Sorting along Business Cycles*

Paweł Gola[†] Haozhou Tang[‡]

February 14, 2026

Abstract

We develop an analytically tractable model featuring heterogeneous workers and firms, where labor markets clear through a one-to-many sorting mechanism. Firms determine both the number and composition of their employees, shaping (1) the income distribution among workers and (2) the productivity distribution across firms. We study business cycles driven by market efficiency shocks that disproportionately benefit more productive firms. The model's implications are consistent with empirical regularities on the cyclical behavior of wage and productivity distributions.

Keywords: sorting, wage inequality, productivity, business cycles
JEL Classification: C78, J21, D24, D31, E32

*We thank seminar and conference participants at Dongbei University of Economics and Finance, and Liaoning University. Gola gratefully acknowledges the support of the UKRI Horizon Europe Guarantee (Grant Ref: EP/Y028074/1). The usual disclaim applies.

[†]University of Edinburgh, pawel.gola at ed.ac.uk
[‡]Dongbei University of Finance and Economics, haozhoutang@dufe.edu.cn.

1 Introduction

A prevailing convention in business-cycle studies is to treat firm-level productivity as exogenous, determined outside the firm's recruitment decisions. Firms are typically modeled as passive takers of their productivity draws, making hiring and investment decisions conditional on those primitives. Yet this view abstracts from a fundamental aspect of how firms actually operate: productivity is shaped by whom firms hire. A firm's capabilities hinge on the talents it brings in—whether a CEO who sets strategic direction, a CFO who manages financial structure, R&D personnel who drive innovation, or sales teams who determine market reach. In many cases, hiring decisions are not responses to productivity; they are determinants of it. This perspective highlights that productivity is an equilibrium outcome of how heterogeneous firms and workers sort in the labor market.

Recognizing productivity as an equilibrium sorting outcome raises natural questions: how are the distribution of wages and firm-level productivity connected, and how do they evolve over the business cycle? A salient feature of the productivity distribution is that its dispersion is countercyclical—widening in recessions and narrowing in booms.[1] This variation is typically taken as exogenous, often attributed to uncertainty or volatility shocks. In contrast, wage inequality, measured as the variance of log wages, exhibits a distinct cyclical pattern. As shown in Figure 1, the left panel reveals a long-run upward trend in wage dispersion from Song et al. (2019), along with clear cyclical movements: wage inequality tends to rise in economic expansions and contract in recessions, highlighted by the NBER-dated shaded regions. The right panel, which plots detrended wage inequality and detrended GDP, makes the comovement more transparent—inequality tends to track the business cycle. Notably, the patterns differ across episodes. For instance, in the late 1990s, this comovement was less pronounced: wage inequality remained below its own trend despite the growth of real GDP. By contrast, wage inequality dropped sharply during the Great Recession.[2]

Overall, the data suggest a potential contrast between the two dispersions: while the

[1]See, for example, Bachmann and Bayer (2014), Kehrig (2015), Bloom et al. (2018), and Cunningham et al. (2023).

[2]Throughout this paper, we focus on wage inequality rather than income inequality, which includes all sources of income. The figure uses wage income, i.e., labor earnings, and thus only accounts for employed individuals, whereas income inequality also includes the unemployed—potentially leading to very different patterns. Additionally, different definitions and measures of income inequality can produce substantially different results. See, for example, Piketty and Saez (2003) and Heathcote et al. (2020).

dispersion of wages—reflecting worker performance—tend to increase in expansions, the dispersion of firm-level productivity—reflecting firm performance—tends to rise in downturns. These patterns highlight an important feature for understanding firm-worker interactions over the business cycle.

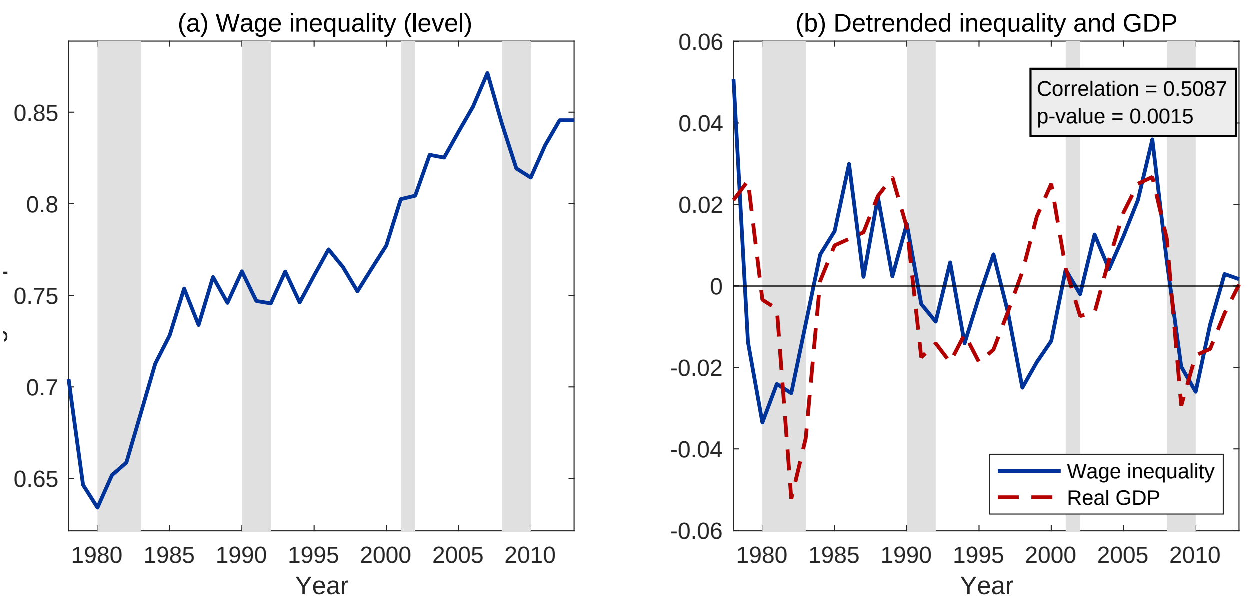

Figure 1: Wage inequality over business cycles

Notes: Shaded areas indicate NBER recessions. Wage inequality is measured by variance of individual log wages. Both series in panel (b) are detrended.
Source: Song et al. (2019) and authors' calculations.

Motivated by these patterns, we introduce a framework with heterogeneous firms and workers in which labor markets clear through a one-to-many matching mechanism. Firms differ in their underlying types, and workers differ in their skill levels. Firms choose both the number and type of their employees, shaping the distributions of wages and productivity, which exhibit Pareto tails in equilibrium. Apart from the introduction of worker heterogeneity and one-to-many sorting, the model is entirely standard, featuring CES demand and decreasing-returns Cobb-Douglas production technology. Crucially, productivity emerges from the equilibrium sorting between firms and workers rather than being imposed exogenously. In this sense, the model generalizes the standard heterogeneous-firm framework: without worker heterogeneity or sorting, it reduces to the canonical setting where productivity is given.

Sorting between firms and workers is sensitive to aggregate conditions: changes in the macro environment can shift the composition of matches and thus alter the distributions of wages and productivity. We introduce a reduced-form market efficiency shock that operates through time-varying firm-specific wedges on labor and capital prices, representing distortions that fluctuate over the business cycle. This reduced-form shock may capture the compound effects of several underlying structural shocks and disproportionately affects high-type firms. It is designed to generate business-cycle fluctuations that reproduce procyclical reallocation of labor to more advanced firms, a key cyclical pattern observed in the data.[3]

We analytically characterize how the reduced-form market efficiency shocks shape the cross-sectional distributions of wages and firm-level productivity. A positive shock induces high-type firms to expand more aggressively, shifting the composition of jobs toward higher-quality positions and thereby widening wage dispersion. Simultaneously, due to this relative expansion, these firms are matched with relatively lower-type workers, pulling their realized productivity closer to that of low-type firms, which shrink in relative size, and thus dampening the dispersion of firm-level productivity. This mechanism demonstrates how shifts in the allocation of talent across firms can generate opposing cyclical movements in wages and productivity, even when underlying technologies are unchanged. It is consistent with the findings in Crane et al. (2023), who document that during downturns high-rank workers are more likely to work at low-rank firms, while low-rank workers are less likely to be employed by high-rank firms.

While the mechanism described above can generate opposing cyclical movements in wage and productivity dispersions, not necessarily all recessions exhibit this disparity. To understand the broader range of patterns, we also examine the effects of several structural shocks. In particular, we consider aggregate productivity shocks and second-moment shocks that alter the variance of firm types and firm-level distortions. This allows us to analyze how different shocks shape the joint dynamics of wage and productivity dispersions over the business cycle.

Literature review The main conceptual contribution of this work is the finding that, under reasonable assumptions, changes in the matching function have opposing effects on the distributions of wages and measured productivity. In particular, the result that improvements in workers' matches decrease the inequality of measured productivity is, to the best of our knowledge, completely novel. Combined with the well-known fact that,

[3] See, for example, Moscarini and Postel-Vinay (2012), Haltiwanger et al. (2018), Crane et al. (2023), and Haltiwanger et al. (2025).

in a model with worker-firm type complementarities, improvements in workers matches increase wage inequality (Costinot and Vogel, 2010), this implies our main result: cyclical fluctuations in sorting cause the cyclical fluctuations of wage inequality and the countercyclical behavior of productivity dispersion.

Our contributions to the literature on labour market sorting extend beyond these novel comparative statics. First, while frictionless one-to-many sorting models (see Eeckhout and Kircher (2018) for the most comprehensive theoretical treatment) have been used to study a wide array of applied questions in labour economics (e.g., Sattinger, 1975; Teulings, 1995, 2005) and international trade (e.g., Costinot, 2009; Costinot and Vogel, 2010; Sampson, 2014; Grossman et al., 2017; Choi, 2023), we are the first to embed this framework into an otherwise canonical business cycle model with aggregate uncertainty and capital accumulation. Second, we develop a novel, tractable specification of the one-to-many sorting model, which admits a fully analytical solution to the sorting problem. While this specification draws inspiration from Gola and Zhao (2023), notably by adopting a similar production function, it improves on their approach considerably by allowing for exponentially distributed firm and worker types, which is precisely what produces Pareto tails of wage, productivity, and firm size distributions in equilibrium.

Finally, our paper contributes to the literature on heterogeneous firms in several ways. First, unlike canonical models that treat firm-level productivity as exogenous, we study the endogeneity of productivity through equilibrium sorting between heterogeneous firms and workers. This mechanism generates a rich mapping from firm-worker matches to the cross-sectional distributions of productivity and wages. In particular, it offers a novel interpretation of uncertainty or volatility shocks. While pioneering studies such as Bloom (2009), Bachmann and Bayer (2013, 2014), Schaal (2017), Bloom et al. (2018), and Arellano et al. (2019) introduce these shocks into heterogeneous-firm frameworks as exogenous drivers of fluctuations in productivity dispersion and aggregate dynamics, we show they can be partially reconciled as an endogenous outcome of time-varying worker-firm sorting. Second, by incorporating worker heterogeneity, we extend the analysis to wage inequality and its cyclical behavior. Finally, the model remains highly tractable, allowing for analytical characterization of a broad range of cross-sectional and business-cycle implications.

The paper is organized as follows. Section 2 introduces the model, and Section 3 presents the equilibrium analysis. In Section 4, we show how aggregate conditions shape the cross-sectional distributions of wages and productivity. Section 5 discusses alternative drivers of business cycles, and Section 6 concludes.

2 Model

In this section, we develop a model that incorporates one-to-many sorting between heterogeneous firms and workers. Firms make decisions regarding both the quantity of labor, i.e., the number of workers to hire, and the quality of labor, i.e., the skill level or type of workers. In this framework, firm-level productivity is determined by not only firms' own type, but also their match with employee types. We demonstrate that our model nests a standard heterogeneous-firm model with exogenous idiosyncratic productivity.

Each firm is subject to firm-specific distortions, represented by labor and capital wedges. The distortions are i.i.d. across firms but are correlated with firm types. Dynamically, we introduce shocks to market allocative efficiency, modeled as a time-varying correlation between firm types and firm-specific distortions.

2.1 Household

Time is discrete and runs from zero to infinity. A representative household consists of a unit continuum of members, indexed by $i \in [0,1]$, each endowed with one unit of labor and skill type follows exponential distribution $x_i \sim \text{Exp}(\lambda_x)$. The household maximizes lifetime utility function

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \log C_t$$

where C_t denotes the aggregate consumption, β represents the discount factor. The household owns all firms, as well as capital and labor inputs. It is subject to the following budget constraint:

$$C_t + K_{t+1} - (1-\delta)K_t = \int_0^1 w_{it} di + R_t K_t + D_t, \tag{1}$$

where K_t denotes the aggregate capital stock, δ represents its depreciation rate, w_{it} denotes the wage received by household member i, R_t represents the capital price, and D_t represents operation profits from firms. Each family member works independently to earn wages, but their income is pooled together when making decisions about consumption and savings.

2.2 Production

Final good producer There is a competitive sector of final good producers, all of which produce the homogeneous numeraire good by combining differentiated intermediate goods $j \in [0,1]$

$$Y_t = \left(\int_0^1 Y_{jt}^{1-\frac{1}{\xi}} dj \right)^{\frac{\xi}{\xi-1}}$$

The price of final good is normalized to one.

Intermediate good producers There is a unit continuum of intermediate firms. Production function follows[4]

$$Q_t(l,k,x,\theta) = A_t q(x,\theta) k^{\alpha} l^{\gamma}, \tag{2}$$

where θ and x respectively denote the type of the firm and the type of worker that is matched with the firm, k and l represent the capital and labor input. Firm type θ follows an exponential distribution $\theta \sim \text{Exp}(\lambda_\theta)$. We assume that $\alpha + \gamma < 1$. A_t denotes a common productivity component that is identical across every firm. The idiosyncratic productivity component $q(x,\theta)$ follows

$$q(x,\theta) = \exp\left\{ x^{\psi} \theta^{1-\psi} \right\}, \tag{3}$$

where $\psi \in [0,1]$. Now, productivity depends not only on the firm's own type θ but also on the type of workers x it employs. As discussed below, aggregate conditions affect the firm's hiring decisions, which in turn shape productivity.

Later we will discuss the economic interpretations for each of the functional parameters. For now, it is worth noting that a standard Cobb-Doulas function can be viewed as a special case of production function (2). Specifically, if $\psi = 0$, idiosyncratic productivity is equal to e^{θ}, independent of worker types. In this case, the production function (2) is mathematically equivalent to a standard Cobb-Douglas function with decreasing-returns-to-scale (henceforth DRS): $e^{\theta} k^{\alpha} l^{\gamma}$, where e^{θ} can be interpreted as firm-level idiosyncratic productivity and is itself Pareto distributed. Work type x does not matter for the production, and all workers receive the same wage: no wage heterogeneity, just as in a standard heterogeneous-firm model.

[4]For notation ease, we omit time and firm-specific subscripts from the functional inputs in the following analysis.

Distortions and market efficiency shcoks Intermediate producers face demand schedule

$$Y_{jt} = (P_{jt})^{-\xi} Y_t. \tag{4}$$

At time t, firms compete in the labor market by offering wage $w_t(x)$ to workers of type x. We follow Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) by incorporating market frictions in the form of reduced-form, firm-specific wedges in capital and labor prices. These wedges introduce distortions that capture inefficiencies in resource allocation across firms. Given demand y, firms minimize cost by solving the following Lagrangian problem

$$\min_{l,k,m} \tau_1 w_t(m) l + \tau_2 R_t k - \chi \left[Q_t(l,k,m,\theta) - y \right]. \tag{5}$$

The firm-specific wedges τ_1 and τ_2 represent distortions arising from market frictions. Lagrangian multiplier χ denotes the marginal cost.

Throughout the paper, we focus on labor market frictions and fluctuations in labor market efficiency. Specifically, we assume that labor wedge is correlated with firm type and follows

$$\tau_1 = \exp(z_t \theta + \epsilon_1). \tag{6}$$

where z_t is constant across firms, while $\epsilon_1 \sim N\left(0, \sigma_1^2\right)$ represents a firm-specific component that is i.i.d. across firms. Thus, z reflects the relationship between the (log of) distortion τ_1 and firm type θ. In terms of capital market, we assume that capital wedge follows $\tau_2 = \exp(\epsilon_2)$, where $\epsilon_2 \sim N\left(0, \sigma_2^2\right)$ is i.i.d. across firms. For simplicity, we omit the correlation between capital wedges and firm types. However, the analysis can be readily extended to include type-dependent capital market distortions by modeling capital wedges, similar to (6), as being correlated with firm types.

Throughout the paper, we assume that $z_t > 0$, $\forall t$. This assumption has two considerations. First, as discussed in Section 3.2, it implies that larger firms tend to allocate a smaller share of their revenue to wages, a pattern that aligns with well-documented empirical findings.[5] Second, this assumption is consistent with empirical evidence on the correlation between distortions and firm size and productivity, which shows that high-

[5] See Autor et al. (2020).

productivity firms tend to face higher marginal input costs and distortions.[6] In particular, Bento and Restuccia (2017) estimate the elasticity of distortions with respect to productivity across countries and find values ranging from 0.22 to 0.74.

We follow Chari et al. (2007) by introducing time-varying labor wedges into the model. Specifically, we assume z_t follows a Markov process. Fluctuations in z_t affect firms asymmetrically depending on their type θ, with higher-θ firms being more exposed to these shocks. We intentionally refrain from microfounding this reduced-form shock; instead, one may interpret z_t as capturing the combined effect of multiple underlying structural forces that vary over the business cycle.

This modeling choice is consistent with empirical evidence showing that labor reallocation toward more productive firms is procyclical. Recent studies by Crane et al. (2023) and Haltiwanger et al. (2025), using a variety of firm-level productivity measures, document that during expansions high-type firms expand relative to low-type firms by recruiting workers who would otherwise be employed by low-type firms. Firm size can also serve as a proxy for productivity: Moscarini and Postel-Vinay (2012) documents that employment growth in large firms is more strongly correlated with aggregate economic activity than growth in smaller firms. Furthermore, some evidence suggests that these larger, more productive firms are more sensitive to structural shocks, such as monetary policy shocks.[7] Finally, Buera and Moll (2015) develop models showing that credit crunches can be manifested as time-varying input wedges and cause the labor share of the most productive firms to shrink.

Note that the purpose of this paper is not to identify the structural sources of aggregate fluctuations. Rather, we examine what happens when better firms expand more during booms—specifically, how such asymmetric responses shape the distribution of wages and firm-level outcomes over the business cycle.

3 Equilibrium Analysis

An equilibrium consists of: (1) a consumption (and saving) path that maximizes the representative household's lifetime utility and satisfies its intertemporal budget constraint;

[6] See, for example, Hsieh and Klenow (2014), Gourio and Roys (2014), Garicano et al. (2016), and Bento and Restuccia (2017).

[7] See, for example, Ottonello and Winberry (2020), Morlacco and Zeke (2021), and Kroen et al. (2021). However, this evidence is subject to debate, as some studies document opposite effects; for instance, Gertler and Gilchrist (1994) and Crouzet and Mehrotra (2020).

2) optimal capital inputs k_{jt}, labor quantities l_{jt}, and labor types x_{jt} that maximize each firm j's profits; and (3) prices for intermediate goods P_{jt}, the rental rate of capital, R_t, and wage schedule for each skill type $w_t(x)$ that clear markets for all t.

To sum up, the equilibrium definition above is analogous to that in a standard stochastic neoclassical growth model with heterogeneous firms, except that we incorporate one-to-many sorting: firms' recruiting decisions now involve both the quantity and the quality of labor. Labor market clearing requires that the supply of each skill type equals its demand.

This section presents the equilibrium analysis. We defer the detailed analysis of cross-sectional distributions to Section 4.

3.1 One-to-many Sorting

We first define a *job offer* as $h = \theta$. The type of job offers is determined by the underlying firm type θ. However, we introduce a separate notation h to distinguish job offers from firm types, since the distribution of job offers generally differs from that of θ. In particular, the distribution of h also depends on firm size:

$$f_t(h) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l_t(h,\epsilon_1,\epsilon_2) * \lambda_\theta \exp(-\lambda_\theta h) \mathrm{d}\Phi(\frac{\epsilon_1}{\sigma_1}) \mathrm{d}\Phi(\frac{\epsilon_2}{\sigma_2})$$

where $f_t(h)$ denotes the probability density function (PDF) of h, and $l_t(\theta,\epsilon_1,\epsilon_2)$ represents firm size. It is immediate that h and θ may follow different distributions due to the incorporation of firm size.

Our framework thus features a *one-to-many* sorting structure, in which firms endogenously determine their size and post multiple job offers, in contrast to a one-to-one sorting setup where each firm corresponds to a single job. To see this more clearly, consider a conventional one-to-one sorting framework in which all firms have equal unit size, i.e., $l_t(\theta,\epsilon_1,\epsilon_2) = 1$. In that case, we have $f_t(h) = \lambda_\theta e^{-\lambda_\theta h}$, implying that h and θ follow the same distribution.

To facilitate the analysis, we conjecture that

$$f_t(h) = \lambda_t \exp(-\lambda_t h),$$

that is, the number of job offers follow an exponential distribution with a time-varying parameter λ_t. Throughout most of this section, we take this as given. In Section 3.3, we verify this conjecture and characterize the equilibrium value of λ_t.

The function $f_t(h)$ can also be interpreted as the *labor demand* generated by firms of type $\theta = h$. On the other hand, the supply of type-x workers is assumed to follow an exponential distribution, $x_i \sim \text{Exp}(\lambda_x)$. The corresponding labor supply function is

$$S(x) \equiv -\lambda_x \exp(-\lambda_x x).$$

Let $h = \mu_t(x)$ denote the matching function between a worker of type x and a job offer h. Labor market clearing requires that $\forall x \in [0, \infty)$

$$S(x) = D(x) \equiv -\lambda_t \mu_t'(x) \exp(-\lambda_t \mu_t(x)). \tag{7}$$

From (7) and that $\int_0^\infty D_t(x)dx = 1$, we obtain

$$\mu_t(x) = \frac{\lambda_x}{\lambda_t} x. \tag{8}$$

The matching function (8) reveals a positive assortative matching (PAM) pattern between workers and job offers (and thus firms): workers with higher x are matched with firms offering higher h (and equivalently higher θ).

3.2 Firms' problem

The firm level state variables can be summarized by $(\theta, \epsilon_1, \epsilon_2)$. Firm-level variables can be expressed as functions of these state variables. Demand schedule (4) can be rewritten into

$$Q_t(\theta, \epsilon_1, \epsilon_2) = (P_t(\theta, \epsilon_1, \epsilon_2))^{-\xi} Y_t, \tag{9}$$

where the price is a constant markup over the marginal cost

$$P_t(\theta, \epsilon_1, \epsilon_2) = \frac{\xi}{\xi - 1} \chi_t(\theta, \epsilon_1, \epsilon_2). \tag{10}$$

Next we proceed to analyze the input choices of intermediate firms.

First-order conditions and wage schedule As to intermediate firms, the firm's first-order condition (FOC) with respect to labor input l is given by

$$\chi A_t \gamma k^\alpha l^{\gamma-1} \exp(x^\psi h^{1-\psi}) = \tau_1 w(x), \tag{11}$$

where χ denotes the associated marginal cost. Note that equation (11) implies that the labor share of a company is equal to $\frac{\xi-1}{\xi\tau_1}$, which is *negatively* correlated with the type of firm θ.

Similarly, the FOC with respect to worker type x yields

$$\chi A_t \psi k^\alpha l^\gamma \exp(x^\psi h^{1-\psi}) x^{\psi-1} h^{1-\psi} = \tau_1 w_t'(x) l. \tag{12}$$

Dividing the second condition by (11) gives

$$\frac{w_t'(x)}{w_t(x)} = \frac{\psi}{\gamma} x^{\psi-1} h^{1-\psi}, \tag{13}$$

Applying the fundamental theorem of calculus, we obtain

$$w_t(x) = w_t(0) \exp(\frac{\psi}{\gamma}(\frac{\lambda_x}{\lambda_t})^{1-\psi} x) \tag{14}$$

Taking logarithms gives

$$\ln(w_t(x)) = \frac{\psi}{\gamma}(\frac{\lambda_x}{\lambda_t})^{1-\psi} x + \ln(w_t(0)), \tag{15}$$

a result that reveals log wages follow an exponential distribution—or, equivalently, that raw wage levels conform to a Pareto distribution.

Equation (15) characterizes how worker types are mapped to their wages. Workers with higher type θ consistently receive higher wages. The steepness of this wage-worker type relationship hinges on two components, each capturing distinct economic tradeoffs.

First, the gradient increases with the ratio $\frac{\psi}{\gamma}$, a ratio determined by production technology. Here, ψ quantifies the *skill intensity* of worker contributions to firm productivity. Whereas γ reflects *scale intensity*, i.e. gains from expanding the workforce. Their ratio encapsulates the core tradeoff firms face: prioritizing worker quality (skill) versus quantity (headcount). A high $\frac{\psi}{\gamma}$ ratio signals that quality dominates this tradeoff, forcing firms to compete more fiercely for high-type workers by offering a steeper wage premium tied to worker ability.

Second, the gradient is increasing in $(\frac{\lambda_x}{\lambda_t})^{1-\psi}$, a term that reflects the relative thickness of the job-supply and job-demand tails. The parameter λ_t governs the tail thinness of the labor-demand distribution—smaller values imply a thicker upper tail and thus a larger mass of high-type job offers, while λ_x plays the analogous role for labor supply, capturing the prevalence of high-skill workers. The ratio $\frac{\lambda_x}{\lambda_t}$ thus measures the relative scarcity of high-skill labor compared with high-type job offers: a higher value means firms face fiercer competition for top talent, pushing the wage gradient steeper to attract skilled workers.

From equilibrium matching (8), we derive the sorting relationship that links worker type to firm type:

$$h = \mu_t(x) = \frac{\lambda_x}{\lambda_t} x. \tag{16}$$

Combining (15) with (16), equation (15) can be equivalently rewritten as

$$\ln(w_t(\mu_t^{-1}(h))) = \frac{\psi}{\gamma}(\frac{\lambda_t}{\lambda_x})^{\psi} h + \ln(w_t(0)), \tag{17}$$

which characterizes how wages vary across different firms.

It is informative to examine firms' problem given capital stock k and wedge τ_1. Substituting (17) into (11) yields the optimal labor input:

$$l_t^*(\theta, k, \chi, \tau_1) = [\frac{w_t(0)\tau_1}{\gamma \chi A_t k^{\alpha}}]^{\frac{1}{\gamma-1}} \exp(\frac{\gamma-\psi}{\gamma(1-\gamma)}(\frac{\lambda_t}{\lambda_x})^{\psi}\theta) \tag{18}$$

Note that in our model k and τ_1 depend on state variables $(\theta, \epsilon_1, \epsilon_2)$. However, (18) applies also to a scenario where firm-level capital stock and distortions are prefixed. The difference $\gamma - \psi$ captures the trade-off between worker quality and quantity. If $\gamma < \psi$, the demand for quality overwhelms the demand for scale: The wage schedule (15) is so steep that firms must shrink their size to afford higher worker quality, holding k and τ_1 fixed. In this case, high-type firms optimally operate with a small team of elite workers. This stands in contrast to a standard framework without worker heterogeneity (a special case of our model with $\psi = 0$), where higher-type firms always hire more workers.

With (17) and firms' optimality conditions, we can readily express our variables of interests as functions of firm-level state variables.

Lemma 3.1. *Suppose that in period t jobs h are exponentially distributed with parameter λ_t, that is $h \sim Exp(\lambda_t)$. Then the equilibrium output Q, rented capital k, marginal cost χ and labor hired l at firm of type $(\theta, \epsilon_1, \epsilon_2)$ are as follows:*

$$Q_t(\theta,\epsilon_1,\epsilon_2) = \bar{Q}_t \exp\left(\eta^Q\left(\eta^Q_{\theta t}\theta - \gamma\epsilon_1 - \alpha\epsilon_2\right)\right), \tag{19}$$

$$k_t(\theta,\epsilon_1,\epsilon_2) = \bar{k}_t \exp\left(\frac{\xi-1}{\xi}\eta^Q\left(\eta^Q_{\theta t}\theta - \gamma\epsilon_1 - (\alpha + \frac{\xi}{\xi-1}\frac{1}{\eta^Q})\epsilon_2\right)\right), \tag{20}$$

$$\chi_t(\theta,\epsilon_1,\epsilon_2) = \bar{\chi}_t \exp\left(-\frac{\eta^Q}{\xi}\left(\eta^Q_{\theta t}\theta - \gamma\epsilon_1 - \alpha\epsilon_2\right)\right), \tag{21}$$

$$l_t(\theta,\epsilon_1,\epsilon_2) = \bar{l}_t \exp\left(\eta^l_{\theta t}\theta - \left(\frac{\xi-1}{\xi}\eta^Q\gamma + 1\right)\epsilon_1 - \frac{\xi-1}{\xi}\eta^Q\alpha\epsilon_2\right), \tag{22}$$

where

$$\eta^Q \equiv \frac{\xi}{1+(1-\alpha-\gamma)(\xi-1)}, \quad \eta^Q_{\theta t} \equiv -\gamma z_t + (1-\psi)(\frac{\lambda_t}{\lambda_x})^{\psi}, \quad \eta^l_{\theta t} \equiv \frac{\xi-1}{\xi}\eta^Q\eta^Q_{\theta t} - z_t - \frac{\psi}{\gamma}(\frac{\lambda_t}{\lambda_x})^{\psi}.$$

The time-dependent but type-independent constants $\bar{Q}_t, \bar{k}_t, \bar{\chi}_t, \bar{l}_t$ are defined in the proof of Lemma 3.1 (Appendix A.1).

In equilibrium, if $\epsilon_1 = \epsilon_2 = 0$, the firm-level variables above, such as firm size l, follow a Pareto distribution, since the underlying firm type θ is exponentially distributed. When ϵ_1 and ϵ_2 are drawn from the assumed zero-mean normal distributions with nonzero variance, these variables follow a convolution of a Pareto distribution and a log-normal distribution. Importantly, because the log-normal component decays faster in the upper tail than the Pareto component, the resulting distributions retain Pareto tails, consistent with the fat-tailed firm size distributions observed in the data.

3.3 Equilibrium Job Distribution

Next we proceed to characterize λ_t. The PDF of h follows

$$
\begin{aligned}
f_t(h) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} l_t(h,\epsilon_1,\epsilon_2) * \lambda_\theta \exp(-\lambda_\theta h) \mathrm{d}\Phi(\frac{\epsilon_1}{\sigma_1})\mathrm{d}\Phi(\frac{\epsilon_2}{\sigma_2}) \\
&= \lambda_\theta \bar{l}_t \int_{-\infty}^{\infty} \exp\left(-\left(\frac{\xi-1}{\xi}\eta^Q\gamma+1\right)\epsilon_1\right)\mathrm{d}\Phi(\frac{\epsilon_1}{\sigma_1}) \int_{-\infty}^{\infty} \exp\left(-\frac{\xi-1}{\xi}\eta^Q\alpha\epsilon_2\right)\mathrm{d}\Phi(\frac{\epsilon_2}{\sigma_2}) \\
&\times \exp\left(\eta^l_{\theta t}h - \lambda_\theta h\right) \\
&= \tilde{l}_t * \exp\left(\eta^l_{\theta t}h - \lambda_\theta h\right),
\end{aligned} \tag{23}
$$

where $\tilde{l}_t$ is a constant independent to h.

Therefore λ_t solves the following equation

$$
\begin{aligned}
-\lambda_t &= \eta^l_{\theta t} - \lambda_\theta \\
&= \frac{\big((\xi-1)\big(\gamma-\psi(1-\alpha)\big)-\psi\big)}{\gamma\big(1+(1-\alpha-\gamma)(\xi-1)\big)}\left(\frac{\lambda_t}{\lambda_x}\right)^\psi - \frac{1+(\xi-1)(1-\alpha)}{1+(1-\alpha-\gamma)(\xi-1)}z_t - \lambda_\theta.
\end{aligned} \tag{24}
$$

Lemma 3.2. *There exists a unique solution $\lambda_t \in [0,\infty)$ to equation (24).*

Proof. See Appendix A.2. □

The existence of this solution confirms our conjecture that an equilibrium with an exponential job offer distribution, $h \sim \text{Exp}(\lambda_t)$, indeed exists. Our model falls into the generic class studied by Eeckhout and Kircher (2018), who establish the uniqueness of the equilibrium matching function and firm size distribution in this environment. Taken together with our result, their uniqueness theorem implies that the exponential distribution we conjecture and verify is the only distribution consistent with equilibrium in our setting.

Figure 2 graphically depicts the two sides of equation (24) as functions of λ_t, represented by the two curves. Their intersection identifies a unique equilibrium value $\lambda_t > 0$. Combined with the parameter governing the distribution of worker skills, λ_x, this equilibrium allows us to pin down the assortative matching pattern between firms and workers described in equation (8).

Note that we must still impose $\int_0^\infty f_t(h)\mathrm{d}h = 1$, or equivalently, $\tilde{l}_t = \lambda_t$. Together with

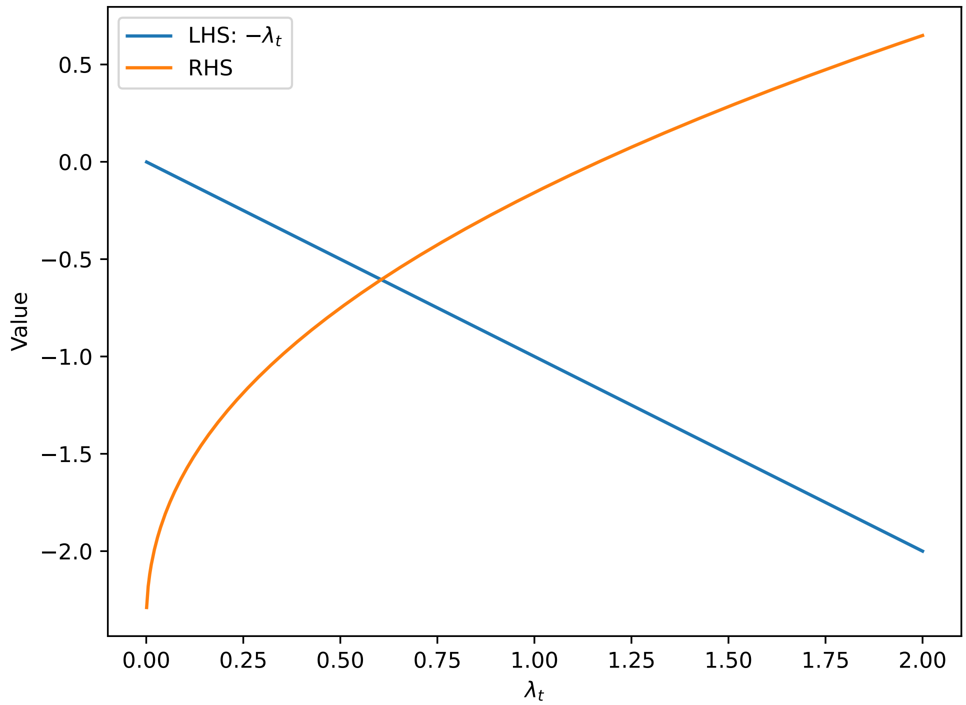

Figure 2: Equilibrium determination of λ_t.

equation (24), this identity completes the labor-market clearing condition. Recall that

$$\begin{aligned}\tilde{l}_t &\equiv \lambda_\theta \bar{l}_t \int_{-\infty}^{\infty} \exp\left(-\left(\frac{\xi-1}{\xi}\eta^Q\gamma+1\right)\epsilon_1\right) \mathrm{d}\Phi(\frac{\epsilon_1}{\sigma_1}) \int_{-\infty}^{\infty} \exp\left(-\frac{\xi-1}{\xi}\eta^Q\alpha\epsilon_2\right) \mathrm{d}\Phi(\frac{\epsilon_2}{\sigma_2}) \\ &= \lambda_\theta \bar{l}_t \exp\Big(\frac{1}{2}\big[(\frac{\xi-1}{\xi}\eta^Q\gamma+1)^2\sigma_1^2 + (\frac{\xi-1}{\xi}\eta^Q\alpha)^2\sigma_2^2\big]\Big).\end{aligned}$$

where $\bar{l}_t$, derived in Appendix A.1, is given by

$$\bar{l}_t = \left[\left(\frac{\gamma A_t \left(\frac{\bar{Q}_t}{Y_t}\right)^{-\frac{1}{\xi}} \frac{\xi-1}{\xi}}{w_t(0))}\right) \left(\frac{\alpha \left(\frac{\bar{Q}_t}{Y_t}\right)^{-\frac{1}{\xi}} \frac{\xi-1}{\xi} \bar{Q}_t}{R_t}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}}.$$

$\bar{l}_t$ is a function of aggregate variables $w_t(0)$, R_t, and Y_t. We next move to characterize these objects.

3.4 Aggregate Variables

Prices and aggregate output We can use market clearing conditions to characterize the remaining equilibrium variables. First, as noted above, the labor market clearing condition $\int_0^\infty f_t(h)\mathrm{d}h = 1$, implies that

$$\lambda_\theta B_1 \left[\left(\frac{\gamma A_t \left(\frac{\bar{Q}_t}{Y_t}\right)^{-\frac{1}{\xi}} \frac{\xi-1}{\xi}}{w_t(0))} \right) \left(\frac{\alpha \left(\frac{\bar{Q}_t}{Y_t}\right)^{-\frac{1}{\xi}} \frac{\xi-1}{\xi} \bar{Q}_t}{R_t} \right)^{\alpha} \right]^{\frac{1}{1-\gamma}} = \lambda_t, \tag{25}$$

where

$$B_1 \equiv \exp\Big(\frac{1}{2}\big[(\frac{\xi-1}{\xi}\eta^Q\gamma+1)^2\sigma_1^2 + (\frac{\xi-1}{\xi}\eta^Q\alpha)^2\sigma_2^2\big]\Big).$$

Equation (25) provides a relationship among $w_t(0)$, R_t, and Y_t. Importantly, λ_t is separately characterized from the job distribution equation (24), which does not involve $w_t(0)$, R_t, and Y_t.

Next, capital market clearing requires aggregate capital supply to equal aggregate capital demand:

$$K_t = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^{\infty} k_t(\theta,\epsilon_1,\epsilon_2) * \lambda_\theta \exp(-\lambda_\theta\theta)\mathrm{d}\theta \mathrm{d}\Phi(\frac{\epsilon_1}{\sigma_1})\mathrm{d}\Phi(\frac{\epsilon_2}{\sigma_2}),$$

where the LHS, K_t, represents capital supply, an aggregate state variable predetermined at t, while the RHS aggregates firms' capital demand. Substituting equation (20) and the expression of $\bar{k}_t$ yields

$$\begin{aligned} K_t &= \lambda_\theta \bar{k}_t \times \int_0^\infty \exp\left(\frac{\xi-1}{\xi}\eta^Q\eta^Q_{\theta t}\theta - \lambda_\theta\theta\right)\mathrm{d}\theta \\ &\times \int_{-\infty}^{\infty} \exp\left(-\frac{\xi-1}{\xi}\eta^Q\gamma\epsilon_1\right)\mathrm{d}\Phi(\frac{\epsilon_1}{\sigma_1}) \times \int_{-\infty}^{\infty}\exp\left(-\frac{\xi-1}{\xi}\eta^Q(\alpha+\frac{\xi}{\xi-1}\frac{1}{\eta^Q})\epsilon_2\right)\mathrm{d}\Phi(\frac{\epsilon_2}{\sigma_2}) \\ &= \lambda_\theta B_2 \alpha \frac{\xi-1}{\xi}\frac{\bar{Q}_t^{\frac{\xi-1}{\xi}} Y_t^{\frac{1}{\xi}}}{R_t} \end{aligned} \tag{26}$$

where

$$B_2 \equiv \frac{\exp\left[\frac{1}{2}\left(\left(\frac{\xi-1}{\xi}\eta^Q\gamma\right)^2\sigma_1^2 + \left(\frac{\xi-1}{\xi}\alpha\eta^Q+1\right)^2\sigma_2^2\right)\right]}{\lambda_\theta - \frac{\xi-1}{\xi}\eta^Q\eta^Q_{\theta t}}$$

provided that $\lambda_\theta - \frac{\xi-1}{\xi}\eta^Q\eta^Q_{\theta t} > 0$, otherwise capital demand becomes unbounded. Equation (26) therefore delivers a second equilibrium condition linking $w_t(0)$, R_t, and Y_t. Combining it with equation (25), we can represent prices $w_t(0)$ (and the entire wage schedule) and R_t as functions of aggregate state variables and aggregate output Y_t.

Finally, the aggregate output Y_t follows

$$Y_t = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{\xi}{\xi-1}\chi_t^*(\theta,\epsilon_1,\epsilon_2)Q_t^*(\theta,\epsilon_1,\epsilon_2)\lambda_\theta \exp(-\lambda_\theta\theta)\mathrm{d}\theta\mathrm{d}\Phi(\frac{\epsilon_1}{\sigma_1})\mathrm{d}\Phi(\frac{\epsilon_2}{\sigma_2}). \tag{27}$$

The supply of each intermediate good is equal to its demand

$$Q_t^*(\theta,\epsilon_1,\epsilon_2) = (\frac{\xi}{\xi-1}\chi_t^*(\theta,\epsilon_1,\epsilon_2))^{-\xi}Y_t, \tag{28}$$

which be substituted into (27) to get

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \left[\frac{\xi}{\xi-1}\chi_t^*(\theta,\epsilon_1,\epsilon_2)\right]^{1-\xi} \lambda_\theta \exp(-\lambda_\theta\theta)\mathrm{d}\theta\mathrm{d}\Phi(\frac{\epsilon_1}{\sigma_1})\mathrm{d}\Phi(\frac{\epsilon_2}{\sigma_2}). \tag{29}$$

Hence

$$\begin{aligned} 1 &= \lambda_\theta \left(\frac{\xi}{\xi-1}\right)^{1-\xi} \overline{\chi}_t^{1-\xi} \times \int_0^\infty \exp\left(\frac{\xi-1}{\xi}\eta^Q\eta^Q_{\theta t}\theta - \lambda_\theta\theta\right)\mathrm{d}\theta \\ &\times \int_{-\infty}^{\infty} \exp\left(-\frac{\xi-1}{\xi}\eta^Q\gamma\epsilon_1\right)\mathrm{d}\Phi(\frac{\epsilon_1}{\sigma_1}) \times \int_{-\infty}^{\infty} \exp\left(-\frac{\xi-1}{\xi}\eta^Q\alpha\epsilon_2\right)\mathrm{d}\Phi(\frac{\epsilon_2}{\sigma_2}) \\ &= \lambda_\theta B_3 \bar{Q}_t^{\frac{\xi-1}{\xi}} Y_t^{\frac{1-\xi}{\xi}}, \end{aligned} \tag{30}$$

where

$$B_3 \equiv \frac{\exp\left[\frac{1}{2}\left(\left(\frac{\xi-1}{\xi}\eta^Q\gamma\right)^2\sigma_1^2 + \left(\frac{\xi-1}{\xi}\alpha\eta^Q\right)^2\sigma_2^2\right)\right]}{\lambda_\theta - \frac{\xi-1}{\xi}\eta^Q\eta^Q_{\theta t}}.$$

Equation (30) is therefore another equilibrium condition of $w_t(0)$, R_t, and Y_t. Together with (25) and (26), the three equations determine the three unknown variables $\{w_t(0), R_t, Y_t\}$ uniquely for given aggregate state variables. The results are presented in Appendix A.3.

Consumption and investment Finally, we aggregate our previous results to characterize the dynamics of consumption, and ultimately the evolution of the endogenous aggregate state variable K_t. At the aggregate level, the model behaves analogously to a stochastic neoclassical growth model. Aggregate consumption satisfies the representative household's Euler equation

$$\beta E_t \frac{C_t}{C_{t+1}}[R_{t+1} + (1-\delta)] = 1, \tag{31}$$

together with the intertemporal budget constraint

$$C_t + K_{t+1} = (1-\delta)K_t + Y_t^l + Y_t^k + Y_t^d. \tag{32}$$

where Y_t^l, Y_t^k, and Y_t^d respectively denote aggregate labor income, capital income, and distributed profits. These objects are given by

$$Y_t^l = \frac{\gamma \lambda_\theta B_1 \bar{\chi}_t \bar{Q}_t}{\lambda_\theta - \frac{\xi-1}{\xi}\eta^Q \eta_{\theta t}^Q + z_t}, \tag{33}$$

$$Y_t^k = \alpha \lambda_\theta B_2 \bar{\chi}_t \bar{Q}_t, \tag{34}$$

$$Y_t^d = (\frac{\xi}{\xi - 1} - \gamma - \alpha)\lambda_\theta B_3 \bar{\chi}_t \bar{Q}_t. \tag{35}$$

The terms $\bar{\chi}_t$ and $\bar{Q}_t$ are obtained from our previous analysis. Note that Y_t^l, Y_t^k, and Y_t^d do not add up to the aggregate output Y_t, because wedges from market distortions generate resource losses that drive a gap between factor payments and aggregate output.

4 Micro Implications

The framework developed above enables an analytical examination of cyclical fluctuations in micro-level distributions. In this section, we focus on two key margins: wages, which serve as a quantitative measure of employee performance, and productivity, which reflects employer effectiveness. Specifically, we study the dispersion of these variables and analyze how it evolves over the business cycle.

We study business cycles driven by fluctuations in z_t. Higher-type firms are more sensitive to these changes, since the shocks have a larger impact on firms with greater θ. These changes induce shifts in the composition of labor demand, which in turn alter the sorting pattern between firms and workers.

To begin with, consider how a decline in z_t affects the equilibrium job distribution parameter λ_t. Recall that λ_t is determined by equation

$$\begin{aligned} -\lambda_t &= g(\lambda_t, z_t) \\ &\equiv \frac{\big((\xi-1)\big(\gamma-\psi(1-\alpha)\big)-\psi\big)}{\gamma\big(1+(1-\alpha-\gamma)(\xi-1)\big)}\left(\frac{\lambda_t}{\lambda_x}\right)^{\psi} - \frac{1+(\xi-1)(1-\alpha)}{1+(1-\alpha-\gamma)(\xi-1)} z_t - \lambda_\theta. \end{aligned} \tag{36}$$

Figure 3 illustrates the comparative statics following a decline in z_t: the function $g(\lambda_t, z_t)$ shifts upward, leading to a lower new equilibrium λ_t. A formal proof is provided in Appendix A.4. The main result is summarized as follows.

Lemma 4.1. *When z_t decreases, the equilibrium job distribution parameter λ_t decreases.*

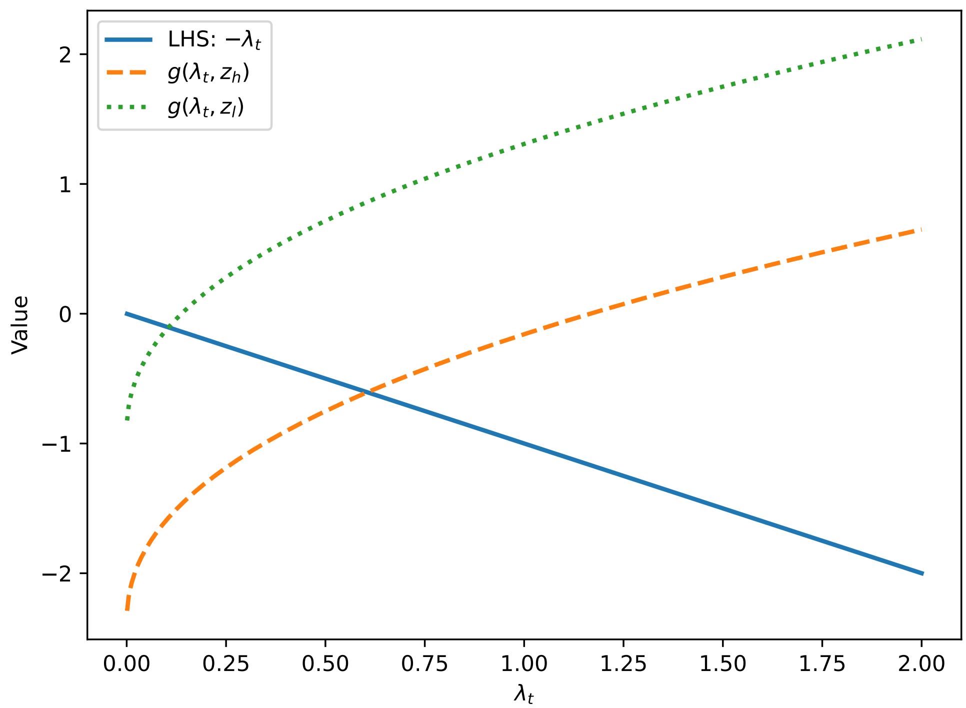

Figure 3: Equilibrium determination of λ_t following an decrease of z_t

Notes: The figure plots a decrease of z_t from z_h to z_l. The RHS of equation (36), $g(\lambda_t, z_t)$ shifts upward.

We can infer from the matching function (8) that during episodes of high z_t—which correspond to lower aggregate allocative efficiency—worker–firm sorting deteriorates, with workers being matched to inferior firms. This mechanism is reminiscent of the findings in Crane et al. (2023), who document that during downturns high-rank workers are more likely to work at low-rank firms, while low-rank workers are less likely to be employed by high-rank firms.

Note that z_t is a reduced-form shock, designed to abstract from the micro-foundations of market frictions and their interactions with aggregate shocks. A decline in z_t generates aggregate expansions in which better (or higher-type) firms expand relatively more. In reality, however, business cycles can arise from a variety of structural shocks, each potentially producing different patterns in micro-level distributions. In Section 5, we use our theoretical framework to examine the effects of some of these structural drivers on labor market and productivity distributions.

4.1 Wage Inequality

Recall that the wage schedule is given by

$$\ln(w_t(x)) = \frac{\psi}{\gamma}(\frac{\lambda_x}{\lambda_t})^{1-\psi}x + \ln(w_t(0)). \tag{37}$$

Hence the dispersion of wages is

$$\text{Var}_t(\log w_{it}) = (\frac{\psi}{\gamma})^2(\frac{\lambda_x}{\lambda_t})^{2-2\psi}\lambda_x^{-2} = (\frac{\psi}{\gamma})^2\lambda_x^{-2\psi}\lambda_t^{2\psi-2}. \tag{38}$$

Since a decline in z_t reduces λ_t, we obtain the following result.

Proposition 1. *When z_t decreases, wage inequality,* $\text{Var}_t(\log w_{it})$*, increases.*

Intuitively, a decline in z_t enables high-type firms—those offering high-type jobs—to expand more aggressively. This shifts the composition of job offers toward the upper tail, increasing the relative mass of high-type offers in the labor market. Consequently, the wage schedule steepens as competition intensifies, even in the absence of any change in labor supply. This results in greater wage dispersion.

4.2 Firm-level Productivity

It is well established in macroeconomics that firms exhibit heterogeneous responses to aggregate shocks, particularly in terms of output and investment. What sets our framework apart is that this heterogeneity extends beyond output and investment to measured productivity, which also reacts differently across firms.

There are various ways to measure firms' productivity:

$$\log TFP_{jt} := \log y_{jt} - \alpha \log k_{jt} - \gamma \log l_{jt} \tag{39}$$

where y_{jt} can refer to either the output quantity or revenue of firm j, depending on whether we aim to measure quantity-based productivity (TFPQ), or revenue-based productivity (TFPR).

Let us first consider the physical productivity, i.e. TFPQ. Using (3) and (8), we obtain

$$\log TFPQ(\theta) = \left(\mu_t^{-1}(\theta)\right)^{\psi} \theta^{1-\psi} = (\frac{\lambda_t}{\lambda_x})^{\psi}\theta. \tag{40}$$

Thus, TFPQ reflects only the intrinsic productivity of the firm and does not depend on firm-level distortions ϵ_1 or ϵ_2, which capture external market frictions. Its dependence on the aggregate state z_t is only indirect, operating through the effect on the job-type distribution parameter λ_t. Similar to the distribution of wages, the distribution of TFPQ across firms is also Pareto.

The dispersion of $\log TFPQ_{jt}$ at time t is therefore

$$\text{Var}_t(\log TFPQ_{jt}) = (\frac{\lambda_t}{\lambda_x})^{2\psi}\text{Var}\left(\theta^{1-\psi}\right) = (\frac{\lambda_t}{\lambda_x})^{2\psi}\lambda_\theta^{-2}$$

We already know that a decline in z_t reduces λ_t. This immediately yields the following result.

Proposition 2. *When z_t decreases, the dispersion of firm-level $\log TFPQ$, $\text{Var}_t(\log TFPQ_{jt})$, decreases.*

Intuitively, when z_t falls, high-type firms expand more aggressively than their low-type counterparts. This expansion shifts the composition of the labor market: high-type firms must increasingly hire lower-skill workers. As a result, their physical productivity is pulled down relative to that of low-type firms, compressing the overall dispersion of TFPQ.

Next, we turn to revenue-based productivity, TFPR. By definition, TFPR is the product of TFPQ and the price of the intermediate good. From (21), we obtain

$$\log TFPR_t(\theta, \epsilon_1, \epsilon_2) = \log \frac{\xi}{\xi - 1} - \frac{\eta^Q}{\xi}\left(\eta^Q_{\theta t}\theta - \gamma\epsilon_1 - \alpha\epsilon_2\right) + (\frac{\lambda_t}{\lambda_x})^{\psi}\theta$$
$$= \log \frac{\xi}{\xi - 1} + \left[(\frac{\lambda_t}{\lambda_x})^{\psi} - \frac{\eta^Q \eta^Q_{\theta t}}{\xi}\right]\theta + \frac{\eta^Q}{\xi}(\gamma\epsilon_1 + \alpha\epsilon_2) \qquad (41)$$

Unlike TFPQ, TFPR responds directly to firm-level distortions, as these distortions shift marginal costs and thus affect pricing decisions. Through its price component, TFPR therefore reflects not only "pure" production efficiency but also the impact of distortions and demand conditions. Importantly, higher TFPR may actually be associated with greater distortions, since distortions can raise marginal costs and thus inflate prices. As with firm sizes, TFPR also displays a Pareto tail.

The dispersion of $\log \text{TFPR}_{jt}$ at time t is given by

$$\text{Var}_t(\log TFPR_{jt}) = \left[(\frac{\lambda_t}{\lambda_x})^{\psi} - \frac{\eta^Q \eta^Q_{\theta t}}{\xi}\right]^2 \lambda_\theta^{-2} + (\frac{\eta^Q}{\xi})^2\left(\gamma^2\sigma_1^2 + \alpha^2\sigma_2^2\right) \qquad (42)$$

We can now state the main result.

Proposition 3.

$$(\frac{\lambda_t}{\lambda_x})^{\psi} - \frac{\eta^Q \eta^Q_{\theta t}}{\xi} > 0.$$

Moreover, when z_t decreases, the dispersion of firm-level revenue productivity, $\text{Var}_t(\log TFPR_{jt})$, *decreases.*

Proof. We have

$$(\frac{\lambda_t}{\lambda_x})^{\psi} - \frac{\eta^Q \eta^Q_{\theta t}}{\xi} = (\frac{\lambda_t}{\lambda_x})^{\psi} + \frac{1}{1 + (1 - \alpha - \gamma)(\xi - 1)}\left[\gamma z_t - (1 - \psi)(\frac{\lambda_t}{\lambda_x})^{\psi}\right]$$
$$= \left[1 - \frac{1 - \psi}{1 + (1 - \alpha - \gamma)(\xi - 1)}\right](\frac{\lambda_t}{\lambda_x})^{\psi} + \frac{\gamma}{1 + (1 - \alpha - \gamma)(\xi - 1)} z_t.$$

Both terms on the RHS are strictly positive. Moreover, a reduction in z_t reduces each term, and therefore reduces the expression as a whole. Q.E.D. □

The change in TFPR dispersion reflects shifts in both the dispersion of TFPQ and the dispersion of prices (or, equivalently, marginal costs). Mathematically, the impact of z_t

operates through its effect on the term

$$(\frac{\lambda_t}{\lambda_x})^{\psi} - \frac{\eta^Q \eta^Q_{\theta t}}{\xi},$$

The first component, $(\lambda_t/\lambda_x)^{\psi}$, always decreases when z_t falls, as shown earlier. The second component,

$$-\frac{\eta^Q \eta^Q_{\theta t}}{\xi} = \frac{1}{1+(1-\alpha-\gamma)(\xi-1)}\left[\gamma z_t - (1-\psi)(\frac{\lambda_t}{\lambda_x})^{\psi}\right], \tag{43}$$

embeds two opposing forces. A lower z_t reduces the dispersion of firm-level distortions, compressing the dispersion of marginal costs and therefore prices—this corresponds to the first term on the RHS. At the same time, the change in the matching pattern increases marginal costs through $-(\lambda_t/\lambda_x)^{\psi}$, a force increasing the dispersion of marginal costs, as indicated by the second term on the RHS. However, as shown in the previous proof, the forces is outweighed by the decline in $(\lambda_t/\lambda_x)^{\psi}$, i.e. the direct effect operating through TFPQ. Consequently, the overall term unambiguously decreases when z_t falls, implying that the dispersion of TFPR decreases as well.

Robustness As the response of the productivity distribution to changes in the matching functions is the main result of this article, it is natural to wonder about the extent to which it is driven by our functional form assumptions. In our specification improvements in firms' matches increase log TFPQ multiplicatively. While this very particular form of dependence is unlikely to hold in more general models, it is also clearly not necessary for the larger conclusion to hold. Indeed, what suffices is that improvements in firms' matches improve the distribution of TFPQ in a strong enough sense, specifically in the hazard rate order (HRO) sense (see Section 1.B in Shaked and Shanthikumar, 2007), and that the equilibrium distribution of log TFPQ has a decreasing hazard rate (DHR). These conditions ensure, by Theorem 3.B.20 in Shaked and Shanthikumar (2007), that the distribution of log TFPQ becomes more unequal in the sense of the dispersive order (see Section 3.B in Shaked and Shanthikumar, 2007), which in turn implies an increase in the variance of log TFPQ. Both of these conditions are satisfied in our specification, because the exponential distribution has a constant hazard rate and a proportional improvement in log TFPQ implies an improvement in the hazard rate order sense for any distribution with a decreasing hazard rate.

Parameter	Description	Value
α	Capital intensity	0.3
γ	Labor intensity	0.6
δ	Depreciation rate	0.10
β	Discount factor	0.96
ξ	CES elasticity of substitution	9
p_l	Prob of staying in a boom	0.977
p_h	Prob of staying in a recession	0.688
σ_2	Std of ϵ_2	0

Table 1: Fixed Parameters

It remains an open question what general conditions would ensure that the equilibrium log TFPQ distribution is log-convex and that aggregate shocks improve said distribution in HRO sense.

4.3 Numerical Exercise

Next we move to a numerical illustration. Consider an environment in which aggregate fluctuations are driven exclusively by the reduced-form market efficiency shock z_t. The model is simulated by feeding in a stochastic process for z_t, which follows a two-state Markov chain taking values $\{0, z_h\}$, corresponding to boom and crisis states. The transition probabilities (p_l, p_h) govern the persistence of each state, representing the probability of remaining in a boom or a crisis, respectively.

Calibration We calibrate the model at an annual frequency. Table 1 reports the parameters fixed a priori, including standard macroeconomic values such as capital intensity $\alpha = 0.3$, labor intensity $\gamma = 0.6$, depreciation $\delta = 0.10$, discount factor $\beta = 0.96$, and CES parameter $\xi = 9$. The Markov transition probabilities (p_l, p_h) are taken from Khan and Thomas (2013) to match the durations of booms and crises. For simplicity, we set $\sigma_2 = 0$. As established earlier in this section, σ_2 is isomorphic to σ_1 in the computation for the second moments of our interests.

The remaining parameters are calibrated by minimizing the sum of squared deviations between model-implied moments and their empirical counterparts. The model is simulated for 10^4 periods, and all moments are computed using the final 9,900 observations to eliminate the influence of initial conditions.

Panel A: Calibrated Parameters

Parameter	Description	Value
ψ	intensity of worker type x	0.4022
z_h	z in a recession	0.3984
λ_θ	distribution of θ (type of firms)	2.6160
λ_x	distribution of x (type of workers)	0.8681
σ_1	Std of ϵ_1	0.2293

Panel B: Targets and Model Fit

Moment	Data	Model
Labor share	0.6097	0.6102
Wage inequality	0.7666	0.7666
Share of revenue of top 10% firms	0.9074	0.8906
Share of revenue of top 50%-top 10% firms	0.0842	0.0840
Std of TFP	0.0090	0.0090

Table 2: Calibration and Model Fit

Notes: Labor share data and TFP data are retrieved from FRED. Wage inequality is from Song et al. (2019). Shares of revenue are from Kwon et al. (2024). All the moments are calculated over year 1978-2013.

We target three classes of moments. First, we discipline the cross-sectional distributions of workers and firms by matching overall wage inequality—measured as the standard deviation of log wages—as well as the concentration of firm revenues, captured by the revenue shares of the top 10 percent of firms and of firms between the 50th and 90th percentiles. Second, we target the aggregate labor share, which in our framework reflects distortions arising from labor market wedges, as discussed in Section 3.2. Finally, we target the volatility of aggregate productivity, measured by the standard deviation of aggregate TFP. Wage inequality moments are taken from Song et al. (2019), firm revenue distribution moments from Kwon et al. (2024), and aggregate labor share and TFP from FRED. All empirical moments are computed over the 1978-2013 sample. Table 2 reports the calibration results: Panel A presents the calibrated parameter values, while Panel B compares the targeted moments with their model-implied counterparts.

Impulse responses Figure 4 reports the average impulse responses to an increase in the market-efficiency shock, $z_t = z_h$, which corresponds to a deterioration in allocative efficiency and hence a recessionary episode. The top-left panel shows that aggregate TFP declines on impact, reflecting the worsening allocation of labor across heterogeneous firms.

Aggregate output falls by an even larger magnitude, driven both by lower TFP and by a gradual contraction in the capital stock.

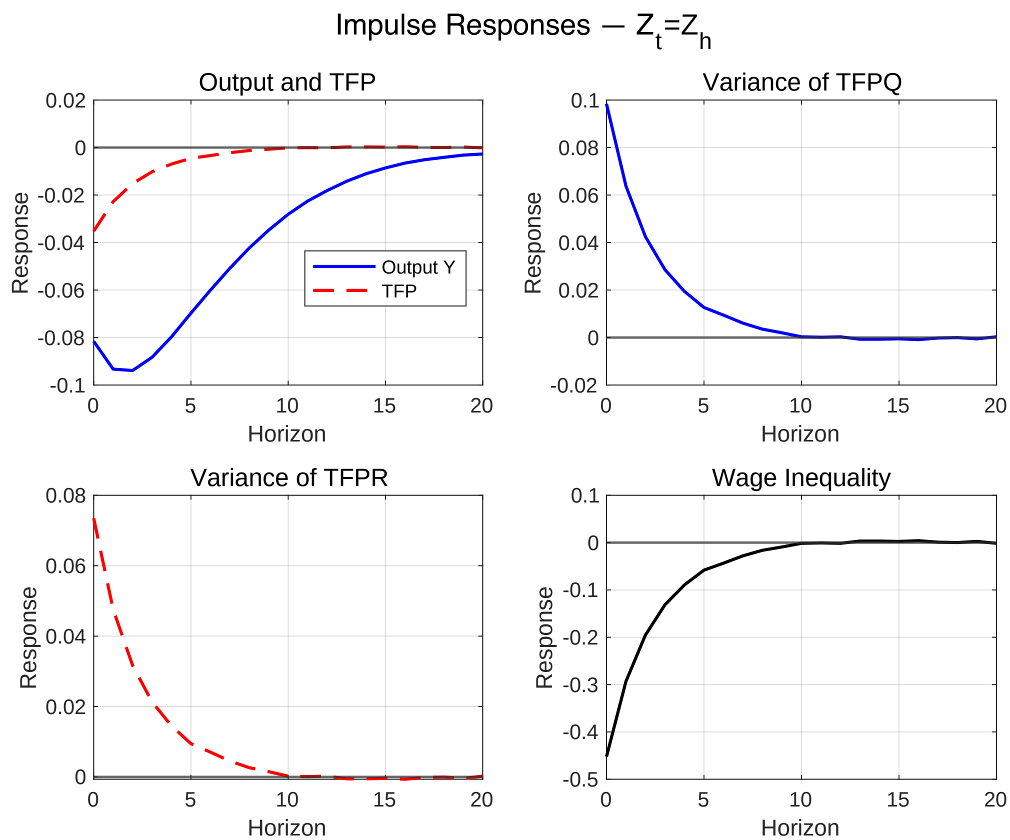

Figure 4: Impulse responses to a market-efficiency shock

The remaining panels illustrate how the shock reshapes cross-sectional distributions. The variance of firm-level quantity productivity (log TFPQ) rises sharply on impact and remains elevated for several periods, indicating an increase in productivity dispersion. Intuitively, as market efficiency deteriorates, high-type firms contract disproportionately and concentrate employment on their most productive workers, amplifying differences in measured productivity across firms. A similar dynamic is observed for the dispersion of firm-level revenue productivity (log TFPR), which also increases before gradually converging back to its steady state.

In contrast, wage inequality declines persistently following the shock. As allocative efficiency worsens, high-productivity firms shrink and high-type positions become scarcer, flattening the wage schedule across worker types and compressing earnings dispersion.

Taken together, the impulse responses highlight the central mechanism of the model: market-efficiency shocks generate opposing cyclical movements in productivity dispersion and wage inequality—productivity dispersion rises while wage dispersion falls during downturns—mirroring the key empirical regularities.

The magnitudes of the impulse responses are economically significant. The responses of GDP and aggregate TFP are reported in log differences. Following a simulated crisis—that is, a transition to the low-efficiency state—GDP contracts by more than 9 percent on impact, reflecting both a decline in allocative efficiency and a gradual reduction in the capital stock. The remaining panels report level deviations from their unconditional means. When $z_t = 0$, the variances of log firm-level productivity (TFPQ) and log firm-level revenue (TFPR) are 0.1203 and 0.0546, respectively; when $z_t = z_h$, these variances increase to 0.2254 and 0.1330, indicating a substantial widening of productivity and revenue dispersion during crises. In contrast, wage inequality declines sharply, falling from 0.7901 in booms to 0.3132 in recessions.

While the quantitative responses are large relative to available empirical estimates,[8] they reflect the stylized nature of the experiment and serve to highlight the strength of the model's underlying mechanisms. When aggregate TFP fluctuations are driven jointly by the reduced-form market efficiency shock z_t and the aggregate productivity shock A_t, the implied volatility of z_t is smaller than in our numerical experiment. As implied by the analysis in Section 5, the resulting fluctuations of other endogenous variables are closer in magnitude to those observed in the data.

5 Structural Drivers of Business Cycles

For simplicity and clarity in both exposition and notation, we have focused exclusively on aggregate fluctuations driven by time-varying z_t. As noted earlier, we do not model market frictions explicitly, so z_t should be interpreted as a reduced-form shock capturing the combined influence of multiple structural shocks. The benefit of this approach is that it delivers closed-form implications for the distributions we care about.

That said, no two booms or recessions are necessarily alike, since they may arise from different underlying forces. Macroeconomists have proposed a range of structural shocks to account for business-cycle dynamics. We now examine the implications of several such shocks within our framework.

[8] See, for example, Kehrig (2015) and Cunningham et al. (2023).

Productivity shock One of the most common shocks in business cycle models is the aggregate productivity shock. Recall that the production function is

$$Q_t(l,k,x,\theta) = A_t q(x,\theta) k^{\alpha} l^{\gamma},$$

where A_t denotes an aggregate productivity shock. Our earlier analysis shows that λ_t is independent of A_t, and that A_t does not affect the sorting equation (8). Consequently, in our framework, productivity shocks do not influence the dispersion of wage or productivity.

It is important to emphasize, however, that we have treated the micro-foundations of market distortions as a black box. In a model with explicit micro-founded frictions, firms of different types may respond heterogeneously to productivity shocks. Therefore, our result should not be interpreted as a general claim that productivity shocks never alter sorting patterns or dispersion moments. Rather, within the current framework, productivity shocks leave sorting and dispersion unchanged precisely because they do not generate any relative compositional shifts across firm types in the labor market. Without such shifts, the underlying worker-firm matching and the associated dispersion of wages or productivity remain unaffected.

Second-moment shocks Thus far, we have not imposed any specific structure or assumptions on the idiosyncratic process of firm type θ. We have only assumed that $\theta \sim \text{Exp}(\lambda_\theta)$. Consider the following simple idiosyncratic process

$$\theta_{j,t+1} = \begin{cases} \rho\theta_{j,t}, & \text{with probability } \rho, \\ \rho\theta_{j,t} + \varepsilon_{t+1}, & \text{with probability } 1-\rho, \end{cases}$$

where $\varepsilon_{t+1} \sim \text{Exp}(\lambda_\theta)$ is i.i.d across firms and time. The process ensures that θ follows an exponential distribution with fixed λ_θ.

We can extend this framework to allow for a time-varying distribution of θ by considering the following process:

$$\theta_{j,t+1} = \begin{cases} \rho\theta_{j,t}, & \text{with probability } p_{t+1}, \\ \rho\theta_{j,t} + \varepsilon_{t+1}, & \text{with probability } 1-p_{t+1}, \end{cases}$$

where

$$\varepsilon_{t+1} \sim \text{Exp}(\lambda_{\theta,t+1}), \qquad p_{t+1} = \frac{\rho\,\lambda_{\theta,t+1}}{\lambda_{\theta,t}}.$$

For all t, $\lambda_{\theta t} \in \left\{\lambda_\theta^L, \lambda_\theta^H\right\}$ and follows a Markov process, with $\lambda_\theta^L < \lambda_\theta^H < \frac{\lambda_\theta^L}{\rho}$. This process guarantees that $\theta \sim \text{Exp}\left(\lambda_{\theta t}\right)$ with a time-varying $\lambda_{\theta t}$ evolving according to a Markov process.

Note that the shock to λ_θ is not purely a second-moment shock: a decrease in λ_θ increases both the variance and the mean of θ.

How does a fall in $\lambda_{\theta t}$ affect the distribution of jobs? Using equation (36), we see that, similar to a decline in z_t, the decrease in $\lambda_{\theta t}$ reduces λ_t. From equation (38), it then follows immediately that wage dispersion increases.

However, unlike the reduced-form shock in z_t, a decrease in λ_θ also increases the dispersions of firm-level TFPQ and TFPR. As shown in Appendix A.5, this type of shock shifts wage inequality and productivity dispersion in the same direction.

Another way to introduce second-moment shocks is through time-varying variances of ϵ_1 and ϵ_2, the labor and capital wedges that are independent to firm types. For example, one can consider the following processes:

$$\log \sigma_{1t} = \rho_1 \log \sigma_{1t-1} + \varepsilon_{1t}$$
$$\log \sigma_{2t} = \rho_2 \log \sigma_{2t-1} + \varepsilon_{2t}$$

where $\varepsilon_{1t} \sim N\left(0, \sigma_l^2\right)$, and $\varepsilon_{2t} \sim N\left(0, \sigma_k^2\right)$.

From equation (36), it follows that the variances of ϵ_1 and ϵ_2—distortions uncorrelated with firm type θ—do not affect λ_t. Consequently, shocks to these variances have no impact on the dispersion of wages or TFPQ, nor do they alter coefficients such as η^Q or $\eta^Q_{\theta t}$. However, as indicated by equation (42), such shocks can influence the dispersion of TFPR: Increases in σ_{1t} or σ_{2t} raise the dispersion of revenue productivity across firms.

6 Concluding Remarks

This paper develops a tractable model in which heterogeneous firms and workers sort in the labor market, allowing productivity and wages to emerge endogenously from the allocation of talent across firms. We introduce a reduced-form market efficiency shock that affects high-type firms more strongly and show analytically how it generates opposing cyclical movements in wage and productivity dispersions—wage dispersion rises in booms while productivity dispersion narrows—mirroring key empirical patterns. We further explore how alternative structural shocks, including aggregate productivity and second-moment shocks, shape these distributions and their joint dynamics. Taken together, the framework provides a unified way to interpret business-cycle fluctuations in both wage inequality and productivity dispersion, reconciling several empirical regularities within a single, highly tractable theoretical framework.

A limitation of our approach is that market distortions enter in reduced-form through firm-specific wedges on labor and capital prices, while market efficiency shocks are modeled as time variation in these wedges. This modeling choice preserves analytical tractability and allows us to derive sharp, closed-form implications for the joint behavior of wage and productivity dispersions. An important direction for future work is to embed the same sorting mechanism into a fully micro-founded environment with explicit market frictions, structural shocks, and firm entry and exit. Such a framework would necessarily rely on numerical solution methods but would permit a richer quantitative assessment of how alternative frictions and shocks shape the co-movement of wages, productivity, and other variables over the business cycle.

The analysis in this paper abstracts from unemployment and matching frictions: all workers are employed, and matches form without search. Incorporating unemployment and search-and-matching frictions would again require numerical solutions, but doing so would open an important avenue for future work. In the current framework, the focus is naturally on wage inequality, which reflects dispersion in *labor income* among *employed* workers. While wage inequality is an important and widely studied object, introducing unemployment would allow us to study an additional—and empirically crucial—dimension of income inequality. Recent evidence suggests that income inequality is countercyclical, with unemployment playing a central role.[9] Understanding the sources of disparity between wage inequality and income inequality is a promising direction for future research.

[9]See, for example, Heathcote et al. (2020).

References

Arellano, C., Y. Bai, and P. J. Kehoe (2019). Financial Frictions and Fluctuations in Volatility. *Journal of Political Economy 127*(5), 2049–2103.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. V. Reenen (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics 135*(2), 645–709.

Bachmann, R. and C. Bayer (2013). "Wait-and-See Business Cycles?". *Journal of Monetary Economics 60*(6), 704–719.

Bachmann, R. and C. Bayer (2014). Investment Dispersion and the Business Cycle. *American Economic Review 104*(4), 1392–1416.

Bento, P. and D. Restuccia (2017). Misallocation, Establishment Size, and Productivity. *American Economic Journal: Macroeconomics 9*(3), 267–303. Published version of NBER Working Paper 22809.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica 77*(3), 623–685.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018). Really Uncertain Business Cycles. *Econometrica 86*(3), 1031–1065. Published version of NBER Working Paper 18245.

Buera, F. J. and B. Moll (2015). Aggregate Implications of a Credit Crunch: The Importance of Heterogeneity. *American Economic Journal: Macroeconomics 7*(3), 1–42.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2007). Business Cycle Accounting. *Econometrica 75*(3), 781–836.

Choi, J. (2023). Offshoring, matching, and income inequality. *Journal of Mathematical Economics 106*, 102846.

Costinot, A. (2009). An elementary theory of comparative advantage. *Econometrica 77*(4), 1165–1192.

Costinot, A. and J. Vogel (2010, 08). Matching and Inequality in the World Economy. *Journal of Political Economy 118*(4), 747–786.

Crane, L. D., H. R. Hyatt, and S. M. Murray (2023). Cyclical Labor Market Sorting. *Journal of Econometrics 233*(2), 524–543.

Crouzet, N. and N. R. Mehrotra (2020). Small and Large Firms over the Business Cycle. *American Economic Review 110*(11), 3549–3601.

Cunningham, C., L. Foster, C. Grim, J. C. Haltiwanger, S. W. Pabilonia, J. Stewart, and Z. Wolf (2023). Dispersion in Dispersion: Measuring Establishment-Level Differences in Productivity. *Review of Income and Wealth 69*(4), 999–1032.

Eeckhout, J. and P. Kircher (2018). Assortative Matching with Large Firms. *Econometrica 86*(1), 85–132.

Garicano, L., C. Lelarge, and J. V. Reenen (2016). Firm Size Distortions and the Productivity Distribution: Evidence from France. *The American Economic Review 106*(11), 3439–3479.

Gertler, M. and S. Gilchrist (1994). Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms. *The Quarterly Journal of Economics 109*(2), 309–340.

Gola, P. and Y. Zhao (2023). A Firm Link: Overall, Between- and Within-Firm Inequality Through the Lens of a Sorting Model. Technical report. working paper.

Gourio, F. and N. Roys (2014). Size-Dependent Regulations, Firm Size Distribution, and Reallocation. *Quantitative Economics 5*(2), 377–416. Published version of NBER Working Paper No. 18657.

Grossman, G. M., E. Helpman, and P. Kircher (2017). Matching, Sorting, and the Distributional Impacts of International Trade. *Journal of Political Economy 124*(1), 224–264.

Haltiwanger, J. C., H. R. Hyatt, L. B. Kahn, and E. L. McEntarfer (2018). Cyclical Job Ladders by Firm Size and Firm Wage. *American Economic Journal: Macroeconomics 10*(2), 52–85. Published version of NBER Working Paper 23485.

Haltiwanger, J. C., H. R. Hyatt, E. McEntarfer, and M. Staiger (2025). Cyclical Worker Flows: Cleansing vs. Sullying. *Review of Economic Dynamics 55*, 101252. Published version of NBER Working Paper 28802.

Heathcote, J., F. Perri, and G. L. Violante (2020). The Rise of US Earnings Inequality: Does the Cycle Drive the Trend? *Review of Economic Dynamics 37*, S181–S204.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics 124*(4), 1403–1448.

Hsieh, C.-T. and P. J. Klenow (2014). The Life Cycle of Plants in India and Mexico. *The Quarterly Journal of Economics 129*(3), 1035–1084.

Kehrig, M. (2015). The Cyclical Nature of the Productivity Distribution. Technical report. Working Paper.

Khan, A. and J. K. Thomas (2013). Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity. *Journal of Political Economy 121*(6), 1055–1107.

Kroen, T., E. Liu, A. R. Mian, and A. Sufi (2021). Falling Rates and Rising Superstars. Technical report. Working Paper.

Kwon, S. Y., Y. Ma, and K. Zimmermann (2024). 100 Years of Rising Corporate Concentration. *American Economic Review 114*(7), 2111–2140.

Morlacco, M. and D. L. Zeke (2021). Monetary Policy, Customer Capital, and Market Power. *Journal of Monetary Economics 121*(C), 116–134.

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

Ottonello, P. and T. Winberry (2020). Financial Heterogeneity and the Investment Channel of Monetary Policy. *Econometrica 88*(6), 2473–2502.

Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics 118*(1), 1–41.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics 11*(4), 707–720.

Sampson, T. (2014). Selection into Trade and Wage Inequality. *American Economic Journal: Microeconomics 6*(3), 157–202.

Sattinger, M. (1975). Comparative Advantage and the Distributions of Earnings and Abilities. *Econometrica 43*(3), 455–468.

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

Shaked, M. and G. Shanthikumar (2007). *Stochastic Orders* (1 ed.). Springer Series in Statistics. Springer-Verlag New York.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019). Firming Up Inequality. *The Quarterly Journal of Economics 134*(1), 1–50. Published version of NBER Working Paper 21199.

Teulings, C. (2005). Comparative Advantage, Relative Wages, and the Accumulation of Human Capital. *Journal of Political Economy 113*(2), 425–461.

Teulings, C. N. (1995, April). The Wage Distribution in a Model of the Assignment of Skills to Jobs. *Journal of Political Economy 103*(2), 280–315.

Online Appendix

A Omitted Proofs and Derivations

A.1 Proof of Lemma 3.1

First, given $\bar{Q}_t$, we have

$$\bar{k}_t = \alpha \frac{\xi - 1}{\xi} \frac{\bar{Q}_t^{\frac{\xi-1}{\xi}} Y_t^{\frac{1}{\xi}}}{R_t}, \qquad \bar{\chi}_t = \frac{\xi - 1}{\xi} \bar{Q}_t^{-\frac{1}{\xi}} Y_t^{\frac{1}{\xi}}, \qquad \bar{l}_t = \left[\bar{\chi}_t A_t \bar{k}_t^{\alpha} \gamma / w_t(0)\right]^{\frac{1}{1-\gamma}}.$$

Using the identity

$$\bar{Q}_t = A_t \bar{k}_t^{\alpha} \bar{l}_t,$$

we obtain

$$\bar{Q}_t = \left[A_t \left(\frac{\gamma}{w_t(0)} \right)^{\gamma} \left(\frac{\alpha}{R_t} \right)^{\alpha} \left(\frac{\xi - 1}{\xi} Y_t^{\frac{1}{\xi}} \right)^{\alpha+\gamma} \right]^{\eta^Q}.$$

The markup for an intermediate good is a constant $\frac{\xi}{\xi-1}$. From (4),

$$\chi_{jt} = \left(\frac{Q_{jt}}{Y_t} \right)^{-\frac{1}{\xi}} \frac{\xi - 1}{\xi} \tag{44}$$

The first-order condition with respect to capital yields:

$$k_{jt} = \frac{\alpha \left(\frac{Q_{jt}}{Y_t} \right)^{-\frac{1}{\xi}} \frac{\xi-1}{\xi} Q_{jt}}{\tau_{2jt} R_t}. \tag{45}$$

Further, substituting (18) and (17) into (11) gives

$$Q_{jt} = \chi_{jt}^{-1} \tau_{1jt} w_t(0) \exp\left(\frac{1-\psi}{1-\gamma} (\frac{\lambda_t}{\lambda_x})^{\psi} \theta_{jt} \right) \left[\left(\frac{\gamma A_t \chi_{jt}}{w_t(0) \tau_{1jt})} \right) k_{jt}^{\alpha} \right]^{\frac{1}{1-\gamma}}. \tag{46}$$

Next, substituting for χ from (44) and for k from (45) results in

$$Q_{jt} = w_t(0)\tau_{1jt}\exp\left(\frac{1-\psi}{1-\gamma}(\frac{\lambda_t}{\lambda_x})^{\psi}\theta_{jt}\right)$$
$$\times\left[\left(\frac{Q_{jt}}{Y_t}\right)^{-\frac{1}{\xi}}\frac{\xi-1}{\xi}\right]^{-1}\left[\left(\frac{\gamma A_t\left(\frac{Q_{jt}}{Y_t}\right)^{-\frac{1}{\xi}}\frac{\xi-1}{\xi}}{w_t(0)\tau_{1jt})}\right)\left(\frac{\alpha\left(\frac{Q_{jt}}{Y_t}\right)^{-\frac{1}{\xi}}\frac{\xi-1}{\xi}Q_{jt}}{\tau_{2jt}R_t}\right)^{\alpha}\right]^{\frac{1}{1-\gamma}},$$

where

$$\tau_1 = \exp(z_t\theta+\epsilon_1),$$
$$\tau_2 = \exp(\epsilon_2),$$

and after simple rearrangements, (19) follows. Similarly, (21) and (20) follow from substituting (19) into (44) and (45), respectively. Finally, (22) is the result of substituting (6), (20) and (21) into (18).

A.2 Proof of Lemma 3.2

The LHS of equation (24) is decreasing λ_t and ranges from 0 to $-\infty$. We denote the RHS as $g(\lambda_t)$. Our objective is to prove that the equation $G(\lambda_t) \equiv g(\lambda_t) + \lambda_t = 0$ has a unique solution $\lambda_t \in [0,\infty)$.

First, let us consider the scenario where

$$(\xi-1)\big(\gamma-\psi(1-\alpha)\big)-\psi > 0,$$

i.e. $g'(x) \geq 0$. It is immediate $G(0) < 0$, $\lim_{x\to\infty} G(x) \to \infty$, and $G'(x) > 0$ for $x \in [0,\infty)$. It is straightforward $G(x) \equiv g(x) + x = 0$ has a unique solution $x \in [0,\infty)$.

Second, let us consider the case

$$(\xi-1)\big(\gamma-\psi(1-\alpha)\big)-\psi < 0.$$

Now we have $G''(x) > 0$ for $x \in [0,\infty)$. Again, we have $G(0) < 0$, $\lim_{x\to\infty} G(x) \to \infty$. Since G is continuous on $[0,\infty)$, the conditions $G(0) < 0$ and $\lim_{x\to\infty} G(x) > 0$ imply, by the Intermediate Value Theorem, the existence of a point $x^* > 0$ such that $G(x^*) = 0$. Hence a solution exists.

Because $G''(x) > 0$ for all $x \geq 0$, the function f is strictly convex on $[0,\infty)$. Strict convexity implies that there exists a unique point $m \in \mathbb{R}$ at which G attains its global minimum, and that $G'(m) = 0$. We distinguish two cases according to the location of m.

Case 1: $m \leq 0$. For any $x > m$, strict convexity implies $G'(x) > G'(m) = 0$. In particular, $G'(x) > 0$ for all $x \geq 0$, and therefore G is strictly increasing on $[0, \infty)$. Since $G(0) < 0$ and G is strictly increasing, the equation $G(x) = 0$ has at most one solution in $[0, \infty)$. Combined with existence, this solution is unique.

Case 2: $m > 0$. Strict convexity implies that G is strictly decreasing on $[0, m]$ and strictly increasing on $[m, \infty)$. Since m is the unique minimizer and $G(0) < 0$, we have $G(m) \leq G(0) < 0$. Consequently,

$$G(x) < 0 \qquad \text{for all } x \in [0, m], \tag{47}$$

so no solution to $G(x) = 0$ can lie in this interval. Any solution must lie in (m, ∞), where G is strictly increasing. Thus at most one such solution exists. Together with existence, the solution is unique.

In both cases, the equation $G(x) = 0$ has exactly one solution on $[0, \infty)$, which completes the proof.

A.3 Derivation of $\{w_t(0), R_t, Y_t\}$

For notation ease, define

$$\kappa \equiv \frac{\xi - 1}{\xi}.$$

We have

$$\lambda_\theta B_3 \bar{Q}_t^{\frac{\xi-1}{\xi}} Y_t^{\frac{1-\xi}{\xi}} = 1.$$

Using $\kappa = \frac{\xi-1}{\xi}$:

$$\lambda_\theta B_3 \bar{Q}_t^{\kappa} Y_t^{-\kappa} = 1.$$

Thus,

$$\left(\frac{Y_t}{\bar{Q}_t}\right)^\kappa = \lambda_\theta B_3 \quad \Rightarrow \quad Y_t = \bar{Q}_t (\lambda_\theta B_3)^{1/\kappa} = \bar{Q}_t (\lambda_\theta B_3)^{\frac{\xi}{\xi-1}}.$$

Define

$$M \equiv (\lambda_\theta B_3)^{1/\kappa},$$

so

$$Y_t = M \bar{Q}_t.$$

Next, we have

$$\lambda_\theta B_2 \alpha\kappa \frac{\bar{Q}_t^\kappa Y_t^{1/\xi}}{R_t} = K_t.$$

Substitute $Y_t = M\bar{Q}_t$:

$$\bar{Q}_t^\kappa Y_t^{1/\xi} = \bar{Q}_t^\kappa (M\bar{Q}_t)^{1/\xi} = M^{1/\xi}\bar{Q}_t^{\kappa+1/\xi}.$$

Thus,

$$R_t = \frac{\lambda_\theta B_2 \alpha\kappa\, M^{1/\xi}\bar{Q}_t^{\kappa+1/\xi}}{K_t}.$$

Since $\kappa + \frac{1}{\xi} = 1$:

$$R_t = \frac{\lambda_\theta B_2 \alpha\kappa\, M^{1/\xi}\bar{Q}_t}{K_t}.$$

Using $M^{1/\xi} = (\lambda_\theta B_3)^{\frac{1}{\xi-1}}$:

$$R_t = \frac{\lambda_\theta B_2 \alpha\kappa\; \bar{Q}_t (\lambda_\theta B_3)^{\frac{1}{\xi-1}}}{K_t}.$$

The labor market clearing equation is

$$\lambda_\theta B_1 \left[\left(\frac{\gamma A_t (\frac{\bar{Q}_t}{Y_t})^{-1/\xi}\kappa}{w_t(0)} \right) \left(\frac{\alpha (\frac{\bar{Q}_t}{Y_t})^{-1/\xi}\kappa \bar{Q}_t}{R_t} \right)^\alpha \right]^{\frac{1}{1-\gamma}} = \lambda_t.$$

Since

$$\left(\frac{\bar{Q}_t}{Y_t}\right)^{-1/\xi} = M^{1/\xi},$$

we get

$$\lambda_\theta B_1 \left[\frac{\gamma A_t M^{1/\xi}\kappa}{w_t(0)} \left(\frac{\alpha M^{1/\xi}\kappa \bar{Q}_t}{R_t} \right)^\alpha \right]^{\frac{1}{1-\gamma}} = \lambda_t.$$

Now use

$$\frac{\alpha M^{1/\xi}\kappa \bar{Q}_t}{R_t} = \frac{K_t}{\lambda_\theta B_2}.$$

Thus,

$$\lambda_\theta B_1 \left[\frac{\gamma A_t M^{1/\xi}\kappa}{w_t(0)} \left(\frac{K_t}{\lambda_\theta B_2} \right)^\alpha \right]^{\frac{1}{1-\gamma}} = \lambda_t.$$

Raise both sides to power $(1-\gamma)$:

$$\frac{\gamma A_t M^{1/\xi}\kappa}{w_t(0)}\left(\frac{K_t}{\lambda_\theta B_2}\right)^{\alpha}=\left(\frac{\lambda_t}{\lambda_\theta B_1}\right)^{1-\gamma}.$$

Solve for $w_t(0)$:

$$w_t(0)=\gamma A_t M^{1/\xi}\kappa\left(\frac{K_t}{\lambda_\theta B_2}\right)^{\alpha}\left(\frac{\lambda_t}{\lambda_\theta B_1}\right)^{\gamma-1}.$$

Since $M^{1/\xi}=(\lambda_\theta B_3)^{\frac{1}{\xi-1}}$:

$$w_t(0)=A_t\gamma\kappa(\lambda_\theta B_3)^{\frac{1}{\xi-1}}\left(\frac{\lambda_t}{\lambda_\theta B_1}\right)^{\gamma-1}\left(\frac{K_t}{\lambda_\theta B_2}\right)^{\alpha}.$$

We are now ready to solve $\bar{Q}_t$. We have:

$$\bar{Q}_t=\left[A_t\left(\frac{\gamma}{w_t(0)}\right)^{\gamma}\left(\frac{\alpha}{R_t}\right)^{\alpha}\left(\kappa Y_t^{1/\xi}\right)^{\alpha+\gamma}\right]^{\eta^Q}.$$

Use $Y_t=M\bar{Q}_t$:

$$Y_t^{1/\xi}=(M\bar{Q}_t)^{1/\xi}=M^{1/\xi}\bar{Q}_t^{1/\xi}.$$

Thus

$$(\kappa Y_t^{1/\xi})^{\alpha+\gamma}=\kappa^{\alpha+\gamma}M^{\frac{\alpha+\gamma}{\xi}}\bar{Q}_t^{\frac{\alpha+\gamma}{\xi}}.$$

Also

$$\left(\frac{\alpha}{R_t}\right)^{\alpha}=\left(\frac{K_t}{\lambda_\theta B_2\kappa}\right)^{\alpha}\bar{Q}_t^{-\alpha}M^{-\alpha/\xi}.$$

Combine terms:

$$\bar{Q}_t=\left[A_t K_t^{\alpha}\kappa^{\gamma}\lambda_\theta^{-\alpha}B_2^{-\alpha}M^{\gamma/\xi}(\gamma/w_t(0))^{\gamma}\bar{Q}_t^{-\alpha+\frac{\alpha+\gamma}{\xi}}\right]^{\eta^Q},$$

and define

$$C_{\text{in}}\equiv A_t K_t^{\alpha}\kappa^{\gamma}\lambda_\theta^{-\alpha}B_2^{-\alpha}M^{\gamma/\xi}\left(\frac{\gamma}{w_t(0)}\right)^{\gamma}.$$

Then the fixed point is

$$\bar{Q}_t=\left[C_{\text{in}}\bar{Q}_t^{-\alpha+\frac{\alpha+\gamma}{\xi}}\right]^{\eta^Q}.$$

Take logs to get:

$$\log \bar{Q}_t = \eta^Q \log C_{\text{in}} + \eta^Q \left(-\alpha + \frac{\alpha + \gamma}{\xi} \right) \log \bar{Q}_t.$$

Then rearrange terms:

$$\left[1 - \eta^Q \left(-\alpha + \frac{\alpha + \gamma}{\xi} \right) \right] \log \bar{Q}_t = \eta^Q \log C_{\text{in}}.$$

Thus

$$\log \bar{Q}_t = \frac{\eta^Q}{1 - \eta^Q \left(-\alpha + \frac{\alpha+\gamma}{\xi} \right)} \log C_{\text{in}}.$$

Finally, exponentiate:

$$\bar{Q}_t = C_{\text{in}}^{\frac{\eta^Q}{1 - \eta^Q \left(-\alpha + \frac{\alpha+\gamma}{\xi} \right)}}.$$

The final solutions can be represented as follows

$$w_t(0) = A_t \gamma \kappa (\lambda_\theta B_3)^{\frac{1}{\xi - 1}} \left(\frac{\lambda_t}{\lambda_\theta B_1} \right)^{\gamma - 1} \left(\frac{K_t}{\lambda_\theta B_2} \right)^{\alpha},$$

$$\bar{Q}_t = C_{\text{in}}^{\frac{\eta^Q}{1 - \eta^Q \left(-\alpha + \frac{\alpha+\gamma}{\xi} \right)}},$$

$$Y_t = M \bar{Q}_t,$$

$$R_t = \frac{\lambda_\theta B_2 \alpha \kappa \, \bar{Q}_t \, (\lambda_\theta B_3)^{\frac{1}{\xi - 1}}}{K_t}.$$

A.4 Proof of Lemma 4.1

Consider the equation

$$G(x; z) \equiv b\, x^{\psi} + x - z = 0, \qquad x \in [0, \infty),$$

with $0 < \psi < 1$ and $z > 0$. Define

$$H(x) \equiv b\, x^{\psi} + x,$$

so that the solution $x(z)$ satisfies $H(x) = z$. $G(x;z)$ corresponds to equation (36) if we denote

$$b \equiv \frac{\big((\xi-1)\big(\gamma-\psi(1-\alpha)\big)-\psi\big)}{\gamma\big(1+(1-\alpha-\gamma)(\xi-1)\big)}\left(\frac{1}{\lambda_x}\right)^{\psi},$$

$$z \equiv \frac{1+(\xi-1)(1-\alpha)}{1+(1-\alpha-\gamma)(\xi-1)} z_t + \lambda_\theta.$$

Our objective is to prove the solution of $G(x;z) = 0$, $x(z)$ increases with z.

Case 1: $b \geq 0$**.** Then

$$H'(x) = 1 + b\psi x^{\psi-1} > 0 \quad \text{for all } x > 0,$$

so H is strictly increasing. By the implicit function theorem,

$$\frac{dx}{dz} = \frac{1}{1+b\psi x^{\psi-1}} > 0.$$

Hence $x(z)$ is strictly increasing in z, and a decrease in z necessarily decreases $x(z)$.

Case 2: $b < 0$**.** Then

$$H'(x) = 1 + b\psi x^{\psi-1}, \qquad H''(x) = b\psi(\psi-1)x^{\psi-2} > 0 \text{ for } x > 0,$$

so H' is strictly increasing. There exists a unique $x^* > 0$ such that $H'(x^*) = 0$. Therefore, H is strictly decreasing on $(0, x^*)$ and strictly increasing on (x^*, ∞). Since $H(0) = 0$ and $z > 0$, the solution $x(z)$ must lie on the increasing branch, i.e. $x(z) > x^*$, where $H'(x(z)) > 0$. Applying the implicit function theorem again gives

$$\frac{dx}{dz} = \frac{1}{1+b\psi x^{\psi-1}} > 0.$$

Conclusion. For all real b, under $0 < \psi < 1$ and $z > 0$, the equation $b\,x^{\psi} + x - z = 0$ has a unique solution $x(z) \in [0, \infty)$, and $x(z)$ is strictly increasing in z. In particular, a decrease in z necessarily decreases $x(z)$. ■

A.5 The effects of shocks to λ_θ

Consider the equilibrium condition

$$-\lambda_t = b\left(\frac{\lambda_t}{\lambda_x}\right)^\psi - dz_t - \lambda_\theta, \tag{48}$$

where all variables are strictly positive, $0 < \psi < 1$, $d > 0$, and $z_t > 0$. This equation implicitly defines $\lambda_t = \lambda_t(\lambda_\theta)$.

Step 1. Comparative statics for $\left(\frac{\lambda_t}{\lambda_x}\right)^{2\psi}\lambda_\theta^{-2}$.

Define

$$A \equiv \left(\frac{\lambda_t}{\lambda_x}\right)^{2\psi}, \qquad B \equiv A\lambda_\theta^{-2}.$$

By the implicit function theorem, differentiating (48) with respect to λ_θ gives

$$\frac{\partial\lambda_t}{\partial\lambda_\theta} = \frac{1}{1 + b\psi\lambda_x^{-\psi}\lambda_t^{\psi-1}} > 0.$$

Taking logarithms of B and differentiating:

$$\frac{1}{B}\frac{\partial B}{\partial\lambda_\theta} = 2\psi\frac{1}{\lambda_t}\frac{\partial\lambda_t}{\partial\lambda_\theta} - \frac{2}{\lambda_\theta} = 2\left(\frac{\psi}{\lambda_t(1 + b\psi\lambda_x^{-\psi}\lambda_t^{\psi-1})} - \frac{1}{\lambda_\theta}\right).$$

From (48), we can write

$$\lambda_\theta = \lambda_t\left(1 + b\lambda_x^{-\psi}\lambda_t^{\psi-1}\right) - dz_t,$$

so that

$$\frac{\psi}{\lambda_t(1 + b\psi\lambda_x^{-\psi}\lambda_t^{\psi-1})} < \frac{1}{\lambda_\theta}.$$

Hence

$$\frac{\partial B}{\partial\lambda_\theta} < 0.$$

This shows that $B = (\lambda_t/\lambda_x)^{2\psi}\lambda_\theta^{-2}$ increases when λ_θ decreases.

Step 2. Comparative statics for $\left(e(\lambda_t/\lambda_x)^{2\psi} + mz_t\right)^2\lambda_\theta^{-2}$.

Here we have $e > 0$ and $m > 0$. Define

$$A \equiv \left(\frac{\lambda_t}{\lambda_x}\right)^{2\psi}, \qquad S \equiv \left(eA + mz_t\right)^2 \lambda_\theta^{-2}.$$

Take logarithms and differentiate with respect to λ_θ:

$$\frac{1}{S}\frac{\partial S}{\partial \lambda_\theta} = 2\frac{e\partial_{\lambda_\theta} A}{eA + mz_t} - \frac{2}{\lambda_\theta}.$$

Since $A = (\lambda_t/\lambda_x)^{2\psi}$ and λ_x is constant, the chain rule gives

$$\partial_{\lambda_\theta} A = 2\psi(\lambda_t)^{2\psi-1}\frac{1}{\lambda_x^{2\psi}}\frac{\partial \lambda_t}{\partial \lambda_\theta} = 2\psi\frac{A}{\lambda_t}\frac{\partial \lambda_t}{\partial \lambda_\theta}.$$

Here we used the identity $A = (\lambda_t)^{2\psi}/(\lambda_x)^{2\psi}$.

From Step 1, we already have

$$\frac{\partial \lambda_t}{\partial \lambda_\theta} = \frac{1}{1 + b\psi\lambda_x^{-\psi}\lambda_t^{\psi-1}} > 0.$$

Substituting into the derivative of S:

$$\frac{e\partial_{\lambda_\theta} A}{eA + mz_t} = \frac{2\psi eA}{(eA + mz_t)\lambda_t(1 + b\psi\lambda_x^{-\psi}\lambda_t^{\psi-1})} \leq \frac{2\psi}{\lambda_t(1 + b\psi\lambda_x^{-\psi}\lambda_t^{\psi-1})}.$$

Finally, using the inequality established in Step 1, we conclude

$$\frac{\partial S}{\partial \lambda_\theta} < 0,$$

so S increases when λ_θ decreases.

Conclusion. Both quantities

$$B = \left(\frac{\lambda_t}{\lambda_x}\right)^{2\psi}\lambda_\theta^{-2}, \quad S = \left(e(\lambda_t/\lambda_x)^{2\psi} + mz_t\right)^2 \lambda_\theta^{-2}$$

are strictly decreasing in λ_θ. Section 5 shows that the TFPQ dispersion can be expressed as B, and the TFPR dispersion can be expressed as

$$\text{Var}_t(\log TFPR_{jt}) = S + (\frac{\eta^Q}{\xi})^2 \left(\gamma^2\sigma_1^2 + \alpha^2\sigma_2^2\right)$$

if

$$e = 1 - \frac{1-\psi}{1+(1-\alpha-\gamma)(\xi-1)}, \qquad m = \frac{\gamma}{1+(1-\alpha-\gamma)(\xi-1)}.$$

Therefore, a decrease in λ_θ increases both the dispersions of TFPQ and TFPR.