterization of the equilibrium. In the limit of $dt \to 0$, the stochastic process of z_t described by (4), is governed by the following diffusion process:

$$dz = \mu(z)dt + \sigma(z)dZ_t, \tag{18}$$

where dZ_t is an increment of a standard Brownian motion.

The following problem is the continuous time limit of (14), which we derive in Appendix A.2.

Problem 1. *In the continuous time limit, the surplus function $S(z,n)$ solves the following Hamilton-Jacobi-Bellman (HJB) equation:*

$$\begin{aligned}\rho S(z,n) = \max_{v,V} y(z,n) - c(v)n - rUn - V\lambda^F G(V)vn + \lambda^E \int_V^\infty \tilde{V} dF(\tilde{V})n \\ + \mathcal{L}(z,n,v,V)S(z,n),\end{aligned}$$

where $\rho \equiv r + \varkappa$, $\mathcal{L}(z,n,v,V)$ is the generator defined for an arbitrary function $S(z,n)$:

$$\begin{aligned}\mathcal{L}(z,n,v,V)S(z,n) = \mu(z)\partial_z S(z,n) + \frac{\sigma^2(z)}{2}\partial_{zz}S(z,n) \\ + \left[\lambda^F G(V)v - s - \lambda^E(1 - F(V))\right] n\partial_n S(z,n),\end{aligned} \tag{19}$$

with the following boundary conditions:

$$S(z,n) \geq 0, \quad S_n(z,n) \geq 0. \tag{20}$$

An important object throughout the paper is what we call marginal surplus, defined as $S_n \equiv \partial_n S(z,n)$. This corresponds to the marginal contribution of a worker to the joint surplus. This is a dynamic counterpart of the marginal product of labor in static models.

We denote $v(z,n)$ and $V(z,n)$ as the policy functions associated with the above HJB equation. Taking expectations of (17) over the productivity shocks while imposing (16) and taking the continuous time limit, we have the wage offered by firm (z,n) as

$$w(z,n) = (r + s + \varkappa)V(z,n) + rU - \lambda^E \int_{V(z,n)}^\infty (\tilde{V} - V(z,n))dF(\tilde{V}) - \mathcal{L}V(z,n), \tag{21}$$

where we used the shorthand notation $\mathcal{L}V(z,n) \equiv \mathcal{L}(z,n,v(z,n),V(z,n))V(z,n)$, and this expression is for regions in which $S(z,n)>0$ and $S_n(z,n)>0$. For other regions, $V(z,n)=0$. The first term $(r+s+\varkappa)V(z,n)$ accounts for the flow value of the contracts taking into account the separation risk. The second term rU compensates for the unemployment value. The third term $-\lambda^E\int_{V(z,n)}^{\infty}(\tilde{V}-V(z,n))dF(\tilde{V})$ adjusts for the option value from outside offers. If the workers are expected to receive attractive outside offers, the firm needs to pay less in the form of wages to deliver the same value. The last term $\mathcal{L}V(z,n)$ is the compensating differential from the changes in the contract value. For example, if the contract is expected to worsen in the future, the firm needs to pay more in the form of wages to deliver the same value.

2.3 Aggregation and Equilibrium Definition

Entrants draw the initial productivity z_0 and employment n_0 from a cumulative distribution function $\Psi(z_0,n_0)$ with associated probability density function $\psi(z_0,n_0)$. We assume that entrants hire directly from the unemployed pool. This clearly implies that the entrants choose to offer zero promised utility. As a result, the free-entry condition is given by

$$\int S(z,n)d\Psi(z,n)=c^e. \tag{22}$$

Let $\phi(z,n)$ denote the measure of firms with productivity z and firm size n. It satisfies the following steady state Kolmogorov forward equation (KFE):

$$0=\mathcal{L}^{\dagger}(z,n,v(z,n),V(z,n))\phi(z,n)-\varkappa\phi(z,n)+m^0\psi(z,n), \tag{23}$$

for regions with $S(z,n)>0$ and $S_n(z,n)>0$, where m^0 is the mass of entrants. Here, $\mathcal{L}^{\dagger}$ is the adjoint of the generator $\mathcal{L}$. At the endogenous exit boundary, firms leave the economy. At the firing boundary, firms instantaneously reduce employment to the post-firing state.

Each vacancy randomly meets a worker. With probability $\frac{u}{u+\zeta(1-u)}$ vacancies meet unemployed workers who always accept the offer. With the remaining probability, the worker is employed and accepts it whenever the offered value is higher than the value that they are currently offered. Therefore the probability that a vacancy at a firm offering V hires a worker is given by

$$G(V)=\frac{1}{\tilde{u}}\left[u+\zeta\int_{(z,n):V(z,n)\leq V} n\,d\phi(z,n)\right], \tag{24}$$

where $\tilde{u}=u+\zeta(1-u)$ is the total efficiency of search by workers.

The aggregate vacancy posting is given by

$$\tilde{v} = \int v(z,n)n \, d\phi(z,n). \tag{25}$$

The vacancy-weighted distribution of continuation value is given by

$$F(V) = \frac{1}{\tilde{v}} \left[\int_{(z,n):V(z,n)\leq V} v(z,n)n \, d\phi(z,n) \right]. \tag{26}$$

The value of unemployment, U, solves

$$rU = b + \lambda^U \int_0^\infty V dF(V). \tag{27}$$

Note that the above equation already incorporates the possibility of unemployed workers hired by entrants, as entrants offer exactly the unemployment value. The steady state unemployment satisfies

$$u = 1 - \int n \, d\phi(z,n). \tag{28}$$

We define the steady state equilibrium of this economy as follows. A steady state equilibrium consists of the value and policy functions of firms $\{S(z,n), v(z,n), V(z,n), w(z,n)\}$, the two distributions of continuation value $G(V)$ and $F(V)$, the distribution of firms $\{\phi(z,n)\}$, the mass of entrants m^0, the meeting rates $\{\lambda^U, \lambda^E, \lambda^F\}$, the unemployment value U, the aggregate vacancies $\tilde{v}$ and unemployment u such that: (i) Given $\{G(V), F(V), \lambda^E, U\}$, value and policy functions $\{S(z,n), v(z,n), V(z,n), w(z,n)\}$ solve Problem 1, and wages $w(z,n)$ are given by (21). (ii) The steady state firm distribution $\{\phi(z,n)\}$ solves (23). (iii) The two distributions of continuation value $G(V)$ and $F(V)$ satisfy (24) and (26). (iv) The unemployment value U satisfies (27). (v) The aggregate vacancies $\tilde{v}$ and unemployment u are given by (25) and (28). (vi) The meeting rates $\{\lambda^U, \lambda^E, \lambda^F\}$ are given by (7). (vii) The free-entry condition is satisfied, (22).

3 Equilibrium Characterization

We first show that the equilibrium distributions cannot have a mass point or a gap in the support.

Proposition 2. *In any equilibrium, the distributions $F(V)$ and $G(V)$ have no mass point at any $V > 0$, and $\operatorname{supp} F = \operatorname{supp} G$ is a connected interval containing zero. In particular, neither support has a gap.*

The underlying logic of Proposition 2 is the same as BM. If there were a mass point in the promised utility distribution, the firm at the mass point could improve profits by slightly changing the promised utility. Likewise, if there is a gap, then the firm just above the gap could improve profits by slightly lowering the promised utility.

Proposition 2 may come as a surprise to some readers. While it is well-known since BM that the mass point in wage distribution cannot arise under constant returns to scale production function, both Burdett and Vishwanath (1988) and Menzio (2024b) point out that the mass point is rather a feature under decreasing returns to scale production function, which is the case here. To see this, suppose that in a model like BM, there is a mass point in wages that are equal to the marginal product of labor of firms offering the mass point wage. If the production function is constant returns to scale, then these firms earn zero profits, so slightly reducing the wage must be a profitable deviation. Under decreasing returns to scale, however, these firms earn positive profits because, while the profit from the marginal worker is zero, firms profit from the inframarginal workers. In this case, slightly reducing the wage is not necessarily a profitable deviation, as the gain in profit from the marginal worker may be overwhelmed by the loss in profit from the inframarginal workers leaving the firm.

The key difference between the above argument and here is that, while models in the tradition of BM restrict wages to be constant over time, firms are allowed to offer time-varying wage contracts here. In our environment, in a situation like the above, firms are allowed to deviate to offer a slightly worse contract only for today. In continuous time, such a deviation affects only the marginal workers, who are hired today and poached today, without affecting the inframarginal workers. Put differently, the dynamic contract allows firms to price discriminate marginal workers from the inframarginal workers, which eliminates the mass point.[10]

Proposition 2 implies that the distributions F and G are continuous at every $V > 0$ with a common connected support. In what follows, we further assume that they are continuously differentiable at every $V > 0$, a property we verify in our constructed equilibrium, and let $f(V)$ and $g(V)$ be the densities of F and G, respectively. The first-order condition with respect to V is given by

$$\left[\lambda^F g(V)v + \lambda^E f(V)\right] (\partial_n S(z,n) - V) = \lambda^F G(V)v, \tag{29}$$

for $V > 0$. The left-hand side is the marginal value of offering higher promised utility per worker. Firms can hire at an additional rate $\lambda^F g(V)v$ of workers and prevent workers from being poached at an additional rate of $\lambda^E f(V)$. These are multiplied by the marginal net value of a worker,

[10]No part of the proof relies on firm heterogeneity or on the presence of the idiosyncratic productivity shocks. This implies that even when all firms are symmetric, the equilibrium offer distribution cannot feature a mass point, suggesting oscillatory behavior. This is precisely the conjecture of Burdett and Vishwanath (1988) (see their footnote 10).

$(\partial_n S - V)$. The right-hand side is the cost of offering higher promised utility per worker, which corresponds to the hiring rate. This reflects that the cost of offering higher promised utility is that firms have to pay more to new hires, with whom firms do not have existing commitments. At the optimum, the marginal benefit and cost must be equal.

The first-order condition (29) is related, but distinct from, the familiar Lerner formula for wage markdowns. To see this, we can rewrite the expression as

$$\underbrace{\frac{\partial_n S(z,n) - V}{V}}_{\text{firm's surplus share}} = \underbrace{\lambda^F G(V) v}_{\text{hiring rate}} \times \underbrace{\frac{1}{\epsilon_{dn/n,V}}}_{\substack{\text{inverse semi-elasticity} \\ \text{of growth rate}}}, \tag{30}$$

where $\epsilon_{dn/n,V} \equiv \frac{\partial(dn/n)}{\partial \ln V}$ is the semi-elasticity of the firm growth rate with respect to the contract value. The left-hand side is the dynamic analogue of the static wage markdown. We call it the firm's surplus share. The right-hand side is the inverse semi-elasticity of the firm *growth* rate with respect to the contract value, multiplied by the hiring rate. When the growth rate is perfectly elastic to the contract value ($\epsilon_{dn/n,V} \to \infty$), the firm optimally sets the markdown to zero, which resembles the competitive outcome. To the extent that the growth rate is inelastic, the firm optimally charges a positive markdown. Unlike the standard Lerner formula, the elasticity is no longer with respect to the size or the level of the labor supply to the firm. This hints at the idea that, in our model, monopsony power is determined by firm growth rather than firm size. The inverse semi-elasticity is multiplied by the hiring rate, because for a given semi-elasticity, a higher hiring rate makes it more costly to offer higher value.

Another important observation is that the semi-elasticity $\epsilon_{dn/n,V}$ concerns the *partial* derivative that holds the vacancy rate fixed. This implies that estimates of the semi-elasticity must strip out the part of the firm growth response driven by vacancy creation. We quantify the importance of this correction for inferring the firm's monopsony power in Section 6.8.

The first-order condition with respect to v is

$$\lambda^F G(V) \left(\partial_n S(z,n) - V\right) = c'(v), \tag{31}$$

for the interior solution $v > \underline{v}$. The left-hand side is the marginal benefit of a vacancy. Each vacancy hires a worker with probability $\lambda^F G(V)$, and each worker adds the net value of $(\partial_n S - V)$. The right hand side is the marginal cost of vacancy. At the optimum, these two must be equal.

An important observation from (29) and (31) is that firm's optimal promised utility and vacancy decisions are solely summarized by marginal surplus $S_n \equiv \partial_n S(z,n)$. This follows from the fact that (z,n) enter into the optimality conditions only through S_n. The assumption of a constant returns to scale vacancy cost function plays a crucial role in this result. With constant returns to scale, the vacancy rate depends on firm size only through the marginal surplus, as (31) shows. Otherwise, the vacancy rate in general depends directly on the firm size. With this result in mind, we will often write the policy functions in terms of the marginal surplus, which we denote as $\hat{V}(S_n(z,n)) \equiv V(z,n)$ and $\hat{v}(S_n(z,n)) \equiv v(z,n)$.

Since the marginal benefit of increasing promised utility – the left-hand side of (29) – is strictly increasing in the marginal surplus, it is tempting to think that the firms with higher marginal surplus offer higher promised utility. While this observation is obviously correct whenever the vacancy rate is exogenous, the endogenous vacancy introduces additional complication. With endogenous vacancy creation, firms with higher marginal surplus have two ways of increasing their workforce: offering higher utility or posting more vacancies. If the vacancy creation is sufficiently elastic, firms would primarily use vacancy postings as a way to increase their workforce. This increase in new hires, in turn, discourages firms from offering higher utility because the cost of doing so increases with the relative size of new hires – the right-hand side of (29).

We show, however, that as long as the elasticity of vacancy creation satisfies $\gamma \geq 1$, the promised utility offer is strictly increasing in the marginal surplus.

Proposition 3. *Assume $\gamma \geq 1$. Promised utility offer $\hat{V}(S_n)$ is strictly increasing in S_n, and consequently, workers at firm (z,n) accept the job offer from firm (z',n') if and only if $S_n(z',n') \geq S_n(z,n)$.*

The monotonicity of promised utility implies that the job ladder is ranked according to the marginal surplus. This rank-preserving property is precisely what makes the wage posting models tractable. At any point in time, we can define two rankings of the job ladder at the firm-level, differing in terms of how the ranking is weighted. The first is the ranking of the marginal surplus weighted by the mass of search efficiency of workers:

$$\hat{G}(S_n) \equiv \frac{1}{\tilde{u}} \left[u + \zeta \int_{S_n(z,n) \leq S_n} n d\phi(z,n) \right], \tag{32}$$

which corresponds to the probability that a vacancy meets a worker with marginal surplus below S_n. The second is the vacancy-weighted ranking of the marginal surplus

$$\hat{F}(S_n) \equiv \frac{1}{\tilde{v}} \left[\int_{S_n(z,n) \leq S_n} v(z,n) n d\phi(z,n) \right], \tag{33}$$

which corresponds to the probability that a worker meets a vacancy with marginal surplus below S_n.

The monotonicity of the promised utility offer implies

$$G(\hat{V}(S_n)) = \hat{G}(S_n), \quad F(\hat{V}(S_n)) = \hat{F}(S_n). \tag{34}$$

Exploiting the rank-preserving property, we have a tight analytical characterization of the equilibrium policies as a function of the distribution of the marginal surplus.

Proposition 4. *Assume $\gamma \geq 1$. The promised utility each firm with marginal surplus S_n offers is given by*

$$\hat{V}(S_n) = \frac{1}{\hat{G}^2(S_n)} \int_0^{S_n} \tilde{S}_n d\hat{G}^2(\tilde{S}_n) \tag{35}$$

$$\equiv \mathbb{E}_{\hat{G}^2}\left[\tilde{S}_n | \tilde{S}_n \leq S_n\right], \tag{36}$$

where $\hat{G}^2(S_n)$ denotes the quadratic of $\hat{G}(S_n)$ and $\mathbb{E}_{\hat{G}^2}[\tilde{S}_n|\tilde{S}_n \leq S_n]$ is the weighted average of marginal surplus conditional on being below S_n based on cdf $\hat{G}^2(\tilde{S}_n|\tilde{S}_n \leq S_n)$.

While wage posting models are often considered challenging to solve due to the infinite-dimensional fixed point problem, we obtain a closed-form solution to the firm's policies conditional on the distribution of the marginal surplus. The expression (36) draws a strong parallel to the wage expression in BM, (1), with two differences. First, the static wage in BM is replaced by the contract value, which corresponds to the present discounted value of the wage payments that workers are expected to receive. Second, the firm productivity is replaced by the marginal surplus, which is a dynamic counterpart of the static marginal product of labor.

The rest of the features have many commonalities with BM. First, the equilibrium contract is fully endogenously determined as a function of job-ladder distribution. For example, even if a firm has an extremely high marginal surplus, $S_n \to \infty$, the contract value converges to the unconditional mean, $\lim_{S_n \to \infty} \hat{V}(S_n) = \mathbb{E}_{\hat{G}^2}[\tilde{S}_n]$. This widening gap between the marginal surplus and the contract value implies that the high marginal surplus firms endogenously obtain extreme monopsony power. Second, as the equilibrium contract depends on the entire job-ladder distribution, it limits the influence of the unemployment outside option, in line with the evidence from Jäger et al. (2020).[11]

Interestingly, the distribution of the marginal surplus is weighted by the *quadratic* of $\hat{G}$, as in BM. This quadratic weighting is a direct consequence of the two motives for firms to offer higher wages: to attract more workers and to retain existing workers. In fact, as we later discuss

[11] This draws a sharp distinction from the bargaining model of McCrary (2022). There, the worker value is a function of the unemployment outside option and the average surplus of the employer and does not depend on the job-ladder distribution.

in Section 5.1, in a model only with retention motive, the distribution of the marginal surplus is weighted by $\hat{G}$ instead of $\hat{G}^2$.

The next proposition shows that the firm growth rate is only a function of the marginal surplus as well.

Proposition 5. *Firm growth rate is only a function of the marginal surplus, and is given by*

$$\frac{dn}{n} = \lambda^F \hat{G}(S_n)\hat{v}(S_n) - s - \lambda^E(1 - \hat{F}(S_n)), \tag{37}$$

with the vacancy rate of the form:

$$\hat{v}(S_n) = c'^{-1}\left(\lambda^F \hat{G}(S_n)(S_n - \hat{V}(S_n))\right). \tag{38}$$

As both the job-ladder ranking and vacancy rate only depend on the marginal surplus, the firm growth rate is also only a function of the marginal surplus. Naturally, the firm growth is strictly increasing in the marginal surplus.

Propositions 4 and 5 jointly imply the tight link between firm wages and firm growth, as both are increasing functions of the marginal surplus. This feature stands in sharp contrast to BM, which predicts a strong relationship between firm size and wages, while the firm growth distribution is degenerate. Moreover, to the extent that our calibration is consistent with Gibrat's law, which will be the case, the model further predicts little to no relationship between firm size and wages. As we discuss in more detail below, the strong link between firm growth and wages, together with the *lack* of a relationship between firm size and wages, is broadly consistent with the existing empirical evidence.

4 Efficiency and Optimal Policy

As Proposition 4 shows, our model delivers endogenous monopsony power that is strongly heterogeneous across firms. As discussed in detail by Menzio (2024a), market power stemming from search frictions is not necessarily a sign of inefficiency. To understand the efficiency property of the equilibrium, we study the planning problem and optimal policies that implement the planner's solution.

4.1 Planning Problem

We consider the problem of a social planner who maximizes the expected discounted consumption in the economy subject to the search frictions.

Problem 2. *The constrained planner solves*

$$\max_{\{\phi_t, v_t, m_t^0, dn_t, \varsigma_t^e, \varsigma_t^f, x_{t,zn,z'n'}, x_{t,u,zn}\}} \int_0^\infty e^{-rt} \left\{ \int (y(z,n) - c(v_t(z,n))n)\, d\phi_t + (1 - \textstyle\int n d\phi_t) b - m_t^0 c^e \right\} dt$$

subject to

$$\begin{aligned}
\partial_t \phi_t(z,n) = &-\partial_n \left[dn_t(z,n)\phi_t(z,n)\right] - \partial_z[\mu(z)\phi_t(z,n)] + \frac{1}{2}\partial_{zz}[(\sigma(z))^2 \phi_t(z,n)] \\
&- (\varkappa + \varsigma_t^e(z,n))\phi_t(z,n) + m_t^0 \psi(z,n) \qquad (39)\\
dn_t(z,n) = &\left[\lambda_t^F G_t^p(z,n) v_t(z,n) - s - \lambda_t^E (1 - F_t^p(z,n))\right] n - \varsigma_t^f(z,n) \qquad (40)
\end{aligned}$$

where

$$\begin{aligned}
G_t^p(z,n) &\equiv \frac{1}{\tilde{u}_t}\left[x_{t,u,zn}\left(1 - \int n d\phi_t\right) + \zeta \int x_{t,z'n',zn} n' d\phi_t(z',n')\right] \\
1 - F_t^p(z,n) &\equiv \frac{1}{\tilde{v}_t} \int x_{t,zn,z'n'} v_t(z',n') n' d\phi_t(z',n'),
\end{aligned}$$

λ_t^F *and* λ_t^E *are given by* (7), $\tilde{u}_t$ *and* $\tilde{v}_t$ *are given by* (6) *with* (28) *and* (25), *respectively, and* $x_{t,u,zn} \in [0,1]$, $x_{t,zn,z'n'} \in [0,1]$, $v_t(z,n) \geq 0$, $\varsigma_t^e(z,n) \geq 0$ *and* $\varsigma_t^f(z,n) \geq 0$.

The objective function is the present discounted value of aggregate consumption. The constraints (39)-(40) describe the law of motion of the firm distribution. The terms G_t^p and F_t^p are the distributions of the job ladder, where $x_{t,zn,z'n'}$ describes the probability that a worker at firm (z,n) moves to firm (z',n') conditional on meeting, and $x_{t,u,zn}$ describes the probability that an unemployed worker is hired by a firm (z,n) conditional on meeting. The planner chooses these mobility decisions, vacancy postings, as well as the endogenous firing and exit decisions, ς_t^e and ς_t^f.

We first define two objects. First, let $S_t^p(z,n)$ denote the co-state variable (Lagrange multiplier) associated with the law of motion of $\phi_t(z,n)$. This variable captures the social surplus of the firm with productivity z and employment size n. Second, define

$$\eta^U \equiv -\frac{\partial \ln \lambda^U}{\partial \ln \tilde{u}} = 1 - \frac{\partial \ln M(\tilde{u},\tilde{v})}{\partial \ln \tilde{u}} \qquad (41)$$

$$\eta^F \equiv -\frac{\partial \ln \lambda^F}{\partial \ln \tilde{v}} = 1 - \frac{\partial \ln M(\tilde{u},\tilde{v})}{\partial \ln \tilde{v}} \qquad (42)$$

as the steady-state elasticity of congestion from additional worker search and vacancy posting, respectively.

The following proposition provides a complete characterization of the planner's solution.

Proposition 6. *In the steady state, the efficient allocation has the following properties. The social surplus of a firm solves*

$$\rho S^p_t(z,n) = y(z,n) - c(v^p(z,n))n - bn - \overbrace{\lambda^F \left\{ \hat{G}(S^p_n)\mathbb{E}_{\hat{G}}[\tilde{S}^p_n | \tilde{S}^p_n \leq S^p_n] + \eta^F \Omega \right\} v^p(z,n)n}^{\text{social cost of vacancy}} \tag{43}$$
$$- \underbrace{\lambda^U \left\{ \mathbb{E}_{\hat{F}}[\tilde{S}^p_n] - \eta^U \Omega \right\} n}_{\text{social value of off-the-job search}} + \underbrace{\lambda^E \left\{ (1-\hat{F}(S^p_n))\mathbb{E}_{\hat{F}}[\tilde{S}^p_n | \tilde{S}^p_n \geq S^p_n] - \eta^U \Omega \right\} n}_{\text{social value of on-the-job search}} + \mathcal{L}^p S^p(z,n)$$

with the boundary conditions $S^p(z,n) \geq 0$ and $S^p_n(z,n) \geq 0$, where

$$\mathcal{L}^p S^p(z,n) \equiv dn_t(z,n) S^p_n(z,n) + \mu(z) S^p_z(z,n) + \frac{1}{2}(\sigma(z))^2 S^p_{zz}(z,n)$$

and Ω is the average social gains from a meeting in the steady state defined as

$$\Omega \equiv \int G^p(z',n') \left(S^p_n(z',n') - \mathbb{E}_{\hat{G}}[\tilde{S}^p_n | \tilde{S}^p_n \leq S^p_n(z',n')] \right) \frac{1}{\tilde{v}} v^p(z',n') n' d\phi_t(z',n'). \tag{44}$$

The distribution functions $\hat{F}$ and $\hat{G}$ are defined as in (33) and (32), respectively. The vacancy rate is

$$v^p(z,n) = c'^{-1} \left(\lambda^F \left\{ \hat{G}(S^p_n) \left(S^p_n(z,n) - \mathbb{E}_{\hat{G}}[\tilde{S}^p_n | \tilde{S}^p_n \leq S^p_n] \right) - \eta^F \Omega \right\} \right). \tag{45}$$

The unemployed workers accept the job as long as $S^p_n(z,n) \geq 0$, and the employed workers at firm (z',n') accept the job offer from firm (z,n) if and only if $S^p_n(z,n) \geq S^p_n(z',n')$. The entry satisfies

$$\int S^p(z,n) d\Psi(z,n) = c^e.$$

The structure of the above characterization mirrors that of the equilibrium. The social surplus of a firm $S^p(z,n)$ is the analogue of the equilibrium surplus $S(z,n)$. Conditional on the surplus, (45) describes the efficient vacancy creation, and the same boundary conditions dictate the efficient firm exit and firing decisions. The job ladder is ranked according to the planner's marginal surplus, S^p_n. The free entry condition is the same as in the equilibrium, except that the surplus is evaluated at the planner's marginal surplus.

The important differences lie inside the HJB equation that defines the surplus function. To see this clearly, recall the equilibrium surplus solves the following HJB equation:

$$\begin{aligned}\rho S(z,n) = y(z,n) - c(v(z,n))n - bn - \overbrace{\lambda^F \hat{G}(S_n)\hat{V}(S_n)v(z,n)n}^{\text{private cost of vacancy}} \\ - \underbrace{\lambda^U \mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n)]n}_{\text{private value of off-the-job search}} + \underbrace{\lambda^E(1-\hat{F}(S_n))\mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n)|\tilde{S}_n \geq S_n]n}_{\text{private value of on-the-job search}} + \mathcal{L}S(z,n)\end{aligned}. \tag{46}$$

Below, we provide a heuristic discussion of the differences in each term in the HJB equation. Later, we formally prove how these differences translate into the systematic pattern of optimal policies.

The first difference between (43) and (46) is the last term in the first line, which captures the difference in the social and the private cost of vacancy posting and is the novel source of inefficiency. In the planner's solution, the term consists of the business stealing effect, $\mathbb{E}_{\hat{G}}[\tilde{S}_n^p|\tilde{S}_n^p \leq S_n^p]$ and the congestion effect $\eta^F\Omega$. The business stealing effect corresponds to the loss in the surplus from poaching, which is the average of the marginal loss among the firms ranked below in the job ladder. The congestion effect captures the expected loss in the surplus from a unit vacancy posting, which corresponds to the average meeting gains. By contrast, in equilibrium, the analogous term is $\hat{V}(S_n) \equiv \mathbb{E}_{\hat{G}^2}[\tilde{S}_n|\tilde{S}_n \leq S_n]$, which corresponds to the contract value offered and is the private cost of hiring a worker.

Interestingly, while the private cost is the weighted average with distribution $\hat{G}^2$, the social cost uses the distribution $\hat{G}$. As $\hat{G}^2$ first order stochastically dominates $\hat{G}$, the gap between $\mathbb{E}_{\hat{G}^2}[\tilde{S}_n|\tilde{S}_n \leq S_n]$ and $\mathbb{E}_{\hat{G}}[\tilde{S}_n|\tilde{S}_n \leq S_n]$ grows in marginal surplus S_n. Consequently, the wages offered by firms at the top of the job ladder are excessively high relative to other firms. This discourages the vacancy creation by the top firms in the job ladder, and as a result, top firms under-post vacancies relative to other firms, as the difference between (38) and (45) shows.

This result contrasts with the normative implications of monopsony models based on workers' idiosyncratic taste differences (e.g., Card et al. 2018, Berger et al. 2022, Kline 2025). There, wages of high labor market power firms are relatively too low. Here, wages of high labor market power firms (high marginal surplus firms) are relatively too high. Excessive labor market competition among the firms at the top of the job ladder is a novel source of inefficiency. This mechanism is absent in the canonical BM model, as we show in Appendix F.2.

The rest of the differences come from workers' search and are similar to the forces emphasized in Fukui and Mukoyama (2025). The term $\mathbb{E}_{\hat{F}}[\tilde{S}_n^p] - \eta^U\Omega$ captures the social value of off-the-job search. The first term $\mathbb{E}_{\hat{F}}[\tilde{S}_n^p]$ corresponds to the direct gain of unemployed workers finding a job. This direct gain is partially offset by the loss due to the congestion effect, $\eta^U\Omega$, captured by the second term. The extent of congestion depends on the elasticity of the matching function, η^U, and the average gain from a meeting in the economy, Ω. Likewise, the term $(1-\hat{F}(S_n^p))\mathbb{E}_{\hat{F}}[\tilde{S}_n^p|\tilde{S}_n^p \geq$

$S_n^p] - \eta^U \Omega$ captures the social value of on-the-job search. The direct gain from a job-to-job transition is $(1 - \hat{F}(S_n^p))\mathbb{E}_{\hat{F}}[\tilde{S}_n^p | \tilde{S}_n^p \geq S_n^p]$ and is partially offset by the loss due to the congestion effect, $\eta^U \Omega$.

The equilibrium analogues of these terms are $\mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n)]$ and $(1 - \hat{F}(S_n))\mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n) | \tilde{S}_n \geq S_n]$, which are the private values of off-the-job search and on-the-job search, respectively. The difference arises because, in equilibrium, workers appreciate only a part of the marginal surplus due to the monopsony power while not internalizing the congestion forces. For both the planner and the equilibrium, the difference between the value of on-the-job search and off-the-job search governs how private agents value the search component of employment. We call this difference the net value of on-the-job search.

As emphasized by Fukui and Mukoyama (2025), the private agents tend to overvalue employment especially at the top of the job ladder. The intuition is easiest to see in the case of $\lambda^U = \lambda^E$. In this case, at the bottom of the job ladder ($S_n = 0$), the net value of on-the-job search is zero in both the planner's solution and the equilibrium. By contrast, at the top of the job ladder ($S_n \to \infty$), the net value of on-the-job search is larger in equilibrium than in the planner's solution. This is because the value of on-the-job search is zero in both cases, while the value of off-the-job search is larger in the planner's solution relative to the equilibrium, $\mathbb{E}_{\hat{F}}[\tilde{S}_n] > \mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n)]$. This implies that the employment is over-valued at the top of the job ladder. Intuitively, the private agents undervalue the value of off-the-job search because workers do not fully appreciate the gains from search due to the monopsony power.

4.2 Optimal Policies

We now show that the efficient allocation characterized in Proposition 6 can be implemented with two policy instruments: a vacancy tax and a labor tax on firms. These taxes depend on the firm's productivity and size, (z, n). Specifically, we assume that a firm (z, n) pays a tax of $\tau^v(z, n)$ for each unit of vacancy they post, and a tax of $\tau^n(z, n)$ for each worker they employ. These taxes are financed by lump-sum tax on workers. With these taxes, the HJB equation for the surplus function $S(z, n)$ becomes

$$\begin{aligned}\rho S(z, n) = \max_{v,V} \; & y(z, n) - c(v)n - rUn - \tau^n(z, n)n - \left[V \lambda^F G(V) + \tau^v(z, n)\right] vn \\ & + \lambda^E \int_V^\infty \tilde{V} dF(\tilde{V})n + \mathcal{L}(z, n, v, V)S(z, n),\end{aligned}$$

which replaces Problem 1. The rest of the equilibrium conditions remain unchanged.

The following proposition characterizes the taxes that implement the efficient allocation.

Proposition 7. *Assume $\gamma \geq 1$. Suppose that the taxes are given by*

$$\tau^v(z,n) = \lambda^F \left\{ \hat{G}(S_n^p)\mathbb{E}_{\hat{G}}[\tilde{S}_n^p | \tilde{S}_n^p \leq S_n^p] + \eta^F \Omega \right\} - \lambda^F \hat{G}(S_n^p)\hat{V}(S_n^p) \tag{47}$$

$$\begin{aligned} \tau^n(z,n) = \lambda^U \left\{ \mathbb{E}_{\hat{F}}[\tilde{S}_n^p] - \eta^U \Omega \right\} - \lambda^E \left\{ (1-\hat{F}(S_n^p))\mathbb{E}_{\hat{F}}[\tilde{S}_n^p | \tilde{S}_n^p \geq S_n^p] - \eta^U \Omega \right\} \\ - \lambda^U \mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n^p)] + \lambda^E (1-\hat{F}(S_n^p))\mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n^p) | \tilde{S}_n^p \geq S_n^p], \end{aligned} \tag{48}$$

where we used short-hand notation of $S_n^p \equiv S_n^p(z,n)$, and all the variables are evaluated at the efficient allocation including $\lambda^U, \lambda^F, \lambda^E, \hat{G}, \hat{F}, \Omega, \eta^U$, and η^F. Then, the equilibrium with taxes implements the efficient allocation characterized in Proposition 6. Moreover, in such an equilibrium, $S(z,n) = S^p(z,n)$ for all (z,n).

The vacancy tax $\tau^v(z,n)$ corrects for the difference between the social cost of vacancy and the private cost of vacancy, evaluated at the efficient allocation. The labor tax $\tau^n(z,n)$ corrects for the difference between the net social value of on-the-job search and the net private value of on-the-job search, evaluated at the efficient allocation. As these two margins are the only differences between the planner's problem and the equilibrium, correcting them naturally implements the efficient allocation.

Importantly, we do not need entry taxes, firing taxes, or job mobility taxes to implement the efficient allocation. It is worth emphasizing that our assumption on the vacancy technology is crucial for this result. If the vacancy technology is not constant returns to scale or $\gamma < 1$, then the equilibrium job ladder is no longer ranked according to the marginal surplus, while the planner's job-ladder is still ranked by the marginal surplus.[12]

The following proposition tightly characterizes the cross-sectional properties of the taxes.

Proposition 8. *The optimal taxes given in Proposition 7 satisfy the following properties.*

1. *The vacancy tax τ^v is strictly decreasing in the marginal surplus S_n^p, i.e., for any two firms (z,n) and (z',n') with $S_n^p(z,n) > S_n^p(z',n')$, we have $\tau^v(z,n) < \tau^v(z',n')$. Furthermore, $\tau^v(z,n) > 0$ for firms with sufficiently low marginal surplus $S_n^p(z,n)$.*

2. *The labor tax τ^n is strictly increasing in the marginal surplus S_n^p, i.e., for any two firms (z,n) and (z',n') with $S_n^p(z,n) > S_n^p(z',n')$, we have $\tau^n(z,n) > \tau^n(z',n')$.*

The vacancy tax declines with the job-ladder rank because firms face excessive competition at the top of the job ladder, as discussed earlier. It is, therefore, optimal to subsidize vacancy creation by the top firms relative to the others. We can further show that, at the bottom of the job

[12]See Talesara and Wickard (2026) for a detailed discussion on the distorted job ladder under decreasing returns to scale vacancy technology.

ladder, the vacancy creation must face a positive tax to alleviate congestion. The employment tax increases with the job-ladder rank because employment is over-valued at the top of the job ladder, as discussed earlier and in Fukui and Mukoyama (2025). Interestingly, we show in Appendix F.3 that in BM, the optimal vacancy and employment taxes are both constant across all firms. This suggests the distinct role of firm dynamics in drawing policy implications.

5 Extensions

To transparently convey the model mechanism, the baseline model deliberately abstracts from some important features of the real world labor market. We now discuss several extensions and modifications to the baseline model.

5.1 Hiring Costs and Costless Vacancy

In the baseline model, vacancy posting is costly but hiring worker *per se* is free. Some existing work in the literature reverses these assumptions: vacancy posting is free, but hiring is costly (e.g., Coles and Mortensen 2016, Elsby and Gottfries 2022, Audoly 2026).

Appendix G studies this alternative version of the model with hiring costs and costless vacancy. We show that, in an equilibrium with on-the-job search,[13] the equilibrium contracts and the job ladder are ranked by the marginal surplus, as in the baseline model. The equilibrium contract is given by

$$\hat{V}(S_n) = \mathbb{E}_{\hat{G}}\left[\tilde{S}_n \mid \tilde{S}_n \leq S_n\right]. \tag{49}$$

This expression resembles (36) in the baseline model. In both models, the contract value is given by the weighted average of the marginal surplus among firms below them on the job ladder.

The difference is that, while in the baseline model, the weight is given by the quadratic of the search intensity-weighted employment distribution, $\hat{G}^2$, in (49), the weight is given by the non-quadratic of $\hat{G}$. This result is intuitive. In the baseline model, firms have dual motives to raise wages: to raise retention and to hire more workers. This dual motive is reflected in quadratic weighting. In an environment with free vacancy posting, the only reason for firms to raise wages is to raise retention because firms can always hire more by posting more vacancies. This single motive is reflected in non-quadratic weighting.[14]

[13] In the absence of vacancy costs and thereby search frictions, there always exists an equilibrium without on-the-job search, in which the labor market is perfectly competitive with no unemployment and no wage dispersion. This equilibrium is basically the same as the one in Hopenhayn (1992) and Hopenhayn and Rogerson (1993). In Appendix G, we focus on the equilibrium with on-the-job search.

[14] This result generalizes Audoly (2026) to decreasing-returns-to-scale production.

We furthermore show that, unlike in the baseline model, the equilibrium with hiring costs and costless vacancies is constrained efficient in the sense made precise in Appendix G. Even without search frictions, business-stealing effects are a potential source of inefficiency: when firms hire workers, they do not necessarily internalize the loss in surplus incurred by the firms they poach from. This business-stealing effect is captured by the average loss in marginal surplus among firms ranked below them on the job ladder. In equilibrium, (49) shows that firms offer exactly this average loss as promised utility. Therefore, private firms fully internalize the business-stealing effect, and the equilibrium is constrained efficient.

5.2 Ex-ante Worker Heterogeneity

In the baseline model, all the workers are ex-ante homogeneous. Appendix H extends the baseline economy to allow for many ex-ante heterogeneous worker types, $a \in \mathcal{A}$, which may represent gender, race, or skill levels. Types can differ in flow values of unemployment, search efficiencies, separation rates, matching technologies, and their contribution to production. The production at a firm with productivity z and workforce $\{n^a\}_{a\in\mathcal{A}}$ is given by a general function $y(z, \{n^a\}_{a\in\mathcal{A}})$, which allows for arbitrary complementarity or substitutability among worker types within a firm. With segmented search by worker type, each type has its own job ladder and promised-utility distribution. The baseline characterization applies type by type: contracts are ranked by the type-specific marginal surplus and the mapping from marginal surplus to promised utility is unchanged within each worker type.

5.3 Permanent Firm Heterogeneity

In the baseline model, all the firm heterogeneity is persistent but not permanent. There is some evidence that permanent firm heterogeneity is important to account for firm growth (Sterk et al. 2021). Appendix I allows firms to draw permanent types ϑ at entry. Firm types can differ in technologies, vacancy costs, productivity processes, separation rates, and exogenous exit rates. With permanent firm heterogeneity, each firm is indexed by (z, n, ϑ), but the equilibrium ranking continues to be summarized by marginal surplus. The equilibrium contracts take exactly the same form as in the baseline model, while the distribution of the marginal surplus is pooled across the permanent types ϑ.

6 Quantitative Exploration

We quantitatively explore both the positive and normative implications of our model.

6.1 Parametric Assumptions

The following proposition shows that we can reduce the dimensionality of the state space under commonly used assumptions.

Proposition 9. *Suppose that $y(z,n)$ is homogeneous of degree one in (z,n) and z follows a geometric Brownian motion, $\mu(z) = \mu z$ and $\sigma(z) = \sigma z$. Then, the surplus function $S(z,n)$ takes the form of $S(z,n) = z\hat{S}(\hat{n})$, where $\hat{n} \equiv n/z$ and $\hat{S}(\hat{n})$ solves*

$$\hat{\rho}\,\hat{S}(\hat{n}) = \max_{v,V}\ \hat{y}(\hat{n}) - c(v)\hat{n} - rU\hat{n} - V\lambda^F G(V)v\hat{n} + \lambda^E \int_V \tilde{V} dF(\tilde{V})\,\hat{n} + \hat{\mathcal{L}}(\hat{n},v,V)\hat{S}, \tag{50}$$

where $\hat{\rho} \equiv r + \varkappa - \mu$ and the generator is

$$\hat{\mathcal{L}}(\hat{n},v,V)\hat{S} = \frac{\sigma^2\hat{n}^2}{2}\hat{S}''(\hat{n}) + \left[\lambda^F G(V)v - s - \lambda^E(1-F(V)) - \mu\right]\hat{n}\,\hat{S}'(\hat{n}), \tag{51}$$

with boundary conditions $\hat{S}(\hat{n}) \geq 0$ and $\hat{S}'(\hat{n}) \geq 0$.

Under a homogeneous production function and geometric Brownian motion, we only need to keep track of employment-to-productivity ratio, $\hat{n} \equiv n/z$, which reduces the dimensionality of the state space from two to one. This brings about a large computational advantage. While solving the steady state is feasible even without this reduction, solving the transition dynamics with two state variables is prohibitively expensive, as we discuss in Section 7. The assumption of geometric Brownian motion is natural to impose in order to be consistent with Gibrat's law and the power law in the firm size distribution, as explained in Gabaix (1999) and Luttmer (2007). The homogeneity of production function is more restrictive, as it rules out fixed costs of operation and thereby endogenous exit.

In line with Proposition 9, we parametrize the production function as

$$y(z,n) = z^{1-\alpha}n^{\alpha}, \tag{52}$$

where $\alpha \in (0,1)$ is the returns to scale of production, and we assume the productivity process is given by a geometric Brownian motion: $\mu(z) = \mu z$ and $\sigma(z) = \sigma z$.

We assume the Cobb-Douglas matching function:

$$M(\tilde{u},\tilde{v}) = A\tilde{u}^{\eta}\tilde{v}^{1-\eta}, \tag{53}$$

where A is the matching efficiency parameter, and η is the elasticity of matching function with respect to the job search by workers. The distribution of entrants, Ψ, is assumed to be the mass point distribution at (z^0, n^0).

6.2 Calibration

Armed with the functional form assumptions, we choose the parameter values as follows. One period in the model corresponds to a year. We first set a subset of parameters from external sources. We set the discount rate to 5% at the annual frequency, $r = 0.05$. The matching function elasticity parameter is set to $\eta = 0.5$, as suggested by Petrongolo and Pissarides (2001). The returns to scale parameter, α, is set to 0.64, a standard value in the literature (e.g., Hopenhayn and Rogerson 1993, Elsby and Gottfries 2022). The vacancy cost elasticity, γ, is set to 3.5, as suggested by Bilal et al. (2022).

We set the exit rate to 9%, $\varkappa = 0.09$, which corresponds to the overall exit rate at the establishment level in the Business Dynamics Statistics (BDS) in 2023. We measure the units of vacancy so that the steady state market tightness is 1. The matching efficiency parameter, A, is then set to 4.8, which implies the steady state monthly unemployment to employment transition rate is 40% (=4.8/12). We show in Appendix D that the value of $\bar{c}^v$ can be normalized without loss of generality and set it to $\bar{c}^v = 60$. We then set the relative efficiency of on-the-job search to $\zeta = 0.2$, following Borovičková and Shimer (2024). We also simply set $b = 0$ to reduce the number of free parameters.

The remaining six parameters, $\{\mu, \sigma, s, z^0, n^0, c^e\}$, are internally calibrated to match six moments: (i) a 16% employment share at establishments with more than 1,000 employees; (ii) a standard deviation of firm growth of 0.41, from Elsby and Michaels (2013); (iii) a non-employment rate of 7%; (iv) an average establishment size of 18.3; (v) an average age-zero establishment size of 7.5; and (vi) market tightness equal to 1, which normalizes the units of vacancies. Moments (i), (iv), and (v) are from BDS. Heuristically, these moments identify the parameters as follows. The drift μ governs the thickness of the right tail of the firm size distribution, while volatility σ controls firm growth dispersion. The separation rate s primarily pins down the non-employment rate. The entrant productivity z^0 and employment n^0 discipline average establishment size and entrant size, respectively, and the entry cost c^e clears the market at the normalized tightness.

Note that our calibration does not target any moments related to wages. This reflects the fact that our model has no free parameters that independently control wages. All the parameters are tightly disciplined by standard firm dynamics moments and labor market flows. We then ask what our model says about firm wage dynamics.

6.3 Steady State Computational Algorithm

The equilibrium has the following structure. The general equilibrium (GE) objects are the two distributions $F(V)$ and $G(V)$ and the market tightness θ. Given these objects, firms solve the HJB equation and the KFE gives the steady-state firm distribution. The GE objects, in turn, must be

Parameter	Value	Description	Source/Target
A. Externally assigned parameters			
r	0.05	Discount rate	Annual 5% interest rate
α	0.64	Returns to scale	Standard
$\varkappa$	0.09	Exit rate	BDS 2023
γ	3.5	Vacancy cost elasticity	Bilal et al. (2022)
η	0.5	Matching elasticity	Petrongolo and Pissarides (2001)
$\bar{c}^v$	60	Vacancy cost level	Normalization
A	4.8	Matching efficiency	Monthly UE rate 40%
ζ	0.2	On-the-job search efficiency	Borovičková and Shimer (2024)
b	0.0	Flow value of leisure	Baseline
B. Internally calibrated parameters			
μ	0.04	Productivity growth drift	Emp. share 1,000+ firms 16%
σ	0.4	Productivity growth variance	std($\Delta \ln(n)$) of 0.41
s	0.25	Exogenous separation rate	Non-employment rate 7%
c^e	131	Entry cost	Market tightness 1
z^0	310	Entrants initial productivity	Average firm size 18.3
n^0	3.5	Entrants initial employmet	Average firm size age 0

Table 1: Parameter Values

Note: The table reports the parameter values used in our quantitative exercise.

consistent with the outcomes of the HJB and KFE. Since both $F(V)$ and $G(V)$ are distributions, solving for equilibrium amounts to finding an infinite-dimensional fixed point, which poses a significant computational challenge. This is why the wage posting models like BM have not been studied outside of the analytically solvable cases.

Our analytical characterization, in particular Proposition 4, helps us overcome this computational challenge. Since equilibrium contracts are explicitly characterized by the surplus function and the job-ladder distributions, the equilibrium system collapses to a system of nonlinear differential equations. We solve these equations through an iterative procedure.

We briefly describe the computational algorithm and relegate the details to Appendix B. We start by guessing the GE objects $\{F(V), G(V), \theta\}$. Given them, we solve the HJB equation to obtain the value and policy functions of firms. We then solve the KFE to obtain the steady-state distribution. Given the value function and the steady-state distribution, we use Proposition 4 to update the guess of $\{F(V), G(V)\}$. We repeat this process until we find fixed points in $\{F(V), G(V)\}$, while iterating over θ to ensure that the free-entry condition is satisfied. In practice, we find the equilibrium in a relatively small number of iterations, and the total computational time is tens of seconds on our laptop.

We discuss the existence of the steady state equilibrium in Appendix K under weak regularity conditions. Although we do not have any theoretical results on the uniqueness of the steady state

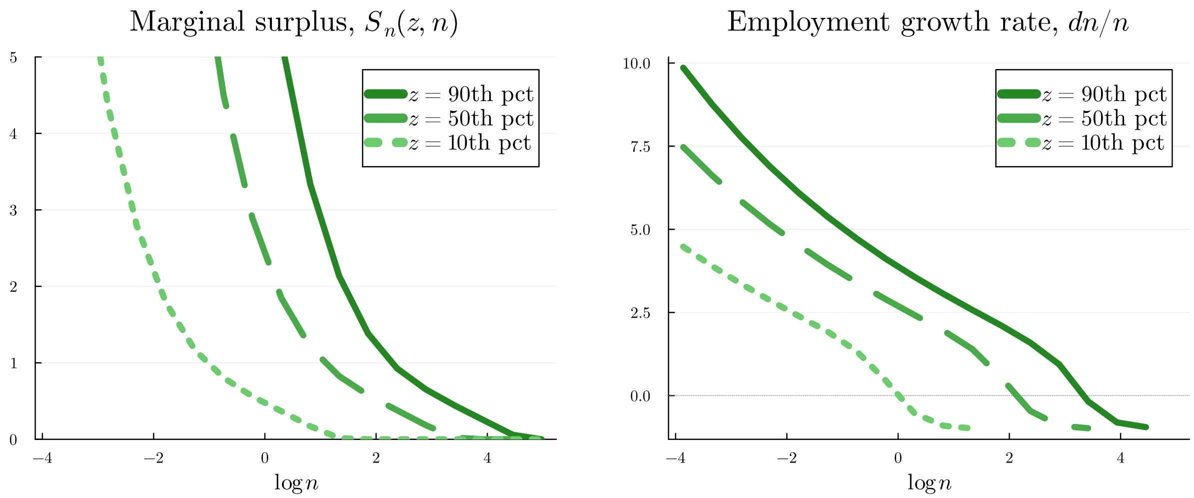

Figure 1: Marginal surplus and employment growth rate

Note: The figure plots the marginal surplus and employment growth rate of firms against the log employment size for three different values of productivity, 90th percentile, median, and 10th percentile of the steady state distribution. The left panel plots the marginal surplus and the right panel plots the employment growth rate.

equilibrium, we did not find any sign of multiplicity in our extensive exploration. Specifically, for a wide range of initial guesses of the GE objects, we always find the same equilibrium.

6.4 Equilibrium Properties

We first illustrate the equilibrium properties of the model.

Figure 1 plots the marginal surplus and employment growth rate of firms against the log employment size for three different values of productivity. The marginal surplus on the left panel shows that, for a given productivity, the marginal surplus is strictly decreasing in the employment size, which comes from the diminishing marginal product of labor. Eventually, the marginal surplus reaches zero when the firm size is large enough, above which the firm will fire workers. Naturally, the marginal surplus is strictly increasing in the productivity level.

As Proposition 5 demonstrates, the marginal surplus closely maps into the employment growth rate on the right panel of Figure 1. When the marginal surplus is higher, the firm grows faster driven by both the higher retention rate and higher hiring rate. This implies that, for given productivity, smaller firms grow faster, and for given employment size, firms with higher productivity grow faster. When the firms are unproductive or too big, so that the marginal surplus is low, the firm growth turns negative. As the marginal surplus decreases with firm size, the employment growth ensures the mean-reverting property of the marginal surplus.

Figure 2a plots the steady state distribution of marginal surplus. Due to the idiosyncratic productivity shocks and the search frictions, the distribution of marginal surplus is highly dispersed even in the steady state. The distribution of marginal surplus plays a crucial role for all the subsequent results. As Proposition 3 highlights, the ranking in this distribution pins down the ranking of the wage contracts and the job ladder. Proposition 4 further shows that the equilibrium value of contracts is given by the weighted average of the marginal surplus of the firms below them on the job ladder.

Figure 2b plots the equilibrium promised utility and wage against the log employment size for three different values of productivity. For a given productivity, the promised utility decreases with employment size, as the marginal surplus falls. When the employment size is small enough, the value of contract becomes flat because the firm sits at the top of the job ladder, in which case the firm offers the unconditional weighted average of the marginal surplus in the entire economy. For any given firm size, the contract value increases with productivity, as the marginal surplus is higher for higher productivity firms.

The right panel shows the wage. While the wage mostly tracks a similar pattern to the promised utility, interestingly, the wage is not necessarily monotone in the employment size or in the productivity. For example, starting from the very small firm size, the wage goes up with employment size, even though the contract value is decreasing. This is due to the compensating differential. As the firm is expected to grow rapidly with small firm size, the firms are also expected to offer less attractive contracts in the near future. In order to deliver the given contract value, the firms have to offer higher wages to compensate for the deteriorating contract value in the future. At the other extreme, when the firm is large enough, the wage goes up with employment size again, as there is a higher firing risk, and the firm has to offer higher wages to compensate for it.

Figure 3 expresses the contract value and wage as a function of the marginal surplus and the marginal product of labor, respectively. The left panel plots $V(z,n)/S_n(z,n)$ against the firm size for three different values of productivity. In many existing models with bargaining under marginal surplus sharing rule (Stole and Zwiebel 1996, Brügemann et al. 2019), such as Elsby and Michaels (2013) and Acemoglu and Hawkins (2014), this share is given by an exogenous bargaining power parameter.[15] Our model, by contrast, gives rise to an endogenous bargaining power, as in the tradition of BM. In particular, the worker has relatively stronger bargaining power when there are many firms near them on the job ladder. Since the marginal surplus distribution is unimodal (Figure 2a), the competition is strongest when the marginal surplus is in the middle of the distribution. The right panel shows the wage markdown, which is the ratio of the wage to the marginal product of labor, $w(z,n)/y_n(z,n)$. Overall, the wage markdown tends to rise with the firm size and fall with the productivity. The non-monotonicity of the wage translates into the non-monotonicity of the wage markdown.

[15]In McCrary (2022), the share of worker value in *average* surplus, $\frac{V}{S(z,n)/n}$, is a constant. In our model, this object endogenously varies and closely tracks the pattern in Figure 3 as well.

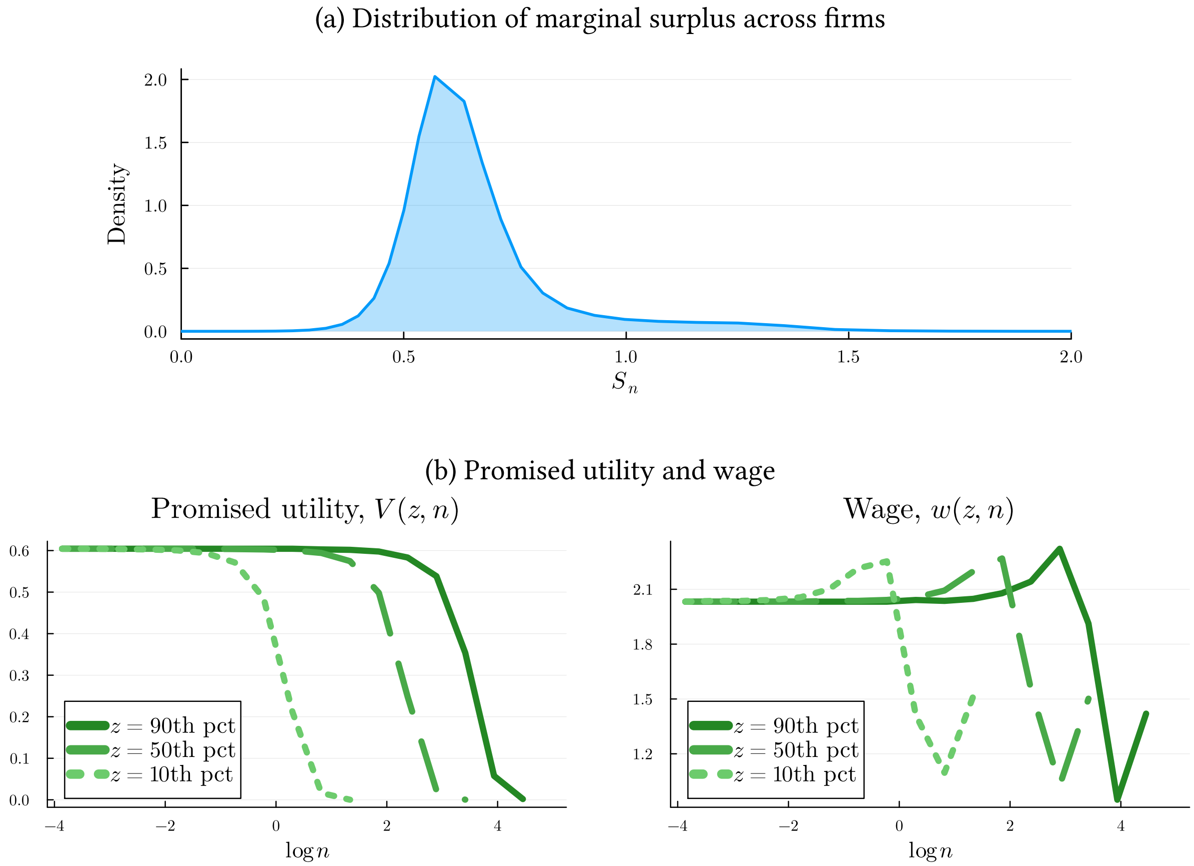

Figure 2: Distribution of marginal surplus and equilibrium promised utility and wage

Note: Figure 2a plots the distribution of marginal surplus across firms. Figure 2b plots the equilibrium promised utility and wage against the log employment size for three different values of productivity, 90th percentile, median, and 10th percentile of the steady state distribution.

6.5 Firm Wage Correlates with Firm Growth, but not with Firm Size

Having understood the conditional relationships, we now study how the firm wage varies unconditionally by observable characteristics in order to confront the model predictions with the data counterparts.

The left panel of Figure 4 shows that the firm wage is strongly positively correlated with the firm growth rate. Moving from the firms with growth rate of less than -50% to more than 50% is associated with a roughly 60% increase in the firm wage. This is a natural consequence of Propositions 4 and 5. As shown there, both the contract value and firm growth increase with the marginal surplus. The result here demonstrates that this relationship is quantitatively large and that the cross-sectional variation in the firm growth rate is an important source of frictional wage dispersion.

By contrast, the right panel of Figure 4 shows that firm wages are unrelated to firm size. As

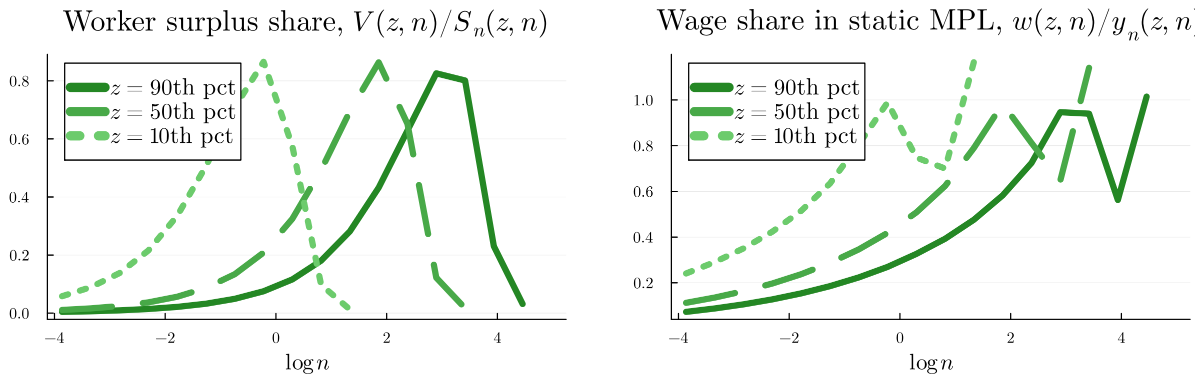

Figure 3: Wage markdown

Note: The left panel plots $V(z,n)/S_n(z,n)$ against log employment size for three productivity values. The right panel plots the wage markdown, $w(z,n)/y_n(z,n)$.

explained earlier, this is the natural consequence of Gibrat's law, which is satisfied in our model. Appendix Figure E.11 shows that the firm growth is uncorrelated with the firm size except for the very small firms. If the wage is strongly related to firm growth, and firm growth is unrelated to firm size, then the model naturally predicts firm wage to be unrelated to firm size. We can alternatively understand the result as an outcome of two offsetting forces. On one hand, holding productivity fixed, larger firms have lower marginal surplus and thus pay lower wages. On the other hand, larger firms tend to be more productive, as shown in Appendix Figure E.9, which results in higher marginal surplus. In our calibration, these two forces cancel each other out, resulting in a flat relationship between the wage and the firm size.

The strong relationship between firm wages and firm growth, together with the *lack* of a relationship between firm wages and firm size, is broadly consistent with the existing empirical evidence. Bloom et al. (2018) document that firm wage is no longer strongly related to firm size in the recent US data. Diegmann et al. (2026) show that within the same firm, the wage does not vary with establishment size in Germany. Haltiwanger et al. (2018) find evidence in support of the wage ladder but not the size ladder in the US. Instead, the firm wage is strongly related to the firm growth rate (net job flows) (Haltiwanger et al. 2018, Tanaka et al. 2023). These empirical findings are hard to square from the perspective of the canonical BM model. In BM, the firm wage must be strictly monotone in the firm size, and there is no variation in the firm growth rate. Our finding here is that once firm dynamics is incorporated, the model provides a natural interpretation of these empirical regularities.

6.6 Firm Wage Declines over the Firm Life Cycle

Figure 5 plots the firm size and firm wage over the firm life cycle. The left panel shows the evolution of the firm size, which grows over the life cycle. As the firms are born small, their marginal

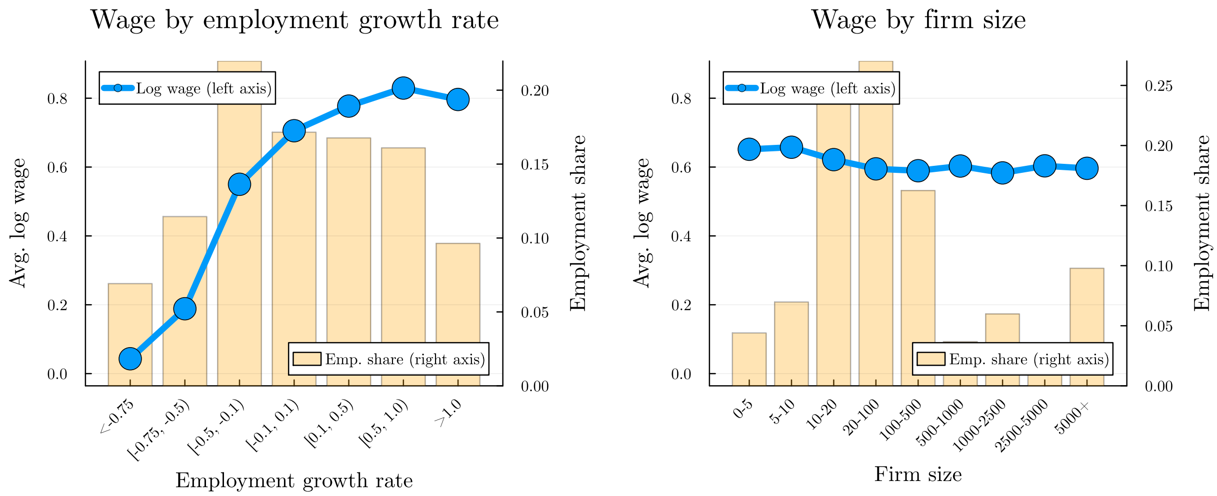

Figure 4: Firm wage by size and growth

Note: The figure plots the average log wage of firms by growth and size. The blue line in the left panel plots the average log wage of firms for each employment growth rate bin. The blue line in the right panel plots the employment-weighted average log wage of firms for each size bin. In both panels, the bars show the employment share of firms in each bin.

surplus is initially high, leading to the initial rapid growth. As firms get larger, their marginal surplus falls and the growth becomes more modest. The right panel shows the evolution of the firm wage. Since the marginal surplus decreases over the life cycle on average, the firm wage also decreases over the life cycle.[16] The decline of firm wage over the life cycle is quantitatively large, with firms with age 0 paying over 10% higher wage than the firms with age 10.

The life-cycle pattern of the firm wage is not only qualitatively but also quantitatively consistent with the empirical findings in Schmieder (2023). Using the German data, he documents that, after controlling for the worker and firm fixed effects, the firm wage monotonically decreases over the firm life cycle with a decline of around 10% from age 0 to age 10. In a similar vein, Babina et al. (2019) use the US data to document that the younger firms pay higher wages after controlling for the worker and firm fixed effects. They find that firms with age 0 pay around 4 log points higher wage than the firms with age 10, somewhat smaller in magnitude than in Schmieder (2023).

6.7 Pass-Through of Productivity Shocks to Wage

A large literature uses the pass-through of productivity shocks to wages and employment to infer firms' monopsony power, as we discuss below. Most existing studies interpret the pass-through through the lens of models that are either static or do not have a notion of firms. Since our model is explicitly dynamic and has a well-defined notion of firms, it provides an ideal laboratory for

[16] Appendix Figure E.10 shows that the productivity z grows over the life cycle on average. Therefore, the decline in the firm wage and the marginal surplus are not due to the decline in the productivity.

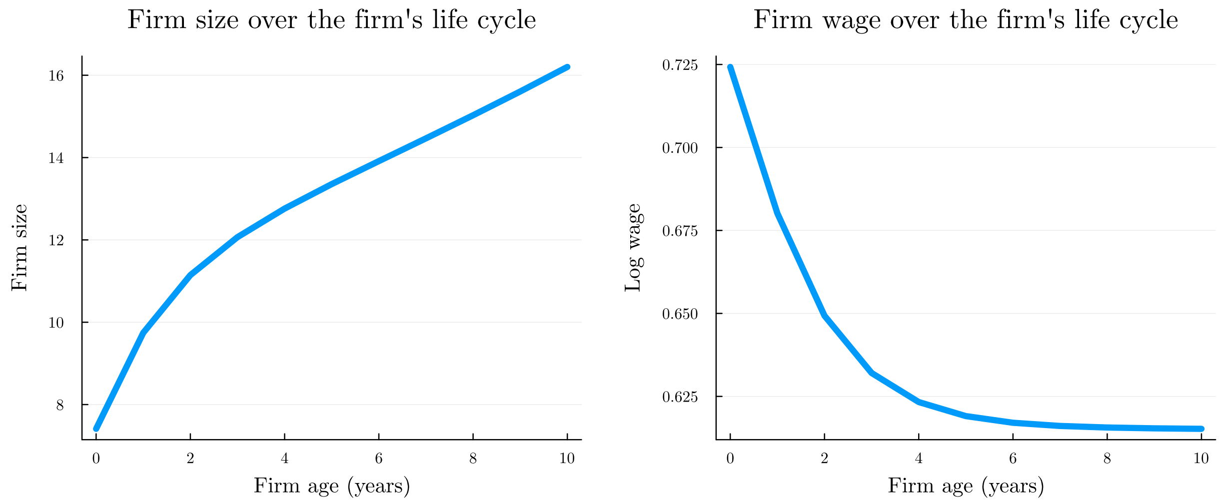

Figure 5: Firm wage and firm age

Note: The figure plots the average log wage and firm size over the firm life cycle.

studying and interpreting the pass-through of productivity shocks to wages and employment.

Figure 6 plots the impulse response of the wage and employment to a 10% permanent positive productivity shock at the firm level. We shock each firm in our economy and average the impulse response across firms. The top-left panel plots the shock path. The top-right panel plots the impulse response of the wage. First, for the entire impulse, the pass-through is far below one. Upon impact, the pass-through to wage is around 0.3. The wage pass-through monotonically declines over time and fades out in the long run. The bottom-left panel plots the response of employment and the bottom-right panel plots the response of employment growth. These two figures show that the employment grows rapidly initially and then slows down over time. In the long run, the employment rises one for one with the productivity, at which point the growth rate converges back to zero.

What underlies these patterns is the dynamics of the marginal surplus. Upon impact, the marginal surplus jumps up, which leads to a higher contract value, higher wage, and higher employment growth. As the firms grow in size, the marginal surplus diminishes over time, which leads to a decline in the wage and employment growth. Eventually, in the long run, the marginal surplus converges back to the original level on average, in which case the wage and firm growth go back to the original level as well.

Existing evidence is supportive of these mechanisms. Seegmiller (2025) exploits shocks to firms' stock market value to estimate wage and employment responses that are in line with our model's predictions. In particular, he finds strong short-run responses of employment growth and wages, both of which fade out at longer horizons. Likewise, Bloesch et al. (2026) and Talesara and Wickard (2026) exploit export and value-added shocks, respectively, to document broadly similar patterns.

We decompose the response of employment growth into three components: (i) the response of vacancy yield; (ii) the response of retention rate; and (iii) the response of vacancy rate. Formally, we have that $dn/n = \lambda^F G(V)v - s - \lambda^E(1 - F(V))$. To a first-order approximation, we have

$$\Delta\left(\frac{dn}{n}\right) \approx \underbrace{v \times \Delta(\lambda^F G(V))}_{\text{(i) vacancy yield}} - \underbrace{\Delta(\lambda^E(1 - F(V)))}_{\text{(ii) retention rate}} + \underbrace{\lambda^F G(V) \times \Delta v}_{\text{(iii) vacancy rate}}, \tag{54}$$

where Δx denotes the change in x relative to the pre-shock level. The vacancy yield is the change in the offer acceptance probability per unit vacancy, as a result of offering a higher contract value. The retention rate is the change in the probability that a worker rejects the outside offer, again as a result of offering a higher contract value. The vacancy rate is the change in the vacancy rate holding the vacancy yield fixed.

The bottom-right panel shows the decomposition of employment growth into each of the three components in (54).[17] We find roughly 40% of the employment growth comes from the vacancy yield, another 40% comes from the retention rate, and the remaining 20% comes from the vacancy rate. This decomposition is important for the inference of monopsony power, which we discuss next.

6.8 Inferring Monopsony Power from Pass-Through

Inferring monopsony power from the data is challenging precisely because we rarely observe the marginal product of labor. To overcome this challenge, it has been common in the literature to infer the monopsony power using the pass-through estimates (see Kline (2025) for a review). In particular, one can take the ratio of the employment response to wage response to form an estimate of the labor supply elasticity. In many models, the labor supply elasticity is informative of wage markdowns through the Lerner formula. As explained in the context of (30), our model does not fall into the class of the standard monopsony framework. Therefore, the natural question is: how and what can we learn about the monopsony power from the pass-through estimates if our model is the data generating process?

Formula (30) highlights that the right notion of "labor supply elasticity" here is the semi-elasticity of the firm growth rate with respect to the contract value, as opposed to the elasticity of the firm employment level with respect to the wage. In constructing this, we can devise a ratio of growth rate changes to the contract value changes using Figure 6 with two caveats. First, the change in contract value must be recovered from the present-value worker equation, using the expected path of wages and adjusting for separation, outside-offer option values, and changes in future continuation values. Second, the change in growth rate must exclude the vacancy rate

[17] While (54) relies on a first-order approximation and is not exact, the errors are negligible in our exercise.

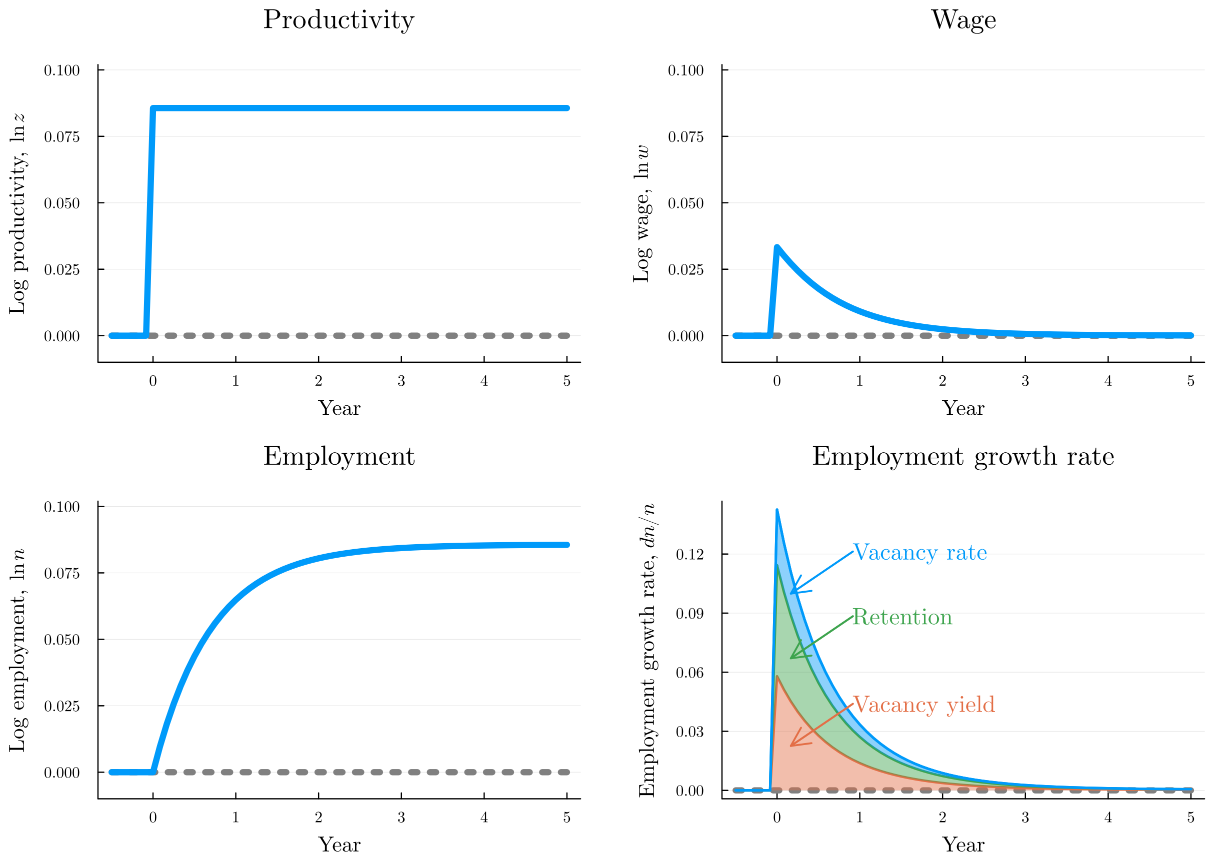

Figure 6: Pass-Through of Productivity Shocks to Wage and Employment

Note: The figure plots the impulse responses to a 10% permanent positive productivity shock at the firm level, averaged across firms. The top-left panel plots the shock path, the top-right panel plots the wage response, the bottom-left panel plots the employment response, and the bottom-right panel plots the employment growth response together with its decomposition in (54).

response, as the semi-elasticity is the *partial* derivative with respect to the contract value, which holds the vacancy rate fixed. We focus on the importance of the second caveat. Although this issue has been discussed in Caldwell et al. (2026), its quantitative relevance has remained an open question, and our model provides an ideal laboratory to address it.

Table 2 highlights the importance of excluding the vacancy response when calculating the labor supply elasticity. The first row reports the impact response of employment growth to a 10% permanent productivity shock, as in Figure 6. This response is 15% when the vacancy rate response is included, but 12% when it is excluded, so that firm growth comes only from vacancy yield and retention (see (54)). Dividing each by the contract value response yields an estimate of the semi-elasticity in (30), and hence of the implied surplus share of firms, which we report as $1 - V/S_n$.[18] Column three uses the response including the vacancy response (incorrectly),

[18] While the Lerner formula (30) expresses the surplus share as $(S_n - V)/V$, here and in Table 2 we report the monotone transformation $1 - V/S_n = \frac{(S_n-V)/V}{1+(S_n-V)/V}$.

	Emp. growth response		Implied surplus share		True
	w/ Δv	w/o Δv	w/ Δv	w/o Δv	surplus share
Aggregate	0.15	0.12	0.26	0.34	0.38
By job-ladder rank					
1st quartile (top)	0.12	0.05	0.11	0.29	0.34
2nd quartile	0.16	0.12	0.17	0.24	0.27
3rd quartile	0.19	0.16	0.2	0.24	0.27
4th quartile (bottom)	0.14	0.13	0.41	0.44	0.5

Table 2: Inferring Firm's Surplus Share ($1 - V/S_n$) from Pass-Through

Note: The two columns under "Emp. growth response" report the impact response of employment growth to a 10% permanent positive productivity shock, as in Figure 6. Column "w/ Δv" shows the employment growth response that includes all components in (54). Column "w/o Δv" shows the employment growth that only accounts for (i) vacancy yield and (ii) retention rate and excludes (iii) vacancy rate. Implied surplus share computes the firm's surplus share using (30). Column "w/ Δv" (incorrectly) computes the semi-elasticity with $\epsilon_{dn/n,V} \equiv \frac{\Delta(dn/n)}{\Delta \ln V}$. Column "w/o Δv" (correctly) computes the semi-elasticity excluding the vacancy rate response with $\epsilon_{dn/n,V} \equiv \frac{\Delta(dn/n) - \lambda^F G(V)\Delta v}{\Delta \ln V}$. The true surplus share, as well as all the other surplus share measures, refer to $1 - V/S_n$. The first row aggregates the results across all firms. Rows under "By job-ladder rank" show the results by the first quartile, second quartile, third quartile, and fourth quartile of the job-ladder ranking (equivalently, the marginal surplus rank).

while column four excludes it (correctly). Failing to exclude the vacancy response substantially underestimates the firm's monopsony power. The final column reports the true surplus share, which is aligned with the implied surplus share in column four.[19]

The bias in monopsony power estimates is even more problematic when we look into the heterogeneity along the job-ladder ranking. The second to fifth rows of Table 2 repeat the same exercise after splitting the sample into four quartiles of the job-ladder rank. The key observation is that at the top of the job ladder, the employment growth response comes predominantly from the vacancy response (1st quartile). This is because, since firms are already offering the most attractive contracts at the top, there is no reason for them to offer higher value even when marginal surplus is higher. Instead, they increase the workforce by increasing the vacancy postings. The opposite is true at the bottom of the job ladder (4th quartile). This heterogeneity in the importance of vacancy response across the job-ladder ranking implies that the underestimation of monopsony power is more severe at the top of the job ladder. In fact, the failure to exclude the vacancy response leads the researchers to conclude that the firms at the top of the job ladder have the lowest monopsony power, when they actually have the second highest monopsony power.

[19]They are not perfectly aligned as a combination of the grid spacing and the numerical differentiation in our semi-elasticity calculation.

6.9 Optimal Policies

Building on the analytical characterization in Section 4, we now quantitatively evaluate the optimal policies.

Figure 7 plots the optimal vacancy and optimal employment taxes. The left panel shows that the optimal vacancy tax increases with firm size and decreases with productivity, in line with Proposition 8. The magnitude is substantial. Firms at the top of the job ladder (smaller or more productive firms) receive a vacancy subsidy of around 10% of the average wage per vacancy, whereas firms at the bottom (larger or less productive firms) pay a tax of around 40%. The right panel shows the employment tax, which decreases with firm size and increases with productivity, consistent with Proposition 8. The variation is much more modest than for the vacancy tax. Moving from the top to the bottom of the job ladder shifts the employment tax from 22% to 13% of the average wage.

Table 3 summarizes the aggregate steady-state outcomes in the equilibrium without policy and with optimal policy. The first two rows show that both aggregate output and consumption increase by around 1% under the optimal policy. Market tightness decreases substantially, since the employment tax corrects for the over-valuation of jobs, as in Fukui and Mukoyama (2025). This naturally leads to declines in both the unemployment-to-employment (UE) and employment-to-employment (EE) rates, though the decline in the EE rate is substantially more muted than that in the UE rate. As a result, the acceptance probability, defined as the ratio of the EE rate to the UE rate, increases by around 10%. Since the composition of vacancy shifts more toward high marginal surplus firms, workers are more likely to take the outside offer. In line with this, the reallocation rate, defined as the size-weighted mean of absolute employment growth rates, also increases by around 8%. The average firm size is smaller, meaning that firm entry increases. However, these rises in business dynamism substantially raise wage dispersion.

Overall, the optimal policy subsidizes the growth of high marginal surplus firms and results in elevated business dynamism, at the expense of increasing wage dispersion. Consistent with the idea that the optimal policy increases business dynamism, Appendix Figure E.12 shows that the optimal policy speeds up the firm growth for smaller firms and slows down the firm growth for larger firms conditional on the same productivity. Appendix Figure E.13 shows that the optimal policy speeds up the firm growth at the early stage of firm life cycle but slows it down at the later stage.

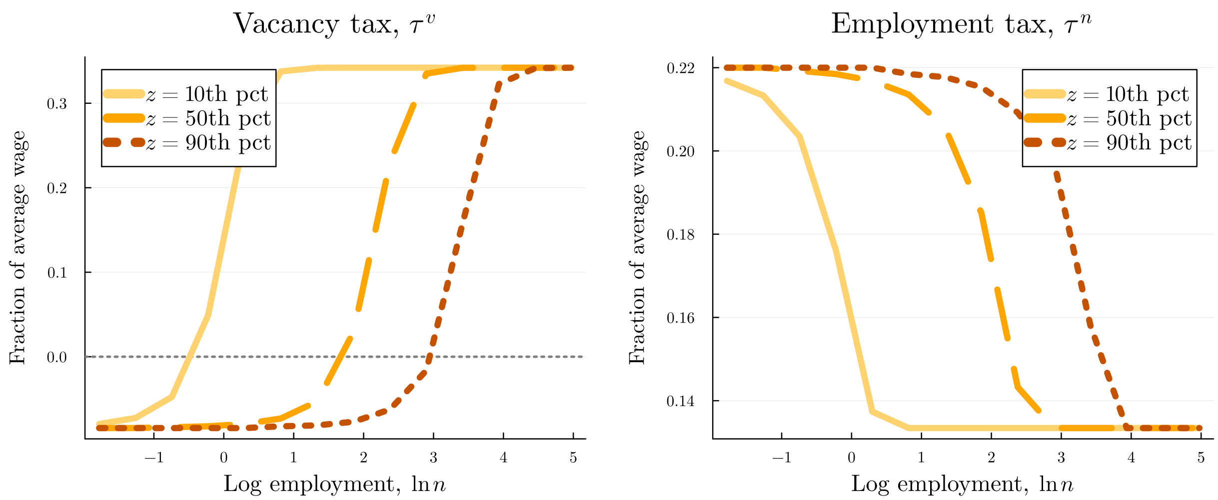

Figure 7: Optimal Taxes

Note: The figure plots the optimal taxes that implement the efficient allocation. The left panel plots the optimal vacancy tax τ^v against the log employment for three different values of productivity, 90th percentile, median, and 10th percentile of the steady state distribution. The right panel plots the optimal labor tax τ^n analogously to the vacancy tax.

7 Transition Dynamics

So far, our focus has been on the steady state of the economy. While all our analytical characterizations straightforwardly extend to the transition dynamics, with everything indexed by time, computing the transition dynamics is challenging because agents in our economy interact through the distribution of contracts, F and G. This means we have to find a fixed point of infinite-dimensional objects period by period.

To overcome this challenge, we build on Bilal (2023) to study the first-order approximation of transition dynamics around the steady state. We explain the key ideas here and relegate the details to Appendix J. The first step is to write the HJB equation for the transition dynamics in recursive form by carrying the distribution ϕ as a state variable. This is the master equation, in which the firm's surplus function is given by $S(x, \phi)$, where $x \equiv (z, n)$. Then, we consider the first-order approximation of the surplus function $S(x, \phi)$ around the steady state $\phi = \bar{\phi}$:

$$S(x, \phi) \approx S(x, \bar{\phi}) + \int q(x, \xi)\big(\phi(\xi) - \bar{\phi}(\xi)\big)d\xi, \tag{55}$$

where $q(x, \xi)$ is the four-dimensional function that solves a version of HJB equation, and $\xi \equiv (z_\xi, n_\xi)$. This is the "FAME" in Bilal (2023).

While solving for $q(x, \xi)$ is feasible in principle, computing the four-dimensional value function with a reasonable number of grids on (z, n) is still prohibitively expensive. We show, how-

	Equilibrium	Optimal policy	% change
Aggregate output	3.11	3.14	+0.8%
Aggregate consumption	2.47	2.49	+1.0%
Market tightness θ	1.0	0.778	-22.2%
UE rate	4.8	4.23	-11.8%
EE rate	0.519	0.511	-1.6%
Acceptance prob (EE/UE)	0.108	0.121	+11.6%
Reallocation rate	0.3	0.325	+8.4%
Average firm size	18.3	17.8	-2.7%
Var($\ln w$)	0.0649	0.0788	+21.3%

Table 3: Equilibrium vs Optimal Policy

Note: The table summarizes the steady state statistics of the equilibrium without policy and with the optimal policy. Aggregate output refers to $\int y(z,n)d\phi$, while aggregate consumption subtracts the vacancy cost and the entry cost from the aggregate output, $\int \{y(z,n) - c(v)n\}\, d\phi - m^0 c^e$. UE rate is the rate at which an unemployed worker finds a job, and EE rate is the rate at which an employed worker moves employers without experiencing an unemployment spell. Acceptance probability is defined as the ratio of EE rate to UE rate, following Moscarini and Postel-Vinay (2025). Reallocation rate is the size-weighted mean of absolute employment growth rates among firms, $\frac{1}{\int n d\phi} \int |dn/n| n d\phi$, following Davis and Haltiwanger (1999).

ever, that under the assumption of Proposition 9, we can write

$$q(x,\xi) = z\, z_\xi\, \hat{q}(\hat{n}, \hat{n}_\xi), \tag{56}$$

with $\hat{n} \equiv n/z$ and $\hat{n}_\xi \equiv n_\xi / z_\xi$, where $\hat{q}$ solves a version of HJB equation. This factorization allows us to reduce the four-dimensional problem to a two-dimensional problem, which speeds up the computation by several orders of magnitude.

We apply the above methodology to compute the transition dynamics following the implementation of the optimal policy.[20] This is important because comparisons based solely on steady states between the laissez-faire equilibrium and the equilibrium with the optimal policy can be misleading.

Figure 8 plots the transition dynamics following the implementation of the optimal policy. The left panel shows the path of aggregate output, which approximately reaches the new steady state within the first few years. The middle panel shows the mass of firms, which surges following the implementation of the optimal policy and gradually decays to the new steady state. The right panel shows the path of the reallocation rate, which also overshoots in the short run and then converges to the new steady state. The surge in entry and the reallocation rate in the

[20] More specifically, we solve the transition dynamics of the planner's problem. Therefore, the optimal policy here means the policy is optimal at every point in time along the transition path, not only at the steady state.

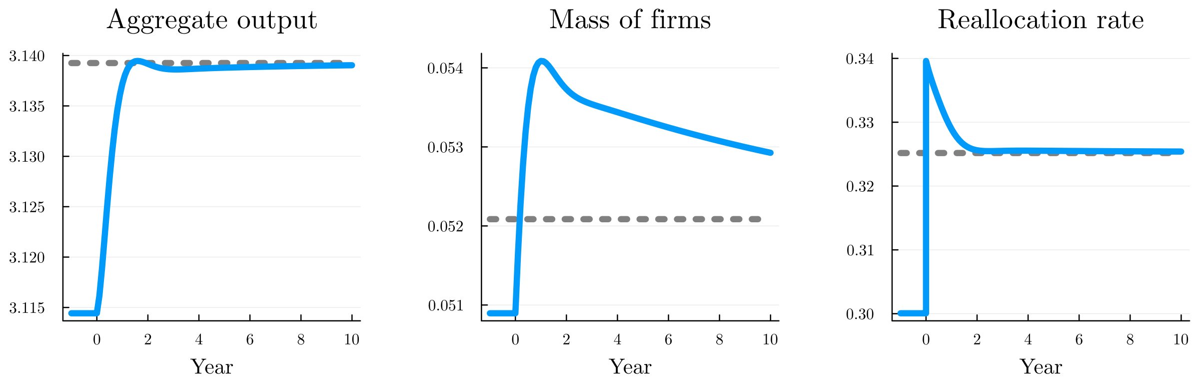

Figure 8: Transition Dynamics

Note: The figure plots the transition dynamics of the economy following the implementation of the optimal policy. The left panel plots the path of aggregate output. The middle panel plots the path of the mass of firms, m_t. The right panel plots the path of the reallocation rate, $\frac{1}{\int n d\phi_t} \int |dn/n| n d\phi_t$, defined following Davis and Haltiwanger (1999). The grey dashed line indicates the new steady state level.

short run suggests that the optimal policy elevates business dynamism even more than what the steady-state analysis suggests. Intuitively, the transition starts from the steady-state equilibrium distribution, where misallocation is most severe, so the gains from elevating business dynamism are largest in the short run.

8 Conclusion

We develop a theory of firm wage dynamics that integrates the wage-posting model à la BM with firm dynamics à la Hopenhayn (1992). In contrast to the canonical wage-posting model, which ties firm wage with firm size, our model links firm wage with firm growth. Consequently, the model naturally explains (i) why the firm wage is strongly related to firm growth but not to firm size; (ii) why the firm wage declines over the firm life cycle; and (iii) why the pass-through of productivity shocks to wages is higher in the short run than in the long run. Furthermore, the theory offers a new perspective on what we can learn about the monopsony power from the pass-through of productivity into wages and employment.

On the normative side, we identify a novel source of inefficiency: the monopsony power of firms at the top of the job ladder is too low, even though it is among the highest in the economy. The optimal policy thus subsidizes their vacancy creation. This encourages business dynamism, increasing worker reallocation and firm entry in the steady state, and even more so along the transition.

References

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

Acemoglu, Daron, "Good jobs versus bad jobs," *Journal of Labor Economics*, 2001, *19* (1), 1–21.

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

Addario, Sabrina Di, Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten, "It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages," *Journal of Econometrics*, 2023, *233* (2), 340–374.

Audoly, Richard, "Firm dynamics and random search over the business cycle," *Journal of Monetary Economics*, 2026, p. 103902.

Babina, Tania, Wenting Ma, Christian Moser, Paige Ouimet, and Rebecca Zarutskie, "Pay, employment, and dynamics of young firms," *Kenan Institute of Private Enterprise Research Paper*, 2019, (19-25).

Berger, David, Kyle Herkenhoff, and Simon Mongey, "Labor market power," *American Economic Review*, 2022, *112* (4), 1147–1193.

Bilal, Adrien, "Solving heterogeneous agent models with the master equation," *Working Paper*, 2023.

_ **and Hugo Lhuillier**, "Outsourcing, inequality and aggregate output," Technical Report, National Bureau of Economic Research 2021.

_ **, Niklas Engbom, Simon Mongey, and Giovanni L Violante**, "Firm and worker dynamics in a frictional labor market," *Econometrica*, 2022, *90* (4), 1425–1462.

Bloesch, Justin, Birthe Larsen, and Anders Yding, "Monopsony with Recruiting," *Available at SSRN*, 2026.

Bloom, Nicholas, Fatih Guvenen, Benjamin S Smith, Jae Song, and Till von Wachter, "The disappearing large-firm wage premium," in "AEA Papers and Proceedings," Vol. 108 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2018, pp. 317–322.

Borovičková, Katarína and Robert Shimer, "Assortative Matching and Wages: The Role of Selection," Technical Report, National Bureau of Economic Research 2024.

Brügemann, Björn, Pieter Gautier, and Guido Menzio, "Intra firm bargaining and Shapley values," *The Review of Economic Studies*, 2019, *86* (2), 564–592.

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

_ **and Tara Vishwanath**, "Balanced matching and labor market equilibrium," *Journal of Political Economy*, 1988, *96* (5), 1048–1065.

Cai, Xiaoming, “Efficiency of Wage Bargaining with on-the-job search,” *International Economic Review*, 2020, *61* (4), 1749–1775.

Caldwell, Sydnee, Arindrajit Dube, and Suresh Naidu, “Monopsony makes it big,” *Journal of Economic Literature*, 2026.

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

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

Davis, Steven J and John Haltiwanger, “Gross job flows,” *Handbook of labor economics*, 1999, *3*, 2711–2805.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger, “The Establishment-Level Behavior of Vacancies and Hiring,” *The Quarterly Journal of Economics*, 2013, *128* (2), 581–622.

Diegmann, André, Steffen Müller, and Benjamin Schoefer, “Off the Labor Supply Curve: The Zero Employer Size Wage Effect Within Large Firms,” Technical Report, National Bureau of Economic Research 2026.

Elsby, Michael WL and Axel Gottfries, “Firm dynamics, on-the-job search, and labor market fluctuations,” *The Review of Economic Studies*, 2022, *89* (3), 1370–1419.

— **and Ryan Michaels**, “Marginal jobs, heterogeneous firms, and unemployment flows,” *American Economic Journal: Macroeconomics*, 2013, *5* (1), 1–48.

Engbom, Niklas and Christian Moser, “Earnings inequality and the minimum wage: Evidence from Brazil,” *American Economic Review*, 2022, *112* (12), 3803–3847.

Fukui, Masao, “A theory of wage rigidity and unemployment fluctuations with on-the-job search,” *Available at SSRN 5389577*, 2020.

— **and Toshihiko Mukoyama**, “Efficiency in job-ladder models,” Technical Report, National Bureau of Economic Research 2025.

— **, Emi Nakamura, and Jón Steinsson**, “The commoditization of labor,” Technical Report, Working papery 2026.

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

Gautier, Pieter A, Coen N Teulings, and Aico Van Vuuren, “On-the-job search, mismatch and efficiency,” *The Review of Economic Studies*, 2010, *77* (1), 245–272.

Gottfries, Axel and Gregor Jarosch, “Dynamic monopsony with large firms and noncompetes,” 2023, (31965).

Gouin-Bonenfant, Émilien, “Productivity dispersion, between-firm competition, and the labor share,” *Econometrica*, 2022, *90* (6), 2755–2793.

Hall, Robert E and Alan B Krueger, "Evidence on the incidence of wage posting, wage bargaining, and on-the-job search," *American Economic Journal: Macroeconomics*, 2012, *4* (4), 56–67.

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

Hazell, Jonathon, Christina Patterson, Heather Sarsons, and Bledi Taska, "National wage setting," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2022, (2022-150).

Hopenhayn, Hugo, "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica: Journal of the Econometric Society*, 1992, pp. 1127–1150.

__ **and Richard Rogerson**, "Job turnover and policy evaluation: A general equilibrium analysis," *Journal of political Economy*, 1993, *101* (5), 915–938.

Hosios, Arthur J, "On the efficiency of matching and related models of search and unemployment," *The Review of Economic Studies*, 1990, *57* (2), 279–298.

Jäger, Simon, Benjamin Schoefer, Samuel Young, and Josef Zweimüller, "Wages and the Value of Nonemployment," *The Quarterly Journal of Economics*, 2020, *135* (4), 1905–1963.

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

Kline, Patrick, "Firm wage effects," *Handbook of Labor Economics*, 2024, *5*, 115–181.

__ , "Labor market monopsony: Fundamentals and frontiers," *Handbook of Labor Economics*, 2025, *6*, 655–728.

Luttmer, Erzo GJ, "Selection, growth, and the size distribution of firms," *The Quarterly Journal of Economics*, 2007, *122* (3), 1103–1144.

McCrary, Sean, "A Job Ladder Model of Firm, Worker, and Earnings Dynamics," *Worker, and Earnings Dynamics (November 4, 2022)*, 2022.

Menzio, Guido, "Markups: A search-theoretic perspective," Technical Report, National Bureau of Economic Research 2024.

__ , "Search theory of imperfect competition with decreasing returns to scale," *Journal of Economic Theory*, 2024, *218*, 105827.

Morchio, Iacopo and Christian Moser, "The gender pay gap: Micro sources and macro consequences," *American Economic Review*, 2026, *116* (5), 1765–1810.

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

__ **and** __ , "The job ladder: Inflation vs. reallocation," Technical Report, National Bureau of Economic Research 2025.

Petrongolo, Barbara and Christopher A Pissarides, "Looking into the black box: A survey of the matching function," *Journal of Economic literature*, 2001, *39* (2), 390–431.

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

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

Schmieder, Johannes F, "Establishment age and wages," *Journal of Econometrics*, 2023, *233* (2), 424–442.

Seegmiller, Bryan, "Valuing labor market power: The role of productivity advantages," Technical Report, Working paper 2025.

Spear, Stephen E and Sanjay Srivastava, "On repeated moral hazard with discounting," *The Review of Economic Studies*, 1987, *54* (4), 599–617.

Sterk, Vincent, Petr Sedláček, and Benjamin Pugsley, "The nature of firm growth," *American Economic Review*, 2021, *111* (2), 547–579.

Stole, Lars A and Jeffrey Zwiebel, "Organizational design and technology choice under intrafirm bargaining," *The American Economic Review*, 1996, pp. 195–222.

Talesara, Ishaana and Arthur Wickard, "Monopsony with Dynamic Wage Contracts," 2026.

Tanaka, Satoshi, Lawrence Warren, and David Wiczer, "Earnings growth, job flows and churn," *Journal of Monetary Economics*, 2023, *135*, 86–98.

Online Appendix for:

A Theory of Firm Wage Dynamics

Marc de la Barrera   Masao Fukui

Contents

Appendix Contents

A Derivations and Proofs

A.1 Proof of Proposition 1

The original Bellman equation is

$$\begin{aligned}
J(z_t, n_t, W_t) = \max_{w_t, \upsilon, \varsigma, n_{t+dt}, W_{t+dt}, \chi} & \; y(z_t, n_t)dt - w_t n_t dt - c(\upsilon) n_t dt \\
& + e^{-(r+\varkappa)dt}(1-\chi)\mathbb{E}J(z_{t+dt}, n_{t+dt}, W_{t+dt}(z_{t+dt})) \\
\text{s.t.} \qquad W_t = w_t dt + e^{-rdt} & \left[e^{-\lambda^E dt} V_{t+dt} + (1 - e^{-\lambda^E dt}) \int \max\{V_{t+dt}, \tilde{V}\} dF(\tilde{V}) \right] - (1 - e^{-rdt}) U \\
n_{t+dt} = e^{-sdt}(1-\varsigma) & \left[\left(e^{-\lambda^E dt} + (1 - e^{-\lambda^E dt}) F(V_{t+dt}) \right) n_t + (1 - e^{-\lambda^F dt}) G(V_{t+dt}) \upsilon n_t \right]
\end{aligned}$$

where $V_{t+dt} = e^{-(s+\varkappa)dt}(1-\varsigma)(1-\chi)\mathbb{E}[W_{t+dt}(z_{t+dt}; n, W_t)]$. Substituting the promise-keeping constraint to eliminate w, we have

$$\begin{aligned}
J(z_t, n_t, W_t) = \max_{\upsilon, \varsigma, n_{t+dt}, W_{t+dt}, \chi} & \; y(z_t, n_t)dt - c(\upsilon) n_t dt - W_t n_t \\
& + e^{-rdt} \left[e^{-\lambda^E dt} V_{t+dt} + (1 - e^{-\lambda^E dt}) \int \max\{V_{t+dt}, \tilde{V}\} dF(\tilde{V}) \right] n_t - (1 - e^{-rdt}) U n_t \\
& + e^{-(r+\varkappa)dt}(1-\chi)\mathbb{E}J(z_{t+dt}, n_{t+dt}, W_{t+dt}(z_{t+dt})) \\
\text{s.t.} \qquad n_{t+dt} = e^{-sdt}(1-\varsigma) & \left[\left(e^{-\lambda^E dt} + (1 - e^{-\lambda^E dt}) F(V_{t+dt}) \right) n_t + (1 - e^{-\lambda^F dt}) G(V_{t+dt}) \upsilon n_t \right]
\end{aligned}$$

We now guess and verify that the value function is of the form

$$J(z_t, n_t, W_t) = S(z_t, n_t) - W_t n_t.$$

Substituting this into the Bellman equation, we get

$$\begin{aligned}
S(z_t, n_t) = \max_{\upsilon, \varsigma, n_{t+dt}, W_{t+dt}, \chi} & \; y(z_t, n_t)dt - c(\upsilon) n_t dt \\
& + e^{-rdt} \left[e^{-\lambda^E dt} V_{t+dt} + (1 - e^{-\lambda^E dt}) \int \max\{V_{t+dt}, \tilde{V}\} dF(\tilde{V}) \right] n_t - (1 - e^{-rdt}) U n_t \\
& + e^{-(r+\varkappa)dt}(1-\chi)\mathbb{E}S(z_{t+dt}, n_{t+dt}) - e^{-(r+\varkappa)dt}(1-\chi)\mathbb{E}[W_{t+dt}(z_{t+dt})] n_{t+dt} \\
\text{s.t.} \qquad n_{t+dt} = e^{-sdt}(1-\varsigma) & \left[\left(e^{-\lambda^E dt} + (1 - e^{-\lambda^E dt}) F(V_{t+dt}) \right) n_t + (1 - e^{-\lambda^F dt}) G(V_{t+dt}) \upsilon n_t \right]
\end{aligned}$$

Using the employment evolution equation, we can rewrite the above equation as

$$\begin{aligned}
S(z_t, n_t) = & \max_{v,\varsigma,n_{t+dt},V_{t+dt},\chi} y(z_t, n_t)dt - c(v)n_t dt \\
& + e^{-rdt}\left[e^{-\lambda^E dt}V_{t+dt} + (1-e^{-\lambda^E dt})\int \max\{V_{t+dt}, \tilde{V}\}dF(\tilde{V})\right] n_t - (1-e^{-rdt})Un_t \\
& + e^{-(r+\varkappa)dt}(1-\chi)\mathbb{E}S(z_{t+dt}, n_{t+dt}) \\
& - e^{-rdt}V_{t+dt}\left[\left(e^{-\lambda^E dt} + (1-e^{-\lambda^E dt})F(V_{t+dt})\right) n_t + (1-e^{-\lambda^F dt})G(V_{t+dt})vn_t\right] \\
\text{s.t.} \quad & n_{t+dt} = e^{-sdt}(1-\varsigma)\left[\left(e^{-\lambda^E dt} + (1-e^{-\lambda^E dt})F(V_{t+dt})\right) n_t + (1-e^{-\lambda^F dt})G(V_{t+dt})vn_t\right],
\end{aligned}$$

which further simplifies to

$$\begin{aligned}
S(z_t, n_t) = & \max_{v,\varsigma,n_{t+dt},V_{t+dt},\chi} y(z_t, n_t)dt - c(v)n_t dt - (1-e^{-rdt})Un_t - e^{-rdt}V_{t+dt}(1-e^{-\lambda^F dt})G(V_{t+dt})vn_t \\
& + e^{-rdt}(1-e^{-\lambda^E dt})\int_{V_{t+dt}} \tilde{V}dF(\tilde{V})n_t + e^{-(r+\varkappa)dt}(1-\chi)\mathbb{E}S(z_{t+dt}, n_{t+dt}) \\
\text{s.t.} \quad & n_{t+dt} = e^{-sdt}(1-\varsigma)\left[\left(e^{-\lambda^E dt} + (1-e^{-\lambda^E dt})F(V_{t+dt})\right) n_t + (1-e^{-\lambda^F dt})G(V_{t+dt})vn_t\right].
\end{aligned}$$

This coincides with (14). Since the promised utility W_t does not enter the problem, the policy functions depend only on (z, n), which completes the proof.

A.2 Continuous Time Limit

We can rewrite (14) as

$$\begin{aligned}
S(z_t, n_t) = & \max_{v,n_{t+dt},V_{t+dt}} y(z_t, n_t)dt - c(v)n_t dt - (1-e^{-rdt})Un_t \\
& - e^{-rdt}V_{t+dt}(1-e^{-\lambda^F dt})G(V_{t+dt})vn_t + e^{-rdt}(1-e^{-\lambda^E dt})\int_{V_{t+dt}} \tilde{V}dF(\tilde{V})n_t \\
& + e^{-(r+\varkappa)dt}\max\{\mathbb{E}S(z_{t+dt}, n_{t+dt}), S^*(z_t, n_{t+dt})\} \\
\text{s.t.} \quad & n_{t+dt} = e^{-sdt}\left[\left(e^{-\lambda^E dt} + (1-e^{-\lambda^E dt})F(V_{t+dt})\right) n_t + (1-e^{-\lambda^F dt})G(V_{t+dt})vn_t\right],
\end{aligned}$$

where $S^*(z_t, n_t)$ is the value from firing or exiting:

$$S^*(z_t, n_t) = \max\left\{\max_{n^* \leq n_t} \mathbb{E}S(z_{t+dt}, n^*), 0\right\}. \tag{57}$$

Now we will consider a continuous time limit of the above equation. We divide into two cases: (i) firms do not fire or exit ($n_{t+dt} = \arg\max_{n^* \leq n_{t+dt}} \mathbb{E}S(z_{t+dt}, n^*)$ and $\mathbb{E}S(z_{t+dt}, n_{t+dt}) \geq 0$), and

(ii) firms fire or exit (otherwise). If firms do not fire or exit, the continuous time limit (i.e., $dt \to 0$) of the Bellman equation is

$$\rho S(z_t, n_t) = \max_{v,V} y(z_t, n_t) - c(v)n_t - rUn_t - V\lambda^F G(V)vn_t + \lambda^E \int_V \tilde{V} dF(\tilde{V})n_t + \mathcal{L}(z_t, n_t, v, V)S(z_t, n_t), \tag{58}$$

where $\mathcal{L}(z, n, v, V)$ is the generator for (z, n) defined for an arbitrary function $f(z, n)$:

$$\mathcal{L}(z, n, v, V)f(z, n) = \mu(z)\partial_z f(z, n) + \frac{\sigma^2(z)}{2}\partial_{zz} f(z, n) + \left[\lambda^F G(V)v - s - \lambda^E(1 - F(V))\right] n\partial_n f(z, n). \tag{59}$$

Now we turn to the case where firms fire or exit. In the continuous time limit, the value from firing or exiting is given by

$$S^*(z_t, n_t) = \max\{\max_{n^* \le n_t} S(z_t, n^*), 0\}. \tag{60}$$

If $n_t = \arg\max_{n^* \le n_t} S(z_t, n^*)$ and $S(z_t, n_t) > 0$, then firms do not fire or exit and follow the HJB equation in (58). Otherwise, firms fire or exit with $S(z_t, n_t) = S^*(z_t, n_t)$, in which case $S(z_t, n_t)$ would be strictly higher than the value from not adjusting, which is the right-hand side of the HJB equation (58). This can be compactly written as the HJB Quasi-Variational Inequality (QVI), where $S \equiv S(z, n)$ solves

$$\min\Big\{\rho S - \max_{v,V}\Big\{y(z, n) - c(v)n - rUn - V\lambda^F G(V)vn + \lambda^E \int_V \tilde{V} dF(\tilde{V})n + \mathcal{L}(z, n, v, V)S\Big\}, S - S^*(z, n)\Big\} = 0. \tag{61}$$

This problem is equivalent to Problem 1.

A.3 Risk-Averse Workers

In the baseline model, we assumed that the workers are risk-neutral. We relax this assumption by assuming that the worker flow utility is given by

$$u(c) = \frac{c^{1-\gamma_c}}{1 - \gamma_c}. \tag{62}$$

The baseline model is nested as a special case with $\gamma_c = 0$.

With this modification, the Bellman equation for the firm's problem becomes

$$
\begin{aligned}
J(z_t, n_t, W_t) = & \max_{w, v, \varsigma, n_{t+dt}, W_{t+dt}, \chi} y(z_t, n_t)dt - wn_t dt - c(v)n_t dt \\
& + e^{-(r+\varkappa)dt}(1-\chi)\mathbb{E}J(z_{t+dt}, n_{t+dt}, W_{t+dt}(z_{t+dt})) \\
\text{s.t.} \qquad & W_t = u(w)dt + e^{-rdt}\left[e^{-\lambda^E dt}V_{t+dt} + (1-e^{-\lambda^E dt})\int \max\{V_{t+dt}, \tilde{V}\}dF(\tilde{V})\right] - (1-e^{-rdt})U \\
& n_{t+dt} = e^{-sdt}(1-\varsigma)\left[\left(e^{-\lambda^E dt} + (1-e^{-\lambda^E dt})F(V_{t+dt})\right)n_t + (1-e^{-\lambda^F dt})G(V_{t+dt})vn_t\right]
\end{aligned}
$$

and $V_{t+dt} = e^{-(s+\varkappa)dt}(1-\varsigma)(1-\chi)\mathbb{E}[W_{t+dt}(z_{t+dt})]$. We can substitute the promise-keeping constraint into the Bellman equation to eliminate w:

$$
\begin{aligned}
J(z_t, n_t, W_t) = & \max_{w, v, \varsigma, n_{t+dt}, W_{t+dt}, \chi} y(z_t, n_t)dt - \tilde{w}(W_t, V_{t+dt})n_t dt - c(v)n_t dt \\
& + e^{-(r+\varkappa)dt}(1-\chi)\mathbb{E}J(z_{t+dt}, n_{t+dt}, W_{t+dt}(z_{t+dt})) \\
\text{s.t.} \qquad & n_{t+dt} = e^{-sdt}(1-\varsigma)\left[\left(e^{-\lambda^E dt} + (1-e^{-\lambda^E dt})F(V_{t+dt})\right)n_t + (1-e^{-\lambda^F dt})G(V_{t+dt})vn_t\right],
\end{aligned}
$$

where

$$
\begin{aligned}
& \tilde{w}(W_t, V_{t+dt}) \\
& \quad \equiv u^{-1}\left(\frac{1}{dt}\left(W_t - e^{-rdt}\left[e^{-\lambda^E dt}V_{t+dt} + (1-e^{-\lambda^E dt})\int \max\{V_{t+dt}, \tilde{V}\}dF(\tilde{V})\right] + (1-e^{-rdt})U\right)\right).
\end{aligned}
$$

The first-order optimality condition with respect to $W_{t+dt}(z_{t+dt})$ implies that for any z^1_{t+dt} and z^2_{t+dt}, we must have

$$
\partial_W J(z^1_{t+dt}, n_{t+dt}, W_{t+dt}(z^1_{t+dt})) = \partial_W J(z^2_{t+dt}, n_{t+dt}, W_{t+dt}(z^2_{t+dt})). \tag{63}
$$

By the envelope condition, we have

$$
\partial_W J(z_t, n_t, W_t) = \frac{1}{u'(\tilde{w}(W_t, V_{t+dt}(z_t, n_t, W_t)))}n_t, \tag{64}
$$

where $V_{t+dt}(z_t, n_t, W_t)$ is the policy function. Therefore, as long as workers are risk-averse, $W_{t+dt}(z^1_t; n_t, W_t)$ must satisfy that for any z^1_t and z^2_t,

$$
\begin{aligned}
& \frac{1}{u'\left(\tilde{w}(W_{t+dt}(z^1_t), V_{t+dt}(z^1_{t+dt}, n_{t+dt}, W_{t+dt}(z^1_{t+dt})))\right)}n_{t+dt} \qquad (65) \\
= & \frac{1}{u'\left(\tilde{w}(W_{t+dt}(z^2_t), V_{t+dt}(z^2_{t+dt}, n_{t+dt}, W_{t+dt}(z^2_{t+dt})))\right)}n_{t+dt}. \qquad (66)
\end{aligned}
$$

This implies that

$$\tilde{w}(W_{t+dt}(z^1_{t+dt}), V_{t+dt}(z^1_{t+dt}, n_{t+dt}, W_{t+dt}(z^1_{t+dt}))) = \tilde{w}(W_{t+dt}(z^2_{t+dt}), V_{t+dt}(z^2_{t+dt}, n_{t+dt}, W_{t+dt}(z^2_{t+dt}))).$$

Note that this condition must hold for any $\gamma_c > 0$, and hence in the limit of $\gamma_c \to 0$.

Now we derive (21) under $\gamma_c \to 0$ and continuous time limit. Explicitly writing down the dependence in the promise keeping constraint (8),

$$\begin{aligned}W_t(z_{t+dt}; z_t, n_t) &= w_t(z_t, n_t)dt - (1 - e^{-rdt})U \\ &+ e^{-rdt}\left[e^{-\lambda^E dt}V_{t+dt}(z_{t+dt}, n_{t+dt}) + (1 - e^{-\lambda^E dt})\int \max\{V_{t+dt}(z_{t+dt}, n_{t+dt}), \tilde{V}\}dF(\tilde{V})\right],\end{aligned}$$

where we have already imposed that the wage does not depend on z_{t+dt} under $\gamma_c \to 0$. Taking expectations of the above equation over z_{t+dt} conditional on z_t and n_t and multiplying by $e^{-(s+\varkappa)dt}(1 - \varsigma_t(z_t, n_t))(1 - \chi_t(z_t, n_t))$, we have

$$\begin{aligned}V_t(z_t, n_t) = e^{-(s+\varkappa)dt}(1 - \varsigma_t(z_t, n_t))(1 - \chi_t(z_t, n_t))\Bigg\{ w_t(z_t, n_t)dt - (1 - e^{-rdt})U \\ + e^{-rdt}\mathbb{E}_t\left[e^{-\lambda^E dt}V_{t+dt}(z_{t+dt}, n_{t+dt}) + (1 - e^{-\lambda^E dt})\int \max\{V_{t+dt}(z_{t+dt}, n_{t+dt}), \tilde{V}\}dF(\tilde{V})\right]\Bigg\}.\end{aligned}$$

The continuous time limit of the above equation is given by

$$\begin{aligned}(r + s + \varkappa)V(z, n) = w(z, n) - rU + \lambda^E \int_{V(z,n)}^{\infty}(\tilde{V} - V(z, n))dF(\tilde{V}) \\ + \mathcal{L}(z, n, v(z, n), V(z, n))V(z, n),\end{aligned}$$

for regions in which $S(z, n) > 0$ and $S_n(z, n) > 0$. For other regions, $V(z, n) = 0$.

A.4 Proof of Proposition 2

Within this proof we make the tie-breaking convention for worker mobility explicit: an employed worker accepts an outside offer if and only if it strictly exceeds her current value, while unemployed workers accept any offer $V \geq 0$. Accordingly, the hiring probability of a vacancy offering $V > 0$ is $G(V^-) \equiv \lim_{V' \uparrow V} G(V')$, the hiring probability of a vacancy offering $V = 0$ is $u/\tilde{u}$, and the poaching rate of a firm offering V is $\lambda^E(1 - F(V))$. Finally, recall that $u > 0$ in any steady state, so that

$$G(V^-) \;\geq\; \frac{u}{\tilde{u}} \;>\; 0 \qquad \text{for every } V > 0. \tag{67}$$

Fix a firm (z,n) in the continuation region and write $S_n = S_n(z,n)$. Collecting the terms of the HJB equation in Problem 1 that involve the controls (V,v), the equilibrium policy must maximize, per employee,

$$\Pi(V,v) \;\equiv\; \lambda^F v\, G(V^-)\,(S_n - V) \;-\; c(v) \;+\; \lambda^E \int_{(V,\infty)} (\tilde{V} - S_n)\, dF(\tilde{V}), \qquad v \geq \underline{v}. \tag{68}$$

The first two terms net the hiring gain against the value promised to new hires and the vacancy cost; the last term prices poaching: a worker who leaves for an outside value $\tilde{V}$ adds $\tilde{V}$ to the joint surplus but removes her marginal contribution S_n. Note that a worker employed at firm (z,n) currently enjoys exactly the value $V(z,n)$ offered by her employer. Hence an atom of G at $V^0 > 0$ is equivalent to a positive employment mass on $E_{V^0} \equiv \{(z,n) : V(z,n) = V^0\}$, and an atom of F at V^0 is equivalent to a positive vacancy mass $\int_{E_{V^0}} v\, n\, d\phi > 0$. Because $v \geq \underline{v} > 0$ and $n > 0$ for every operating firm, these two notions are tightly linked: for any Borel set B of offer values,

$$\int_{\{V(z,n)\in B\}} v\, n\, d\phi \;\geq\; \underline{v} \int_{\{V(z,n)\in B\}} n\, d\phi, \quad \text{and} \quad \int_{\{V(z,n)\in B\}} v\, n\, d\phi > 0 \Leftrightarrow \int_{\{V(z,n)\in B\}} n\, d\phi > 0, \tag{69}$$

where the equivalence holds because $vn > 0$ if and only if $n > 0$, pointwise.

Step 1: no firm offers above its marginal surplus. Suppose a firm's policy is (V,v) with $V > S_n$. Deviating to (S_n, v), keeping v unchanged, yields

$$\Pi(S_n,v) - \Pi(V,v) \;=\; \lambda^F v\, G(V^-)\,(V - S_n) \;+\; \lambda^E \int_{(S_n,\, V]} (\tilde{V} - S_n)\, dF \;\geq\; \lambda^F\, \underline{v}\, \frac{u}{\tilde{u}}\,(V - S_n) \;>\; 0,$$

where the first term uses (67) and $v \geq \underline{v}$, and the integral is nonnegative because its integrand is positive on $(S_n, V]$. This contradicts optimality. Hence $V(z,n) \leq S_n(z,n)$ for every firm.

Step 2: no firm with $S_n > 0$ offers exactly its marginal surplus. Suppose a firm's policy is (S_n, v) with $S_n > 0$. Deviating to $(S_n - \delta, v)$ for $\delta \in (0, S_n)$, keeping v unchanged, yields

$$\Pi(S_n - \delta, v) - \Pi(S_n, v) \;=\; \lambda^F v\, G((S_n - \delta)^-)\, \delta \;-\; \lambda^E \int_{(S_n - \delta,\, S_n]} (S_n - \tilde{V})\, dF(\tilde{V})$$

$$\geq\ \lambda^F\,\underline{v}\,\frac{u}{\tilde{u}}\,\delta\ -\ \lambda^E\,\delta\,\bigl[F(S_n^-)-F(S_n-\delta)\bigr],$$

where the second line uses (67), the bound $S_n - \tilde{V} \leq \delta$ on the integration region, and the fact that an atom of F at S_n (if any) contributes nothing to the integral since the integrand vanishes at $\tilde{V} = S_n$. Note that $F(S_n^-) - F(S_n - \delta) \to 0$ as $\delta \to 0$, which follows from the definition of a left limit of a monotone function (this does not require continuity of F). Hence the right-hand side is strictly positive for δ small, contradicting optimality. Combining with Step 1: in any equilibrium,

$$V(z,n) < S_n(z,n)\ \text{ whenever } S_n(z,n) > 0, \qquad V(z,n) = 0\ \text{ whenever } S_n(z,n) = 0. \tag{70}$$

Step 3: F has no mass point at any $V^0 > 0$. Suppose F has an atom of mass $\mu_F > 0$ at $V^0 > 0$. By (69), the firms with $V(z,n) = V^0$ have positive employment mass, and their workers all enjoy the value V^0, so G has an atom of mass $\mu_G \geq (\zeta/\tilde{u}) \int_{E_{V^0}} n\, d\phi > 0$ at V^0. By (70), all such firms have $S_n > V^0$. Toward a contradiction, let a firm offering $V^0 = V(z,n)$ deviate from (V^0, v) to $(V^0 + \varepsilon, v)$ with $0 < \varepsilon < S_n - V^0$. Since the hiring pool jumps discretely, $G((V^0+\varepsilon)^-) \geq G(V^0) = G(V^{0-}) + \mu_G$, while each worker retained rather than poached at $\tilde{V} \leq V^0 + \varepsilon < S_n$ contributes $S_n - \tilde{V} > 0$:

$$\begin{aligned}
\Pi(V^0+\varepsilon, v) - \Pi(V^0, v) &= \lambda^F v\bigl[G((V^0+\varepsilon)^-)(S_n - V^0 - \varepsilon) - G(V^{0-})(S_n - V^0)\bigr] \\
&\qquad + \lambda^E \int_{(V^0, V^0+\varepsilon]} (S_n - \tilde{V})\, dF \\
&\geq \lambda^F v\bigl[\mu_G\,(S_n - V^0 - \varepsilon) - G(V^{0-})\,\varepsilon\bigr] \\
&\xrightarrow[\varepsilon\to 0]{} \lambda^F v\,\mu_G\,(S_n - V^0) > 0,
\end{aligned}$$

a strictly profitable deviation for small ε. Hence F is atomless on $(0, \infty)$.

Step 4: G has no mass point at any $V^0 > 0$. An atom of G at $V^0 > 0$ means positive employment mass on E_{V^0}, which by (69) implies positive vacancy mass on E_{V^0}, i.e., an atom of F at V^0—ruled out by Step 3.

Step 5: the supports coincide and have no gap. Call a maximal nonempty open interval (V^l, V^h) with $F(V^h) = F(V^l)$ and $F(V^h + \eta) > F(V^h)$ for every $\eta > 0$ a *gap* of F (since F is atomless on $(0, \infty)$ by Step 3, we need not distinguish endpoints). The strategy is to show that, in the presence of such a gap, a firm offering $V \in [V^h, V^h + \eta)$ can strictly profitably deviate to $V' = V^l + \eta'$ for η and η' sufficiently small.

Step 5.1: A gap contains no offers and no worker values. By (69), $F(V^h) = F(V^l)$ implies that the set of firms offering values in (V^l, V^h) has zero employment mass. Hence G is constant on

$[V^l, V^h)$, and the hiring pool $G(V^-)$ is the same for every offer $V \in (V^l, V^h]$.

Step 5.2: F has no gap. Suppose (V^l, V^h) is a gap. Since F is atomless (Step 3), it places positive mass on $(V^h, V^h + \bar{\eta})$ for any fixed $\bar{\eta} \in (0, V^h - V^l)$; by (69) the corresponding employment mass is positive, so $\beta \equiv (\zeta/\tilde{u}) \int_{\{V(z,n)\in(V^h,V^h+\bar{\eta})\}} n\, d\phi > 0$. For $\eta \in (0, \bar{\eta})$, pick a firm with offer $V \in [V^h, V^h + \eta)$ and positive vacancy mass (it exists by the definition of a gap). Let S_n be the marginal surplus of the firm with offer V. By (70), $S_n > V$.

We first bound its marginal surplus by showing there exists $\bar{S}$ such that $S_n \leq \bar{S}$ for all firms offering $V \in [V^h, V^h + \bar{\eta})$. Set $V'' \equiv V^h + \bar{\eta}$. If $S_n < V''$, the bound below holds trivially. Otherwise $S_n \geq V''$, and optimality against the deviation (V'', v), whose retention term satisfies $\lambda^E \int_{(V,V'']} (S_n - \tilde{V})\, dF \geq 0$, requires $G(V''^-)(S_n - V'') \leq G(V^-)(S_n - V)$. Since $G(V''^-) - G(V^-) \geq \beta/2$ for η sufficiently small, and since $G \leq 1$ and $V'' - V \leq \bar{\eta}$, this yields $S_n \leq V'' + 2\bar{\eta}/\beta$. Either way, $S_n \leq \bar{S} \equiv V^h + \bar{\eta} + 2\bar{\eta}/\beta$, a bound independent of η.

Now if the firm deviates to (V', v) with $V' \equiv V^l + \eta'$, $\eta' \in (0, V^h - V^l)$, the change in Π is given by

$$\Pi(V', v) - \Pi(V, v) = \lambda^F v \Big[G(V'^-)\,(V - V') - \big(G(V^-) - G(V'^-)\big)(S_n - V) \Big] \\ - \lambda^E \int_{(V',V]} (S_n - \tilde{V})\, dF.$$

As $\eta \to 0$: $G(V^-) - G(V'^-) \to 0$, since there is no mass on the gap interior and G is atomless. For the same reason, $F(V) - F(V') = F(V) - F(V^l) \to F(V^h) - F(V^l) = 0$. Using (67), $v \geq \underline{v}$, and $S_n \leq \bar{S}$,

$$\Pi(V', v) - \Pi(V, v) \geq \lambda^F \underline{v} \left[\frac{u}{\tilde{u}} \big(V^h - V^l - \eta'\big) - o(1)\, \bar{S} \right] - \lambda^E\, o(1)\, \bar{S} > 0$$

for η, η' small: a strictly profitable deviation. Hence F has no interior gap.

Step 5.3: The support of F starts at zero. Let $V_m \equiv \inf \operatorname{supp} F$ and suppose $V_m > 0$. Then, repeating the same argument as in *Step 5.2*, a firm offering $V = V_m + \eta$ can profitably deviate to $V' = \eta'$. Hence $V_m = 0$.

Step 5.4: The support of G. By (69), every offer value has positive vacancy mass if and only if it has positive employment mass. Thus, F and G have the same support.

A.5 Proof of Proposition 3

By Proposition 2, F and G have a common connected support, and we maintain that they are differentiable, with densities f and g that are themselves differentiable.

The first-order condition with respect to v is

$$\lambda^F G(V)\,(S_n - V) = c'(v). \tag{71}$$

Since $c'(v) = \bar{c}^v v^\gamma$, we can explicitly solve for v as

$$\begin{aligned} v &= \left(\frac{\lambda^F G(V)(S_n - V)}{\bar{c}^v}\right)^{\frac{1}{\gamma}} && (72)\\ &\equiv \breve{v}(S_n, V) && (73) \end{aligned}$$

The first-order condition with respect to V is given by

$$\left[\lambda^F g(V)\breve{v}(S_n, V) + \lambda^E f(V)\right](S_n - V) = \lambda^F G(V)\breve{v}(S_n, V), \tag{74}$$

which implicitly defines the policy function as a function of the marginal surplus, $V = \hat{V}(S_n)$. The associated second-order condition is given by

$$\begin{aligned} &\left[\lambda^F g'(V)\breve{v}(S_n, V) + \lambda^E f'(V) + \lambda^F g(V)\frac{\partial \breve{v}(S_n, V)}{\partial V}\right](S_n - V)\\ &- \left[\lambda^F g(V)\breve{v}(S_n, V) + \lambda^E f(V)\right] - \lambda^F g(V)\breve{v}(S_n, V) - \lambda^F G(V)\frac{\partial \breve{v}(S_n, V)}{\partial V} && (75)\\ &\equiv \mathcal{G}(S_n, V) < 0. && (76) \end{aligned}$$

Totally differentiating (74) with respect to S_n and V, we get

$$\begin{aligned} &\left[\lambda^F g(V)\breve{v}(S_n, V) + \lambda^E f(V)\right] d(S_n) - \lambda^F G(V)\frac{\partial \breve{v}(S_n, V)}{\partial (S_n)} d(S_n) && (77)\\ &+\lambda^F g(V)(S_n - V)\frac{\partial \breve{v}(S_n, V)}{\partial S_n} dS_n + \mathcal{G}(S_n, V)dV = 0, && (78) \end{aligned}$$

which we can rewrite as

$$\frac{dV}{dS_n} = -\frac{\lambda^E f(V)\left(1 - \frac{1}{\gamma}\right) + \lambda^F g(V)\breve{v}(S_n, V)}{\mathcal{G}(S_n, V)} > 0, \tag{79}$$

where the last inequality follows from the second-order condition $\mathcal{G}(S_n, V) < 0$ as well as our assumption that $\gamma \geq 1$. This shows that the optimal promised utility offer is strictly increasing in the marginal surplus.

A.6 Proof of Proposition 4

The monotonicity of the promised utility offer implies

$$G(\hat{V}(S_n)) = \hat{G}(S_n), \quad F(\hat{V}(S_n)) = \hat{F}(S_n). \tag{80}$$

Differentiating both sides with respect to S_n, we get

$$g(\hat{V}(S_n))\hat{V}'(S_n) = \hat{g}(S_n), \quad f(\hat{V}(S_n))\hat{V}'(S_n) = \hat{f}(S_n), \tag{81}$$

where $\hat{g}(S_n) \equiv \hat{G}'(S_n)$ and $\hat{f}(S_n) \equiv \hat{F}'(S_n)$. Recall the first-order condition for promised utility offer (74) is

$$\left[\lambda^F g(\hat{V}(S_n))\breve{v}(S_n, \hat{V}(S_n)) + \lambda^E f(\hat{V}(S_n))\right]\left(S_n - \hat{V}(S_n)\right) = \lambda^F G(\hat{V}(S_n))\breve{v}(S_n, \hat{V}(S_n)), \tag{82}$$

Substituting (81) into (82), we have

$$\left[\lambda^F \frac{\hat{g}(S_n)}{\hat{V}'(S_n)}\breve{v}(S_n, \hat{V}(S_n)) + \lambda^E \frac{\hat{f}(S_n)}{\hat{V}'(S_n)}\right]\left(S_n - \hat{V}(S_n)\right) = \lambda^F G(\hat{V}(S_n))\breve{v}(S_n, \hat{V}(S_n)), \tag{83}$$

Note that

$$\hat{f}(S_n) = \frac{1}{\tilde{v}}\breve{v}(S_n, \hat{V}(S_n))\frac{\tilde{u}}{\zeta}\hat{g}(S_n), \tag{84}$$

where $h(S_n) \equiv \hat{H}'(S_n)$ denotes the density of workers at a firm with marginal surplus S_n. We can rewrite (83) as

$$\left[\lambda^F \hat{g}(S_n) + \lambda^U \frac{\tilde{u}}{\tilde{v}}\hat{g}(S_n)\right]\left(S_n - \hat{V}(S_n)\right) = \lambda^F \hat{G}(S_n)\hat{V}'(S_n), \tag{85}$$

Note that

$$\frac{\lambda^U}{\lambda^F} = \frac{M(\tilde{u}, \tilde{v})/\tilde{u}}{M(\tilde{u}, \tilde{v})/\tilde{v}} \tag{86}$$

$$= \frac{\tilde{v}}{\tilde{u}}. \tag{87}$$

Substituting this into (85), we get

$$2\hat{g}(S_n)\left(S_n - \hat{V}(S_n)\right) = \hat{G}(S_n)\hat{V}'(S_n), \tag{88}$$

which is an ODE in terms of $\hat{V}(S_n)$. With the boundary condition $\hat{V}(0) = 0$, the solution is

$$\hat{V}(S_n) = \frac{1}{\hat{G}^2(S_n)} \int_0^{S_n} \tilde{S}_n \, d\hat{G}^2(\tilde{S}_n) \tag{89}$$

$$\equiv \mathbb{E}_{\hat{G}^2} \left[\tilde{S}_n | \tilde{S}_n \leq S_n \right]. \tag{90}$$

The vacancy policy function is given by

$$\hat{v}(S_n) = \breve{v}(S_n, \hat{V}(S_n)) \tag{91}$$

$$= \left(\frac{\lambda^F G(\hat{V}(S_n))(S_n - \hat{V}(S_n))}{\overline{c}^v} \right)^{\frac{1}{\gamma}} \tag{92}$$

Now we verify that the second-order condition, (76), is satisfied. Further differentiating (81),

$$\hat{f}'(S_n) = f'(\hat{V}(S_n))\hat{V}'(S_n)^2 + f(\hat{V}(S_n))\hat{V}''(S_n) \tag{93}$$

$$\hat{g}'(S_n) = g'(\hat{V}(S_n))\hat{V}'(S_n)^2 + g(\hat{V}(S_n))\hat{V}''(S_n) \tag{94}$$

Rearranging (83), the first-order condition with respect to V is

$$\left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n) \right] \left(S_n - \hat{V}(S_n) \right) = \lambda^F \hat{G}(S_n)\hat{v}(S_n)\hat{V}'(S_n), \tag{95}$$

Taking the derivative w.r.t. S_n,

$$\left[\lambda^F \hat{g}'(S_n)\hat{v}(S_n) + \lambda^F \hat{g}(S_n)\hat{v}'(S_n) + \lambda^E \hat{f}'(S_n) \right] \left(S_n - \hat{V}(S_n) \right) \tag{96}$$

$$+ \left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n) \right] \left(1 - \hat{V}'(S_n) \right) \tag{97}$$

$$= \lambda^F \hat{g}(S_n)\hat{v}(S_n)\hat{V}'(S_n) + \lambda^F \hat{G}(S_n)\hat{v}'(S_n)\hat{V}'(S_n) + \lambda^F \hat{G}(S_n)\hat{v}(S_n)\hat{V}''(S_n). \tag{98}$$

We can solve for $\hat{V}''(S_n)$ as

$$\begin{aligned}\hat{V}''(S_n) = \frac{1}{\lambda^F \hat{G}(S_n)\hat{v}(S_n)} \times \Bigg\{ &\left[\lambda^F \hat{g}'(S_n)\hat{v}(S_n) + \lambda^F \hat{g}(S_n)\hat{v}'(S_n) + \lambda^E \hat{f}'(S_n) \right] \left(S_n - \hat{V}(S_n) \right) \\ + \left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n) \right] \left(1 - \hat{V}'(S_n) \right) &- \lambda^F \hat{g}(S_n)\hat{v}(S_n)\hat{V}'(S_n) - \lambda^F \hat{G}(S_n)\hat{v}'(S_n)\hat{V}'(S_n) \Bigg\}\end{aligned} \tag{99}$$

Now consider the following expression:

$$
\begin{aligned}
&\lambda^F g'(\hat V(S_n))\breve v(S_n,\hat V(S_n)) + \lambda^E f'(\hat V(S_n)) \\
&= \frac{1}{\hat V'(S_n)^2}\left[\lambda^F \breve v(S_n,\hat V(S_n))(\hat g'(S_n) - g(\hat V(S_n))\hat V''(S_n)) + \lambda^E(\hat f'(S_n) - f(\hat V(S_n))\hat V''(S_n))\right] \\
&= \frac{1}{\hat V'(S_n)^2}\left[\lambda^F \breve v(S_n,\hat V(S_n))\hat g'(S_n) + \lambda^E \hat f'(S_n) - \hat V''(S_n)\left(\lambda^F \breve v(S_n,\hat V(S_n))g(\hat V(S_n)) + \lambda^E f(\hat V(S_n))\right)\right],
\end{aligned}
$$

where the first equality uses (93) and (94). Substituting (99) into the above expression, we get

$$
\begin{aligned}
&\lambda^F g'(\hat V(S_n))\breve v(S_n,\hat V(S_n)) + \lambda^E f'(\hat V(S_n)) \\
=&\frac{1}{\hat V'(S_n)^2}\Bigg[\lambda^F \breve v(S_n,\hat V(S_n))\hat g'(S_n) + \lambda^E \hat f'(S_n) \\
&- \underbrace{\frac{\left(S_n - \hat V(S_n)\right)\left(\lambda^F \breve v(S_n,\hat V(S_n))g(\hat V(S_n)) + \lambda^E f(\hat V(S_n))\right)}{\lambda^F \hat G(S_n)\hat v(S_n)}}_{=1 \text{ by FOC}} \\
&\times\left[\lambda^F \hat g'(S_n)\hat v(S_n) + \lambda^F \hat g(S_n)\hat v'(S_n) + \lambda^E \hat f'(S_n)\right]\Bigg] \\
&- \frac{\left(\lambda^F \breve v(S_n,\hat V(S_n))g(\hat V(S_n)) + \lambda^E f(\hat V(S_n))\right)}{\lambda^F \hat G(S_n)\hat v(S_n)}\Bigg\{\left[\lambda^F \hat g(S_n)\hat v(S_n) + \lambda^E \hat f(S_n)\right]\left(1 - \hat V'(S_n)\right) \\
&- \lambda^F \hat g(S_n)\hat v(S_n)\hat V'(S_n) - \lambda^F \hat G(S_n)\hat v'(S_n)\hat V'(S_n)\Bigg\}\Bigg] \\
=&\frac{1}{\hat V'(S_n)^2}\Bigg[-\lambda^F \hat g(S_n)\hat v'(S_n) \\
&- \frac{\left(\lambda^F \breve v(S_n,\hat V(S_n))g(\hat V(S_n)) + \lambda^E f(\hat V(S_n))\right)}{\lambda^F \hat G(S_n)\hat v(S_n)}\Bigg\{\left[\lambda^F \hat g(S_n)\hat v(S_n) + \lambda^E \hat f(S_n)\right]\left(1 - \hat V'(S_n)\right) \\
&- \lambda^F \hat g(S_n)\hat v(S_n)\hat V'(S_n) - \lambda^F \hat G(S_n)\hat v'(S_n)\hat V'(S_n)\Bigg\}\Bigg] \qquad (100)
\end{aligned}
$$

Now we check the second-order condition, (76). Substituting (100) into the second-order condition, we have

$$
\begin{aligned}
\mathcal{G}(S_n,V) = &\left[\lambda^F g'(V)\breve v(S_n,V) + \lambda^E f'(V)\right](S_n - V) \\
&+ \lambda^F g(V)\frac{\partial \breve v(S_n,V)}{\partial V}(S_n - V)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \left[\lambda^F g(V)\breve{v}(S_n,V) + \lambda^E f(V)\right] \\
&- \lambda^F g(V)\breve{v}(S_n,V) - \lambda^F G(V)\frac{\partial \breve{v}(S_n,V)}{\partial V} \\
=&\frac{1}{\hat{V}'(S_n)^2}\Bigg[- \lambda^F \hat{g}(S_n)\hat{v}'(S_n)(S_n - V) \\
&- \underbrace{\frac{\left(\lambda^F \breve{v}(S_n,\hat{V}(S_n))g(\hat{V}(S_n)) + \lambda^E f(\hat{V}(S_n))\right)(S_n - V)}{\lambda^F \hat{G}(S_n)\hat{v}(S_n)}}_{=1} \\
&\times \Bigg\{ \left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n)\right]\left(1 - \hat{V}'(S_n)\right) \\
&\quad - \lambda^F \hat{g}(S_n)\hat{v}(S_n)\hat{V}'(S_n) - \lambda^F \hat{G}(S_n)\hat{v}'(S_n)\hat{V}'(S_n) \Bigg\} \Bigg] \\
&+ \lambda^F g(V)\frac{\partial \breve{v}(S_n,V)}{\partial V}(S_n - V) \\
&- \left[\lambda^F g(V)\breve{v}(S_n,V) + \lambda^E f(V)\right] \\
&- \lambda^F g(V)\breve{v}(S_n,V) - \lambda^F G(V)\frac{\partial \breve{v}(S_n,V)}{\partial V} \\
=&- \frac{1}{\hat{V}'(S_n)^2}\lambda^F \hat{g}(S_n)\hat{v}'(S_n)(S_n - V) \\
&- \frac{1}{\hat{V}'(S_n)^2}\left\{ \left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n)\right] - \lambda^F \hat{g}(S_n)\hat{v}(S_n)\hat{V}'(S_n) - \lambda^F \hat{G}(S_n)\hat{v}'(S_n)\hat{V}'(S_n) \right\} \\
&+ \lambda^F g(V)\frac{\partial \breve{v}(S_n,V)}{\partial V}(S_n - V) - \lambda^F g(V)\breve{v}(S_n,V) - \lambda^F G(V)\frac{\partial \breve{v}(S_n,V)}{\partial V}
\end{aligned}
$$

Note that the rank-preserving property implies

$$\lambda^F \hat{G}(S_n)\hat{v}(S_n) = \lambda^F G(\hat{V}(S_n))\breve{v}(S_n,\hat{V}(S_n)) \tag{101}$$

Differentiating the above equation with respect to S_n, we have

$$
\begin{aligned}
&\lambda^F \hat{g}(S_n)\hat{v}(S_n)\frac{1}{\hat{V}'(S_n)} + \frac{1}{\hat{V}'(S_n)}\lambda^F \hat{G}(S_n)\hat{v}'(S_n) && (102) \\
&= \lambda^F g(V)\hat{v}(S_n) + \lambda^F G(\hat{V}(S_n))\frac{\partial \breve{v}(S_n,\hat{V}(S_n))}{\partial S_n}\frac{1}{\hat{V}'(S_n)} + \lambda^F G(\hat{V}(S_n))\frac{\partial \breve{v}(S_n,\hat{V}(S_n))}{\partial V} && (103)
\end{aligned}
$$

Using this equation, we can simplify the second-order condition as

$$\mathcal{G}(S_n,V) = -\frac{1}{\hat{V}'(S_n)^2}\lambda^F \hat{g}(S_n)\hat{v}'(S_n)(S_n - V)$$

$$
\begin{aligned}
&- \frac{1}{\hat{V}'(S_n)^2}\left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n)\right] + \lambda^F G(\hat{V}(S_n))\frac{1}{\hat{V}'(S_n)}\frac{\partial \breve{v}(S_n, \hat{V}(S_n))}{\partial S_n} \\
&+ \lambda^F g(V)\frac{\partial \breve{v}(S_n, V)}{\partial V}(S_n - V)
\end{aligned}
$$

We can rewrite the last term as

$$
\begin{aligned}
\lambda^F g(V)\frac{\partial \breve{v}(S_n, V)}{\partial V}(S_n - V) &= \lambda^F \frac{\hat{g}(S_n)}{\hat{V}'(S_n)}\frac{\partial \breve{v}(S_n, V)}{\partial V}(S_n - V) \\
&= \lambda^F \frac{\hat{g}(S_n)}{\hat{V}'(S_n)^2}\left[\hat{v}'(S_n) - \frac{\partial \breve{v}(S_n, V)}{\partial S_n}\right](S_n - V).
\end{aligned}
$$

Plugging this back into the second-order condition, we have

$$
\begin{aligned}
\mathcal{G}(S_n, V) = &- \frac{1}{\hat{V}'(S_n)^2}\left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n)\right] + \lambda^F G(\hat{V}(S_n))\frac{\partial \breve{v}(S_n, \hat{V}(S_n))}{\partial S_n}\frac{1}{\hat{V}'(S_n)} \\
&- \lambda^F \frac{\hat{g}(S_n)}{\hat{V}'(S_n)^2}\frac{\partial \breve{v}(S_n, V)}{\partial S_n}(S_n - V)
\end{aligned}
$$

Note that

$$
\frac{\partial \breve{v}(S_n, \hat{V}(S_n))}{\partial S_n} = \hat{v}(S_n)\frac{1}{\gamma}\frac{1}{S_n - \hat{V}(S_n)}.
$$

Therefore, we have

$$
\begin{aligned}
\mathcal{G}(S_n, V) &= -\frac{1}{\hat{V}'(S_n)^2}\left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n)\right] + \frac{\lambda^F G(\hat{V}(S_n))\hat{v}(S_n)}{S_n - \hat{V}(S_n)}\frac{1}{\gamma}\frac{1}{\hat{V}'(S_n)} \\
&\quad - \lambda^F \frac{\hat{g}(S_n)}{\hat{V}'(S_n)^2}\frac{\partial \breve{v}(S_n, V)}{\partial S_n}(S_n - V) \\
&= -\frac{1}{\hat{V}'(S_n)^2}\left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n)\right] + \left[\lambda^F g(\hat{V}(S_n))\hat{v}(S_n) + \lambda^E f(\hat{V}(S_n))\right]\frac{1}{\gamma}\frac{1}{\hat{V}'(S_n)} \\
&\quad - \lambda^F \frac{\hat{g}(S_n)}{\hat{V}'(S_n)^2}\frac{\partial \breve{v}(S_n, V)}{\partial S_n}(S_n - V) \\
&= -\frac{1}{\hat{V}'(S_n)^2}\left[\lambda^F \hat{g}(S_n)\hat{v}(S_n) + \lambda^E \hat{f}(S_n)\right]\left(1 - \frac{1}{\gamma}\right) \\
&\quad - \lambda^F \frac{\hat{g}(S_n)}{\hat{V}'(S_n)^2}\frac{\partial \breve{v}(S_n, V)}{\partial S_n}(S_n - V) \\
&< 0,
\end{aligned}
$$

where the last inequality follows from the fact that $\gamma \geq 1$. This verifies the second-order condition, (76).

A.7 Proof of Proposition 5

This straightforwardly follows after substituting the equilibrium promised utility conditions and the vacancy policy function (92) into the firm growth rate definition:

$$\frac{dn}{n} = \lambda^F G(\hat{V}(S_n))\hat{v}(S_n) - s - \lambda^E(1 - F(\hat{V}(S_n))) \tag{104}$$

$$= \lambda^F \hat{G}(S_n)\hat{v}(S_n) - s - \lambda^E(1 - \hat{F}(S_n)). \tag{105}$$

A.8 Proof of Proposition 6

Consider the Hamiltonian of Problem 2. We let $S_t^p(z, n)$ denote the co-state variable associated with the law of motion of the distribution. To ease notation, we suppress time subscripts throughout this proof, with all objects understood to be evaluated at time t along the optimal path. The current value Hamiltonian is given by

$$\begin{aligned}
\mathcal{H} = &\int (y(z,n) - c(v(z,n))n)\, d\phi + (1 - \textstyle\int n d\phi)b - m^0 c^e \\
&+ \int\int S^p(z,n)\Big[-\partial_n\left[dn(z,n)\phi(z,n)\right] - \partial_z[\mu(z)\phi(z,n)] + \frac{1}{2}\partial_{zz}[(\sigma(z))^2\phi(z,n)] \\
&\quad - (\varkappa + \varsigma^e(z,n))\phi(z,n) + m^0\psi(z,n)\Big] dzdn \\
= &\int (y(z,n) - c(v(z,n))n)\, d\phi + (1 - \textstyle\int n d\phi)b - m^0 c^e \\
&+ \int\int \Big[S_n^p(z,n)dn(z,n)\phi(z,n) + S_z^p(z,n)\mu(z)\phi(z,n) + \frac{1}{2}S_{zz}^p(z,n)(\sigma(z))^2\phi(z,n) \\
&\quad - (\varkappa + \varsigma^e(z,n))S^p(z,n)\phi(z,n) + S^p(z,n)m^0\psi(z,n)\Big] dzdn,
\end{aligned}$$

where the second line uses integration by parts,

$$dn(z,n) = \left[\lambda^F G^p(z,n)v(z,n) - s - \lambda^E(1 - F^p(z,n))\right] n - \varsigma^f(z,n)$$

and

$$\begin{aligned}
G^p(z,n) &\equiv \frac{1}{\tilde{u}}\left[x_{u,zn}\left(1 - \int n d\phi\right) + \zeta \int x_{z'n',zn} n' d\phi(z',n')\right] \\
1 - F^p(z,n) &\equiv \frac{1}{\tilde{v}} \int x_{zn,z'n'} v(z',n') n' d\phi(z',n')
\end{aligned}$$

$$\lambda^F \equiv \frac{M(\tilde{u},\tilde{v})}{\tilde{v}}, \quad \lambda^E \equiv \zeta \frac{M(\tilde{u},\tilde{v})}{\tilde{u}}, \quad \tilde{u} \equiv 1 - \int n d\phi + \zeta \int n d\phi, \quad \tilde{v} \equiv \int v(z,n) n d\phi$$

and $x_{u,zn} \in [0,1]$, $x_{zn,z'n'} \in [0,1]$, $v(z,n) \geq 0$, $\varsigma^e(z,n) \geq 0$, and $\varsigma^f(z,n) \geq 0$.

The FOC w.r.t. $x_{u,zn}$ implies

$$x_{u,zn} = \begin{cases} 1 & \text{if } S^p_n(z,n) \geq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{106}$$

Likewise, the derivative of the Hamiltonian w.r.t. $x_{z'n',zn}$ is

$$\frac{\partial \mathcal{H}}{\partial x_{z'n',zn}} = \Bigg[\lambda^F \frac{\zeta}{\tilde{u}} v(z,n) n' n d\phi(z',n') d\phi(z,n) S^p_n(z,n) \tag{107}$$

$$- \lambda^E \frac{1}{\tilde{v}} v(z,n) n n' d\phi(z,n) d\phi(z',n') S^p_n(z',n')\Bigg] \tag{108}$$

$$= \zeta \frac{M(\tilde{u},\tilde{v})}{\tilde{v}\tilde{u}} \left[S^p_n(z,n) - S^p_n(z',n')\right], \tag{109}$$

which implies

$$x_{z'n',zn} = \begin{cases} 1 & \text{if } S^p_n(z,n) \geq S^p_n(z',n') \\ 0 & \text{otherwise} \end{cases}. \tag{110}$$

The optimality condition w.r.t. $\varsigma^f(z,n)$ is

$$S^p_n(z,n) \geq 0,$$

with $\varsigma^f(z,n) > 0$ only where $S^p_n(z,n) = 0$. Likewise, since $\varsigma^e(z,n)$ enters the Hamiltonian linearly with coefficient $-S^p(z,n)\phi(z,n)$, the optimality condition w.r.t. $\varsigma^e(z,n)$ is

$$S^p(z,n) \geq 0,$$

with $\varsigma^e(z,n) > 0$ only where $S^p(z,n) = 0$. These two conditions correspond to the boundary conditions in Proposition 6.

The FOC w.r.t. $v(z,n)$ is

$$\begin{aligned}
&\lambda^F G^p(z,n) S_n^p(z,n) - \frac{\lambda^E}{\tilde{v}} \int x_{z'n',zn} S_n^p(z',n') n' d\phi(z',n') \\
&\quad + \frac{\partial \ln \lambda^F}{\partial \tilde{v}} \int \lambda^F G^p(z,n) v(z,n) n S_n^p(z,n) d\phi(z,n) \\
&\quad - \frac{\partial \ln \lambda^F}{\partial \tilde{v}} \int \lambda^E (1 - F^p(z,n)) n S_n^p(z,n) d\phi(z,n) = c'(v(z,n)).
\end{aligned} \tag{111}$$

We seek to rewrite the left-hand side of the above expression as

$$\frac{\partial \ln \lambda^F}{\partial \tilde{v}} \int \lambda^F G^p(z,n) v(z,n) n S_n^p(z,n) d\phi(z,n) \tag{112}$$

$$- \frac{\partial \ln \lambda^F}{\partial \tilde{v}} \int \lambda^E (1 - F^p(z,n)) n S_n^p(z,n) d\phi(z,n) \tag{113}$$

$$= \frac{\partial \ln \lambda^F}{\partial \ln \tilde{v}} \lambda^F \int G^p(z',n') \left(S_n^p(z',n') - \mathbb{E}_{\hat{G}}[\tilde{S}_n | \tilde{S}_n \leq S_n^p(z',n')] \right) \frac{1}{\tilde{v}} v(z',n') n' d\phi(z',n'), \tag{114}$$

where

$$\begin{aligned}
\mathbb{E}_{\hat{G}}[\tilde{S}_n | \tilde{S}_n \leq S_n^p(z,n)] &\equiv \frac{1}{G^p(z,n)} \frac{1}{\tilde{u}} \left[\zeta \int x_{z'n',zn} S_n^p(z',n') n' d\phi(z',n') \right] \\
&= \frac{1}{G^p(z,n)} \frac{1}{\tilde{u}} \left[\zeta \int_{z',n' : S_n^p(z',n') \leq S_n^p(z,n)} S_n^p(z',n') n' d\phi(z',n') \right]
\end{aligned}$$

is the weighted average of the marginal surplus conditional on the marginal surplus being less than $S_n^p(z,n)$, and the weight is given by the search intensity weighted employment distribution. Note that the marginal surplus of an unemployed worker is zero. The second line in the above expression imposes the optimality conditions for $x_{z'n',zn}$ in (110).

Define

$$\Omega \equiv \int G^p(z',n') \left(S_n^p(z',n') - \mathbb{E}_{\hat{G}}[\tilde{S}_n | \tilde{S}_n \leq S_n^p(z',n')] \right) \frac{1}{\tilde{v}} v(z',n') n' d\phi(z',n')$$

as the average gains in the marginal surplus from a meeting. Using this definition, we can rewrite (111) as

$$\lambda^F \left\{ G^p(z,n) \left(S_n^p(z,n) - \mathbb{E}_{\hat{G}}[\tilde{S}_n | \tilde{S}_n \leq S_n^p(z,n)] \right) - \eta^F \Omega \right\} = c'(v(z,n)),$$

where

$$\eta^F \equiv - \frac{\partial \ln \lambda^F}{\partial \ln \tilde{v}}$$

is the congestion effect from additional vacancy posting.

The optimality condition w.r.t. $\phi(z,n)$, evaluated at the steady state, is

$$
\begin{aligned}
(r+\varkappa)S^p(z,n) &= y(z,n) - c(v(z,n))n - bn \\
&\quad + dn(z,n)S^p_n(z,n) + \mu(z)S^p_z(z,n) + \frac{1}{2}(\sigma(z))^2 S^p_{zz}(z,n) \\
&\quad - \frac{1}{\tilde{u}}\int \lambda^F v(z',n')x_{u,z'n'}S^p_n(z',n')n'd\phi(z',n')n \\
&\quad + \frac{\zeta}{\tilde{u}}\int \lambda^F v(z',n')x_{zn,z'n'}S^p_n(z',n')n'd\phi(z',n')n \\
&\quad - \lambda^E\frac{1}{\tilde{v}}v(z,n)\int x_{z'n',zn}n'S^p_n(z',n')d\phi(z',n')n \\
&\quad - (1-\zeta)\frac{\partial \ln \lambda^U}{\partial \tilde{u}}\int \lambda^F G^p(z',n')v(z',n')n'S^p_n(z',n')d\phi(z',n')n \\
&\quad + (1-\zeta)\frac{\partial \ln \lambda^E}{\partial \tilde{u}}\int \lambda^E(1-F^p(z',n'))n'S^p_n(z',n')d\phi(z',n')n \\
&\quad + v(z,n)n\frac{\partial \ln \lambda^F}{\partial \tilde{v}}\int \lambda^F G^p(z,n)v(z,n)nS^p_n(z,n)d\phi(z,n) \\
&\quad - v(z,n)n\frac{\partial \ln \lambda^F}{\partial \tilde{v}}\int \lambda^E(1-F^p(z,n))nS^p_n(z,n)d\phi(z,n) \\
&= y(z,n) - c(v(z,n))n - bn \\
&\quad + dn(z,n)S^p_n(z,n) + \mu(z)S^p_z(z,n) + \frac{1}{2}(\sigma(z))^2 S^p_{zz}(z,n) \\
&\quad - \lambda^U\int \frac{1}{\tilde{v}}v(z',n')x_{u,z'n'}S^p_n(z',n')n'd\phi(z',n')n \\
&\quad + \lambda^E\int \frac{1}{\tilde{v}}v(z',n')x_{zn,z'n'}S^p_n(z',n')n'd\phi(z',n')n \\
&\quad - \lambda^F v(z,n)\int \frac{1}{\tilde{u}}\zeta x_{z'n',zn}n'S^p_n(z',n')d\phi(z',n')n \\
&\quad - (1-\zeta)\frac{\partial \ln \lambda^U}{\partial \ln \tilde{u}}\lambda^U\Omega n \\
&\quad + \frac{\partial \ln \lambda^F}{\partial \ln \tilde{v}}\lambda^F\Omega v(z,n)n \\
&= y(z,n) - c(v(z,n))n - bn \\
&\quad + dn(z,n)S^p_n(z,n) + \mu(z)S^p_z(z,n) + \frac{1}{2}(\sigma(z))^2 S^p_{zz}(z,n) \\
&\quad - \lambda^U\left\{\mathbb{E}_{\hat{F}}[\tilde{S}_n] - \eta^U\Omega\right\}n \\
&\quad + \lambda^E\left\{(1-\hat{F}(S^p_n))\mathbb{E}_{\hat{F}}[\tilde{S}_n|\tilde{S}_n \geq S^p_n] - \eta^U\Omega\right\}n \\
&\quad - \lambda^F\left\{\hat{G}(S^p_n)\mathbb{E}_{\hat{G}}[\tilde{S}_n|\tilde{S}_n \leq S^p_n] + \eta^F\Omega\right\}v(z,n)n
\end{aligned}
$$

A.9 Proof of Proposition 7

We now introduce taxes in our baseline model. We consider two tax instruments on firms. We assume that taxes depend on the firm's productivity and size. The first is a vacancy tax, $\tau^v(z,n)$, which firms have to pay for each unit of vacancy they post. The second is a labor tax, $\tau^n(z,n)$, which firms have to pay for each worker they employ. These taxes are financed by a lump-sum tax on workers. With these taxes, the HJB equation for the surplus function $S(z,n)$ becomes

$$\begin{aligned}\rho S(z,n) = \max_{v,V}\; & y(z,n) - c(v)n - rUn - \tau^n(z,n)n - \left[V\lambda^F G(V) + \tau^v(z,n)\right]vn \\ & + \lambda^E \int_V^\infty \tilde{V} dF(\tilde{V})n + \mathcal{L}(z,n,v,V)S(z,n)\end{aligned}$$

where $\rho \equiv r + \varkappa$, $\mathcal{L}(z,n,v,V)$ is the generator defined for an arbitrary function $S(z,n)$:

$$\begin{aligned}\mathcal{L}(z,n,v,V)S(z,n) = \mu(z)\partial_z S(z,n) + \frac{\sigma^2(z)}{2}\partial_{zz}S(z,n) \\ + \left[\lambda^F G(V)v - s - \lambda^E(1-F(V))\right] n\partial_n S(z,n),\end{aligned} \tag{115}$$

with the following boundary conditions:

$$S(z,n) \geq 0, \quad S_n(z,n) \geq 0. \tag{116}$$

The first-order condition with respect to V is unchanged and given by (29). The first-order condition with respect to v is given by

$$\lambda^F G(V)\left(\partial_n S(z,n) - V\right) - \tau^v(z,n) = c'(v), \tag{117}$$

which we can solve for v as

$$v(z,n) = c'^{-1}\left(\lambda^F G(V)\left(\partial_n S(z,n) - V\right) - \tau^v(z,n)\right). \tag{118}$$

The rest of the equilibrium conditions remain unchanged. Consequently, the equilibrium characterization in Proposition 4 remains the same. Imposing the equilibrium promised utility offer, we have

$$\begin{aligned}\rho S(z,n) = & \; y(z,n) - c(v(z,n))n - bn - \tau^n(z,n)n - \lambda^U \mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n)]n \\ & - \left[\lambda^F \hat{G}(S_n)\hat{V}(S_n) + \tau^v(z,n)\right] v(z,n)n \\ & + \lambda^E(1-\hat{F}(S_n))\mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n)|\tilde{S}_n \geq S_n]n + \mathcal{L}(z,n,v(z,n),\hat{V}(S_n))S(z,n)\end{aligned}. \tag{119}$$

Comparing (119) with (43), the following taxes are needed to implement the efficient allocation:

$$\tau^v(z,n) = \lambda^F \left\{ \hat{G}(S_n^p)\mathbb{E}_{\hat{G}}[\tilde{S}_n^p | \tilde{S}_n^p \leq S_n^p] + \eta^F \Omega \right\} - \lambda^F \hat{G}(S_n^p)\hat{V}(S_n^p) \tag{120}$$

$$\tau^n(z,n) = \lambda^U \left\{ \mathbb{E}_{\hat{F}}[\tilde{S}_n^p] - \eta^U \Omega \right\} - \lambda^E \left\{ (1-\hat{F}(S_n^p))\mathbb{E}_{\hat{F}}[\tilde{S}_n^p | \tilde{S}_n^p \geq S_n^p] - \eta^U \Omega \right\} \tag{121}$$

$$- \lambda^U \mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n^p)] + \lambda^E (1-\hat{F}(S_n^p))\mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n^p) | \tilde{S}_n^p \geq S_n^p]. \tag{122}$$

Note that both $\tau^v(z,n)$ and $\tau^n(z,n)$ depend on (z,n) only through the marginal surplus $S_n^p(z,n)$. For this reason, we can write taxes as a function of the marginal surplus, $\hat{\tau}^v(S_n^p)$ and $\hat{\tau}^n(S_n^p)$.

With these taxes, by construction, the HJB equation (119) and the vacancy policy (118) in the equilibrium with taxes coincide with those of the planner's problem, (43) and (45). Since these two conditions are the only places where the equilibrium and the planner's problem differ, this implies that the equilibrium with taxes implements the efficient allocation.

A.10 Proof of Proposition 8

As we discussed above, we can rewrite the vacancy tax and employment tax as a function of the marginal surplus S_n^p. The vacancy tax is given by

$$\begin{aligned}
\hat{\tau}^v(S_n^p) &= \lambda^F \left\{ \hat{G}(S_n^p)\mathbb{E}_{\hat{G}}[\tilde{S}_n^p | \tilde{S}_n^p \leq S_n^p] + \eta^F \Omega \right\} - \lambda^F \hat{G}(S_n^p)\hat{V}(S_n^p) \\
&= \lambda^F \left[\int_0^{S_n^p} \tilde{S}_n^p \hat{g}(\tilde{S}_n^p) d\tilde{S}_n^p - \hat{G}(S_n^p) \frac{1}{\hat{G}(S_n^p)^2} \int_0^{S_n^p} \tilde{S}_n^p 2\hat{G}(\tilde{S}_n^p)\hat{g}(\tilde{S}_n^p) d\tilde{S}_n^p \right] \\
&\quad + \lambda^F \eta^F \Omega.
\end{aligned}$$

Taking the derivative with respect to S_n^p, we obtain

$$\begin{aligned}
\frac{d\hat{\tau}^v(S_n^p)}{dS_n^p} &= \lambda^F \left[S_n^p \hat{g}(S_n^p) - 2S_n^p \hat{g}(S_n^p) + \frac{\hat{g}(S_n^p)}{\hat{G}(S_n^p)^2} \int_0^{S_n^p} \tilde{S}_n^p 2\hat{G}(\tilde{S}_n^p)\hat{g}(\tilde{S}_n^p) d\tilde{S}_n^p \right] \\
&= \lambda^F \hat{g}(S_n^p) \left[S_n^p - 2S_n^p + \mathbb{E}_{\hat{G}^2}[\tilde{S}_n^p | \tilde{S}_n^p \leq S_n^p] \right] \\
&< 0,
\end{aligned}$$

where the last inequality follows from the fact that $\mathbb{E}_{\hat{G}^2}[\tilde{S}_n^p | \tilde{S}_n^p \leq S_n^p] < S_n^p$. Therefore, the vacancy tax is strictly decreasing in the marginal surplus S_n^p. When the marginal surplus is zero, we have

$$\hat{\tau}^v(0) = \lambda^F \eta^F \Omega > 0.$$

Now we turn to the employment tax. The employment tax is given by

$$
\begin{aligned}
\hat{\tau}^n(S_n^p) &= \lambda^U \left\{ \mathbb{E}_{\hat{F}}[\tilde{S}_n^p] - \eta^U \Omega \right\} - \lambda^E \left\{ (1-\hat{F}(S_n^p))\mathbb{E}_{\hat{F}}[\tilde{S}_n^p | \tilde{S}_n^p \geq S_n^p] - \eta^U \Omega \right\} \\
&\quad - \lambda^U \mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n^p)] + \lambda^E (1-\hat{F}(S_n^p))\mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n^p)|\tilde{S}_n^p \geq S_n^p] \\
&= \lambda^U \left\{ \mathbb{E}_{\hat{F}}[\tilde{S}_n^p] - \eta^U \Omega \right\} - \lambda^E \left\{ \int_{S_n^p}^{\infty} \tilde{S}_n^p \hat{f}(\tilde{S}_n^p) d\tilde{S}_n^p - \eta^U \Omega \right\} \\
&\quad - \lambda^U \mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n^p)] + \lambda^E \int_{S_n^p}^{\infty} \hat{V}(\tilde{S}_n^p) \hat{f}(\tilde{S}_n^p) d\tilde{S}_n^p
\end{aligned}
$$

Taking the derivative with respect to S_n^p, we have

$$
\begin{aligned}
\frac{d\hat{\tau}^n(S_n^p)}{dS_n^p} &= \lambda^E S_n^p \hat{f}(S_n^p) - \lambda^E \hat{V}(S_n^p)\hat{f}(S_n^p) \\
&= \lambda^E \hat{f}(S_n^p) \left(S_n^p - \mathbb{E}_{\hat{G}^2}[\tilde{S}_n^p | \tilde{S}_n^p \leq S_n^p] \right) \\
&> 0,
\end{aligned}
$$

where the last inequality follows from the fact that $\mathbb{E}_{\hat{G}^2}[\tilde{S}_n^p | \tilde{S}_n^p \leq S_n^p] < S_n^p$. Therefore, the employment tax is strictly increasing in the marginal surplus S_n^p.

A.11 Proof of Proposition 9

We impose the following two assumptions. First, we assume that the production function is constant returns to scale in (z, n):

$$
y(z,n) = z\,\hat{y}(\hat{n}), \quad \hat{n} \equiv \frac{n}{z}, \tag{123}
$$

where $\hat{y}(\hat{n}) \equiv y(1, \hat{n})$. Second, we assume that the diffusion process of z is geometric Brownian motion:

$$
\mu(z) = \mu z, \quad \sigma(z) = \sigma z, \tag{124}
$$

where μ and σ are constants.

With these two assumptions, the HJB equation in Problem 1 can be rewritten as

$$
\rho S = \max_{v,V} z\,\hat{y}(\hat{n}) - c(v)n - rUn - V\lambda^F G(V)vn + \lambda^E \int_V \tilde{V} dF(\tilde{V})n + \mathcal{L}(z,n,v,V)S \tag{125}
$$

with

$$\mathcal{L}(z,n,v,V)S(z,n) = \mu z\,\partial_z S(z,n) + \frac{(\sigma z)^2}{2}\partial_{zz}S(z,n) \\ + \left[\lambda^F G(V)v - s - \lambda^E(1-F(V))\right] n\partial_n S(z,n), \tag{126}$$

We now guess that the value function takes the form:

$$S(z,n) = z\,\hat{S}(\hat{n}). \tag{127}$$

With the guess, we have

$$\partial_z S(z,n) = \hat{S}(\hat{n}) - \hat{n}\,\hat{S}'(\hat{n}), \tag{128}$$

$$\partial_{zz}S(z,n) = \frac{\hat{n}^2}{z}\,\hat{S}''(\hat{n}), \tag{129}$$

$$\partial_n S(z,n) = \hat{S}'(\hat{n}). \tag{130}$$

Substituting (127)-(130) into (125), we get

$$\rho z\hat{S}(\hat{n}) = \max_{v,V}\ z\hat{y}(\hat{n}) - c(v)n - rUn - V\lambda^F G(V)vn + \lambda^E \int_V \tilde{V}dF(\tilde{V})\,n \\ + \mu z\left[\hat{S}(\hat{n}) - \hat{n}\,\hat{S}'(\hat{n})\right] + \frac{\sigma^2 z\hat{n}^2}{2}\hat{S}''(\hat{n}) \\ + \left[\lambda^F G(V)v - s - \lambda^E(1-F(V))\right] n\,\hat{S}'(\hat{n}).$$

Dividing both sides by z and using $n = z\hat{n}$, the factor z cancels entirely, yielding a one-dimensional HJB in $\hat{n}$:

$$\hat{\rho}\,\hat{S}(\hat{n}) = \max_{v,V}\ \hat{y}(\hat{n}) - c(v)\hat{n} - rU\hat{n} - V\lambda^F G(V)v\hat{n} + \lambda^E \int_V \tilde{V}dF(\tilde{V})\,\hat{n} + \hat{\mathcal{L}}(\hat{n},v,V)\hat{S}(\hat{n}), \tag{131}$$

where $\hat{\rho} \equiv \rho - \mu$ and the one-dimensional generator is

$$\hat{\mathcal{L}}(\hat{n},v,V)\hat{S}(\hat{n}) = \frac{\sigma^2\hat{n}^2}{2}\hat{S}''(\hat{n}) + \left[\lambda^F G(V)v - s - \lambda^E(1-F(V)) - \mu\right]\hat{n}\,\hat{S}'(\hat{n}), \tag{132}$$

with boundary conditions $\hat{S}(\hat{n}) \geq 0$ and $\hat{S}'(\hat{n}) \geq 0$.

The law of motion for $\hat{n}$ is given by

$$d\hat{n} = \underbrace{\left[\lambda^F G(V(\hat{n}))v(\hat{n}) - s - \lambda^E(1 - F(V(\hat{n}))) - \mu + \sigma^2\right]\hat{n}}_{\equiv \mu_{\hat{n}}(\hat{n})}\, dt - \sigma\hat{n}\, dZ_t. \tag{133}$$

B Computation Algorithm

We divide the computation algorithm into two parts: (i) calibration and (ii) counterfactual. The former fixes θ at the target and calibrates the entry cost to match the target. The latter fixes the entry cost and calibrates θ to ensure the free-entry condition is satisfied.

B.1 Algorithm for Calibration

In calibrating our model, we employ the following algorithm, as summarized in Algorithm 1. Throughout, we fix the value of $\theta = 1$, which is our normalization. Using θ, we can compute the meeting rates, λ^U, λ^E, and λ^F using (7). We start by guessing the GE objects $\{F(V), G(V)\}$. This also allows us to compute U using (27). With these objects, we solve for the HJB equation described in Problem 1 (which can be simplified to the one in Proposition 9 under homogeneity).

Algorithm for Solving the HJB Equation. In solving Problem 1, we use its equivalent HJB quasi-variational inequality (QVI) form. Define the intervention operator as

$$(\mathcal{M}S)(z,n) \equiv \max\left\{0, \sup_{\tilde{n}\leq n} S(z,\tilde{n})\right\}. \tag{134}$$

The first term in the maximum is the value of endogenous exit, while the supremum over $\tilde{n} \leq n$ is the value of instantaneously firing workers and continuing at a smaller employment level. The QVI is

$$0 = \min\left\{\begin{aligned}&\rho S(z,n) - \max_{v\geq 0, V\geq 0}\Big[y(z,n) - c(v)n - rUn - V\lambda^F G(V)vn \\ &\qquad\qquad\qquad + \lambda^E n\int_V^\infty \tilde{V}\,dF(\tilde{V}) + \mathcal{L}(z,n,v,V)S(z,n)\Big], \\ &S(z,n) - (\mathcal{M}S)(z,n)\end{aligned}\right\}. \tag{135}$$

On the continuation region, where $S(z,n) > (\mathcal{M}S)(z,n)$, the first term in (135) binds and the firm solves the HJB equation. On the intervention region, $S(z,n) = (\mathcal{M}S)(z,n)$, so the firm either exits or fires down to an employment point that attains the intervention value. This is the numerical counterpart of the boundary conditions $S(z,n) \geq 0$ and $S_n(z,n) \geq 0$ in Problem 1.

We solve (135) on a log-spaced tensor grid for (z,n) using Howard policy iteration. Starting from an initial guess $\mathbf{S}^{(0)} = \mathbf{0}$, the policy-improvement step searches over a grid for the promised utility V. For each grid on the state space $i \equiv (i_z, i_n)$ and candidate V, the vacancy policy is computed from the first-order condition using the upwind derivative of $\mathbf{S}^{(k)}$ with respect to employment,

$$v_i^{(k)}(V) = c'^{-1}\left(\left[\lambda^F G(V)\left(\mathbf{S}_{n,i}^{(k)} - V\right)\right]_+\right), \tag{136}$$

where $\mathbf{S}_{n,i}^{(k)}$ is the finite difference approximation of $S_n(z_{i_z}, n_{i_n})$.

We use the upwinding scheme described in Achdou et al. (2022). Specifically, let

$$g_{n,i}^F(V) \equiv \lambda^F G(V) v_i^{(k),F}(V) - s - \lambda^E(1 - F(V)) \tag{137}$$
$$g_{n,i}^B(V) \equiv \lambda^F G(V) v_i^{(k),B}(V) - s - \lambda^E(1 - F(V)) \tag{138}$$

denote the employment drift when using the forward difference and backward difference, respectively. We use forward difference when $g_{n,i}^F(V) > 0$ and $g_{n,i}^B(V) > 0$. We use backward difference when $g_{n,i}^F(V) < 0$ and $g_{n,i}^B(V) < 0$. We set $v_i^{(k)}(V) = \frac{s+\lambda^E(1-F(V))}{\lambda^F G(V)}$ so that the employment drift is zero when $g_{n,i}^F(V) > 0$ and $g_{n,i}^B(V) < 0$. If $g_{n,i}^F(V) < 0$ and $g_{n,i}^B(V) > 0$, we pick the one that maximizes the HJB right-hand side. We then choose $V_i^{(k)}$ as the grid point that maximizes the HJB right-hand side and set $v_i^{(k)} = v_i^{(k)}(V_i^{(k)})$.

The same policy-improvement step computes the exit and firing policies by applying $\mathcal{M}$ to $\mathbf{S}^{(k)}$ and comparing the intervention value with the continuation value: at each grid point the firm either continues, exits, or fires down to the grid point $j(i)$ that attains $\max_{\tilde{n} \le n_i} S^{(k)}(z_i, \tilde{n})$. Let $\mathcal{E}^{(k)}$ and $\mathcal{F}^{(k)}$ denote the sets of exit and firing nodes selected in this step. We collect them in the diagonal intervention matrix

$$\mathbf{D}^{I,(k)} \equiv \operatorname{diag}\left(d_i^{I,(k)}\right), \qquad d_i^{I,(k)} \equiv \mathbf{1}\{i \in \mathcal{E}^{(k)} \cup \mathcal{F}^{(k)}\}. \tag{139}$$

The complementary diagonal matrix $\mathbf{D}^{C,(k)} \equiv \mathbf{I} - \mathbf{D}^{I,(k)}$ selects continuation nodes. We also define the fixed-policy intervention matrix $\mathbf{M}^{(k)}$ by

$$M_{i\ell}^{(k)} \equiv \begin{cases} 1, & i \in \mathcal{F}^{(k)} \text{ and } \ell = j(i), \\ 0, & \text{otherwise.} \end{cases} \tag{140}$$

Thus $\mathbf{M}^{(k)}$ has a zero row for continuation and exit nodes, and has a single nonzero entry in row i for firing nodes, selecting the post-firing grid point $j(i)$. In the KFE below, $\mathbf{M}^{(k)\top}$ is the corresponding map for moving mass from pre-intervention firing states to post-firing states.

Given the resulting policy $(v^{(k)}, V^{(k)})$, we can compute $\mathbf{A}^{(k)}$, which denotes the finite-difference discretization of the generator $\mathcal{L}(z, n, v, V)$, using an upwind stencil for the employment drift and the standard second-order stencil for the productivity diffusion. Let $\boldsymbol{\pi}^{(k)}$ collect the static payoff terms,

$$\pi_i^{(k)} = y(x_i) - c(v_i^{(k)})n_i - rUn_i - V_i^{(k)}\lambda^F G(V_i^{(k)})v_i^{(k)}n_i + \lambda^E n_i \int_{V_i^{(k)}}^{\infty} \tilde{V}\, dF(\tilde{V}). \tag{141}$$

The policy-evaluation step updates $\mathbf{S}$ by solving the linear system implied by the fixed continuation and intervention policies. In matrix form, the updated $\mathbf{S}^{(k+1)}$ solves

$$\left[\mathbf{D}^{I,(k)} + \mathbf{D}^{C,(k)}\left(\rho\mathbf{I} - \mathbf{A}^{(k)}\right)\right]\mathbf{S}^{(k+1)} = \mathbf{D}^{I,(k)}\mathbf{M}^{(k)}\mathbf{S}^{(k)} + \mathbf{D}^{C,(k)}\boldsymbol{\pi}^{(k)}, \tag{142}$$

which we can invert to obtain

$$\mathbf{S}^{(k+1)} = \left[\mathbf{D}^{I,(k)} + \mathbf{D}^{C,(k)}\left(\rho\mathbf{I} - \mathbf{A}^{(k)}\right)\right]^{-1}\left[\mathbf{D}^{I,(k)}\mathbf{M}^{(k)}\mathbf{S}^{(k)} + \mathbf{D}^{C,(k)}\boldsymbol{\pi}^{(k)}\right]. \tag{143}$$

We then repeat the above process until both the value function and the policies have converged.

Algorithm for Solving the KFE Equation. Since the distribution ϕ is homogeneous in the entry mass m^0, we first solve for the normalized distribution ϕ^0 by setting $m^0 = 1$. Let $\mathbf{M}$ denote the converged intervention matrix defined above. Given the converged policies and given (z_0, n_0) that forms $\boldsymbol{\psi}$, the normalized distribution is obtained from the discrete KFE

$$0 = \left[\mathbf{D}^C\left[(\mathbf{AM})^\top - \varkappa\mathbf{I}\right] + \mathbf{D}^I\right]\boldsymbol{\phi}^0 + \mathbf{M}\boldsymbol{\psi}, \tag{144}$$

which we can invert to obtain

$$\boldsymbol{\phi}^0 = -\left[\mathbf{D}^C\left[(\mathbf{AM})^\top - \varkappa\mathbf{I}\right] + \mathbf{D}^I\right]^{-1}\mathbf{M}\boldsymbol{\psi}. \tag{145}$$

Let

$$N^0 \equiv \sum_i n_i\phi_i^0, \qquad \mathcal{V}^0 \equiv \sum_i v_i n_i \phi_i^0 \tag{146}$$

denote the employment and vacancy aggregates generated by one unit of entry. Since the actual distribution is $\boldsymbol{\phi} = m^0\boldsymbol{\phi}^0$, aggregate vacancies and search effort are

$$\tilde{v} = m^0\mathcal{V}^0, \qquad u = 1 - m^0N^0, \qquad \tilde{u} = 1 - (1-\zeta)m^0N^0. \tag{147}$$

We then recover the entry mass m^0 in order for the implied market tightness to be consistent

Algorithm 1 Calibration Algorithm

Set the target market tightness $\theta = 1$ and compute $(\lambda^U, \lambda^E, \lambda^F)$ from (7)
Fix tolerance $\varepsilon_{FG} > 0$
Guess $(F^{(0)}, G^{(0)})$ and set $\ell \leftarrow 0$
repeat
    Compute $U^{(\ell)}$ from (27)
    Given $(F^{(\ell)}, G^{(\ell)}, U^{(\ell)})$, solve the HJB-QVI by Howard iteration
    Obtain $\mathbf{S}^{(\ell)}$, policies $(v^{(\ell)}, V^{(\ell)})$, and intervention matrices $(\mathbf{D}^{I,(\ell)}, \mathbf{D}^{C,(\ell)}, \mathbf{M}^{(\ell)})$
    Given entrant distribution $\boldsymbol{\psi}$, solve the normalized KFE for $\boldsymbol{\phi}^0$ with $m^0 = 1$
    Compute $N^0 = \sum_i n_i \phi_i^0$ and $\mathcal{V}^0 = \sum_i v_i^{(\ell)} n_i \phi_i^0$
    Set $m^0 = \theta/[\mathcal{V}^0 + \theta(1-\zeta)N^0]$ and $\boldsymbol{\phi} = m^0 \boldsymbol{\phi}^0$
    Use Proposition 4 to obtain V_i^{new} from $\mathbf{S}^{(\ell)}$
    Update $G^{(\ell+1)}$ using (24)
    Update $F^{(\ell+1)}$ using (26)
    $\ell \leftarrow \ell + 1$
until $\|(F^{(\ell)}, G^{(\ell)}) - (F^{(\ell-1)}, G^{(\ell-1)})\| \leq \varepsilon_{FG}$

with our target. Imposing $\theta = \tilde{v}/\tilde{u}$ gives

$$m^0 = \frac{\theta}{\mathcal{V}^0 + \theta(1-\zeta)N^0}. \tag{148}$$

We then set $\boldsymbol{\phi} = m^0 \boldsymbol{\phi}^0$.

Updating the job-ladder distributions. Finally, given the surplus function $\mathbf{S}$ and distribution $\boldsymbol{\phi}$, we use Proposition 4 to obtain the implied V^{new} on the grid. We then update the job-ladder distributions using the discrete analogues of (24) and (26) in the main text.

Recovering the entry cost. Finally, we can recover the entry cost that is consistent with the market tightness θ. Specifically, we can recover the entry cost so that the free-entry condition (22) holds.

Approximation of the second derivative. Throughout, we follow Achdou et al. (2022) to approximate the second derivative of an arbitrary function $f(x_i, \ldots)$ on the non-uniform grid as

$$\frac{\partial^2}{\partial x^2} f(x_i, \ldots) \approx \frac{\Delta x_{-,i} f_{i+1} - (\Delta x_{+,i} + \Delta x_{-,i}) f_i + \Delta x_{+,i} f_{i-1}}{\frac{1}{2}(\Delta x_{+,i} + \Delta x_{-,i}) \Delta x_{+,i} \Delta x_{-,i}}, \tag{149}$$

where $f_i \equiv f(x_i, \ldots)$ and $\Delta x_{+,i} \equiv x_{i+1} - x_i$, $\Delta x_{-,i} \equiv x_i - x_{i-1}$.

Algorithm 2 Steady State Counterfactual Algorithm

Guess $(F^{(0)}, G^{(0)})$ and $\theta^{(0,0)}$
Fix tolerances $\varepsilon_\theta, \varepsilon_{FG} > 0$
Set outer iteration counter $\ell \leftarrow 0$
repeat
  Initialize the tightness loop with $\theta^{(\ell,0)}$ and set $k \leftarrow 0$
  repeat
    Compute $(\lambda^{U,(\ell,k)}, \lambda^{E,(\ell,k)}, \lambda^{F,(\ell,k)})$ from (7)
    Compute $U^{(\ell,k)}$ from (27)
    Given $(F^{(\ell)}, G^{(\ell)}, \theta^{(\ell,k)}, U^{(\ell,k)})$, solve the HJB-QVI by Howard iteration
    Obtain $\mathbf{S}^{(\ell,k)}$, policies $(v^{(\ell,k)}, V^{(\ell,k)})$, and $(\mathbf{D}^{I,(\ell,k)}, \mathbf{D}^{C,(\ell,k)}, \mathbf{M}^{(\ell,k)})$
    Compute the free-entry residual $\Delta_\theta^{(\ell,k)} = \sum_i S_i^{(\ell,k)} \psi_i - c^e$
    Update $\theta^{(\ell,k+1)}$, increasing it if $\Delta_\theta^{(\ell,k)} > 0$ and decreasing it if $\Delta_\theta^{(\ell,k)} < 0$
    $k \leftarrow k+1$
  until $|\Delta_\theta^{(\ell,k-1)}| \leq \varepsilon_\theta$
  Set $\theta^{(\ell)} = \theta^{(\ell,k-1)}$ and keep the converged HJB-QVI objects from iteration $(\ell, k-1)$
  Given entrant distribution $\boldsymbol{\psi}$, solve the normalized KFE for $\boldsymbol{\phi}^{0,(\ell)}$ with $m^0 = 1$
  Compute $N^{0,(\ell)} = \sum_i n_i \phi_i^{0,(\ell)}$ and $\mathcal{V}^{0,(\ell)} = \sum_i v_i^{(\ell)} n_i \phi_i^{0,(\ell)}$
  Set $m^{0,(\ell)} = \theta^{(\ell)} / [\mathcal{V}^{0,(\ell)} + \theta^{(\ell)}(1-\zeta) N^{0,(\ell)}]$ and $\boldsymbol{\phi}^{(\ell)} = m^{0,(\ell)} \boldsymbol{\phi}^{0,(\ell)}$
  Use Proposition 4 to obtain $V_i^{new,(\ell)}$ from $\mathbf{S}^{(\ell)}$
  Update $G^{(\ell+1)}$ using the discrete analogue of (24)
  Update $F^{(\ell+1)}$ using the discrete analogue of (26)
  $\ell \leftarrow \ell + 1$
until $\|(F^{(\ell)}, G^{(\ell)}) - (F^{(\ell-1)}, G^{(\ell-1)})\| \leq \varepsilon_{FG}$

B.2 Algorithm for Counterfactual

For the counterfactual, we do not fix θ; instead, θ adjusts so that the free-entry condition holds. Relative to the calibration algorithm, the only modification is to add a one-dimensional loop over θ inside the HJB-QVI block. For a given guess of θ, we solve the HJB-QVI and compute the implied value of entry. If the value of entry is higher than the entry cost, we increase θ; if it is lower than the entry cost, we decrease θ. Once (22) is satisfied up to a tolerance, we proceed to the KFE step and update the job-ladder distributions as before. We summarize the algorithm in Algorithm 2.

B.3 Steady State Age Distribution

In studying the firm's life cycle, we are interested in various moments conditional on firm age. For this purpose, we need to compute the joint distribution in (z, n, a), where a denotes firm age, which we denote as $\phi(z, n, a)$. We can obtain it by solving the following KFE:

$$0 = \mathcal{L}^\dagger(z, n, v(z, n), V(z, n), a)\phi(z, n, a) - \varkappa\phi(z, n, a) + m^0 \psi(z, n)\delta_{a=0}, \tag{150}$$

where $\delta_{a=0}$ is a Dirac mass at $a = 0$ and $\mathcal{L}^{\dagger}(z, n, a)$ is the adjoint operator of $\mathcal{L}(z, n, v, V)$ defined for an arbitrary function $f(z, n, a)$:

$$\begin{aligned} \mathcal{L}(z,n,v,V,a)f(z,n,a) &= \mu(z)\partial_z f(z,n,a) + \frac{\sigma^2(z)}{2}\partial_{zz} f(z,n,a) \\ &\quad + \left[\lambda^F G(V)v - s - \lambda^E(1-F(V))\right] n\partial_n f(z,n,a) + \partial_a f(z,n,a), \end{aligned} . \tag{151}$$

C An Alternative Model with Vacancy Posting by Entrants

In the baseline model, we assumed that entrants hire directly from the unemployed pool. In this appendix, we consider an alternative in which entrants post vacancies to hire workers, as incumbent firms do.

Specifically, we assume that the entrants are endowed with as many vacancies as needed to achieve the initial level of employment. This implies that the initial vacancy is given by

$$v^0(z_0, n_0) = \frac{n_0}{\lambda^F G(W_0)}, \tag{152}$$

where W_0 is the contract value offered by the entrants. The entrants choose the initial promised utility to maximize the entry value: $\max_{W_0 \geq 0} J(z_0, n_0, W_0) = \max_{W_0 \geq 0} S(z_0, n_0) - W_0 n_0$. This clearly implies that the entrants choose $W_0 = 0$ and hire only from unemployed workers.

This modification changes the aggregate vacancy posting to

$$\tilde{v} = \int v(z, n) n \, d\phi(z, n) + m^0 \int v^0(z_0, n_0) d\Psi(z_0, n_0). \tag{153}$$

The vacancy-weighted distribution of continuation value becomes

$$F(V) = \frac{1}{\tilde{v}} \left[\int_{V_{t+dt}(z,n) \leq V} v(z, n) n \, d\phi(z, n) + m^0 \int v^0(z_0, n_0) d\Psi(z_0, n_0) \right], \tag{154}$$

where we have imposed that the entrants post the continuation value of $V = 0$. Likewise, the vacancy-weighted ranking of the marginal surplus becomes

$$\hat{F}(S_n) \equiv \frac{1}{\tilde{v}} \left[\int_{S_n(z,n) \leq S_n} v(z, n) n \, d\phi(z, n) + m^0 \int v^0(z_0, n_0) d\Psi(z_0, n_0) \right]. \tag{155}$$

The rest of the equilibrium conditions remain unchanged.

D Scale Invariance of the Economy

The goal is to show that the economy is invariant to the scale of the technology level.

Let Λ index the economy with different technology level. Specifically, economy with Λ features the vacancy cost, $\Lambda \times \bar{c}^v$, entry cost, $\Lambda \times c^e$, the flow value of unemployment, $\Lambda \times b$, and the productivity upon entry $\Lambda^{\frac{1}{1-\alpha}} z^0$:

$$\psi(\Lambda^{\frac{1}{1-\alpha}} z^0, n^0; \Lambda) = \psi(z^0, n^0; 1). \tag{156}$$

The matching technology and preferences are invariant to Λ. The baseline economy corresponds to $\Lambda = 1$.

The following lemma is a useful intermediate result.

Lemma 1. *Assume that the production function is given by*

$$y(z, n) = z^{1-\alpha} n^{\alpha}, \tag{157}$$

and the diffusion process of z is given by the geometric Brownian motion:

$$\mu(z) = \mu z, \quad \sigma(z) = \sigma z. \tag{158}$$

Let $S(z, n; \Lambda)$ be the surplus function of the economy with technology level Λ. Let $F(W; \Lambda)$ and $G(V; \Lambda)$ be the distribution of the promised utility of the economy with technology level Λ. Suppose that

$$F(V; \Lambda) = F(V/\Lambda; 1), \quad G(V; \Lambda) = G(V/\Lambda; 1), \tag{159}$$

and all the meeting rates, $\lambda^U, \lambda^E, \lambda^F$, are invariant to Λ. Then, the value functions satisfy

$$S(\Lambda^{\frac{1}{1-\alpha}} z, n; \Lambda) = \Lambda\, S(z, n; 1), \tag{160}$$

and the policy functions satisfy

$$v(\Lambda^{\frac{1}{1-\alpha}} z, n; \Lambda) = v(z, n; 1), \quad V(\Lambda^{\frac{1}{1-\alpha}} z, n; \Lambda) = \Lambda V(z, n; 1). \tag{161}$$

Proof. We guess and verify that (160) holds. Let $\tilde{z} \equiv \Lambda^{\frac{1}{1-\alpha}} z$. By definition, $S(\tilde{z}, n; \Lambda)$ solves the following HJB equation:

$$\rho S(\tilde{z}, n; \Lambda) = \max_{v,V} y(\tilde{z}, n) - \Lambda\, c(v) n - \Lambda\, bn - \lambda^U \int \tilde{V} dF(\tilde{V}; \Lambda) n$$

$$-V\lambda^F G(V;\Lambda)vn+\lambda^E\int_V \tilde{V}dF(\tilde{V};\Lambda)n+\mathcal{L}(\tilde{z},n,v,V;\Lambda)S(\tilde{z},n;\Lambda)$$

where $\mathcal{L}$ is the generator defined through

$$\begin{aligned}\mathcal{L}(\tilde{z},n,v,V;\Lambda)S(\tilde{z},n;\Lambda)&=\mu\,\tilde{z}\,\partial_{\tilde{z}}S(\tilde{z},n;\Lambda)+\frac{(\sigma\,\tilde{z})^2}{2}\partial_{\tilde{z}\tilde{z}}S(\tilde{z},n;\Lambda)\\&\quad+\left[\lambda^F G(V;\Lambda)v-s-\lambda^E(1-F(V;\Lambda))\right]n\partial_n S(\tilde{z},n),\end{aligned}, \tag{162}$$

with the boundary conditions, $S(\tilde{z},n;\Lambda)\geq 0$ and $S_n(\tilde{z},n;\Lambda)\geq 0$.

The guess of (160) implies

$$\Lambda^{\frac{1}{1-\alpha}}\partial_{\tilde{z}}S(\tilde{z},n;\Lambda)=\Lambda\partial_z S(z,n;1) \tag{163}$$

$$(\Lambda^{\frac{1}{1-\alpha}})^2\partial_{\tilde{z}\tilde{z}}S(\tilde{z},n;\Lambda)=\Lambda\partial_{zz}S(z,n;1) \tag{164}$$

$$\partial_n S(\tilde{z},n;\Lambda)=\Lambda\partial_n S(z,n;1). \tag{165}$$

Consequently,

$$\begin{aligned}\mathcal{L}(\tilde{z},n,v,V;\Lambda)S(\tilde{z},n;\Lambda)&=\Lambda\,\mu\,z\,\partial_z S(z,n;1)+\Lambda\,\frac{(\sigma\,z)^2}{2}\partial_{zz}S(z,n;1)\\&\quad+\Lambda\left[\lambda^F G(V/\Lambda;1)v-s-\lambda^E(1-F(V/\Lambda;1))\right]n\partial_n S(z,n;1),\\&=\Lambda\,\mathcal{L}(z,n,v,V;1)S(z,n;1)\end{aligned}$$

Therefore, plugging the guess into the HJB equation, we obtain

$$\begin{aligned}\rho\Lambda S(z,n;1)=\max_{v,V}\Lambda y(z,n)-\Lambda\,c(v)n-\Lambda\,bn-\Lambda\lambda^U\int\tilde{V}dF(\tilde{V};1)n\\-\Lambda\,(V/\Lambda)\lambda^F G(V/\Lambda;1)vn+\Lambda\,\lambda^E\int_V\tilde{V}dF(\tilde{V};1)n+\Lambda\,\mathcal{L}(z,n,v,V;1)S(z,n;1)\end{aligned}$$

Cancelling the factor Λ, the above equation corresponds to the HJB equation for the economy with $\Lambda=1$, which must hold. The boundary conditions remain the same: $S(z,n;1)\geq 0$ and $S_n(z,n;1)\geq 0$. This confirms the guess of (160). Let $v(z,n;\Lambda)$ and $V(z,n;\Lambda)$ be the optimal vacancy posting rate and the optimal promised utility of the economy with technology level Λ. Then the policy functions are related through

$$v(\Lambda^{\frac{1}{1-\alpha}}z,n;\Lambda)=v(z,n;1),\quad V(\Lambda^{\frac{1}{1-\alpha}}z,n;\Lambda)=\Lambda V(z,n;1). \tag{166}$$

□

The next proposition shows that the entire economy scales with Λ:

Proposition 10. *Consider an economy with technology level Λ. Consider another economy with $\Lambda = 1$. Then, the value functions satisfy:*

$$S(\Lambda^{\frac{1}{1-\alpha}} z, n; \Lambda) = \Lambda\, S(z, n; 1), \tag{167}$$

and the policy functions satisfy:

$$v(\Lambda^{\frac{1}{1-\alpha}} z, n; \Lambda) = v(z, n; 1), \quad V(\Lambda^{\frac{1}{1-\alpha}} z, n; \Lambda) = \Lambda V(z, n; 1). \tag{168}$$

The distributions satisfy

$$g(\Lambda^{\frac{1}{1-\alpha}} z, n; \Lambda) = g(z, n; 1), \tag{169}$$

and

$$F(V; \Lambda) = F(V/\Lambda; 1), \quad G(V; \Lambda) = G(V/\Lambda; 1). \tag{170}$$

Finally, the two economies share the same entry rates, the same market tightness and the meeting rates.

Proof. Lemma 1 implies that, with (170) and the meeting rates being the same, the value functions and the policy functions satisfy (160) and (168). Therefore, it remains to show that the distributions satisfy (170) and that the meeting rates are the same. The Kolmogorov forward equation for the joint distribution of $(\tilde{z}, n)$ under the technology level Λ, where $\tilde{z} \equiv \Lambda^{\frac{1}{1-\alpha}} z$, is given by

$$0 = \mathcal{L}^{\dagger}\left(\tilde{z}, n, v(z, n), V(z, n); \Lambda\right) g(\tilde{z}, n; \Lambda) - \varkappa g(\tilde{z}, n; \Lambda) + m^0 \psi(\tilde{z}, n; \Lambda), \tag{171}$$

where

$$\mathcal{L}^{\dagger}\left(\tilde{z}, n, v(\tilde{z}, n; \Lambda), V(\tilde{z}, n; \Lambda); \Lambda\right) g(\tilde{z}, n; \Lambda) \tag{172}$$

$$= \partial_z \left[-\mu \tilde{z} g(\tilde{z}, n)\right] + \partial_{zz}\left[\frac{\sigma^2 \tilde{z}^2}{2} g(\tilde{z}, n)\right] \tag{173}$$

$$- \partial_n \left[\left\{\lambda^F G(V(\tilde{z}, n; \Lambda)) v(\tilde{z}, n; \Lambda) - s - \lambda^E (1 - F(V(\tilde{z}, n; \Lambda)))\right\} n g(\tilde{z}, n)\right]. \tag{174}$$

Plugging (156) and (168) into the above equation and imposing that the meeting rates are the same, we get

$$0 = \mathcal{L}^{\dagger}\left(\tilde{z}, n, v(z, n; 1), V(z, n; 1); 1\right) g(\tilde{z}, n; 1) - \varkappa g(\tilde{z}, n; 1) + m^0 \psi(z, n; 1), \tag{175}$$

where

$$\mathcal{L}^{\dagger}\left(\tilde{z}, n, v(z,n), V(z,n); 1\right) g(\tilde{z}, n; 1) \tag{176}$$
$$= \partial_z\left[-\mu\tilde{z}g(\tilde{z},n;1)\right] + \partial_{zz}\left[\frac{\sigma^2\tilde{z}^2}{2}g(\tilde{z},n;1)\right] \tag{177}$$
$$-\partial_n\left[\left\{\lambda^F G(V(z,n;1))v(z,n;1) - s - \lambda^E(1-F(V(z,n;1)))\right\} ng(\tilde{z},n;1)\right]. \tag{178}$$

Note that the exit and the firing thresholds for $(\tilde{z}, n)$ under Λ are the same as the ones for (z, n) under $\Lambda = 1$. Since the Kolmogorov forward equation is the same, we must have

$$g(\tilde{z}, n; \Lambda) = g(z, n; 1). \tag{179}$$

This also implies that the two economies share the same entry rates. Now we confirm that (170) holds. Under (167) and (169), Proposition 4 implies that

$$V(z, n; \Lambda) = \Lambda V(z, n; 1), \tag{180}$$

and consequently,

$$F(V; \Lambda) = F(V/\Lambda; 1), \quad G(V; \Lambda) = G(V/\Lambda; 1). \tag{181}$$

Finally, since the employment and vacancy distributions are invariant to Λ, the market tightness and the meeting rates are the same. □

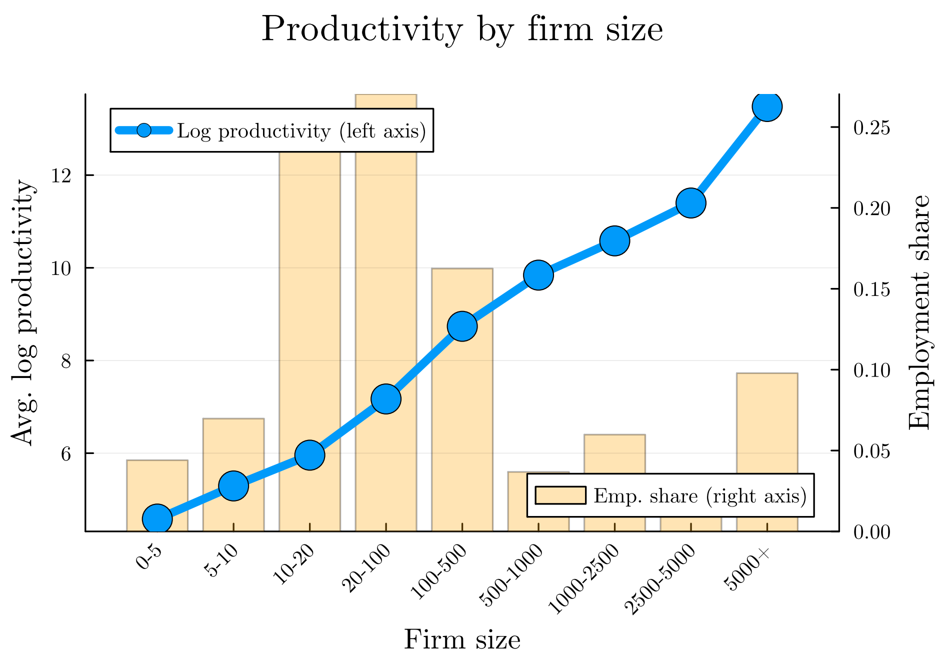

Figure E.9: Firm productivity by size

Note: The figure plots the average log productivity ($\ln z$) by firm size. The blue line plots the employment-weighted average log productivity of firms for each size bin. The bars show the employment share of firms in each size bin.

E Quantitative Appendix

E.1 Additional Figures

Figure E.9 plots the average log productivity ($\ln z$) by firm size in the steady state equilibrium. The productivity is strongly positively correlated with the firm size.

Figure E.10 plots the average log productivity ($\ln z$) over the firm life cycle. The productivity grows over the firm life cycle, on average.

Figure E.12 plots the policy function for employment growth rate, $dn(z, n)/n$, for the equilibrium without policy (red dashdot line) and with optimal policy (blue solid line). The figure plots $dn(z, n)/n$ against the log employment size ($\ln n$) for the firm with the median productivity in the stationary distribution. We see that the optimal policy speeds up the firm growth for smaller firms and slows down the firm growth for larger firms.

Figure E.13 plots the average firm size over the firm life cycle in equilibrium without policy (blue dash line) and with optimal policy (green solid line). We find that the optimal policy speeds up the firm growth for younger firms and slows down the firm growth for older firms.

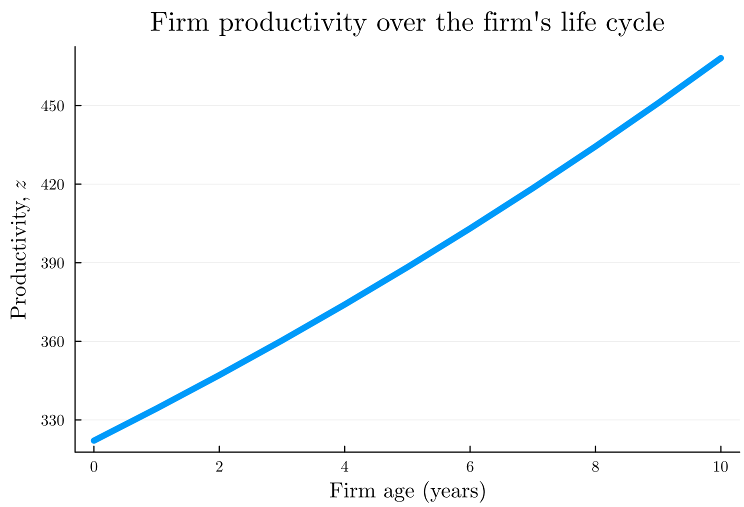

Figure E.10: Firm productivity over the firm life cycle

Note: The figure plots the average log productivity ($\ln z$) over the firm life cycle.

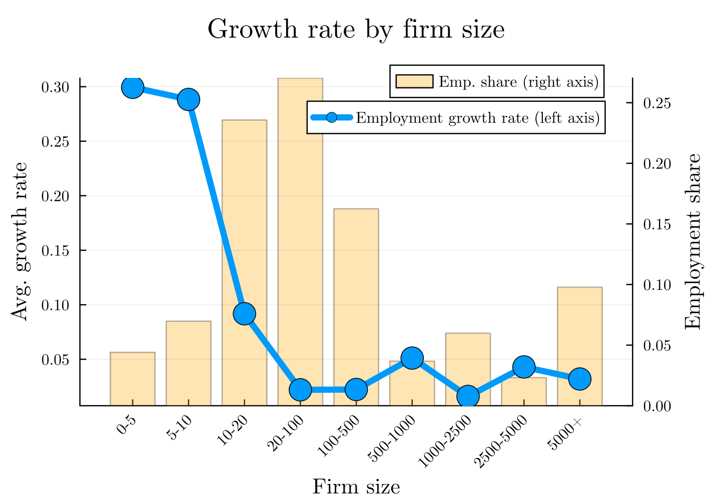

Figure E.11: Firm growth rate by size

Note: The figure plots the average growth rate (dn/n) by firm size. The blue line plots the employment-weighted average growth rate of firms for each size bin. The bars show the employment share of firms in each size bin.

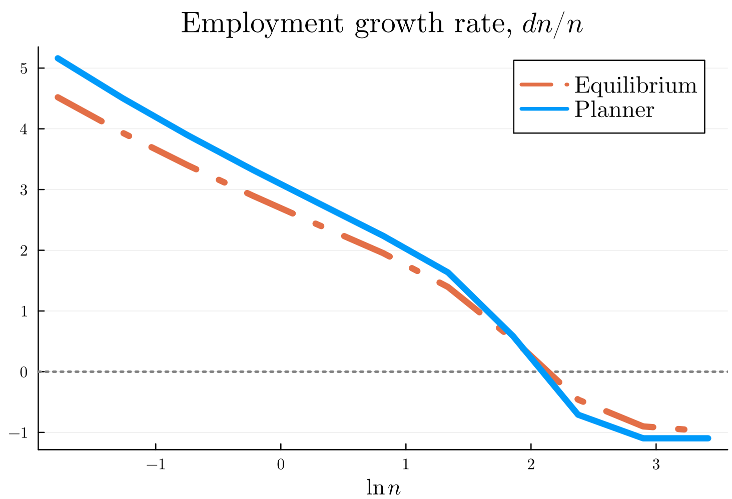

Figure E.12: Policy function for employment growth: equilibrium vs. planner

Note: The figure plots the policy function for employment growth rate, $dn(z,n)/n$, for the equilibrium (red dashdot line) and the planner (blue solid line). The figure plots $dn(z,n)/n$ against the log employment size ($\ln n$) for the firm with the median productivity in the stationary distribution.

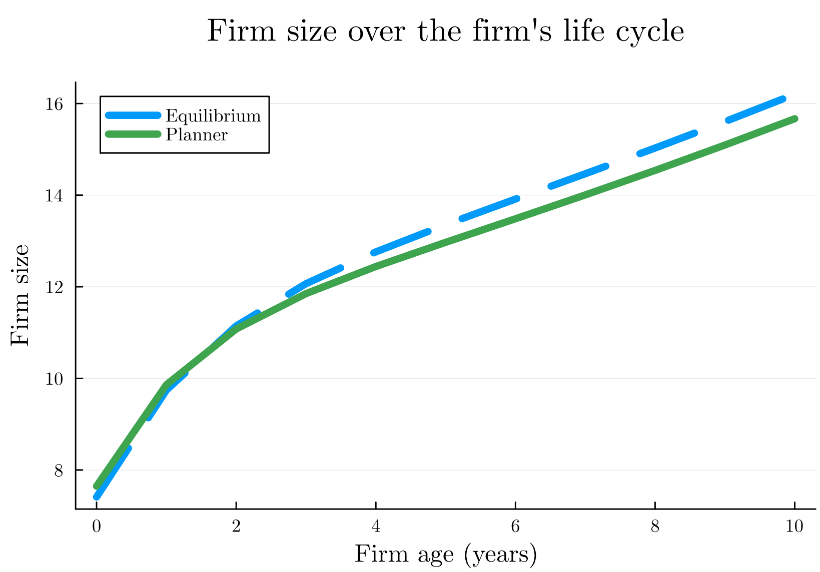

Figure E.13: Firm size over the life cycle: equilibrium vs. planner

Note: The figure plots the average firm size over the firm life cycle in equilibrium without policy (blue dash line) and with optimal policy (green solid line).

F Burdett and Mortensen (1998) Model

F.1 Decentralized Equilibrium

We study the BM model through the lens of our solution approach. To map our model to BM, we make the following assumptions:

1. The vacancy technology exhibits decreasing returns to scale, and the vacancy cost is given by $c(vn)$, instead of $c(v)n$. We treat vn as a choice variable.
2. The production function is linear, $y(z,n) = zn$.
3. There are no changes in productivity, $\mu(z) = 0$ and $\sigma(z) = 0$, and there is no entry or exit. The economy is populated by a fixed mass of firms, and we assume the productivity has a full support on $[0, \bar{z}]$.

We let $\Psi(z)$ be the cdf of the firm productivity distribution.

The HJB equation for the surplus function $S(z,n)$ is

$$\begin{aligned} rS(z,n) = \max_{V,vn} zn - c(vn) - rUn - V\lambda^F G(V)vn + \lambda^E \int_V^\infty \tilde{V} dF(\tilde{V})n \\ + \left[\lambda^F G(V)vn - sn - \lambda^E(1-F(V))n\right] \partial_n S(z,n). \end{aligned} \tag{182}$$

The first-order condition with respect to V is

$$\left[\lambda^F g(V)vn + \lambda^E f(V)n\right](\partial_n S(z,n) - V) = \lambda^F G(V)vn, \tag{183}$$

and the first-order condition with respect to vn is

$$\lambda^F G(V)(\partial_n S(z,n) - V) = c'(vn). \tag{184}$$

As in BM, we focus on the steady state, where each firm z offers a constant wage w and maintains a constant size. Consequently, the offer utility $V(z)$ is constant over time. The steady state distribution $G(V)$ satisfies the following equation:

$$0 = s(1-G(V)) - \lambda^E(1-F(V))G(V), \tag{185}$$

where the first term is the inflow and the second term is the outflow. From the above equation, $F(V)$ and $G(V)$ are related through

$$1 - F(V) = \frac{s}{\lambda^E}\frac{(1-G(V))}{G(V)}. \tag{186}$$

Differentiating both sides with respect to V, we have

$$f(V) = \frac{s}{\lambda^E} \frac{g(V)}{G(V)^2} \tag{187}$$

In the steady state, for firms offering V and posting vn vacancies, the firm size is given by

$$\begin{aligned} n &= \frac{\lambda^F G(V) vn}{s + \lambda^E (1 - F(V))} \\ &= \frac{\lambda^F G(V)^2 vn}{s}, \end{aligned} \tag{188}$$

where we have used (186) in the second line. Applying the envelope theorem to (182) and evaluating at the steady state, we have

$$\partial_n S(z, n) = \frac{z - rU + \lambda^E \int_V^\infty \tilde{V} dF(\tilde{V})}{r + s + \lambda^E (1 - F(V))} \tag{189}$$

Substituting (187), (188), and (189) into (183), we have

$$2g(V) \frac{z - rU - (r+s)V + \lambda^E \int_V^\infty (\tilde{V} - V) dF(\tilde{V})}{r + s + \lambda^E (1 - F(V))} = G(V). \tag{190}$$

Now we will rewrite (190) in terms of wages w instead of promised utility V. These two are related through

$$(r+s)V = w - rU + \lambda^E \int_V^\infty (\tilde{V} - V) dF(\tilde{V}). \tag{191}$$

Let $G^w(w)$ be the cumulative distribution function of wages. Since V and w are one-to-one, we have

$$G^w(w) = G(V(w)). \tag{192}$$

Differentiating both sides with respect to w, we have

$$\begin{aligned} g^w(w) &= g(V(w)) \frac{dV}{dw} \\ &= g(V(w)) \frac{1}{r + s + \lambda^E (1 - F(V(w)))}. \end{aligned} \tag{193}$$

Substituting (191) and (193) into (190), we have

$$2g^w(w)(z-w) = G^w(w). \tag{194}$$

As in BM, $w(z)$ is strictly increasing in z. Let $\hat{G}(z)$ be the employment-weighted cumulative distribution function of productivity. Under the guess,

$$G^w(w(z)) = \hat{G}(z), \tag{195}$$

and by differentiating both sides with respect to z, we have

$$g^w(w(z))w'(z) = \hat{g}(z). \tag{196}$$

Substituting this into (194), we have

$$2\hat{g}(z)(z-w(z)) = \hat{G}(z)w'(z). \tag{197}$$

This is an ODE for $w(z)$ with a boundary condition $w(\underline{z}) = \underline{z}$, where $\underline{z}$ is the reservation wage and is also the lowest productivity level among the active firms. Solving this ODE, we have

$$w(z) = \frac{1}{\hat{G}(z)^2}\left[\hat{G}(\underline{z})^2\underline{z} + \int_{\underline{z}}^{z} 2\hat{G}(\tilde{z})\hat{g}(\tilde{z})\tilde{z}d\tilde{z}\right]. \tag{198}$$

We can rewrite the above expression as

$$w(z) - \underline{z} = \mathbb{E}_{\hat{G}^2}\left[\tilde{z} - \underline{z} | \tilde{z} \leq z\right]. \tag{199}$$

Equation (199) corresponds to (1) once we express both w and z net of the reservation wage, $\underline{z}$.

We can recover the original expression in the BM model using (199). Let $\hat{F}(z)$ denote the vacancy-weighted cumulative distribution function of productivity. From (186), we have

$$\hat{G}(z) = \frac{s}{s + \lambda^E(1-\hat{F}(z))} \tag{200}$$

and

$$\hat{g}(z) = \frac{s\lambda^E\hat{f}(z)}{(s+\lambda^E(1-\hat{F}(z)))^2} \tag{201}$$

Substituting (200) and (201) into (198), we have

$$
\begin{aligned}
w(z) &= \frac{1}{\hat{G}(z)^2}\hat{G}(\underline{z})^2\underline{z} + \frac{1}{\hat{G}(z)^2}\int_{\underline{z}}^{z} 2\frac{s^2\lambda^E \hat{f}(\tilde{z})}{(s+\lambda^E(1-\hat{F}(\tilde{z})))^3}\tilde{z}d\tilde{z} && (202)\\
&= z - \int_{\underline{z}}^{z} \frac{\left(s+\lambda^E(1-\hat{F}(z))\right)^2}{\left(s+\lambda^E(1-\hat{F}(\tilde{z}))\right)^2} d\tilde{z}, && (203)
\end{aligned}
$$

which is exactly the same as equation (47) in BM.[21]

Rewriting (184) using (189), (191), and (200), the optimal vacancy is given by

$$
\lambda^F \hat{G}(z)\frac{z-w(z)}{r+s+\lambda^E(1-\hat{F}(z))} = c'(vn(z)). \tag{204}
$$

The definition of $\hat{F}(z)$ is

$$
\hat{F}(z) = \frac{1}{\tilde{v}}\int_0^z vn(z')d\Psi(z'), \tag{205}
$$

where $\Psi(z)$ is the cdf of the firm productivity distribution.

Finally, the reservation wage $w(\underline{z}) = \underline{z}$ is determined by the worker's indifference condition. The value of unemployment is

$$
rU = b + \lambda^U \int_0^\infty W(z')d\hat{F}(z'), \tag{206}
$$

and the value of employment at firm z in excess of unemployment value, $W(z)$, is

$$
(r+s)W(z) = w(z) - rU + \lambda^E \int_z^\infty (W(z')-W(z))d\hat{F}(z'). \tag{207}
$$

The worker's indifference condition is

$$
W(\underline{z}) = 0. \tag{208}
$$

Imposing this condition, we have

$$
\underline{z} - b = (\lambda^U - \lambda^E)\int_{\underline{z}}^\infty \frac{1-\hat{F}(z')}{r+s+\lambda^E(1-\hat{F}(z'))}dz' \tag{209}
$$

[21] To map notation to BM, k in their model is λ^E/s here, p in their model is z here, b in their model is $\underline{z}$ here, and J in their model is $\hat{G}$ here.

We summarize the steady-state equilibrium conditions in the following proposition under $r \to 0$ (the same limit used in BM).

Proposition 11 (Burdett and Mortensen (1998))**.** *Assume $r \to 0$. The steady-state equilibrium allocation consists of $\{vn(z), w(z), \hat{F}(z), \hat{G}(z), \underline{z}, \tilde{v}\}$ and the meeting rates $\{\lambda^F, \lambda^U, \lambda^E\}$ that solve the following system of equations:*

$$\lambda^F \left\{ \hat{G}(z) \frac{1}{s + \lambda^E(1 - \hat{F}(z))} (z - w(z)) \right\} = c'(vn(z)), \tag{210}$$

$$w(z) = z - \int_{\underline{z}}^{z} \frac{\left(s + \lambda^E(1 - \hat{F}(z))\right)^2}{\left(s + \lambda^E(1 - \hat{F}(\tilde{z}))\right)^2} d\tilde{z}, \tag{211}$$

$$\hat{F}(z) = \frac{1}{\tilde{v}} \int_0^z vn(z') d\Psi(z'), \tag{212}$$

$$\tilde{v} = \int vn(z) d\Psi(z), \tag{213}$$

$$\hat{G}(z) = \frac{s}{s + \lambda^E(1 - \hat{F}(z))}, \tag{214}$$

$$\underline{z} - b = (\lambda^U - \lambda^E) \left\{ \int_{\underline{z}}^{\bar{z}} \frac{1 - \hat{F}(\tilde{z})}{s + \lambda^E(1 - \hat{F}(\tilde{z}))} d\tilde{z} \right\}, \tag{215}$$

and $\lambda^F = \frac{M(\tilde{u},\tilde{v})}{\tilde{v}}$, $\lambda^U = \frac{M(\tilde{u},\tilde{v})}{\tilde{u}}$, $\lambda^E = \zeta \frac{M(\tilde{u},\tilde{v})}{\tilde{u}}$ with $\tilde{u} = \frac{s}{\lambda^U + s} + \zeta \frac{\lambda^U}{\lambda^U + s}$.

F.2 Efficient Allocation

The planning problem for BM is given by the following problem.

Problem 3. *The constrained planner solves*

$$\max_{\{\phi_t, v_t, m_t^0, dn_t, \varsigma_t^e, \varsigma_t^f, x_{t,zn,z'n'}, x_{t,u,zn}\}} \int_0^\infty e^{-rt} \left\{ \int \left(y(z,n) - c(vn_t(z,n))\right) d\phi_t + (1 - \textstyle\int n d\phi_t) b \right\} dt$$

subject to

$$\partial_t \phi_t(z,n) = -\partial_n \left[dn_t(z,n) \phi_t(z,n) \right] \tag{216}$$

$$dn_t(z,n) = \left[\lambda_t^F G^p(z,n) vn_t(z,n) - sn - \lambda_t^E (1 - F^p(z,n)) n \right] - \varsigma_t^f \tag{217}$$

where

$$\begin{aligned}
G^p(z,n) &\equiv \frac{1}{\tilde{u}}\left[x_{t,u,zn}\left(1-\int n d\phi_t\right)+\zeta\int x_{t,z'n',zn}n'd\phi_t(z',n')\right]\\
1-F^p(z,n) &\equiv \frac{1}{\tilde{v}}\int x_{t,zn,z'n'}v(z',n')n'd\phi_t(z',n')\\
\lambda^F &\equiv \frac{M(\tilde{u},\tilde{v})}{\tilde{v}}, \quad \lambda^E \equiv \zeta\frac{M(\tilde{u},\tilde{v})}{\tilde{u}},\\
\tilde{u} &\equiv 1-\int n d\phi_t+\zeta\int n d\phi_t, \quad \tilde{v}\equiv\int v(z,n)n\, d\phi_t(z,n),
\end{aligned}$$

and $x_{t,zn,z'n'} \in [0,1]$, $vn_t(z,n) \geq 0$, *and* $\varsigma_t^f \geq 0$.

The solution to this problem can be characterized in almost the same manner as in Proposition 6. Letting $S_t^p(z,n)$ denote the co-state variable (Lagrange multiplier) associated with the law of motion of $\phi_t(z,n)$, $S_t^p(z,n)$ solves the following HJB equation:

$$\begin{aligned}
rS_t^p(z,n) &= y(z,n)-c(vn_t(z,n))-bn+dn_t(z,n)S_n^p(z,n)\\
&\quad -\lambda^F\left\{G^p(z,n)\mathbb{E}_{\hat{G}}[\tilde{S}_n|\tilde{S}_n\leq S_n^p]+\eta^F\Omega_t\right\}vn_t(z,n)\\
&\quad -\lambda^U\left\{\mathbb{E}_{\hat{F}}[\tilde{S}_n]-\eta^U\Omega\right\}n+\lambda^E\left\{(1-F^p(z,n))\mathbb{E}_{\hat{F}}[\tilde{S}_n|\tilde{S}_n\geq S_n^p]-\eta^U\Omega_t\right\}n\\
&\quad +\partial_t S_t^p(z,n)
\end{aligned} \tag{218}$$

Efficient firing implies that $S_n^p(z,n) \geq 0$. The efficient vacancy creation solves

$$\lambda_t^F\left\{G^p(z,n)\left(S_n^p(z,n)-\mathbb{E}_{\hat{G}}[\tilde{S}_n|\tilde{S}_n\leq S_n^p]\right)-\eta_t^F\Omega_t\right\}=c'(vn_t(z,n)). \tag{219}$$

The average meeting gain, Ω_t, is given by (44).

Steady State. In the steady state, the planner's marginal surplus is given by

$$S_n^p(z)=\frac{z-b-\lambda^U\left\{\int_{\underline{z}}^{\infty}S_n^p(\tilde{z})d\hat{F}(\tilde{z})-\eta^U\Omega_t\right\}+\lambda^E\left\{\int_z^{\infty}S_n^p(\tilde{z})d\hat{F}(\tilde{z})-\eta^U\Omega_t\right\}}{r+s+\lambda^E(1-\hat{F}(z))}, \tag{220}$$

where we have already imposed the fact that the marginal surplus is independent of n in the steady state and rank-preserving in z. Differentiating both sides with respect to z and solving the resulting ODE with the boundary condition $S_n^p(\underline{z})=0$, we have

$$S_n^p(z)=\int_{\underline{z}}^{z}\frac{1}{r+s+\lambda^E(1-\hat{F}(\tilde{z}))}d\tilde{z}. \tag{221}$$

Using the above expression, we can rewrite

$$
\begin{aligned}
& S_n^p(z) - \mathbb{E}_{\hat{G}}[\tilde{S}_n | \tilde{S}_n \leq S_n^p(z)] && (222)\\
&= S_n^p(z) - \frac{1}{\hat{G}(z)} \int_{\underline{z}}^{z} S_n^p(\tilde{z}) \hat{g}(\tilde{z}) d\tilde{z} && (223)\\
&= S_n^p(z) - \frac{1}{\hat{G}(z)} \left[S_n^p(\tilde{z}) \hat{G}(\tilde{z}) \right]_{\underline{z}}^{z} + \frac{1}{\hat{G}(z)} \int_{\underline{z}}^{z} \frac{1}{r+s+\lambda^E(1-\hat{F}(\tilde{z}))} \hat{G}(\tilde{z}) d\tilde{z} && (224)\\
&= \int_{\underline{z}}^{z} \frac{1}{(r+s+\lambda^E(1-\hat{F}(\tilde{z})))} \frac{(s+\lambda^E(1-\hat{F}(z)))}{(s+\lambda^E(1-\hat{F}(\tilde{z})))} d\tilde{z}, && (225)
\end{aligned}
$$

where we have used integration by parts in (224), and substituted (200) in (225). Define

$$
w^p(z) \equiv z - \int_{\underline{z}}^{z} \frac{(r+s+\lambda^E(1-\hat{F}(z)))}{(r+s+\lambda^E(1-\hat{F}(\tilde{z})))} \frac{(s+\lambda^E(1-\hat{F}(z)))}{(s+\lambda^E(1-\hat{F}(\tilde{z})))} d\tilde{z}, \tag{226}
$$

so that

$$
S_n^p(z) - \mathbb{E}_{\hat{G}}[\tilde{S}_n | \tilde{S}_n \leq S_n^p(z)] = \frac{1}{r+s+\lambda^E(1-\hat{F}(z))} \left(z - w^p(z) \right). \tag{227}
$$

The reservation productivity $\underline{z}$ is determined by the boundary condition $S_n^p(\underline{z}) = 0$, which is

$$
\underline{z} - b = (\lambda^U - \lambda^E) \left\{ \int_{\underline{z}}^{\bar{z}} \frac{1-\hat{F}(\tilde{z})}{r+s+\lambda^E(1-\hat{F}(\tilde{z}))} d\tilde{z} - \eta^U \Omega \right\}, \tag{228}
$$

where we have used integration by parts to rewrite $\int_{\underline{z}}^{\bar{z}} S_n^p(\tilde{z}) d\hat{F}(\tilde{z}) = \int_{\underline{z}}^{\bar{z}} \frac{1-\hat{F}(\tilde{z})}{r+s+\lambda^E(1-\hat{F}(\tilde{z}))} d\tilde{z}$.

The average meeting gain can be simplified as

$$
\begin{aligned}
\Omega &= \int \hat{G}(z') \left(S_n^p(z') - \mathbb{E}_{\hat{G}}[S_n^p(\tilde{z}) | \tilde{z} \leq z'] \right) \frac{1}{\tilde{v}} vn(z') d\Psi(z') && (229)\\
&= \int \hat{G}(z') \frac{1}{r+s+\lambda^E(1-\hat{F}(z'))} \left(z' - w^p(z') \right) \frac{1}{\tilde{v}} vn(z') d\Psi(z'). && (230)
\end{aligned}
$$

The following proposition summarizes the efficient allocation with $r \to 0$ (the same limit used in BM).

Proposition 12. *Assume $r \to 0$. The steady-state efficient allocation consists of $\{vn(z), w^p(z), \Omega,$ $\hat{F}(z)$, $\hat{G}(z)$, $\underline{z}$, $\tilde{v}\}$ and the meeting rates $\{\lambda^F, \lambda^U, \lambda^E\}$ that solve the following system of equations:*

$$
\lambda^F \left\{ \hat{G}(z) \frac{1}{s+\lambda^E(1-\hat{F}(z))} \left(z - w^p(z) \right) - \eta^F \Omega \right\} = c'(vn(z)), \tag{231}
$$

$$w^p(z) = z - \int_{\underline{z}}^{z} \frac{\left(s + \lambda^E(1-\hat{F}(z))\right)^2}{\left(s + \lambda^E(1-\hat{F}(\tilde{z}))\right)^2} d\tilde{z}, \tag{232}$$

$$\hat{F}(z) = \frac{1}{\tilde{v}} \int_0^z vn(z')d\Psi(z'), \tag{233}$$

$$\tilde{v} = \int vn(z)d\Psi(z), \tag{234}$$

$$\hat{G}(z) = \frac{s}{s + \lambda^E(1-\hat{F}(z))}, \tag{235}$$

$$\underline{z} - b = (\lambda^U - \lambda^E) \left\{ \int_{\underline{z}}^{\bar{z}} \frac{1-\hat{F}(\tilde{z})}{s + \lambda^E(1-\hat{F}(\tilde{z}))} d\tilde{z} - \eta^U \Omega \right\}, \tag{236}$$

$$\Omega = \int \hat{G}(z') \frac{1}{s + \lambda^E(1-\hat{F}(z'))} \left(z' - w^p(z')\right) \frac{1}{\tilde{v}} vn(z')d\Psi(z'). \tag{237}$$

and $\lambda^F = \frac{M(\tilde{u},\tilde{v})}{\tilde{v}}$, $\lambda^U = \frac{M(\tilde{u},\tilde{v})}{\tilde{u}}$, $\lambda^E = \zeta \frac{M(\tilde{u},\tilde{v})}{\tilde{u}}$ *with* $\tilde{u} = \frac{s}{\lambda^U + s} + \zeta \frac{\lambda^U}{\lambda^U + s}$.

F.3 Decentralized Equilibrium with Taxes

We now consider how the efficient allocation can be decentralized with taxes. As in the baseline model, we consider two taxes on firms: a vacancy tax $\tau^v(z)$ and a labor tax $\tau^n(z)$. The HJB equation for the surplus function $S(z,n)$ becomes

$$\begin{aligned} rS(z,n) = \max_{V,vn} zn - c(vn) - rUn - V\lambda^F G(V)vn + \lambda^E \int_V \tilde{V} dF(\tilde{V})n \\ + \left[\lambda^F G(V)vn - sn - \lambda^E(1-F(V))n\right] \partial_n S(z,n) - \tau^v(z)vn - \tau^n(z)n. \end{aligned} \tag{238}$$

The first-order condition with respect to vn gives

$$\lambda^F G(V) \left(\partial_n S(z,n) - V\right) - \tau^v(z) = c'(vn). \tag{239}$$

The first-order condition with respect to V is unchanged and given by (183). The marginal surplus is modified as

$$\partial_n S(z,n) = \frac{z - \tau^n(z) - rU + \lambda^E \int_V^\infty \tilde{V} dF(\tilde{V})}{r + s + \lambda^E(1-F(V))} \tag{240}$$

All the subsequent analysis remains the same as Appendix F.1 by re-labeling z as $z^\tau \equiv z - \tau^n(z)$. Comparing the equilibrium conditions with the conditions in Proposition 12, we have the following taxes that implement the efficient allocation:

Proposition 13. *Assume $r \to 0$. Suppose that the taxes are given by*

$$\tau^v(z) = \lambda^F \eta^F \Omega \tag{241}$$

$$\tau^n(z) = -(\lambda^U - \lambda^E)\eta^U \Omega, \tag{242}$$

where η^F, η^U and Ω are evaluated at the efficient allocation. Then the decentralized equilibrium with taxes implements the efficient allocation characterized in Proposition 12.

The proof is given in Appendix F.4. The proposition shows that both the vacancy tax and the employment tax are constant across all firms. This result stands in stark contrast to Proposition 7 in the baseline model, in which the taxes must systematically vary with respect to the firm's marginal surplus. This comes from the fact that in BM, the only source of inefficiencies is the congestion in the matching process, which is captured by the average meeting gain Ω.

F.4 Proof of Proposition 13

Since $\tau^n(z)$ does not vary with z, the equilibrium is rank-preserving in z. Then, the wage in the decentralized equilibrium is given by

$$w^\tau(z) = z - \tau^n(z) - \int_{\underline{z}}^{z} \frac{\left(s + \lambda^E(1 - \hat{F}(z))\right)^2}{\left(s + \lambda^E(1 - \hat{F}(\tilde{z}))\right)^2} d\tilde{z}. \tag{243}$$

The optimality condition for the vacancy creation is given by

$$\lambda^F \left\{ \hat{G}(z) \frac{1}{s + \lambda^E(1 - \hat{F}(z))} \left(z - \tau^n(z) - w^\tau(z)\right) \right\} - \tau^v(z) = c'(vn(z)). \tag{244}$$

Combining (243) and (244), we can write the optimality condition for the vacancy creation as

$$\lambda^F \left\{ \hat{G}(z) \frac{1}{s + \lambda^E(1 - \hat{F}(z))} \left(z - w^p(z)\right) \right\} - \tau^v(z) = c'(vn(z)), \tag{245}$$

where $w^p(z) = z - \int_{\underline{z}}^{z} \frac{\left(s+\lambda^E(1-\hat{F}(z))\right)^2}{\left(s+\lambda^E(1-\hat{F}(\tilde{z}))\right)^2} d\tilde{z}$, which is the same as the one defined in Proposition 12.

The reservation match quality is determined by

$$\underline{z} - b = (\lambda^U - \lambda^E) \int_{\underline{z}}^{\bar{z}} \frac{1 - \hat{F}(\tilde{z})}{s + \lambda^E(1 - \hat{F}(\tilde{z}))} d\tilde{z} + \tau^n(z). \tag{246}$$

Comparing (245) and (246) with (231) and (236), we see that when

$$\tau^v(z) = \lambda^F \eta^F \Omega, \quad \tau^n(z) = -(\lambda^U - \lambda^E)\eta^U \Omega, \tag{247}$$

the conditions coincide with each other. The rest of the conditions trivially coincide.

G Model with Hiring Costs and Costless Vacancy

In the baseline economy, firms pay a cost to post vacancies but hire workers for free. Some existing work in the literature reverses these assumptions: vacancy posting is free, but hiring is costly (e.g., Coles and Mortensen 2016, Elsby and Gottfries 2022). In this appendix, we study a version of our model with this alternative assumption.

Suppose that for firm (z, n) to hire hn workers per unit of time, it must pay cost $c^h(h)n$ per unit of time, where h denotes the hiring rate. We assume

$$c^h(h) \equiv \bar{c}^h \frac{h^{1+\gamma}}{1+\gamma}. \tag{248}$$

Vacancy posting is free, and therefore firms post as many vacancies as needed to hire the desired number of workers. The rest of the environment remains unchanged. Throughout, we will assume that $\gamma > 1$.

G.1 Equilibrium

Even in this environment, we can follow the same steps to derive the following HJB equation for the surplus function $S(z, n)$:

$$\begin{aligned}\rho S(z,n) = \max_{h,V} y(z,n) - c^h(h)n - rUn - Vhn + \lambda^E \int_V \tilde{V} dF(\tilde{V})n \\ + \mathcal{L}(z,n,h,V)S(z,n),\end{aligned} \tag{249}$$

where the generator is given by

$$\mathcal{L}(z,n,h,V)S(z,n) = \mu(z)\partial_z S(z,n) + \frac{\sigma^2(z)}{2}\partial_{zz}S(z,n) \tag{250}$$

$$+ \left[h - s - \lambda^E(1 - F(V))\right] n\partial_n S(z,n). \tag{251}$$

The boundary conditions are

$$S(z,n) \geq 0, \quad \partial_n S(z,n) \geq 0. \tag{252}$$

Let $h(z, n)$ be the policy function for the hiring rate. The vacancy rate needed to achieve the hiring rate is given by

$$v(z,n) = \frac{h(z,n)}{\lambda^F G(V(z,n))}. \tag{253}$$

Equations (249)-(253) replace Problem 14 in the baseline model. The rest of the equilibrium definition remains unchanged.

Note that, in this environment, there always exists an equilibrium with a perfectly competitive labor market, in which all the workers are paid the same wage. To see this, suppose that upon receiving the outside offers and whenever indifferent, workers always choose to stay at the current firm. In this case, the equilibrium will be such that all firms offer $V = 0$. Since the only motive to raise V for the firm is to raise the retention rate, no firm has an incentive to raise V above zero because firms are retaining all the workers to begin with. In this equilibrium, there is no on-the-job search, all the workers are paid the same wage w, and wage w adjusts to ensure there is no unemployment. The equilibrium is isomorphic to that in Hopenhayn (1992) and Hopenhayn and Rogerson (1993).

To make the analysis interesting, in what follows, we focus on an equilibrium in which workers accept the poaching firm's offer when indifferent. Under this worker strategy, all firms offering $V = 0$ is no longer an equilibrium because firms have an incentive to raise V to raise the retention rate. In fact, by the same argument as Proposition 2, the equilibrium offer distribution $F(V)$ cannot have a mass point at any $V \geq 0$. This allows us to take the first-order condition with respect to V, which is given by

$$\lambda^E f(V)(\partial_n S(z,n) - V) = h. \tag{254}$$

The first-order condition with respect to h is

$$\partial_n S(z,n) - V = c^{h\prime}(h) \tag{255}$$

These optimality conditions imply the following, which is analogous to Proposition 3.

Proposition 14. *Assume $\gamma > 1$. Policy functions for promised utility and hiring rate are functions only of the marginal surplus, $V(z,n) = \hat{V}(S_n(z,n))$ and $h(z,n) = \hat{h}(S_n(z,n))$. Furthermore, $\hat{V}(S_n)$ is strictly increasing in S_n.*

All the proofs in this section are collected in Appendix G.3. Proposition 14 implies that the equilibrium job ladder is ranked according to the marginal surplus, as in the baseline model. Therefore we can write

$$\hat{G}(S_n) = G(\hat{V}(S_n)), \quad \hat{F}(S_n) = F(\hat{V}(S_n)). \tag{256}$$

Using this rank-preserving property, we obtain a tight analytical characterization of the equilibrium contracts, which is analogous to Proposition 4.

Proposition 15. *Assume $\gamma > 1$. The promised utility each firm with marginal surplus S_n offers is given by*

$$\hat{V}(S_n) = \mathbb{E}_{\hat{G}}\left[s_n | s_n \leq S_n\right], \tag{257}$$

where $\mathbb{E}_{\hat{G}}\left[s_n | s_n \leq S_n\right]$ is the weighted average of the marginal surplus conditional on being below S_n.

There are both similarities and differences between the equilibrium contracts in the baseline model and this model. The similarity is that the equilibrium contracts are still ranked according to the marginal surplus, and firms offer the weighted average of the marginal surplus among firms below them on the job ladder. The difference is that the weight is no longer quadratic in the search intensity-weighted employment distribution, $\hat{G}^2$. Instead, it is the search intensity-weighted employment distribution, $\hat{G}$. This difference is intuitive because in this model, the only reason for firms to offer higher wages is to raise retention, while in the baseline model, firms have dual motives to raise wages: to raise retention and to hire more workers. This dual motive is reflected in quadratic weighting, while the single motive results in non-quadratic weighting.

We can summarize the equilibrium conditions in a form that will be useful for comparison with the planner's problem below. Let S_n denote the marginal surplus and write $V(z, n) = \hat{V}(\partial_n S(z, n))$. The equilibrium HJB equation can be written as

$$\begin{aligned} \rho S(z,n) = y(z,n) &- c^h(h(z,n))n - rUn - V(z,n)h(z,n)n \\ &+ \lambda^E \int_{\partial_n S(z,n)}^{\infty} \hat{V}(s_n) d\hat{F}(s_n) n \\ &+ dn(z,n)\partial_n S(z,n) + \mu(z)\partial_z S(z,n) + \frac{1}{2}\sigma(z)^2 \partial_{zz} S(z,n), \end{aligned} \tag{258}$$

where $\rho \equiv r + \varkappa$ and

$$dn(z,n) = \Big[h(z,n) - s - \lambda^E(1 - \hat{F}(\partial_n S(z,n)))\Big] n. \tag{259}$$

The steady-state firm density $\phi(z, n)$ satisfies the KFE

$$0 = -\partial_n \left[dn(z,n)\phi(z,n)\right] - \partial_z \left[\mu(z)\phi(z,n)\right] + \frac{1}{2}\partial_{zz}\left[\sigma(z)^2\phi(z,n)\right] - \varkappa\phi(z,n) + m^0\psi(z,n). \tag{260}$$

The hiring and promised-utility conditions are

$$\partial_n S(z,n) - V(z,n) = c^{h\prime}(h(z,n)), \tag{261}$$

$$V(z,n) = \hat{V}(\partial_n S(z,n)) = \mathbb{E}_{\hat{G}}\left[s_n \mid s_n \leq \partial_n S(z,n)\right]. \tag{262}$$

The entry, firing, and exit conditions are

$$\int S(z,n) d\Psi(z,n) = c^e, \qquad \partial_n S(z,n) \geq 0, \quad S(z,n) \geq 0. \tag{263}$$

G.2 Efficiency

We now study the efficiency of the equilibrium. As mentioned earlier, in this model, if the planner could ban on-the-job search, the planner achieves the competitive labor market outcome as in Hopenhayn (1992) and Hopenhayn and Rogerson (1993). This is clearly the first-best outcome. This is why Coles and Mortensen (2016) and Elsby and Gottfries (2022) find that on-the-job search is a source of distortion in an economy with hiring costs and costless vacancy.

At the same time, banning on-the-job search is arguably hard to implement in practice. For example, upon employed workers receiving the outside offer, they could temporarily quit the current job and take the outside offer, which would technically avoid being recorded as on-the-job search and caught by the government.

For this reason, we focus on the set of allocations with on-the-job search, where the planner can choose the ranking of the job ladder. The planner assigns the ranking of the job ladder across firms (z,n), where workers have to follow the ranking to accept the job offer from the firm below them on the job ladder. The planner cannot control other dimensions of the worker mobility decisions, such as always staying at the current firm. Note that the equilibrium allocation has a ranking of the job ladder based on the marginal surplus, so the equilibrium is a subset of the implementable allocations by the planner. One can also interpret our planning problem as a Ramsey problem where the planner has access to a limited set of policy instruments. For example, if the planner only has access to employment and vacancy taxes, the planner is unable to ban on-the-job search.

We define the constrained planner's problem as follows.

Problem 4. *The constrained planner solves*

$$\max_{\{\phi_t, h_t, m_t^0, dn_t, \varsigma_t, x_{t,zn,z'n'}\}} \int_0^\infty e^{-rt} \left\{ \int \left(y(z,n) - c^h(h_t(z,n))n\right) d\phi_t + (1 - \textstyle\int n d\phi_t) b - m_t^0 c^e \right\} dt$$

subject to

$$\begin{aligned}\partial_t \phi_t(n,z) = &-\partial_n \left[dn_t(n,z)\phi_t(n,z)\right] - \partial_z[\mu(z)\phi_t(n,z)] + \frac{1}{2}\partial_{zz}[\sigma(z)^2\phi_t(n,z)] \\ &- \varsigma_t^e(z,n)\phi_t(n,z) + m_t^0 \psi(z,n)\end{aligned}$$

$$dn_t(z,n) = \left(h(z,n) - s - \zeta \int x_{t,zn,z'n'} \frac{h(z',n')n'}{1 - \int \tilde{n} d\phi_t(\tilde{z},\tilde{n}) + \zeta \int x_{t,\tilde{z}\tilde{n},z'n'} \tilde{n} d\phi_t(\tilde{z},\tilde{n})} d\phi_t(z',n')\right) n$$
$$- \varsigma_t^f(z,n)$$
$$1 = x_{t,zn,z'n'} + x_{t,z'n',zn} \tag{264}$$

and $x_{t,zn,z'n'} \in [0,1]$ is the probability that a worker employed at firm (z,n) moves to a firm (z',n') in case these two firms meet, $h_t(z,n) \geq 0$ is the hiring rate of firm (z,n), $\varsigma_t^e(z,n) \geq 0$ is the endogenous exit rate of firm (z,n), and $\varsigma_t^f(z,n) \geq 0$ is the firing rate of worker employed at firm (z,n).

We say the economy is constrained efficient if the allocation solves the constrained planning problem. The planner can choose firm hiring rates $h(z,n)$, job-to-job transition decision probabilities $x_{t,zn,z'n'}$, the entry rate of new firms m_t^0, the firing rate of workers ς_t^f, and the endogenous exit rate of firms ς_t^e. The last constraint, (264), says that the planner cannot ban on-the-job search, and all what the planner can do is to choose the ranking of the job ladder across firms (z,n).

Proposition 16. *Assume $\gamma > 1$. The equilibrium with hiring costs and costless vacancy is constrained efficient.*

Somewhat surprisingly, the equilibrium is constrained efficient even through the equilibrium features unemployment and wage dispersion. Unlike the baseline model, there is no search friction here. One possible source of the inefficiency is business stealing externalities. When firms hire workers, they do not internalize the loss in surplus of the firms being poached. The business stealing effect is captured by the average loss in the marginal surplus among firms ranked below them on the job ladder. As we have seen, this is exactly what the firms offer in the form of promised utility in equilibrium (Proposition 15). Therefore, the private firms fully internalize the business stealing effect, and the equilibrium is constrained efficient.

G.3 Proofs

G.3.1 Proof of Proposition 14

Let

$$\tilde{h}(S_n, V) \equiv (c^{h\prime})^{-1}(S_n - V) \tag{265}$$

be the optimal hiring rate given the marginal surplus S_n and the promised utility V. Using this, we write (254) as

$$(S_n - V)\lambda^E f(V) = \tilde{h}(S_n, V), \tag{266}$$

which implicitly determines the optimal promised utility V as a function of the marginal surplus S_n, which we write as $V = \hat{V}(S_n)$. Consequently, the optimal hiring rate is also a function only of the marginal surplus, $h = \hat{h}(S_n) \equiv \tilde{h}(S_n, \hat{V}(S_n))$. The associated second-order condition is given by

$$(S_n - \hat{V}(S_n))\lambda^E f'(\hat{V}(S_n)) - \lambda^E f(\hat{V}(S_n)) - \frac{\partial \tilde{h}(S_n, \hat{V}(S_n))}{\partial V} < 0. \tag{267}$$

We will verify this condition holds later. Totally differentiating (266) with respect to S_n and V, we have

$$\begin{aligned} &\left[\lambda^E f(\hat{V}(S_n)) - \frac{\partial \tilde{h}(S_n, \hat{V}(S_n))}{\partial S_n}\right] dS_n \\ &\quad + \left[-\lambda^E f(\hat{V}(S_n)) + (S_n - \hat{V}(S_n))\lambda^E f'(\hat{V}(S_n)) - \frac{\partial \tilde{h}(S_n, \hat{V}(S_n))}{\partial V}\right] dV = 0, \end{aligned} \tag{268}$$

which results in

$$\hat{V}'(S_n) = -\frac{\lambda^E f(\hat{V}(S_n)) - \frac{\tilde{h}(S_n,\hat{V}(S_n))}{S_n - \hat{V}(S_n)} \frac{\partial \ln \tilde{h}(S_n,\hat{V}(S_n))}{\partial \ln(S_n - \hat{V}(S_n))}}{\left[-\lambda^E f(\hat{V}(S_n)) + (S_n - \hat{V}(S_n))\lambda^E f'(\hat{V}(S_n)) - \frac{\partial \tilde{h}(S_n,\hat{V}(S_n))}{\partial V}\right]} \tag{269}$$

$$= -\lambda^E f(\hat{V}(S_n)) \frac{1 - \frac{1}{\gamma}}{\left[-\lambda^E f(\hat{V}(S_n)) + (S_n - \hat{V}(S_n))\lambda^E f'(\hat{V}(S_n)) - \frac{\partial \tilde{h}(S_n,\hat{V}(S_n))}{\partial V}\right]} \tag{270}$$

$$> 0, \tag{271}$$

under our assumption that $\gamma > 1$.

G.3.2 Proof of Proposition 15

Since $\hat{V}'(S_n) > 0$, the job ladder is inversely ranked by the employment-to-productivity ratio, $\hat{n}$. Using (253), we can rewrite the vacancy-weighted distribution of marginal surplus as

$$\hat{F}(S_n) \equiv \frac{1}{\tilde{v}} \left[\int_{S_n(z,n) \le S_n} v(z,n) n d\phi(z,n)\right] \tag{272}$$

$$= \frac{1}{\tilde{v}} \left[\int_{S_n(z,n) \le S_n} \frac{\hat{h}(S_n(z,n))}{\lambda^F \hat{G}(S_n(z,n))} n d\phi(z,n)\right] \tag{273}$$

$$= \frac{1}{\tilde{v}} \left[\int_0^{S_n} \frac{\hat{h}(s_n)}{\lambda^F \hat{G}(s_n)} \frac{\tilde{u}}{\zeta} \hat{g}(s_n) ds_n\right] \tag{274}$$

Differentiating both sides with respect to S_n, we have

$$\hat{F}'(S_n) = \frac{1}{\tilde{v}} \frac{\hat{h}(S_n)}{\lambda^F \hat{G}(S_n)} \frac{\tilde{u}}{\zeta} \hat{g}(S_n). \tag{275}$$

By the rank-preserving property,

$$\hat{F}(S_n) = F(\hat{V}(S_n)). \tag{276}$$

Differentiating both sides by S_n,

$$\hat{F}'(S_n) = f(\hat{V}(S_n))\hat{V}'(S_n), \tag{277}$$

which can be rewritten as

$$\frac{1}{\tilde{v}} \frac{\hat{h}(S_n)}{\lambda^F \hat{G}(S_n)} \frac{\tilde{u}}{\zeta} \hat{g}(S_n) = f(\hat{V}(S_n))\hat{V}'(S_n). \tag{278}$$

Plugging (278) into (266), we have

$$(S_n - \hat{V}(S_n))\lambda^E \frac{1}{\tilde{v}} \frac{\hat{h}(S_n)}{\lambda^F \hat{G}(S_n)} \frac{\tilde{u}}{\zeta} \hat{g}(S_n) = \hat{h}(S_n)\hat{V}'(S_n). \tag{279}$$

Note that

$$\lambda^E = \zeta \lambda^U = \zeta \frac{\tilde{v}}{\tilde{u}} \lambda^F. \tag{280}$$

Using this, we can rewrite (279) as

$$(S_n - \hat{V}(S_n)) \frac{\hat{g}(S_n)}{\hat{G}(S_n)} = \hat{V}'(S_n). \tag{281}$$

Solving the above ODE with the boundary condition $\hat{V}(0) = 0$,

$$\hat{V}(S_n) = \frac{1}{\hat{G}(S_n)} \int_0^{S_n} \hat{g}(s_n) s_n ds_n \tag{282}$$

$$= \mathbb{E}_{\hat{G}} \left[s_n | s_n \leq S_n \right], \tag{283}$$

as desired.

Now we verify that the second-order condition, (267), is satisfied. Further differentiating (277),

$$\hat{F}''(S_n) = f'(\hat{V}(S_n))\hat{V}'(S_n)^2 + f(\hat{V}(S_n))\hat{V}''(S_n). \tag{284}$$

Recall that (279) can be written as

$$(S_n - \hat{V}(S_n))\lambda^E \hat{F}'(S_n) = \hat{h}(S_n)\hat{V}'(S_n). \tag{285}$$

Taking the derivative with respect to S_n,

$$(1 - \hat{V}'(S_n))\lambda^E \hat{F}'(S_n) + (S_n - \hat{V}(S_n))\lambda^E \hat{F}''(S_n) = \hat{h}'(S_n)\hat{V}'(S_n) + \hat{h}(S_n)\hat{V}''(S_n). \tag{286}$$

Eliminating $\hat{V}''(S_n)$ from (284) and (286), and using the first-order condition

$$(S_n - \hat{V}(S_n))\lambda^E f(\hat{V}(S_n)) = \hat{h}(S_n),$$

we have

$$\begin{aligned} 0 = {} & f'(\hat{V}(S_n))\hat{V}'(S_n)^2 \\ & + f(\hat{V}(S_n))\frac{(1 - \hat{V}'(S_n))\lambda^E \hat{F}'(S_n) - \hat{h}'(S_n)\hat{V}'(S_n)}{\hat{h}(S_n)}. \end{aligned}$$

Equivalently,

$$\hat{h}(S_n)\frac{f'(\hat{V}(S_n))}{f(\hat{V}(S_n))} = -\frac{(1 - \hat{V}'(S_n))\lambda^E \hat{F}'(S_n) - \hat{h}'(S_n)\hat{V}'(S_n)}{\hat{V}'(S_n)^2}. \tag{287}$$

Using the first-order condition, the second-order condition in (267) is equivalent to

$$\hat{h}(S_n)\frac{f'(\hat{V}(S_n))}{f(\hat{V}(S_n))} - \lambda^E f(\hat{V}(S_n)) - \frac{\partial \tilde{h}(S_n, \hat{V}(S_n))}{\partial V} < 0. \tag{288}$$

Plugging (287) into the above expression,

$$\begin{aligned} & -\frac{(1 - \hat{V}'(S_n))\lambda^E \hat{F}'(S_n) - \hat{h}'(S_n)\hat{V}'(S_n)}{\hat{V}'(S_n)^2} - \lambda^E f(\hat{V}(S_n)) - \frac{\partial \tilde{h}(S_n, \hat{V}(S_n))}{\partial V} \\ = {} & -\frac{\lambda^E \hat{F}'(S_n) - \hat{h}'(S_n)\hat{V}'(S_n)}{\hat{V}'(S_n)^2} - \frac{\partial \tilde{h}(S_n, \hat{V}(S_n))}{\partial V} \\ = {} & -\frac{\lambda^E f(\hat{V}(S_n)) - \frac{\partial \tilde{h}(S_n, \hat{V}(S_n))}{\partial S_n}}{\hat{V}'(S_n)} \end{aligned}$$

$$= -\frac{\lambda^E f(\hat{V}(S_n))}{\hat{V}'(S_n)}\left(1-\frac{1}{\gamma}\right) < 0, \tag{289}$$

where the second line uses $\hat{F}'(S_n) = f(\hat{V}(S_n))\hat{V}'(S_n)$, the third line uses $\hat{h}'(S_n) = \frac{\partial \tilde{h}}{\partial S_n} + \frac{\partial \tilde{h}}{\partial V}\hat{V}'(S_n)$, and the final line uses $\frac{\partial \tilde{h}(S_n,\hat{V}(S_n))}{\partial S_n} = \lambda^E f(\hat{V}(S_n))/\gamma$. The inequality follows from $\hat{V}'(S_n) > 0$ and $\gamma > 1$.

G.3.3 Proof of Proposition 16

For notational convenience, let ϕ_t denote the firm measure over (z, n), and let $S_t^p(z, n)$ denote the costate variable associated with ϕ_t. The current-value Hamiltonian is

$$\mathcal{H} = \int \big(y(z,n) - c^h(h_t(z,n))n\big)\, d\phi_t + (1 - \textstyle\int n\, d\phi_t)b - m_t^0 c^e \tag{290}$$

$$+ \iint S_t^p(z,n)\Big[-\partial_n\left[dn_t(z,n)\phi_t(z,n)\right] - \partial_z[\mu z\phi_t(z,n)] + \frac{1}{2}\partial_{zz}[(\sigma z)^2\phi_t(z,n)] \tag{291}$$

$$- \varkappa\phi_t(z,n) + m_t^0\psi(z,n)\Big]\, dzdn \tag{292}$$

$$= \int \big(y(z,n) - c^h(h_t(z,n))n\big)\, d\phi_t + (1 - \textstyle\int n\, d\phi_t)b - m_t^0 c^e \tag{293}$$

$$+ \iint \Big[\partial_n S_t^p(z,n) dn_t(z,n)\phi_t(z,n) + \partial_z S_t^p(z,n)\mu z\phi_t(z,n) + \frac{1}{2}\partial_{zz}S_t^p(z,n)(\sigma z)^2\phi_t(z,n) \tag{294}$$

$$- \varkappa S_t^p(z,n)\phi_t(z,n) + m_t^0 S_t^p(z,n)\psi(z,n)\Big]\, dzdn, \tag{295}$$

where the second equality uses integration by parts and

$$\begin{aligned} dn_t(z,n) \equiv\ & \Big(h_t(z,n) - s \\ & - \zeta \int x_{t,zn,z'n'} \frac{h_t(z',n')n'}{1 - \int \tilde{n}\, d\phi_t(\tilde{z},\tilde{n}) + \zeta\int x_{t,\tilde{z}\tilde{n},z'n'}\tilde{n}\, d\phi_t(\tilde{z},\tilde{n})} d\phi_t(z',n')\Big)\, n \\ & - \varsigma_t^f(z,n). \end{aligned} \tag{296}$$

Define

$$\Delta_t(z,n) \equiv 1 - \int \tilde{n}\, d\phi_t(\tilde{z},\tilde{n}) + \zeta \int x_{t,\tilde{z}\tilde{n},zn}\tilde{n}\, d\phi_t(\tilde{z},\tilde{n}),$$

$$V_t^p(z,n) \equiv \int \frac{\zeta x_{t,z'n',zn}n'}{\Delta_t(z,n)} S_{t,n}^p(z',n')d\phi_t(z',n').$$

Then the first-order condition with respect to $h_t(z, n)$ is

$$S^p_{t,n}(z, n) - V^p_t(z, n) = c^{h\prime}(h_t(z, n)). \tag{297}$$

The first-order condition with respect to the firing rate $\varsigma^f_t(z, n)$ is

$$S^p_{t,n}(z, n) \geq 0, \tag{298}$$

which identifies the bottom rung of the planner's job ladder.

Next fix t and suppress the time subscript. After substituting the adding-up constraint (264), the derivative of the Hamiltonian with respect to $x_{zn,z'n'}$ is

$$\begin{aligned}\frac{\partial \mathcal{H}}{\partial x_{zn,z'n'}} = \zeta\phi(z, n)\phi(z', n')n'n\Bigg[&\frac{h(z, n)}{\Delta(z, n)}\left(S^p_n(z', n') - V^p(z, n)\right)\\ &-\frac{h(z', n')}{\Delta(z', n')}\left(S^p_n(z, n) - V^p(z', n')\right)\Bigg].\end{aligned}$$

We will show that

$$x_{t,zn,z'n'} = 1 \quad \text{iff} \quad S^p_n(z', n') > S^p_n(z, n). \tag{299}$$

We proceed by induction on the planner's job ladder. Consider first any (z, n) such that $S^p_n(z, n) = 0$. The firing condition and (297) then imply $h(z, n) = V^p(z, n) = 0$. Hence, for any (z', n') with $S^p_n(z', n') > 0$,

$$\frac{\partial \mathcal{H}}{\partial x_{zn,z'n'}} = \zeta\phi(z, n)\phi(z', n')n'n\frac{h(z', n')V^p(z', n')}{\Delta(z', n')} \geq 0.$$

Equality can hold only if $V^p(z', n') = 0$, which means that $x_{\breve{z}\breve{n},z'n'} = 0$ for almost all $(\breve{z}, \breve{n})$ and therefore that (z', n') is also at the bottom of the job ladder. By the pairwise feasibility constraint $x_{zn,z'n'} + x_{z'n',zn} = 1$, the set of such states can have at most measure zero. Therefore $\partial\mathcal{H}/\partial x_{zn,z'n'} > 0$ for almost all (z', n') with $S^p_n(z', n') > 0$, which implies $x_{t,zn,z'n'} = 1$ for almost all such (z', n').

Now suppose that all rungs strictly below (z, n) are already ranked by marginal surplus, in the sense that

$$x_{t,\breve{z}\breve{n},z'n'} = \begin{cases} 1 & \text{if} \quad S^p_n(z', n') > S^p_n(\breve{z}, \breve{n}), \\ 0 & \text{otherwise} \end{cases} \tag{300}$$

for every (z', n') and every $(\breve{z}, \breve{n})$ such that $S_n^p(\breve{z}, \breve{n}) < S_n^p(z, n)$. We will show that the same ranking holds at (z, n):

$$x_{t,zn,z'n'} = \begin{cases} 1 & \text{if} \quad S_n^p(z', n') > S_n^p(z, n), \\ 0 & \text{otherwise.} \end{cases} \tag{301}$$

Take any (z', n') such that $S_n^p(z', n') > S_n^p(z, n)$. Under the induction hypothesis, firms at (z', n') are ranked weakly above firms at (z, n), so $\Delta(z', n') \geq \Delta(z, n)$ and $V^p(z', n') \geq V^p(z, n)$. Therefore

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial x_{zn,z'n'}} \geq \zeta \phi(z, n) \phi(z', n') n' n \frac{1}{\Delta(z, n)} \\ \times \Big[h(z, n) \left(S_n^p(z', n') - V^p(z, n) \right) - h(z', n') \left(S_n^p(z, n) - V^p(z, n) \right) \Big]. \end{aligned}$$

Let

$$\Xi(z, n) \equiv S_n^p(z, n) - V^p(z, n), \qquad dS \equiv S_n^p(z', n') - S_n^p(z, n) > 0.$$

Using (297) and $V^p(z', n') \geq V^p(z, n)$, the bracketed term is bounded below by

$$(c^{h\prime})^{-1}(\Xi(z, n)) \left(\Xi(z, n) + dS \right) - (c^{h\prime})^{-1}(\Xi(z, n) + dS) \, \Xi(z, n).$$

Since $c^h(h) = \bar{c}^h h^{1+\gamma}/(1+\gamma)$, we have $(c^{h\prime})^{-1}(x)/x \propto x^{1/\gamma - 1}$, which is strictly decreasing when $\gamma > 1$. It follows that

$$\frac{(c^{h\prime})^{-1}(\Xi(z, n))}{\Xi(z, n)} > \frac{(c^{h\prime})^{-1}(\Xi(z, n) + dS)}{\Xi(z, n) + dS},$$

and therefore

$$(c^{h\prime})^{-1}(\Xi(z, n)) \left(\Xi(z, n) + dS \right) - (c^{h\prime})^{-1}(\Xi(z, n) + dS) \, \Xi(z, n) > 0.$$

Hence $\partial \mathcal{H} / \partial x_{zn,z'n'} > 0$, so the first line of (301) holds. If instead $S_n^p(z', n') < S_n^p(z, n)$, the same argument with the two states reversed and the constraint $x_{t,zn,z'n'} + x_{t,z'n',zn} = 1$ implies $x_{t,zn,z'n'} = 0$. Ties can be broken arbitrarily subject to the pairwise constraint and are immaterial for the allocation. Therefore, by induction, the planner's job ladder is ranked by marginal surplus:

$$x_{t,zn,z'n'} = \begin{cases} 1 & \text{if} \quad S_n^p(z', n') > S_n^p(z, n), \\ 0 & \text{otherwise.} \end{cases} \tag{302}$$

for all (z,n) and (z',n'). Since $x_{t,zn,z'n'} \in \{0,1\}$ for all pairs, the solution to the relaxed problem also solves the original problem.

The optimality condition with respect to $\phi_t(z,n)$, evaluated at the steady state, is

$$\begin{aligned}\rho S^p(z,n) &= y(z,n) - c^h(h(z,n))n - rU^p n - V^p(z,n)h(z,n)n \\ &\quad + \zeta \int x_{zn,z'n'} \frac{h(z',n')n'}{\Delta(z',n')} V^p(z',n')\, d\phi(z',n')\, n \\ &\quad + dn(z,n)\,\partial_n S^p(z,n) + \mu(z)\,\partial_z S^p(z,n) + \frac{1}{2}\sigma(z)^2 \partial_{zz} S^p(z,n)\end{aligned} \tag{303}$$

where $\rho \equiv r + \varkappa$,

$$rU^p \equiv b + \int \frac{h(z',n')n'}{\Delta(z',n')} V^p(z',n')\, d\phi(z',n')$$

is the planner analogue of the flow value of unemployment, and

$$dn(z,n) = \left[h(z,n) - s - \zeta \int x_{zn,z'n'} \frac{h(z',n')n'}{\Delta(z',n')} d\phi(z',n')\right] n.$$

The terms on the second line of (303) and in rU^p capture the marginal surplus gains that workers employed at (z,n) and unemployed workers, respectively, obtain when poached by other firms; the last term in $dn(z,n)$ is the corresponding outflow of workers from (z,n).

As in the equilibrium, we denote the vacancy-weighted distribution of marginal surplus as

$$\begin{aligned}\hat{F}(S_n) &= \frac{1}{\tilde{v}} \int_0^{S_n} v(z',n')n' d\phi(z',n') \\ &= \frac{1}{\tilde{v}} \int_0^{S_n} \frac{h(z',n')n'}{\frac{\lambda^F}{\tilde{u}} \Delta(z',n')} d\phi(z',n') \\ &= \frac{1}{\lambda^U} \int_0^{S_n} \frac{h(z',n')n'}{\Delta(z',n')} d\phi(z',n').\end{aligned}$$

We also let

$$\hat{V}^p(S_n) = \mathbb{E}_{\hat{G}}\left[s_n | s_n \leq S_n\right] \tag{304}$$

so that $\hat{V}^p(\partial_n S^p(z,n)) = V^p(z,n)$. Using these notations, we can rewrite (303) as

$$\begin{aligned}\rho S^p(z,n) = y(z,n) - c^h(h(z,n))n - rU^p n - V^p(z,n)h(z,n)n\\ + \lambda^E \int_{S_n^p(z,n)}^{\infty} \hat{V}^p(s_n)d\hat{F}(s_n)n\\ + dn(z,n)\,\partial_n S^p(z,n) + \mu(z)\,\partial_z S^p(z,n) + \frac{1}{2}\sigma(z)^2\partial_{zz}S^p(z,n)\end{aligned}, \tag{305}$$

with

$$rU^p = b + \lambda^U \int_{S_n^p(z,n)}^{\infty} \hat{V}^p(s_n)d\hat{F}(s_n)$$

and

$$dn(z,n) = \left[h(z,n) - s - \lambda^E(1 - \hat{F}(S_n^p(z,n)))\right]n.$$

The planner's steady-state firm density solves the KFE

$$0 = -\partial_n\left[dn(z,n)\phi(z,n)\right] - \partial_z\left[\mu(z)\phi(z,n)\right] + \frac{1}{2}\partial_{zz}\left[\sigma(z)^2\phi(z,n)\right] - \varkappa\phi(z,n) + m^0\psi(z,n). \tag{306}$$

The hiring condition is (297):

$$\partial_n S^p(z,n) - V^p(z,n) = c^{h\prime}(h(z,n)). \tag{307}$$

The first-order condition with respect to m_t^0 is

$$\int S^p(z,n)d\Psi(z,n) = c^e, \tag{308}$$

and the firing and exit conditions are

$$\partial_n S^p(z,n) \geq 0, \quad S^p(z,n) \geq 0. \tag{309}$$

The planner's solution, $\{S^p(z,n), h(z,n), V^p(z,n), \phi(z,n), m_t^0\}$, therefore solves (304)–(309), together with (306) and the rank-preserving property. These conditions are identical to the equilibrium conditions: (303) coincides with the equilibrium HJB equation (249) once $V^p(z,n)$ is identified with the equilibrium promised utility offer $V(z,n)$; (307) coincides with the equilibrium first-order condition for the hiring rate; (304) coincides with the equilibrium promised utility characterization in Proposition 15; and (306) coincides with the equilibrium KFE (260). Therefore, the equilibrium is constrained efficient.

H Ex-Ante Worker Heterogeneity

In the baseline model, all the workers are ex-ante homogeneous. We extend our baseline economy to incorporate ex-ante worker heterogeneity. There are finitely many types of workers, indexed by $a \in \mathcal{A} \equiv \{a_1, a_2, \ldots, a_N\}$. Let ℓ^a denote the exogenous mass of workers of type a. Each worker type differs in its flow utility from unemployment, b^a, its relative on-the-job search efficiency, ζ^a, separation rate, s^a, and how they enter into the firm's production function. The firm's production from its workforce $\{n^a\}_a$ is given by

$$y(z, \{n^a\}_{a\in\mathcal{A}}), \tag{310}$$

which we allow to be general other than the assumption y is strictly increasing, concave, and smooth in n^a. We assume that search is segmented by worker type. There is a matching function $\mathcal{M}^a(\tilde{v}^a, \tilde{u}^a)$ for each worker type a, where $\tilde{v}^a$ is the total number of vacancy postings for type a and $\tilde{u}^a$ is the total efficiency of search by workers of type a. Firms post vacancies for each worker type a, and the total vacancy cost is additively separable across worker types:

$$\sum_{a\in\mathcal{A}} c^a(v^a)n^a = \sum_{a\in\mathcal{A}} \bar{c}^a \frac{(v^a)^{1+\gamma^a}}{1+\gamma^a} n^a, \tag{311}$$

where v^a is the vacancy rate for type a. As in the baseline model, we assume $\gamma^a > 1$ for all a throughout. Finally, the firms offer contracts that are specific to each worker type.

Under these assumptions, the surplus at a firm with productivity z and workforce $\boldsymbol{n} \equiv \{n^a\}_{a\in\mathcal{A}}$ solves

$$\begin{aligned}
\rho S(z,\boldsymbol{n}) = \max_{\{v^a,V^a\}_{a\in\mathcal{A}}} & y(z,\boldsymbol{n}) - \sum_{a\in\mathcal{A}} c^a(v^a)n^a - \sum_{a\in\mathcal{A}} rU^a n^a \\
& - \sum_{a\in\mathcal{A}} V^a \lambda^{F,a} G^a(V^a) v^a n^a + \sum_{a\in\mathcal{A}} \lambda^{E,a} \int_{V^a} \tilde{V}\, dF^a(\tilde{V}) n^a \\
& + \mathcal{L}(z,\boldsymbol{n},\{v^a,V^a\}_{a\in\mathcal{A}}) S(z,\boldsymbol{n}),
\end{aligned} \tag{312}$$

where $\rho \equiv r + \varkappa$ and, for an arbitrary smooth function $f(z,\boldsymbol{n})$,

$$\begin{aligned}
&\mathcal{L}(z,\boldsymbol{n},\{v^a,V^a\}_{a\in\mathcal{A}}) f(z,\boldsymbol{n}) \\
&\quad = \mu(z)\partial_z f(z,\boldsymbol{n}) + \frac{\sigma(z)^2}{2}\partial_{zz} f(z,\boldsymbol{n}) \\
&\qquad + \sum_{a\in\mathcal{A}} \left[\lambda^{F,a} G^a(V^a) v^a - s^a - \lambda^{E,a}(1 - F^a(V^a))\right] n^a \partial_{n^a} f(z,\boldsymbol{n}).
\end{aligned} \tag{313}$$

The boundary conditions are

$$S(z, \boldsymbol{n}) \geq 0, \quad \partial_{n^a} S(z, \boldsymbol{n}) \geq 0 \quad \text{for all } a \in \mathcal{A}. \tag{314}$$

For each worker type a, the first-order conditions are

$$\left[\lambda^{F,a} g^a(V^a) v^a + \lambda^{E,a} f^a(V^a)\right] \left(\partial_{n^a} S(z, \boldsymbol{n}) - V^a\right) = \lambda^{F,a} G^a(V^a) v^a, \tag{315}$$

$$\lambda^{F,a} G^a(V^a) \left(\partial_{n^a} S(z, \boldsymbol{n}) - V^a\right) = c^{a\prime}(v^a). \tag{316}$$

Thus each worker type has its own marginal surplus,

$$S_a(z, \boldsymbol{n}) \equiv \partial_{n^a} S(z, \boldsymbol{n}), \tag{317}$$

and its own type-specific contract schedule $V^a(S_a)$ and vacancy policy $v^a(S_a)$.

Because search is segmented by worker type, the job ladder is also type specific. Let $\phi(z, \boldsymbol{n})$ denote the steady-state measure over firms. The mass of unemployed workers of type a is

$$u^a \equiv \ell^a - \int n^a \, d\phi(z, \boldsymbol{n}). \tag{318}$$

For each worker type a, define

$$\hat{G}^a(S_a) \equiv \frac{1}{\tilde{u}^a} \left[u^a + \zeta^a \int_{S_a(z,\boldsymbol{n}) \leq S_a} n^a \, d\phi(z, \boldsymbol{n}) \right], \tag{319}$$

$$\hat{F}^a(S_a) \equiv \frac{1}{\tilde{v}^a} \int_{S_a(z,\boldsymbol{n}) \leq S_a} v^a(S_a(z, \boldsymbol{n})) n^a \, d\phi(z, \boldsymbol{n}), \tag{320}$$

where

$$u^a \equiv \ell^a - \int n^a \, d\phi(z, \boldsymbol{n}), \tag{321}$$

$$\tilde{u}^a \equiv u^a + \zeta^a \int n^a \, d\phi(z, \boldsymbol{n}), \tag{322}$$

$$\tilde{v}^a \equiv \int v^a(S_a(z, \boldsymbol{n})) n^a \, d\phi(z, \boldsymbol{n}). \tag{323}$$

The meeting rates are

$$\lambda^{F,a} = \frac{\mathcal{M}^a(\tilde{v}^a, \tilde{u}^a)}{\tilde{v}^a}, \qquad \lambda^{E,a} = \zeta^a \frac{\mathcal{M}^a(\tilde{v}^a, \tilde{u}^a)}{\tilde{u}^a}. \tag{324}$$

The equilibrium contract characterization is the same as in the baseline economy, applied separately by worker type:

$$V^a(S_a) = \mathbb{E}_{(\hat{G}^a)^2}\left[\tilde{S}_a \mid \tilde{S}_a \leq S_a\right], \tag{325}$$

and

$$v^a(S_a) = (c^{a\prime})^{-1}\left(\lambda^{F,a} G^a(V^a(S_a))\,(S_a - V^a(S_a))\right). \tag{326}$$

Therefore, each worker type has a separate promised-utility ladder, but within a worker type the mapping from marginal surplus to promised utility is unchanged.

Finally, the steady-state firm density $\phi(z, \boldsymbol{n})$ satisfies the KFE

$$0 = \mathcal{L}^\dagger(z, \boldsymbol{n}, \{v^a, V^a\}_{a\in\mathcal{A}})\phi(z, \boldsymbol{n}) - \varkappa\phi(z, \boldsymbol{n}) + m^0\psi(z, \boldsymbol{n}), \tag{327}$$

where $\mathcal{L}^\dagger$ is the adjoint of the generator in (313), evaluated at the equilibrium policies. Entry satisfies

$$\int S(z, \boldsymbol{n})\, d\Psi(z, \boldsymbol{n}) = c^e. \tag{328}$$

The only substantive change relative to the baseline model is that all ranking and contract objects are indexed by worker type a.

I Permanent Firm Heterogeneity

In this section, we extend the model to allow for permanent firm heterogeneity. Upon entry, firms draw (z, n) as well as their permanent type ϑ from the cumulative distribution function $\Psi(z, n, \vartheta)$. Firms with different permanent types are heterogeneous in their technology, $y^\vartheta(z, n)$, vacancy cost, $c^\vartheta(v)$, the productivity process, $\mu^\vartheta(z)$ and $\sigma^\vartheta(z)$, separation rate, s^ϑ, and the exogenous exit rate, $\varkappa^\vartheta$. We assume that the firm type is permanently fixed, but it is straightforward to allow for a time-varying firm type. We continue to assume that the search is random and workers cannot direct search toward specific firm types. The rest of the environment remains unchanged.

In this environment, the HJB equation for the surplus function $S^\vartheta(z, n)$ is identical to that in the baseline model except that the firm-side primitives are indexed by ϑ:

$$\begin{aligned}\rho^\vartheta S^\vartheta(z, n) = \max_{v,V}\ & y^\vartheta(z, n) - c^\vartheta(v)n - rUn - V\lambda^F G(V)vn + \lambda^E \int_V \tilde{V} dF(\tilde{V})n \\ & + \mathcal{L}^\vartheta(z, n, v, V)S^\vartheta(z, n),\end{aligned} \tag{329}$$

where $\rho^\vartheta \equiv r + \varkappa^\vartheta$, $c^\vartheta(v) \equiv \bar{c}^\vartheta v^{1+\gamma}/(1+\gamma)$, and for an arbitrary smooth function $f(z, n, \vartheta)$,

$$\begin{aligned}\mathcal{L}^\vartheta(z, n, v, V)f(z, n, \vartheta) = {} & \mu^\vartheta(z)\partial_z f(z, n, \vartheta) + \frac{\sigma^\vartheta(z)^2}{2}\partial_{zz} f(z, n, \vartheta) \\ & + \left[\lambda^F G(V)v - s^\vartheta - \lambda^E(1 - F(V))\right] n\partial_n f(z, n, \vartheta).\end{aligned} \tag{330}$$

The boundary conditions are

$$S^\vartheta(z, n) \geq 0, \quad \partial_n S^\vartheta(z, n) \geq 0. \tag{331}$$

The first-order conditions are likewise type-indexed versions of the baseline conditions:

$$\left[\lambda^F g(V)v + \lambda^E f(V)\right]\left(\partial_n S^\vartheta(z, n) - V\right) = \lambda^F G(V)v, \tag{332}$$

$$\lambda^F G(V)\left(\partial_n S^\vartheta(z, n) - V\right) = c^{\vartheta\prime}(v). \tag{333}$$

Thus the relevant state for equilibrium contracts remains the marginal surplus,

$$S_n^\vartheta(z, n) \equiv \partial_n S^\vartheta(z, n), \tag{334}$$

and the policy functions can be written as $V(S_n^\vartheta(z, n))$ and $v(S_n^\vartheta(z, n), \vartheta)$. The promised utility offer is common across types conditional on the marginal surplus, while vacancy posting can differ by type because c^ϑ differs across types.

By following the identical proof strategy as in the baseline model, the job ladder is ranked by the marginal surplus across all types. Let $\phi(z, n, \vartheta)$ denote the steady-state measure over firm states and types. The employment-weighted and vacancy-weighted rankings are

$$\hat{G}(S_n) \equiv \frac{1}{\tilde{u}} \left[u + \zeta \int_{\partial_n S^\vartheta(z,n) \leq S_n} n \, d\phi(z, n, \vartheta) \right], \tag{335}$$

$$\hat{F}(S_n) \equiv \frac{1}{\tilde{v}} \int_{\partial_n S^\vartheta(z,n) \leq S_n} v(S_n^\vartheta(z, n), \vartheta) n \, d\phi(z, n, \vartheta), \tag{336}$$

where

$$\tilde{u} \equiv 1 - \int n \, d\phi(z, n, \vartheta) + \zeta \int n \, d\phi(z, n, \vartheta), \qquad \tilde{v} \equiv \int v(S_n^\vartheta(z, n), \vartheta) n \, d\phi(z, n, \vartheta). \tag{337}$$

The meeting rates are determined by $(\tilde{u}, \tilde{v})$ exactly as in the baseline economy.

The equilibrium contract characterization is unchanged and is given by

$$\hat{V}(S_n) = \mathbb{E}_{\hat{G}^2} \left[\tilde{S}_n \mid \tilde{S}_n \leq S_n \right]. \tag{338}$$

Thus all firms with the same marginal surplus offer the same promised utility, regardless of their permanent type. The vacancy policy, however, remains type-specific through the vacancy cost:

$$v^\vartheta(S_n) = \left(c^{\vartheta\prime}\right)^{-1} \left(\lambda^F G(\hat{V}(S_n)) \left(S_n - \hat{V}(S_n) \right) \right). \tag{339}$$

Therefore, permanent heterogeneity changes the distribution of firms over marginal surplus and the vacancy intensity chosen at a given marginal surplus, but not the mapping from marginal surplus to promised utility.

Finally, for each permanent type ϑ, the steady-state density satisfies the KFE

$$0 = \mathcal{L}^{\vartheta\dagger}(z, n, v, V)\phi(z, n, \vartheta) - \varkappa^\vartheta \phi(z, n, \vartheta) + m^0 \psi(z, n, \vartheta), \tag{340}$$

where $\mathcal{L}^{\vartheta\dagger}$ is the adjoint of the generator in (330), evaluated at the equilibrium policies. Entry satisfies

$$\int S^\vartheta(z, n) \, d\Psi(z, n, \vartheta) = c^e, \tag{341}$$

with entry composition inherited from $\Psi(z, n, \vartheta)$. Therefore, permanent firm heterogeneity only expands the firm state space from (z, n) to (z, n, ϑ); the equilibrium characterization and ranking arguments are otherwise unchanged.

J Transition Dynamics

We now study the model out of the steady state. Out of the steady state, the firm distribution ϕ is a part of the state variables. We formulate the master equation that explicitly carries the distribution as a state variable in the HJB equation. We focus on the planner's problem here. The decentralized equilibrium is analogous.

Recall that the planner's solution consists of the HJB equation (43) together with (39)-(44) and (45). Let $x \equiv (z, n)$. To ease the readability, we assume the matching function is Cobb-Douglas with constant returns to scale so that η^U and η^F are constants and $\eta^U + \eta^F = 1$, as in our calibration. The master equation for the transition dynamics is given by

$$
\begin{aligned}
\rho S^p(x,\phi) = {} & y(x) - c(v(x,\phi))n - bn + \mathcal{L}(x,\phi)S^p(x) \\
& - \lambda^U(\phi)\left\{\Gamma^U(\phi) - \eta^U\Omega(\phi)\right\}n \qquad (342)\\
& + \lambda^E(\phi)\left\{\Gamma^E(x,\phi) - \eta^U\Omega(\phi)\right\}n \\
& - \lambda^F(\phi)\left\{\Gamma^F(x,\phi) + \eta^F\Omega(\phi)\right\}v(x,\phi)n \qquad (343)\\
& + \int \frac{\partial S^p}{\partial \phi}(x,x',\phi)\left[\mathcal{L}^\dagger(x',\phi)\phi - \varkappa\phi(x') + m^0(\phi)\psi(x')\right]dx' \qquad (344)
\end{aligned}
$$

where $\mathcal{L}(x,\phi)$ is the generator defined as

$$
\begin{aligned}
\mathcal{L}(x,\phi)f(x,\phi) = {} & \mu(z)\partial_z f(x,\phi) + \frac{\sigma(z)^2}{2}\partial_{zz}f(x,\phi) \\
& + \left[\lambda^F(\phi)\hat{G}(S_n(x,\phi),\phi)v(x,\phi) - s - \lambda^E(\phi)(1-\hat{F}(S_n(x,\phi),\phi))\right]n\partial_n f(x,\phi),
\end{aligned}
\tag{345}
$$

and the vacancy policy is the first-order condition with respect to $v(x,\phi)$. Holding the aggregate objects fixed, the terms in the Master Equation that depend on the vacancy choice at state x are

$$
-c(v(x,\phi))n + \lambda^F(\phi)\hat{G}(S_n(x,\phi),\phi)v(x,\phi)nS_n(x,\phi) - \lambda^F(\phi)\left\{\Gamma^F(x,\phi) + \eta^F\Omega(\phi)\right\}v(x,\phi)n.
\tag{346}
$$

Thus

$$
c'(v(x,\phi)) = \lambda^F(\phi)\left[\hat{G}(S_n(x,\phi),\phi)S_n(x,\phi) - \Gamma^F(x,\phi) - \eta^F\Omega(\phi)\right],
\tag{347}
$$

so that

$$
\Gamma^U(\phi) \equiv \frac{1}{\tilde{v}(\phi)}\int S_n(x';\phi)v(x';\phi)n'\phi(x')dx'
\tag{348}
$$

$$\Gamma^E(x,\phi) \equiv \frac{1}{\tilde{v}(\phi)} \int_{x':S_n(x')\geq S_n(x)} S_n(x',\phi)v(x',\phi)n'\phi(x')dx' \tag{349}$$

$$\Gamma^F(x,\phi) \equiv \frac{1}{\tilde{u}(\phi)} \int_{x':S_n(x')\leq S_n(x)} S_n(x',\phi)\zeta n'\phi(x')dx' \tag{350}$$

$$\hat{G}(S_n,\phi) \equiv \frac{1}{\tilde{u}(\phi)} \left(1 - \int n'\phi(x')dx' + \int_{x':S_n(x')\leq S_n} \zeta n'\phi(x')dx'\right) \tag{351}$$

$$\hat{F}(S_n,\phi) \equiv \frac{1}{\tilde{v}(\phi)} \int_{x':S_n(x')\leq S_n} v(x')n'\phi(x')dx' \tag{352}$$

$$v(x,\phi) = c'^{-1}\left(\lambda^F(\phi)\left[\hat{G}(S_n(x,\phi),\phi)S_n(x,\phi) - \Gamma^F(x,\phi) - \eta^F\Omega(\phi)\right]\right) \tag{353}$$

$$\lambda^F(\phi) = \frac{\mathcal{M}(\tilde{u}(\phi),\tilde{v}(\phi))}{\tilde{v}(\phi)} \tag{354}$$

$$\lambda^U(\phi) = \frac{\mathcal{M}(\tilde{u}(\phi),\tilde{v}(\phi))}{\tilde{u}(\phi)} \tag{355}$$

$$\lambda^E(\phi) = \zeta\lambda^U(\phi) \tag{356}$$

$$\tilde{u}(\phi) = 1 - \int n'\phi(x')dx' + \int \zeta n'\phi(x')dx' \tag{357}$$

$$\tilde{v}(\phi) = \int v(x',\phi)n'\phi(x')dx' \tag{358}$$

$$\Omega(\phi) = \int \left(\hat{G}(S_n(x',\phi),\phi)S_n^p(x',\phi) - \Gamma^F(x',\phi)\right)\frac{1}{\tilde{v}(\phi)}v(x',\phi)n'\phi(x')dx', \tag{359}$$

and the entry rate $m^0(\phi)$ is pinned down by free entry. An entrant pays c^e, draws (z_0, n_0) from ψ, and immediately becomes an incumbent with value S^p, so the value of entry is $\int S^p(x',\phi)\psi(x')dx'$. With a continuum of potential entrants, free entry must hold at *every* date, i.e. for all ϕ along the transition:

$$\int S^p(x',\phi)\,\psi(x')dx' = c^e. \tag{360}$$

This is the equilibrium condition that determines the functional $m^0(\phi)$. Note that $m^0(\phi)$ enters the Master Equation only through the entry flow $m^0(\phi)\psi(x')$ in the last (Kolmogorov forward) term; firms take it as given.

J.1 First-order Approximation to the Master Equation (FAME)

We now derive the FAME of Bilal (2023) for the problem above. We first write the Master Equation compactly. To this end, collect the flow payoff into

$$\pi(x,\phi) \equiv y(x) - c(v(x,\phi))n - bn - \lambda^U(\phi)\left\{\Gamma^U(\phi) - \eta^U\Omega(\phi)\right\}n$$

$$+\,\lambda^E(\phi)\left\{\Gamma^E(x,\phi)-\eta^U\Omega(\phi)\right\}n-\lambda^F(\phi)\left\{\Gamma^F(x,\phi)+\eta^F\Omega(\phi)\right\}v(x,\phi)n, \tag{361}$$

and collect the Kolmogorov forward flow into the operator

$$\mathcal{A}(x',\phi)\equiv\mathcal{L}^\dagger(x',\phi)\phi-\varkappa\phi(x')+m^0(\phi)\psi(x'). \tag{362}$$

The Master Equation (363) is then

$$\rho S^p(x,\phi)=\pi(x,\phi)+\mathcal{L}(x,\phi)S^p(\cdot,\phi)(x)+\int\frac{\partial S^p}{\partial\phi}(x,x',\phi)\,\mathcal{A}(x',\phi)\,dx'. \tag{363}$$

Stationary equilibrium and the Impulse Value. Let $\bar\phi$ be the stationary distribution, so that $\mathcal{A}(x',\bar\phi)=0$ for all x'. Throughout this subsection, every object without an explicit distribution argument is evaluated at $\bar\phi$; for example, $S^p(x)\equiv S^p(x,\bar\phi)$, $S_n(x)\equiv\partial_n S^p(x,\bar\phi)$, $\pi(x)\equiv\pi(x,\bar\phi)$, $v(x)\equiv v(x,\bar\phi)$, and $\mathcal{L}(x)\equiv\mathcal{L}(x,\bar\phi)$. The steady-state Master Equation is $\rho S^p(x)=\pi(x)+\mathcal{L}(x)S^p(x)$. For a distributional impulse $h\equiv\phi-\bar\phi$ with $\|h\|$ small, the value is, to first order, a linear functional of h:

$$S^p(x,\bar\phi+h)=S^p(x)+\int q(x,\xi)\,h(\xi)\,d\xi+o(\|h\|),\qquad q(x,\xi)\equiv\left.\frac{\partial S^p}{\partial\phi}(x,\xi,\phi)\right|_{\phi=\bar\phi}. \tag{364}$$

We call $q(x,\xi)$ the impulse value: the marginal effect on the surplus of a firm at x of adding an atom of mass at state ξ. Differentiating in own employment defines the marginal impulse value $q_n(x,\xi)\equiv\partial_n q(x,\xi)$, so that the perturbed marginal surplus is

$$S^p_n(x,\bar\phi+h)=S_n(x)+\int q_n(x,\xi)\,h(\xi)\,d\xi+o(\|h\|). \tag{365}$$

Linearization. Differentiating (363) in the direction h, using $\mathcal{A}(\bar\phi)=0$ to kill the term that multiplies $\frac{\partial S^p}{\partial\phi}$ directly, and using the adjoint identity

$$\int q(x,x')\,\mathcal{L}^\dagger[h](x')\,dx'=\int\big(\mathcal{L}(x')\,q(x,\cdot)\big)(x')\,h(x')\,dx', \tag{366}$$

we equate the coefficient of $h(\xi)$ on both sides. This yields the FAME:

$$\rho\,q(x,\xi)=\underbrace{\Pi(x,\xi)}_{\text{direct: aggregates+own rank}}+\underbrace{\mathcal{L}(x)\big[q(\cdot,\xi)\big]}_{\text{PE: own state }x}+\underbrace{\mathcal{L}(\xi)\big[q(x,\cdot)\big]-\varkappa\,q(x,\xi)}_{\text{transport \& exit of impulse}} \tag{367}$$

$$+\underbrace{\int q(x,x')\,\mathcal{G}(x',\xi)\,dx'}_{\text{GE: law of motion}}+\underbrace{m^1(\xi)\int q(x,x')\psi(x')dx'}_{\text{entry}}, \tag{368}$$

where $\mathcal{L}(x)$ acts on the firm's own state (the first argument of q) and $\mathcal{L}(\xi)$ on the impulse location (the second argument), both through the steady-state generator. We unpack each term below.

Direct impact Π**.** Holding $v(x,\cdot) = v(x)$ (envelope), the direct response of the flow payoff and of the incumbent generator to the impulse is

$$\begin{aligned}\Pi(x,\xi) = n\Bigg[&- \frac{\delta\lambda^U}{\delta\phi(\xi)}\big(\Gamma^U - \eta^U\Omega\big) - \lambda^U\Big(\tfrac{\delta\Gamma^U}{\delta\phi(\xi)} - \eta^U\tfrac{\delta\Omega}{\delta\phi(\xi)}\Big) + \frac{\delta\lambda^E}{\delta\phi(\xi)}\big(\Gamma^E(x) - \eta^U\Omega\big)\\ &+ \lambda^E\Big(\tfrac{\delta\Gamma^E(x)}{\delta\phi(\xi)} - \eta^U\tfrac{\delta\Omega}{\delta\phi(\xi)}\Big) - v(x)\frac{\delta\lambda^F}{\delta\phi(\xi)}\big(\Gamma^F(x) + \eta^F\Omega\big) - \lambda^F v(x)\Big(\tfrac{\delta\Gamma^F(x)}{\delta\phi(\xi)} + \eta^F\tfrac{\delta\Omega}{\delta\phi(\xi)}\Big)\Bigg]\\ &+ n\,S_n(x)\left[\frac{\delta\lambda^F}{\delta\phi(\xi)}\hat{G}\,v(x) + \lambda^F v(x)\frac{\delta\hat{G}(x)}{\delta\phi(\xi)} - \frac{\delta\lambda^E}{\delta\phi(\xi)}\big(1-\hat{F}(x)\big) + \lambda^E\frac{\delta\hat{F}(x)}{\delta\phi(\xi)}\right],\end{aligned} \tag{369}$$

where $\hat{G}(x) \equiv \hat{G}(S_n(x),\bar{\phi})$, $\hat{F}(x) \equiv \hat{F}(S_n(x),\bar{\phi})$, and $\frac{\delta\Gamma^E(x)}{\delta\phi(\xi)}$, $\frac{\delta\Gamma^F(x)}{\delta\phi(\xi)}$, $\frac{\delta\hat{G}(x)}{\delta\phi(\xi)}$, $\frac{\delta\hat{F}(x)}{\delta\phi(\xi)}$ are Fréchet derivatives evaluated at the firm's own rank $S_n(x)$.

Aggregate Fréchet derivatives. The search pool and aggregate vacancies perturb as

$$\frac{\delta\tilde{u}}{\delta\phi(\xi)} = -(1-\zeta)\,n_\xi, \qquad \frac{\delta\tilde{v}}{\delta\phi(\xi)} = v(\xi)n_\xi + \int \frac{\delta v(x')}{\delta\phi(\xi)}\,n'\bar{\phi}(x')\,dx', \tag{370}$$

where n_ξ is the employment coordinate of ξ. Under the Cobb-Douglas matching function (so η^U, η^F are constant, as in the calibration),

$$\frac{\delta\lambda^F}{\delta\phi(\xi)} = \eta^F\,\lambda^F\left(\frac{1}{\tilde{u}}\frac{\delta\tilde{u}}{\delta\phi(\xi)} - \frac{1}{\tilde{v}}\frac{\delta\tilde{v}}{\delta\phi(\xi)}\right), \qquad \frac{\delta\lambda^U}{\delta\phi(\xi)} = \eta^U\,\lambda^U\left(\frac{1}{\tilde{v}}\frac{\delta\tilde{v}}{\delta\phi(\xi)} - \frac{1}{\tilde{u}}\frac{\delta\tilde{u}}{\delta\phi(\xi)}\right), \tag{371}$$

and $\frac{\delta\lambda^E}{\delta\phi(\xi)} = \zeta\frac{\delta\lambda^U}{\delta\phi(\xi)}$. The poaching-pool rankings perturb as

$$\frac{\delta\Gamma^U}{\delta\phi(\xi)} = -\frac{\Gamma^U}{\tilde{v}}\frac{\delta\tilde{v}}{\delta\phi(\xi)} + \frac{1}{\tilde{v}}S_n(\xi)v(\xi)n_\xi + \frac{1}{\tilde{v}}\int\Big[q_n(x',\xi)v(x') + S_n(x')\tfrac{\delta v(x')}{\delta\phi(\xi)}\Big]n'\bar{\phi}(x')dx'. \tag{372}$$

For a rank-indexed integral $\mathcal{I}(r) \equiv \int_{S_n(x')\le r} g(x',\phi)\,n'\phi(x')dx'$ evaluated at the own rank $r = S_n(x)$, the chain rule gives the three standard contributions—the direct impulse mass, the interior response of the integrand, and the boundary flux from the moving rank threshold:

$$\begin{aligned}\frac{\delta\mathcal{I}(S_n(x))}{\delta\phi(\xi)} &= g(\xi)\,n_\xi\,\mathbf{1}\{S_n(\xi)\le S_n(x)\} + \int_{S_n(x')\le S_n(x)}\frac{\delta g(x')}{\delta\phi(\xi)}n'\bar{\phi}(x')dx'\\ &+ \int g(x')n'\bar{\phi}(x')\,\delta_{\mathrm{D}}\big(S_n(x') - S_n(x)\big)\big(q_n(x,\xi) - q_n(x',\xi)\big)dx',\end{aligned} \tag{373}$$

where δ_{D} is the Dirac measure and $\mathbf{1}\{\cdot\}$ the indicator. Applying (373) with $g = \zeta$, with $g = S_n\zeta$, with $g = S_n v$ over the upper set, and with $g = v$ yields, respectively, $\frac{\delta \hat{G}}{\delta\phi(\xi)}$, $\frac{\delta\Gamma^F}{\delta\phi(\xi)}$, $\frac{\delta\Gamma^E}{\delta\phi(\xi)}$, and $\frac{\delta\hat{F}}{\delta\phi(\xi)}$ (with the appropriate normalization by $\tilde{u}$ or $\tilde{v}$ and the corresponding $-\frac{\delta\tilde{u}}{\delta\phi}/\tilde{u}$ or $-\frac{\delta\tilde{v}}{\delta\phi}/\tilde{v}$ scale term). Finally

$$\begin{aligned}\frac{\delta\Omega}{\delta\phi(\xi)} &= -\frac{\Omega}{\tilde{v}}\frac{\delta\tilde{v}}{\delta\phi(\xi)} + \frac{1}{\tilde{v}}\Big(\hat{G}(\xi)S_n(\xi) - \Gamma^F(\xi)\Big)v(\xi)n_\xi \\ &\quad + \frac{1}{\tilde{v}}\int \frac{\delta}{\delta\phi(\xi)}\Big[\big(\hat{G}(x')S_n(x') - \Gamma^F(x')\big)v(x')\Big]n'\bar{\phi}(x')dx'. \end{aligned} \tag{374}$$

Vacancy response. Linearizing the vacancy first-order condition gives the response of firm x''s vacancy posting to the impulse:

$$\begin{aligned}\frac{\delta v(x')}{\delta\phi(\xi)} = \frac{1}{c''(v(x'))}\Bigg[&\frac{\delta\lambda^F}{\delta\phi(\xi)}\Big(\hat{G}(x')S_n(x') - \Gamma^F(x') - \eta^F\Omega\Big) \\ &+ \lambda^F\left\{\hat{G}(x')q_n(x',\xi) + S_n(x')\frac{\delta\hat{G}(x')}{\delta\phi(\xi)} - \frac{\delta\Gamma^F(x')}{\delta\phi(\xi)} - \eta^F\frac{\delta\Omega}{\delta\phi(\xi)}\right\}\Bigg],\end{aligned} \tag{375}$$

where $\frac{\delta\hat{G}(x')}{\delta\phi(\xi)}$ is the total derivative of the composite object $\hat{G}(x') \equiv \hat{G}(S_n(x',\phi),\phi)$ evaluated at the steady state, so it includes both the direct effect of the distribution and the movement of firm x''s rank through $q_n(x',\xi)$. Equations (375) and the aggregate derivatives form a linear system that determines all aggregate perturbations as functionals of q.

General-equilibrium kernel $\mathcal{G}$. The impulse changes the employment drift of every firm x', hence the forward flow $\mathcal{L}^\dagger\phi$. Because the law of motion is not subject to the envelope theorem, the own-vacancy response is retained here:

$$\begin{aligned}\mathcal{G}(x',\xi) = -\partial_{n'}\Bigg[\Bigg\{&\frac{\delta\lambda^F}{\delta\phi(\xi)}\hat{G}(x')v(x') + \lambda^F\frac{\delta\hat{G}(x')}{\delta\phi(\xi)}v(x') + \lambda^F\hat{G}(x')\frac{\delta v(x')}{\delta\phi(\xi)} \\ &- \frac{\delta\lambda^E}{\delta\phi(\xi)}\big(1-\hat{F}(x')\big) + \lambda^E\frac{\delta\hat{F}(x')}{\delta\phi(\xi)}\Bigg\}n'\bar{\phi}(x')\Bigg].\end{aligned} \tag{376}$$

Entry response. The entry rate $m^0(\phi)$ is an aggregate that firms take as given: it enters the Master Equation (363) only through the entry flow $m^0(\phi)\psi$ in $\mathcal{A}$, so its perturbation $m^1(\xi) \equiv \frac{\delta m^0}{\delta\phi(\xi)}$ enters the FAME (368) only through the entry term $m^1(\xi)\int q(x,x')\psi(x')dx'$. The response m^1 is not given by a standalone formula; it is pinned down by the requirement that free entry (360) keeps holding along the transition. Because $\int S^p(x',\phi)\psi(x')dx' = c^e$ is an identity in ϕ, its

Fréchet derivative vanishes, giving the linearized free-entry condition

$$\int q(x', \xi)\, \psi(x') dx' = 0, \qquad \forall \xi, \tag{377}$$

where the integral runs over the entrant's landing state x' (the first argument of q). The FAME (368) is linear in q and in the function $m^1(\cdot)$, which enters only through the entry term; (377) supplies exactly one scalar condition per impulse location ξ, matching the single unknown $m^1(\xi)$ per ξ. Thus m^1 acts as the price response that clears the entry margin, and (q, m^1) are solved jointly from (368) and (377).

Evolution of the impulse. Once q is known, the impulse itself evolves, to first order, according to the linearized Kolmogorov forward equation

$$\frac{\partial h_t}{\partial t}(x) = \mathcal{L}^\dagger(x)[h_t] - \varkappa\, h_t(x) + \int \mathcal{G}(x, \xi)\, h_t(\xi)\, d\xi + \psi(x) \int m^1(\xi)\, h_t(\xi)\, d\xi, \tag{378}$$

where the kernel $\mathcal{G}$ coincides with the general-equilibrium kernel in the FAME (368), since both encode the change in firms' vacancy behavior in response to the impulse. Equations (368) and (378) together provide the block-recursive characterization of the transition dynamics around the stationary equilibrium: solve the FAME (368) once for the Impulse Value q, then propagate any initial impulse h_0 forward with (378).

J.2 Computing the Transition Dynamics

We solve the FAME (368) on a grid $\{x_i\}_{i=1}^N$ for the state $x = (z, n)$. Stack the Impulse Value into the matrix $\mathbf{Q} \in \mathbb{R}^{N \times N}$ with $\mathbf{Q}_{ij} = q(x_i, \xi_j)$, where rows index the firm's own state and columns the impulse location. Let $\mathbf{A}$ and $\mathbf{B}$ be the discrete steady-state generators $\mathcal{L}$ acting, respectively, on the own state (first argument of q) and on the impulse location (second argument); let $\mathbf{P}$ with $\mathbf{P}_{ij} = \Pi(x_i, \xi_j)$ be the direct impact (369); let $\mathbf{G}$ with $\mathbf{G}_{ij} = \mathcal{G}(x_i, \xi_j)$ be the discretized GE kernel; and let $\boldsymbol{\psi}$ be the discretized vector with $\boldsymbol{\psi}_i = \psi(x_i)$. The FAME (368) is then the matrix equation

$$\rho\, \mathbf{Q} = \mathbf{P} + \mathbf{A}\mathbf{Q} + \mathbf{Q}\big(\mathbf{B}^\top - \varkappa \mathbf{I}\big) + \mathbf{Q}\,\mathbf{G} + \mathbf{E}\,(\mathbf{m}^1)^\top, \qquad \mathbf{E} \equiv \mathbf{Q}\,\boldsymbol{\psi}, \tag{379}$$

where $\mathbf{E}_i = \int q(x_i, x')\psi(x')dx'$ is the entry value of firm i (the integral over the *second* argument of q) and $\mathbf{m}^1 \in \mathbb{R}^N$ collects $m^1(\xi_j)$. The linearized free-entry condition (377) integrates over the *first* argument of q,

$$(\boldsymbol{\psi})^\top \mathbf{Q} = \mathbf{0}_{1\times N}. \tag{380}$$

If we knew $(\mathbf{P}, \mathbf{G}, \mathbf{E})$, then we could solve the FAME (379): it is a Sylvester equation whose solution is affine in $\mathbf{m}^1$. Writing $\mathbf{C} \equiv \mathbf{B}^\top - \varkappa \mathbf{I} + \mathbf{G} - \rho \mathbf{I}$,

$$\mathbf{A}\mathbf{Q} + \mathbf{Q}\mathbf{C} = -\mathbf{P} - \mathbf{E}(\mathbf{m}^1)^\top \quad \Longrightarrow \quad \mathbf{Q} = \mathbf{Q}^P + \sum_{j=1}^{N} m^1(\xi_j)\, \mathbf{Q}^{E,j}, \tag{381}$$

where $\mathbf{Q}^P$ solves $\mathbf{A}\mathbf{Q}+\mathbf{Q}\mathbf{C} = -\mathbf{P}$ and $\mathbf{Q}^{E,j}$ solves $\mathbf{A}\mathbf{Q}+\mathbf{Q}\mathbf{C} = -\mathbf{E}\,\mathbf{e}_j^\top$. Imposing the free-entry condition (380) then yields a square linear system for the entry response,

$$\mathbf{M}\,\mathbf{m}^1 = -\mathbf{Q}^{P\top}\boldsymbol{\psi}, \qquad \mathbf{M}_{kj} = \left[(\boldsymbol{\psi})^\top \mathbf{Q}^{E,j}\right]_k, \tag{382}$$

which we solve directly for $\mathbf{m}^1$.

However, the coefficients $(\mathbf{P}, \mathbf{G})$ and the entry value $\mathbf{E}$ themselves depend on $\mathbf{Q}$ — through the marginal Impulse Value q_n and the linear system for the aggregate Fréchet derivatives in (375). For this reason, we need to solve the FAME (379) iteratively.

Algorithm 3 provides the overview. We first guess the impulse value $\mathbf{Q}^{(0)}$. For each iteration n, we proceed as follows. Given $\mathbf{Q}^{(n)}$, we can solve the linear system formed by the aggregate Fréchet derivatives of the previous subsection together with the vacancy response (375). Note that after substituting all the aggregate Fréchet derivatives into the vacancy response (375), we have a linear fixed point problem for the vacancy response. Collecting the vacancy responses across impulse columns into $\mathbf{V} \in \mathbb{R}^{N\times N}$ with $\mathbf{V}_{ij} = \frac{\delta v(x_i)}{\delta \phi(\xi_j)}$, the system is affine,

$$\mathbf{V} = \mathbf{V}^0(\mathbf{Q}^{(n)}) + \mathcal{K}\,\mathbf{V}, \tag{383}$$

or equivalently

$$\mathbf{V} = \left(\mathbf{I} - \mathcal{K}\right)^{-1} \mathbf{V}^0(\mathbf{Q}^{(n)}). \tag{384}$$

The term $\mathbf{V}^0(\mathbf{Q}^{(n)})$ gathers all q_n- and direct-impulse terms while the linear operator $\mathcal{K}$ is assembled from steady-state Fréchet derivatives.

Given the vacancy response $\mathbf{V}$, we can assemble the GE kernel $\mathbf{G}$ and the direct impact $\mathbf{P}$ from (369) and (376), respectively. We can then assemble the FAME (379) to solve for the Impulse Value $\mathbf{Q}^{(n+1)}$. If $\mathbf{Q}^{(n+1)}$ is close enough to $\mathbf{Q}^{(n)}$, we stop. Otherwise, we update our guess for the impulse value $\mathbf{Q}^{(n+1)}$ and repeat.

Algorithm 3 First-order Transition Dynamics (FAME)

Input: stationary equilibrium $\{S^p, S_n, v, \lambda^U, \lambda^E, \lambda^F, \Gamma^U, \Gamma^E, \Gamma^F, \Omega, \hat{G}, \hat{F}, \bar{\phi}\}$ on the grid
Assemble generators $\mathbf{A}, \mathbf{B}$, exit rate $\varkappa$, weights $\mathbf{D}$, entry vector $\boldsymbol{\psi}$
Fix tolerances $\varepsilon_Q > 0$; initialize $\mathbf{Q}^{(0)} \leftarrow \mathbf{0}$
repeat
  Given $\mathbf{Q}^{(n)}$, use (384) to obtain $\{\delta\lambda, \delta\Gamma, \delta\Omega, \delta\hat{G}, \delta\hat{F}, \delta v\}/\delta\phi$
  Assemble $\mathbf{P}$ from (369) and $\mathbf{G}$ from the GE kernel; set $\mathbf{C} \leftarrow \mathbf{B}^\top - \varkappa\mathbf{I} + \mathbf{G} - \rho\mathbf{I}$
  Assemble $\mathbf{E} \leftarrow \mathbf{Q}\,\boldsymbol{\psi}$
  Solve the Sylvester equations for $\mathbf{Q}^P$ and $\{\mathbf{Q}^{E,j}\}_{j=1}^N$
  Solve (382) for the entry response $\mathbf{m}^1$
  Update $\mathbf{Q}^{(n+1)}$ using $\mathbf{Q} = \mathbf{Q}^P + \sum_j m^1(\xi_j)\,\mathbf{Q}^{E,j}$
until $\|\mathbf{Q}^{(n+1)} - \mathbf{Q}^{(n)}\| \leq \varepsilon_Q$
Output: Impulse Value $q = \mathbf{Q}$, entry response $\mathbf{m}^1$, GE kernel $\mathbf{G}$
Given an initial impulse h_0, propagate (378), i.e. $\partial_t \mathbf{h}_t = \big(\mathbf{A}^\dagger - \varkappa\mathbf{I} + \mathbf{G} + \boldsymbol{\psi}(\mathbf{m}^1)^\top\big)\mathbf{h}_t$, forward in time

Once the Impulse Value $\mathbf{Q}$ has converged, we obtain the GE kernel $\mathbf{G}$ and the entry response $\mathbf{m}^1$. We can then solve for the first-order approximation of the transition dynamics of the firm distribution following an arbitrary initial impulse h_0 using (378). In matrix form,

$$\partial_t \mathbf{h}_t = \big(\mathbf{A}^\dagger - \varkappa\mathbf{I} + \mathbf{G} + \boldsymbol{\psi}(\mathbf{m}^1)^\top\big)\mathbf{h}_t. \tag{385}$$

J.3 Dimensionality Reduction of the FAME under Homogeneity

The cost of Algorithm 3 is dominated by the Sylvester solves on the $N \times N$ Impulse Value $\mathbf{Q}$, with $N = N_z N_n$, and by the free-entry loop, which solves N Sylvester systems $\{\mathbf{Q}^{E,j}\}_{j=1}^N$. For example, in our application, we have $N_z = 50$ and $N_n = 60$, so $N = 3000$, which is prohibitively expensive.

However, building on the same idea as Proposition 9, we can substantially reduce the dimensionality under the same assumptions. In particular, when the firm's production $y(z,n)$ is homogeneous of degree one and z follows geometric Brownian motion, the entire FAME collapses onto the one-dimensional state $\hat{n} \equiv n/z$, reducing $\mathbf{Q}$ from $N \times N$ to $N_h \times N_h$ where N_h is the number of $\hat{n}$ gridpoints.

Homogeneity of the stationary equilibrium. By Proposition 9, the stationary planner value is homogeneous of degree one, $S^p(z,n) = z\,\hat{S}(\hat{n})$, so the marginal surplus $S_n(z,n) = \hat{S}_n(\hat{n})$ and the vacancy policy $v(z,n) = \hat{v}(\hat{n})$ depend on the state only through $\hat{n}$; every rank-indexed aggregate $\hat{G}, \hat{F}, \Gamma^E, \Gamma^F$ is then a function of $\hat{n}$ alone, while $\lambda^U, \lambda^E, \lambda^F, \Gamma^U, \Omega$ are scalars. We record two consequences used repeatedly below. First, the generator inherits the homogeneity: for any C^2 function f, with $\mu(z) = \mu z$ and $\sigma(z) = \sigma z$,

$$\mathcal{L}(x)\big[z\, f(\hat{n})\big] = z\,\big(\hat{\mathcal{L}} f\big)(\hat{n}), \qquad \hat{\mathcal{L}} f(\hat{n}) \equiv \mu\, f + \hat{n}\big(b_{\hat{n}}(\hat{n}) - \mu\big) f' + \tfrac{\sigma^2}{2}\hat{n}^2 f'', \tag{386}$$

where $b_{\hat{n}}(\hat{n}) = \lambda^F \hat{G}(\hat{n})\hat{v}(\hat{n}) - s - \lambda^E(1 - \hat{F}(\hat{n}))$ is the steady-state employment growth rate. Second, the only feature of the stationary distribution $\bar{\phi}$ that the reduced problem requires is the employment-weighted marginal of $\hat{n}$, which we denote as $\mathsf{m}(\hat{n})$ defined as

$$\int f(n'/z')\, n'\, \bar{\phi}(x')\, dx' \equiv \int f(\hat{n}')\, \mathsf{m}(\hat{n}')\, d\hat{n}' \qquad \text{for all } f, \tag{387}$$

which is again produced by the stationary $\hat{n}$ reduction.

Proposition 17. *Assume $y(z,n)$ is homogeneous of degree one, $y(z,n) = z\hat{y}(\hat{n})$ with $\hat{n} \equiv n/z$, z follows a geometric Brownian motion, and the matching function is Cobb-Douglas with constant returns to scale. Suppose the Impulse Value takes the separable form*

$$q(x,\xi) = z\, z_\xi\, \hat{q}(\hat{n}, \hat{n}_\xi). \tag{388}$$

Then the marginal value and the objects derived from the Impulse Value inherit the factorization

$$q_n(x,\xi) = z_\xi\, \partial_{\hat{n}}\hat{q}(\hat{n}, \hat{n}_\xi), \tag{389}$$

$$\tfrac{\delta a}{\delta\phi(\xi)} = z_\xi\, \widehat{\delta a}(\hat{n}_\xi), \qquad a \in \{\tilde{u}, \tilde{v}, \lambda^U, \lambda^E, \lambda^F, \Gamma^U, \Omega\}, \tag{390}$$

$$\tfrac{\delta v(x')}{\delta\phi(\xi)} = z_\xi\, \widehat{\delta v}(\hat{n}', \hat{n}_\xi), \tag{391}$$

$$\tfrac{\delta\{\hat{G},\hat{F},\Gamma^E,\Gamma^F\}(x')}{\delta\phi(\xi)} = z_\xi\, \widehat{\delta\{\cdot\}}(\hat{n}', \hat{n}_\xi), \tag{392}$$

$$m^1(\xi) = z_\xi\, \hat{m}(\hat{n}_\xi), \tag{393}$$

and the Impulse Value q solves the FAME (368) subject to the linearized free-entry condition (377) if and only if the reduced kernel $\hat{q}$ solves the reduced FAME

$$\rho\, \hat{q}(\hat{n}, \hat{n}_\xi) = \hat{\Pi}(\hat{n}, \hat{n}_\xi) + \hat{\mathcal{L}}_{\hat{n}}\hat{q} + \hat{\mathcal{L}}_{\hat{n}_\xi}\hat{q} - \varkappa\, \hat{q} + (\hat{\mathcal{G}}\hat{q})(\hat{n}, \hat{n}_\xi) + \hat{e}(\hat{n})\, \hat{m}(\hat{n}_\xi), \tag{394}$$

subject to the reduced free-entry condition

$$\hat{e}_2(\hat{n}_\xi) \equiv \int z'\, \hat{q}(\hat{n}', \hat{n}_\xi)\, \psi(x')\, dx' = 0 \qquad \textit{for every } \hat{n}_\xi, \tag{395}$$

where $\hat{\Pi}$, $\hat{\mathcal{L}}_{\hat{n}}$, $\hat{\mathcal{G}}$, $\hat{e}$, and $\hat{m}$ are the z-stripped reduced operators constructed explicitly in the proof.

Proof. Substitute the separable form (388) into the FAME (368). We show that, term by term, every operator on the right-hand side returns $z\,z_\xi$ times a function of $(\hat{n}, \hat{n}_\xi)$ alone, and along the way that the induced objects inherit the stated factorizations. Since $z\,z_\xi > 0$, the FAME then holds for all (z, n, z_ξ, n_ξ) if and only if the common factor $z\,z_\xi$ cancels and the bracketed reduced equation holds for all $(\hat{n}, \hat{n}_\xi)$, which is exactly the reduced FAME (394); the same cancellation applied to (377) gives (395). We make no claim that an arbitrary solution of (368) is separable; we show only that the separable subspace is invariant, so the reduced problem (394) characterizes the separable Impulse Value. Throughout we use the homogeneity of the stationary equilibrium, $S_n(x) = \hat{S}_n(\hat{n})$ and $v(x) = \hat{v}(\hat{n})$, the generator identity (386), and the employment-weighted marginal (387); the marginal of the ansatz is $q_n(x, \xi) = \partial_n[z\,z_\xi\,\hat{q}(\hat{n}, \hat{n}_\xi)] = z_\xi\,\partial_{\hat{n}}\hat{q}(\hat{n}, \hat{n}_\xi)$, which is the first claim of the proposition.

Aggregate Fréchet derivatives. The search pool perturbs in closed form, $\frac{\delta\tilde{u}}{\delta\phi(\xi)} = -(1-\zeta)n_\xi$, which is z_ξ times a function of $\hat{n}_\xi$. Aggregate vacancies are $\frac{\delta\tilde{v}}{\delta\phi(\xi)} = \hat{v}(\hat{n}_\xi)\hat{n}_\xi z_\xi + \int \frac{\delta v(x')}{\delta\phi(\xi)} n'\bar{\phi}(x')dx'$; using the vacancy-response factorization $\frac{\delta v(x')}{\delta\phi(\xi)} = z_\xi\widehat{\delta v}(\hat{n}', \hat{n}_\xi)$ derived below and the employment-weighted marginal (387), the integral equals $z_\xi \sum_{\hat{n}'} \widehat{\delta v}(\hat{n}', \hat{n}_\xi)\mathsf{m}(\hat{n}')$. Hence $\frac{\delta\tilde{u}}{\delta\phi(\xi)} = z_\xi\widehat{\delta\tilde{u}}(\hat{n}_\xi)$ and $\frac{\delta\tilde{v}}{\delta\phi(\xi)} = z_\xi\widehat{\delta\tilde{v}}(\hat{n}_\xi)$, where

$$\widehat{\delta\tilde{u}}(\hat{n}_\xi) = -(1-\zeta)\hat{n}_\xi, \qquad \widehat{\delta\tilde{v}}(\hat{n}_\xi) = \hat{v}(\hat{n}_\xi)\hat{n}_\xi + \sum_{\hat{n}'} \widehat{\delta v}(\hat{n}', \hat{n}_\xi)\,\mathsf{m}(\hat{n}'). \tag{396}$$

The Cobb-Douglas meeting rates inherit the same factor,

$$\widehat{\delta\lambda^F} = \eta^F\lambda^F\left(\frac{\widehat{\delta\tilde{u}}}{\tilde{u}} - \frac{\widehat{\delta\tilde{v}}}{\tilde{v}}\right), \qquad \widehat{\delta\lambda^U} = (1-\eta)\lambda^U\left(\frac{\widehat{\delta\tilde{v}}}{\tilde{v}} - \frac{\widehat{\delta\tilde{u}}}{\tilde{u}}\right), \qquad \widehat{\delta\lambda^E} = \zeta\,\widehat{\delta\lambda^U}, \tag{397}$$

while $\widehat{\delta\Gamma^U}$ and $\widehat{\delta\Omega}$ are the z-stripped forms of their (z, n) counterparts. Thus $\frac{\delta a}{\delta\phi(\xi)} = z_\xi\widehat{\delta a}(\hat{n}_\xi)$ for $a \in \{\tilde{u}, \tilde{v}, \lambda^U, \lambda^E, \lambda^F, \Gamma^U, \Omega\}$.

Rank rule. Since S_n is monotone in $\hat{n}$, the lower set $\{x' : S_n(x') \le S_n(x)\}$ equals $\{\hat{n}' \ge \hat{n}\}$. In (373) the direct-mass term is $g(\xi)n_\xi\mathbf{1} = z_\xi\hat{g}(\hat{n}_\xi)\hat{n}_\xi\mathbf{1}\{\hat{n}_\xi \ge \hat{n}\}$, and the interior term integrates $\frac{\delta g}{\delta\phi(\xi)} = z_\xi\widehat{\delta g}$ against $n'\bar{\phi}$ and reduces through (387). The boundary-flux term of (373) carries $q_n(x, \xi) - q_n(x', \xi)$ evaluated on the rank boundary $S_n(x') = S_n(x)$; under homogeneity rank is monotone in $\hat{n}$, so this boundary is $\hat{n}' = \hat{n}$, where $q_n(x', \xi) = z_\xi\partial_{\hat{n}'}\hat{q}(\hat{n}', \hat{n}_\xi)\big|_{\hat{n}'=\hat{n}} = z_\xi\partial_{\hat{n}}\hat{q}(\hat{n}, \hat{n}_\xi) = q_n(x, \xi)$. The integrand therefore vanishes identically: firms that share a rank also share $\hat{n}$ and move together in rank space, so the moving threshold produces no flux. Each surviving term carries a single factor z_ξ, so writing $\hat{\mathcal{I}}(\hat{n}) = \sum_{\hat{n}' \ge \hat{n}} \hat{g}(\hat{n}')\mathsf{m}(\hat{n}')$ for the z-stripped rank-indexed aggregate,

$$\widehat{\delta\mathcal{I}}(\hat{n},\hat{n}_\xi)=\hat{g}(\hat{n}_\xi)\hat{n}_\xi\,\mathbf{1}\{\hat{n}_\xi\geq\hat{n}\}+\sum_{\hat{n}'\geq\hat{n}}\widehat{\delta g}(\hat{n}',\hat{n}_\xi)\,\mathsf{m}(\hat{n}'). \tag{398}$$

Applying (398) with $\hat{g}=\zeta,\ \zeta\hat{S}_n,\ \hat{S}_n\hat{v}$ (upper set), and $\hat{v}$, and normalizing by $\tilde{u}$ or $\tilde{v}$, gives $\widehat{\delta\hat{G}}, \widehat{\delta\Gamma^F}, \widehat{\delta\Gamma^E}, \widehat{\delta\hat{F}}$, so that $\frac{\delta\{\hat{G},\hat{F},\Gamma^E,\Gamma^F\}(x')}{\delta\phi(\xi)}=z_\xi\widehat{\delta\{\cdot\}}(\hat{n}',\hat{n}_\xi)$.

Vacancy response. In (375), $c''(\hat{v}(\hat{n}'))$ depends only on $\hat{n}'$, the leading bracket multiplies $\frac{\delta\lambda^F}{\delta\phi(\xi)}=z_\xi\widehat{\delta\lambda^F}$, the term $\hat{G}(\hat{n}')q_n(x',\xi)=z_\xi\hat{G}(\hat{n}')\partial_{\hat{n}'}\hat{q}$, and the remaining δ-terms are each $z_\xi\times(\cdot)$. Thus $\frac{\delta v(x')}{\delta\phi(\xi)}=z_\xi\widehat{\delta v}(\hat{n}',\hat{n}_\xi)$ with

$$\widehat{\delta v}(\hat{n}',\hat{n}_\xi)=\frac{1}{c''(\hat{v}(\hat{n}'))}\Big[\widehat{\delta\lambda^F}\big(\hat{G}(\hat{n}')\hat{S}_n(\hat{n}')-\Gamma^F(\hat{n}')-\eta^F\Omega\big)+\lambda^F\big(\hat{G}(\hat{n}')\partial_{\hat{n}'}\hat{q}+\hat{S}_n(\hat{n}')\widehat{\delta\hat{G}}-\widehat{\delta\Gamma^F}-\eta^F\widehat{\delta\Omega}\big)\Big], \tag{399}$$

zero where $\hat{v}(\hat{n}')\leq 0$. Collected across $\hat{n}'$ together with the aggregate vacancies $\widehat{\delta\tilde{v}}$ and $\widehat{\delta\Omega}$, (399) is the affine fixed point $(\mathbf{I}-\hat{\mathcal{K}})\widehat{\mathbf{V}}=\widehat{\mathbf{V}}^0$, which closes the aggregate Fréchet derivatives used above; the operator $\hat{\mathcal{K}}$ is assembled from the steady-state reduced objects.

Direct impact. Each term of (369) carries the firm's own factor $n=z\hat{n}$ times a z_ξ-scaled Fréchet derivative or $\hat{n}$-function, and $S_n(x)=\hat{S}_n(\hat{n})$, so $\Pi(x,\xi)=z\,z_\xi\,\hat{\Pi}(\hat{n},\hat{n}_\xi)$ with

$$\begin{aligned}\hat{\Pi}(\hat{n},\hat{n}_\xi)=\hat{n}\Big[&-\widehat{\delta\lambda^U}\big(\Gamma^U-\eta^U\Omega\big)-\lambda^U\big(\widehat{\delta\Gamma^U}-\eta^U\widehat{\delta\Omega}\big)+\widehat{\delta\lambda^E}\big(\Gamma^E(\hat{n})-\eta^U\Omega\big)+\lambda^E\big(\widehat{\delta\Gamma^E}(\hat{n},\hat{n}_\xi)-\eta^U\widehat{\delta\Omega}\big)\\&-\hat{v}(\hat{n})\,\widehat{\delta\lambda^F}\big(\Gamma^F(\hat{n})+\eta^F\Omega\big)-\lambda^F\hat{v}(\hat{n})\big(\widehat{\delta\Gamma^F}(\hat{n},\hat{n}_\xi)+\eta^F\widehat{\delta\Omega}\big)\Big]\\&+\hat{n}\,\hat{S}_n(\hat{n})\Big[\widehat{\delta\lambda^F}\,\hat{G}(\hat{n})\hat{v}(\hat{n})+\lambda^F\hat{v}(\hat{n})\,\widehat{\delta\hat{G}}(\hat{n},\hat{n}_\xi)-\widehat{\delta\lambda^E}\big(1-\hat{F}(\hat{n})\big)+\lambda^E\,\widehat{\delta\hat{F}}(\hat{n},\hat{n}_\xi)\Big],\end{aligned} \tag{400}$$

where $\widehat{\delta\lambda^\bullet}, \widehat{\delta\Gamma^U}, \widehat{\delta\Omega}$ are functions of $\hat{n}_\xi$ alone and $\widehat{\delta\Gamma^E}, \widehat{\delta\Gamma^F}, \widehat{\delta\hat{G}}, \widehat{\delta\hat{F}}$ are evaluated at the firm's own rank $\hat{n}$.

Generators and exit. By (386), $\mathcal{L}(x)[q(\cdot,\xi)]=zz_\xi\hat{\mathcal{L}}_{\hat{n}}\hat{q}$ and $\mathcal{L}(\xi)[q(x,\cdot)]=zz_\xi\hat{\mathcal{L}}_{\hat{n}_\xi}\hat{q}$, while $-\varkappa q=-\varkappa zz_\xi\hat{q}$.

General-equilibrium term. The drift-response bracket of the GE kernel $\mathcal{G}$ in (376) is the z-stripped law-of-motion response

$$\hat{D}(\hat{n}',\hat{n}_\xi)=\widehat{\delta\lambda^F}\,\hat{G}(\hat{n}')\hat{v}(\hat{n}')+\lambda^F\widehat{\delta\hat{G}}\,\hat{v}(\hat{n}')+\lambda^F\hat{G}(\hat{n}')\,\widehat{\delta v}-\widehat{\delta\lambda^E}\big(1-\hat{F}(\hat{n}')\big)+\lambda^E\,\widehat{\delta\hat{F}}, \tag{401}$$

so that $\mathcal{G}(x',\xi)=-z_\xi\,\partial_{n'}[\hat{D}(\hat{n}',\hat{n}_\xi)n'\bar{\phi}(x')]$. Integrating by parts in n' and using $\partial_{n'}[z'\hat{q}(\hat{n},\hat{n}')]=$

$\partial_{\hat{n}'}\hat{q}(\hat{n},\hat{n}')$ together with (387),

$$\int q(x,x')\,\mathcal{G}(x',\xi)\,dx' = z\,z_\xi \int \partial_{\hat{n}'}\hat{q}(\hat{n},\hat{n}')\,\hat{D}(\hat{n}',\hat{n}_\xi)\,\mathsf{m}(\hat{n}')\,d\hat{n}' \equiv z\,z_\xi\,(\hat{\mathcal{G}}\hat{q})(\hat{n},\hat{n}_\xi). \tag{402}$$

On the grid the reduced GE term $(\hat{\mathcal{G}}\hat{q})(\hat{n},\hat{n}_\xi) = \sum_{\hat{n}'}\partial_{\hat{n}'}\hat{q}(\hat{n},\hat{n}')\,\hat{D}(\hat{n}',\hat{n}_\xi)\,\mathsf{m}(\hat{n}')$ is the right-multiplication $\hat{\mathbf{Q}}\hat{\mathbf{G}}$ with column $\hat{\mathbf{G}}_{\cdot\,\xi} = -\partial_{\hat{n}'}[\hat{D}(\cdot,\hat{n}_\xi)\mathsf{m}]$.

Entry. With $\int q(x,x')\psi(x')dx' = z\,\hat{e}(\hat{n})$, where $\hat{e}(\hat{n}) \equiv \int z'\hat{q}(\hat{n},\hat{n}')\psi(x')dx'$, and $m^1(\xi) = z_\xi\hat{m}(\hat{n}_\xi)$, the entry term is $zz_\xi\hat{e}(\hat{n})\hat{m}(\hat{n}_\xi)$. Because all entrants share $z = z_{\text{entry}}$, both the entry value $\hat{e}(\hat{n}) = \sum_{\hat{n}'}\hat{q}(\hat{n},\hat{n}')\hat{\psi}_z(\hat{n}')$ and the reduced free-entry condition contract the z-weighted entry vector $\hat{\psi}_z(\hat{n}') = z_{\text{entry}}\hat{\psi}(\hat{n}')$. The linearized free-entry condition (377) integrates the separable q against ψ over the entrant's landing state x'; stripping the common z_ξ leaves the reduced free-entry condition (395), i.e. $\sum_{\hat{n}'}\hat{q}(\hat{n}',\hat{n}_\xi)\hat{\psi}_z(\hat{n}') = 0$.

Each contribution to the FAME (368)—the direct impact Π, the own-state and impulse-side generators, the exit term, the general-equilibrium integral, and the entry term—is proportional to $z\,z_\xi$ with a coefficient depending on the states only through $(\hat{n},\hat{n}_\xi)$, and the same is true of the left-hand side $\rho\,q = z\,z_\xi\,\rho\,\hat{q}$. Cancelling the common factor $z\,z_\xi > 0$, the FAME (368) holds for all states if and only if the reduced FAME (394) holds for all $(\hat{n},\hat{n}_\xi)$, and likewise (377) reduces to (395). The factorizations of q_n, the aggregate Fréchet derivatives, $\delta v/\delta\phi(\xi)$, the rank-indexed derivatives, and m^1 asserted in the proposition were established in the corresponding steps above. □

Discretized system and active set. Stack $\hat{\mathbf{Q}}_{ij} = \hat{q}(\hat{n}_i,\hat{n}_j)$. Let $\hat{\mathbf{A}}$ discretize the drift–diffusion part $\hat{n}(b_{\hat{n}}-\mu)\partial_{\hat{n}} + \frac{\sigma^2}{2}\hat{n}^2\partial_{\hat{n}}^2$ of (386); the two level terms μ — one from $\hat{\mathcal{L}}_{\hat{n}}$ and one from $\hat{\mathcal{L}}_{\hat{n}_\xi}$ — combine with $\rho\hat{q}$ into an effective discount $\rho - 2\mu$. With $\hat{\mathbf{C}} = \hat{\mathbf{A}}^\top - \varkappa\mathbf{I} + \hat{\mathbf{G}} - (\rho-2\mu)\mathbf{I}$ and $\hat{\mathbf{E}} = \hat{\mathbf{Q}}\hat{\boldsymbol{\psi}}_z$, the reduced FAME is the Sylvester system

$$\hat{\mathbf{A}}\hat{\mathbf{Q}} + \hat{\mathbf{Q}}\hat{\mathbf{C}} = -\hat{\mathbf{P}} - \hat{\mathbf{E}}(\hat{\mathbf{m}})^\top, \qquad \hat{\boldsymbol{\psi}}_z^\top\hat{\mathbf{Q}} = \mathbf{0}_{1\times N_h}, \tag{403}$$

i.e. the matrix system (379)–(382) with $N_h\times N_h$ $\hat{n}$-indexed blocks and discount $\rho\to\rho-2\mu$, so the per-iteration cost falls from $O(N^3)$ to $O(N_h^3)$ with $N_h \ll N = N_zN_n$.

The remaining steps follow the analogue of Algorithm 3. Solving the Sylvester system (403) for $\hat{q}$, recovering the free-entry response $\hat{\mathbf{m}}$, and iterating the Fréchet derivatives off q_n proceed exactly as in Algorithm 3, with the $N\times N$ blocks replaced by their $N_h\times N_h$ $\hat{n}$-reduced counterparts and the discount shifted to $\rho-2\mu$. After convergence, the GE kernel and entry response are reconstructed on the (z,n) grid from $\hat{D}$ and $\hat{m}$ via $\mathcal{G}(x,\xi) = -z_\xi\,\partial_n[\hat{D}(\hat{n},\hat{n}_\xi)\,n\,\bar{\phi}(x)]$ and $m^1(\xi) = z_\xi\,\hat{m}(\hat{n}_\xi)$. The impulse transition (378) itself does not reduce—firms with the same $\hat{n}$

but different z are transported differently by the idiosyncratic drift—so h_t is propagated on the full (z, n) grid as in Algorithm 3.

K Existence of Steady State Equilibrium

We discuss the existence of a steady state equilibrium.

K.1 Summary of the Equilibrium Conditions

A stationary equilibrium consists of the surplus function $S(z,n)$, the promised-utility and vacancy policies $\hat{V}(S_n)$ and $\hat{v}(S_n)$, the firm measure $\phi(z,n)$ and entry mass m^0, the rankings $\hat{G}$ and $\hat{F}$, the meeting rates $(\lambda^U, \lambda^E, \lambda^F)$, and the aggregates $(\tilde{u}, \tilde{v}, u)$ that jointly satisfy the conditions below.

Job-ladder rankings. The search-effort–weighted and vacancy-weighted rankings of the marginal surplus are

$$\hat{G}(S_n) = \frac{1}{\tilde{u}}\big[\, u + \zeta \int_{S_n(z,n)\le S_n} n\, d\phi(z,n)\,\big], \qquad \hat{F}(S_n) = \frac{1}{\tilde{v}} \int_{S_n(z,n)\le S_n} v(z,n)\, n\, d\phi(z,n). \tag{404}$$

$\hat{G}(S_n)$ is the probability that a vacancy posted by a firm with marginal surplus S_n hires the worker it meets, and $1-\hat{F}(S_n)$ is the probability that one of its workers is poached upon meeting an outside firm.

Firm optimality. The surplus function solves the Hamilton–Jacobi–Bellman equation

$$\begin{aligned}\rho\, S(z,n) = y(z,n) - c\big(\hat{v}(S_n)\big)\, n - b\, n - \lambda^U\, \mathbb{E}_{\hat{F}}\big[\hat{V}(\tilde{S}_n)\big]\, n - \lambda^F\, \hat{G}(S_n)\, \hat{V}(S_n)\, \hat{v}(S_n)\, n \\ + \lambda^E\big(1-\hat{F}(S_n)\big)\, \mathbb{E}_{\hat{F}}\big[\hat{V}(\tilde{S}_n)\,\big|\, \tilde{S}_n \ge S_n\big]\, n + \mathcal{L}\, S(z,n),\end{aligned} \tag{405}$$

where the generator acting on a function $S(z,n)$ is

$$\mathcal{L}\, S(z,n) = \mu(z)\, \partial_z S + \tfrac{1}{2}\sigma^2(z)\, \partial_{zz} S + \Big[\lambda^F \hat{G}(S_n)\, \hat{v}(S_n) - s - \lambda^E\big(1-\hat{F}(S_n)\big)\Big]\, n\, \partial_n S. \tag{406}$$

The optimal promised-utility and vacancy policies are

$$\hat{V}(S_n) = \mathbb{E}_{\hat{G}^2}\big[\tilde{S}_n \,\big|\, \tilde{S}_n \le S_n\big] \tag{407}$$

$$\hat{v}(S_n) = c'^{-1}\Big(\lambda^F\, \hat{G}(S_n)\, \big(S_n - \hat{V}(S_n)\big)\Big), \tag{408}$$

where the expectation $\mathbb{E}_{\hat{G}^2}[\,\cdot \mid \tilde{S}_n \le S_n]$ is taken under the cumulative weighting function $\hat{G}^2(\tilde{S}_n|\tilde{S}_n \le S_n)$. A firm employs a worker only where $S_n \ge 0$ and operates only where $S \ge 0$;

the firing margin is set by $S_n = 0$.

Stationary distribution and free entry. The firm distribution $\phi(z,n)$ and the entry mass m^0 satisfy the Kolmogorov forward equation

$$0 = \mathcal{L}^\dagger \phi(z,n) - \varkappa \phi(z,n) + m^0 \psi(z,n) \tag{409}$$

on the region where $S(z,n) > 0$ and $S_n(z,n) > 0$, where $\mathcal{L}^\dagger$ is the adjoint of the generator $\mathcal{L}$. The entry mass m^0 is pinned down by the free-entry condition

$$\int S(z,n)\, d\Psi(z,n) = c^e. \tag{410}$$

Aggregates and meeting rates. Aggregate search effort, aggregate vacancies, and unemployment are

$$\tilde{u} = u + \zeta(1-u), \qquad \tilde{v} = \int v(z,n)\, n\, d\phi(z,n), \qquad u = 1 - \int n\, d\phi(z,n), \tag{411}$$

and the meeting rates implied by the matching function are

$$\lambda^U = \frac{M(\tilde{u},\tilde{v})}{\tilde{u}}, \qquad \lambda^E = \zeta\,\frac{M(\tilde{u},\tilde{v})}{\tilde{u}}, \qquad \lambda^F = \frac{M(\tilde{u},\tilde{v})}{\tilde{v}}. \tag{412}$$

K.2 Existence of a Stationary Distribution

It is useful to characterize the equilibrium in two steps. First, for any given market tightness θ, we collect all the equilibrium conditions except for the free entry condition and define a mapping from the search-weighted ranking of the marginal surplus $\hat{G}$ to itself, which we write $\mathcal{T}_\theta(\hat{G})$. Then, the steady state equilibrium with positive entry is defined as

$$\hat{G} = \mathcal{T}_\theta(\hat{G}) \tag{413}$$

$$\int S(z,n)\, d\Psi(z,n) = c^e. \tag{414}$$

We now establish the existence of a fixed point of (413) for a given θ under a set of regularity conditions.

We first formally define the mapping $\mathcal{T}_\theta(\hat{G})$. In the definition, we introduce the lower bound on unemployment $\underline{u}$ to regularize the domain of the mapping. This lower bound on unemployment can be set to an arbitrarily small positive value, so that it does not restrict the equilibrium.

Definition 1. *Given a pair* $(\theta, \hat{G}) \in [\underline{\theta}, \bar{\theta}] \times \mathcal{P}(\mathbb{S})$ *with* $\hat{G}(\{0\}) \geq \underline{u}$*, define successively:*

1. *the meeting rates* $\lambda^U(\theta) = M(1,\theta)$, $\lambda^E(\theta) = \zeta\, M(1,\theta)$, $\lambda^F(\theta) = M(1/\theta, 1)$;

2. *the policies, which depend only on* $(\theta, \hat{G})$,

$$\hat{V}_{\hat{G}}(S_n) = \frac{1}{\left(\hat{G}(S_n)\right)^2} \int_0^{S_n} \tilde{S}_n \, d\big(\hat{G}(\tilde{S}_n)^2\big) \tag{415}$$

$$\hat{v}_{\hat{G}}(S_n) = \left(\frac{\lambda^F(\theta)\, \hat{G}(S_n) \left(S_n - \hat{V}_{\hat{G}}(S_n)\right)}{\overline{c}^v} \right)^{1/\gamma}; \tag{416}$$

3. *the vacancy-weighted ranking*

$$\hat{F}_{\hat{G}}(S_n) = \frac{\int_0^{S_n} \hat{v}_{\hat{G}}(\tilde{S}_n)\, d\hat{G}(\tilde{S}_n)}{\int_0^{\bar{S}} \hat{v}_{\hat{G}}(\tilde{S}_n)\, d\hat{G}(\tilde{S}_n)}; \tag{417}$$

4. *the surplus* $S_{\hat{G}}(z,n)$ *and marginal surplus* $S_{\hat{G},n}(z,n)$ *solving the following HJB equation given* $\lambda^U(\theta)$, $\lambda^E(\theta)$, $\lambda^F(\theta)$, $\hat{V}_{\hat{G}}(S_n)$, $\hat{v}_{\hat{G}}(S_n)$, *and* $\hat{F}_{\hat{G}}(S_n)$ *defined above:*

$$\begin{aligned}\rho\, S(z,n) = y(z,n) - c\big(\hat{v}_{\hat{G}}(S_n)\big)\, n - b\, n - \lambda^U\, \mathbb{E}_{\hat{F}_{\hat{G}}}\big[\hat{V}_{\hat{G}}(\tilde{S}_n)\big]\, n - \lambda^F\, \hat{G}(S_n)\, \hat{V}_{\hat{G}}(S_n)\, \hat{v}_{\hat{G}}(S_n)\, n \\ + \lambda^E\big(1 - \hat{F}_{\hat{G}}(S_n)\big)\, \mathbb{E}_{\hat{F}_{\hat{G}}}\big[\hat{V}_{\hat{G}}(\tilde{S}_n) \,\big|\, \tilde{S}_n \geq S_n\big]\, n + \mathcal{L}_{\hat{G}}\, S(z,n),\end{aligned} \tag{418}$$

where the generator acting on a function $S(z,n)$ *is*

$$\mathcal{L}_{\hat{G}}\, S(z,n) = \mu(z)\, \partial_z S + \tfrac{1}{2}\sigma^2(z)\, \partial_{zz} S \tag{419}$$

$$+ \Big[\lambda^F(\theta)\, \hat{G}(S_n)\, \hat{v}_{\hat{G}}(S_n) - s - \lambda^E(\theta)\big(1 - \hat{F}_{\hat{G}}(S_n)\big)\Big]\, n\, \partial_n S \tag{420}$$

with boundary conditions $S(z,n) \geq 0$ *and* $\partial_n S(z,n) \geq 0$.

5. *the unit-entry measure* $\phi^0_{\hat{G}}$ *solving* $0 = \mathcal{L}^\dagger_{\hat{G}} \phi^0_{\hat{G}} - \varkappa \phi^0_{\hat{G}} + \psi$.

6. *The entry mass* $m^0_{\hat{G}}$ *and unemployment* $u_{\hat{G}}$ *that are consistent with the labor market clearing:*

$$m^0_{\hat{G}} = \min\left\{ \frac{\theta}{\int \hat{v}_{\hat{G}}(S_{n,\hat{G}}(z,n)) n d\phi^0_{\hat{G}} + \theta(1-\zeta) \int n d\phi^0_{\hat{G}}}, \quad \frac{1 - \underline{u}}{\int n d\phi^0_{\hat{G}}} \right\} \tag{421}$$

$$u_{\hat{G}} = 1 - m^0_{\hat{G}} \int n d\phi^0_{\hat{G}} \tag{422}$$

$$\tilde{u}_{\hat{G}} = 1 - (1-\zeta)\, m^0_{\hat{G}} \int n d\phi^0_{\hat{G}}, \tag{423}$$

Then the induced mapping $\mathcal{T}_\theta(\hat{G})$ is given by

$$\mathcal{T}_\theta(\hat{G}) = \frac{u_{\hat{G}} + \zeta\, m^0_{\hat{G}} \int_{S_n(z,n)\leq S_n} n\, d\phi^0_{\hat{G}}}{u_{\hat{G}} + \zeta\, m^0_{\hat{G}} \int n d\phi^0_{\hat{G}}}. \tag{424}$$

Our goal is to establish the existence of a fixed point, $\hat{G}^*$, such that $\mathcal{T}_\theta(\hat{G}^*) = \hat{G}^*$. To this end, we need to impose the following regularity conditions.

Assumption 1. *The production function takes the form*

$$y(z,n) = \tilde{y}(z,n) - c_f, \tag{425}$$

where $\tilde{y}(z,n)$ is homogeneous of degree one in (z,n).

This assumption largely defines the units of productivity z and nests almost all the applications considered in the literature. Note that by the homogeneity of $\tilde{y}(z,n)$, we can write $\tilde{y}(z,n) = z\hat{y}(\hat{n})$, where $\hat{y}(\hat{n}) \equiv \tilde{y}(1,\hat{n})$ and $\hat{n} \equiv n/z$.

We restrict the diffusion process as follows.

Assumption 2. *The drift and volatility of the diffusion process satisfy the following conditions. There exists $\bar{\mu} < \varkappa$ such that*

$$\mu(z) < \bar{\mu} z \quad \textit{for all } z, \tag{426}$$

and

$$\lim_{z\to\infty} \frac{1}{z}\mu(z) = \mu, \qquad \lim_{z\to\infty} \frac{1}{z}\sigma(z) = \sigma, \tag{427}$$

with $|\mu| < \infty$ and $|\sigma| < \infty$.

The first condition ensures the existence of a stationary distribution. The second condition nests many commonly used productivity processes such as geometric Brownian motion and Ornstein-Uhlenbeck process.

Assumptions 1 and 2 are useful to bound the firing region. In general, the firing threshold is given by $n^f(z)$ such that $S_{n,\hat{G}}(z, n^f(z)) = 0$. The following lemma characterizes the firing threshold as $z \to \infty$.

Lemma 2. *Suppose Assumptions 1 and 2 hold. Then, as $z \to \infty$, the firing threshold $n^f(z)$ is given by $n^f(z) \to z\hat{n}^f$ for some $\hat{n}^f$ that does not depend on z.*

Proof. As $z \to \infty$, we have $S(z,n) \to z\hat{S}(\hat{n})$, where $\hat{S}(\hat{n})$ solves

$$\hat{\rho}\,\hat{S}(\hat{n}) = \hat{y}(\hat{n}) - c(\hat{v}(S_n))\hat{n} - b\hat{n} - \lambda^U \mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n)]\hat{n} - \lambda^F \hat{G}(S_n)\hat{V}(S_n)\hat{v}(S_n)\hat{n} \tag{428}$$

$$+\lambda^E\big(1-\hat{F}(S_n)\big)\mathbb{E}_{\hat{F}}[\hat{V}(\tilde{S}_n)|\tilde{S}_n \geq S_n]\hat{n} \tag{429}$$

$$+\frac{\sigma^2\hat{n}^2}{2}\hat{S}''(\hat{n}) + \Big[\lambda^F\hat{G}(S_n)\hat{v}(S_n) - s - \lambda^E(1-\hat{F}(S_n)) - \mu\Big]\,\hat{n}\,\hat{S}'(\hat{n}), \tag{430}$$

where $\hat{\rho} \equiv r + \varkappa - \mu$ with boundary conditions $\hat{S}(\hat{n}) \geq 0$ and $\hat{S}'(\hat{n}) \geq 0$. Note that $\hat{S}'(\hat{n}) = S_n$. Consequently, the firing threshold is given by $n^f(z) = z\hat{n}^f$ such that $\hat{S}'(\hat{n}^f) = 0$, as $z \to \infty$. □

Lemma 3. *Suppose Assumptions 1 and 2 hold. Then, $S_{\hat{G}}(z,n)/z$ is uniformly bounded for all (z,n) and $\hat{G}$.*

Proof. Let

$$y^{\sup}(z) \equiv \sup_{n\in[0,n^f(z)]} \tilde{y}(z,n) \tag{431}$$

Note that since $n^f(z) \to z\hat{n}^f$ as $z \to \infty$ from Lemma 2, we have

$$\lim_{z\to\infty} y^{\sup}(z)/z = \hat{y}(\hat{n}^f). \tag{432}$$

As $y^{\sup}(z)$ is asymptotically linear in z, there exists $C_y < \infty$ such that

$$y^{\sup}(z) \leq C_y z. \tag{433}$$

Note that the surplus can be bounded above by the value of a frictionless firm who always chooses the best feasible employment below the threshold:

$$S_{\hat{G}}(z,n) \leq \mathbb{E}_z\left[\int_0^\infty e^{-\rho t} y^{\sup}(z)dt\right] \tag{434}$$

$$\leq C_y \int_0^\infty e^{-\rho t}\mathbb{E}_z\left[z\right]dt \tag{435}$$

$$\leq \frac{C_y}{\rho - \bar{\mu}} z, \tag{436}$$

which is finite since $\rho = r + \varkappa > \bar{\mu}$. □

We now impose an additional regularity assumption so as to ensure positive employment.

Assumption 3. *Assume the entry distribution ψ satisfies the following condition. There exists $\underline{N} >$*

0 *such that, for every* $\hat{G}$*,*

$$\mathbb{E}\left[\int_0^{\tau_f} e^{-(s+\varkappa+\lambda^E(\theta))t} n_0\, dt\right] \geq \underline{N} > 0, \tag{437}$$

where τ_f *is the stopping time at which the initial cohort of workers is first fired.*

This assumption ensures that the workers brought in by entrants stay employed for a strictly positive amount of time. It is satisfied, for example, if entrants are sufficiently small or productive that they enter strictly inside the operating region. To see that it delivers the desired lower bound, note that the unit-entry employment is the occupation measure $\int n\, d\phi^0_{\hat{G}} = \mathbb{E}\big[\int_0^\infty e^{-\varkappa t} n(t)\, dt\big]$. Before the first firing time τ_f the initial cohort is reduced only by separation (rate s) and poaching (rate $\lambda^E(\theta)(1-\hat{F}_{\hat{G}}) \leq \lambda^E(\theta)$), so $n(t) \geq n_0 e^{-(s+\lambda^E(\theta))t}$; firing only reduces n further but is excluded on $[0, \tau_f)$. Hence

$$\int n\, d\phi^0_{\hat{G}} \;\geq\; \mathbb{E}\left[\int_0^{\tau_f} e^{-(s+\varkappa+\lambda^E(\theta))t} n_0\, dt\right] \;\geq\; \underline{N}. \tag{438}$$

Lemma 4. *Suppose Assumptions 1, 2, and 3 hold. Then,* $\mathcal{T}_\theta(\hat{G})$ *has uniformly bounded first moment:*

$$\int s\, d\mathcal{T}_\theta(\hat{G})(s) \leq \bar{S}, \tag{439}$$

Proof. By concavity of $S_{\hat{G}}(z, \cdot)$,

$$S_{\hat{G}}(z,n) - S_{\hat{G}}(z,0) \geq n S_{n,\hat{G}}(z,n) \tag{440}$$

Since $S_{\hat{G}}(z, 0) \geq 0$ on the operating region,

$$n S_{n,\hat{G}}(z,n) \leq S_{\hat{G}}(z,n). \tag{441}$$

Therefore

$$\begin{aligned}
\int s\, d\mathcal{T}_\theta(\hat{G})(s) &= \frac{\zeta m^0_{\hat{G}} \int n S_{n,\hat{G}}(z,n)\, d\phi^0_{\hat{G}}(z,n)}{u_{\hat{G}} + \zeta\, m^0_{\hat{G}} \int n\, d\phi^0_{\hat{G}}(z,n)} && (442)\\
&\leq \frac{\zeta m^0_{\hat{G}} \int S_{\hat{G}}(z,n)\, d\phi^0_{\hat{G}}(z,n)}{u_{\hat{G}} + \zeta\, m^0_{\hat{G}} \int n\, d\phi^0_{\hat{G}}(z,n)} && (443)\\
&\leq \frac{C_y}{\rho - \bar{\mu}} \frac{\zeta m^0_{\hat{G}} \int z\, d\phi^0_{\hat{G}}(z,n)}{u_{\hat{G}} + \zeta\, m^0_{\hat{G}} \int n\, d\phi^0_{\hat{G}}(z,n)} && (444)\\
&\leq \frac{C_y}{\rho - \bar{\mu}} m^0_{\hat{G}} \int z\, d\phi^0_{\hat{G}}(z,n), && (445)
\end{aligned}$$

where (443) uses $nS_{n,\hat{G}}(z,n) \leq S_{\hat{G}}(z,n)$ on the operating region, (444) uses Lemma 3, and the last inequality uses the identity

$$u_{\hat{G}} + \zeta m^0_{\hat{G}} \int n\, d\phi^0_{\hat{G}} = \tilde{u}_{\hat{G}} = u_{\hat{G}} + \zeta(1-u_{\hat{G}}) \geq \zeta, \tag{446}$$

which holds for any admissible allocation and requires no lower bound on the endogenous unemployment rate. Two primitive facts then close the bound. First, because productivity diffuses exogenously and entrants draw z_0 from ψ, the z-marginal of $\phi^0_{\hat{G}}$ solves a Kolmogorov forward equation that does not involve $\hat{G}$; by Assumption 2 it has a finite, $\hat{G}$-independent mean $\int z\, d\phi^0_{\hat{G}} = \bar{Z} < \infty$. Second, the entry mass is uniformly bounded, $m^0_{\hat{G}} \leq \bar{m} \equiv 1/\underline{N}$: since $m^0_{\hat{G}} \int n\, d\phi^0_{\hat{G}} = 1 - u_{\hat{G}} \leq 1$ and cohort employment is bounded away from zero, $\int n\, d\phi^0_{\hat{G}} \geq \underline{N} > 0$ by Assumption 3. Hence

$$\int s\, d\mathcal{T}_\theta(\hat{G})(s) \;\leq\; \frac{C_y\, \bar{m}\, \bar{Z}}{\rho - \bar{\mu}} \;\equiv\; \bar{S}, \tag{447}$$

a constant independent of $\hat{G}$. □

Define

$$\mathcal{D} = \left\{ \hat{G} \in \mathcal{P}(\mathbb{R}_+) : \hat{G}(\{0\}) \geq \underline{u},\ \int s\, d\hat{G}(s) \leq \bar{S} \right\}.$$

The condition $\hat{G}(\{0\}) \geq \underline{u}$ keeps the denominator in $\hat{V}_{\hat{G}}$ away from zero. The moment bound gives tightness.

Lemma 5 (Compactness of the admissible-ranking set)**.** *The set $\mathcal{D}$ is nonempty, convex, and compact under weak convergence.*

Proof. Convexity is immediate because each defining restriction is linear in the measure. The set is nonempty. The moment bound implies tightness. Indeed, for any $R > 0$,

$$\hat{G}([R,\infty)) \leq \frac{1}{R} \int s\, d\hat{G}(s) \leq \frac{\bar{S}}{R}.$$

Therefore $\mathcal{D}$ is relatively compact by Prokhorov's theorem.

It remains to check closedness. If $\hat{G}_k \Rightarrow \hat{G}$, then by the Portmanteau theorem, since $\{0\}$ is closed,

$$\hat{G}(\{0\}) \geq \limsup_k \hat{G}_k(\{0\}) \geq \underline{u}.$$

Finally, because s is nonnegative and lower semicontinuous,

$$\int s\, d\hat{G}(s) \leq \liminf_k \int s\, d\hat{G}_k(s) \leq \bar{S}.$$

Thus $\mathcal{D}$ is weakly closed and relatively compact, hence compact. $\square$

Lemma 6. *Suppose Assumptions 1, 2, and 3 hold. Then $\mathcal{T}_\theta : \mathcal{D} \to \mathcal{D}$.*

Proof. First, $\mathcal{T}_\theta(\hat{G})$ is a probability measure on $\mathbb{R}_+$. By construction of the labor-market clearing (421), $u_{\hat{G}} \geq \underline{u}$, so

$$\mathcal{T}_\theta(\hat{G})(\{0\}) \geq \frac{u_{\hat{G}}}{\tilde{u}_{\hat{G}}} \geq u_{\hat{G}} \geq \underline{u}, \tag{448}$$

since $\tilde{u}_{\hat{G}} \leq 1$. Lemma 4 gives the moment bound $\int s \, d\mathcal{T}_\theta(\hat{G})(s) \leq \bar{S}$. Therefore $\mathcal{T}_\theta(\hat{G}) \in \mathcal{D}$. $\square$

Assumption 4 (Continuity of the ranking map)**.** *For each fixed θ, the map $\mathcal{T}_\theta$ defined in Definition 1 is continuous on $\mathcal{D}$ under weak convergence: if $\hat{G}_k \Rightarrow \hat{G}$ in $\mathcal{D}$, then*

$$\mathcal{T}_\theta(\hat{G}_k) \Rightarrow \mathcal{T}_\theta(\hat{G}).$$

This assumption rules out nonunique HJB–KFE solutions, discontinuous movements of the firing frontier, atoms at moving marginal-surplus thresholds, vanishing vacancy normalizations, and tail behavior that is not controlled by weak convergence. While we view this assumption as effectively imposing sufficient smoothness in the underlying primitives and thus a weak restriction, it is unsatisfactory to directly impose a continuity assumption on the fixed point mapping. Finding the appropriate conditions on the primitives and formally proving continuity of the mapping was extremely challenging for us.

Claim 1. *Suppose Assumptions 1, 2, 3, and 4 hold. Then for a given θ, there exists $\hat{G}^* \in \mathcal{D}$ such that*

$$\mathcal{T}_\theta(\hat{G}^*) = \hat{G}^*.$$

Proof. Fix θ. By Lemma 5, $\mathcal{D}$ is a nonempty, convex, compact subset of the space of finite signed measures on $\mathbb{R}_+$ endowed with the weak topology. By Lemma 6, $\mathcal{T}_\theta : \mathcal{D} \to \mathcal{D}$, and by Assumption 4, $\mathcal{T}_\theta$ is continuous.

Therefore all hypotheses of the Schauder–Tychonoff fixed-point theorem are satisfied. Hence there exists $\hat{G}^* \in \mathcal{D}$ such that

$$\mathcal{T}_\theta(\hat{G}^*) = \hat{G}^*.$$

$\square$

References

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "Income and wealth distribution in macroeconomics: A continuous-time approach," *The review of economic studies*, 2022, *89* (1), 45–86.

Bilal, Adrien, "Solving heterogeneous agent models with the master equation," *Working Paper*, 2023.

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

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

Elsby, Michael WL and Axel Gottfries, "Firm dynamics, on-the-job search, and labor market fluctuations," *The Review of Economic Studies*, 2022, *89* (3), 1370–1419.

Hopenhayn, Hugo, "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica: Journal of the Econometric Society*, 1992, pp. 1127–1150.

— **and Richard Rogerson**, "Job turnover and policy evaluation: A general equilibrium analysis," *Journal of political Economy*, 1993, *101* (5), 915–938.