The "Matthew Effect" and Market Concentration: Search Complementarities and Monopsony Power

Jesús Fernández-Villaverde
University of Pennsylvania

Federico Mandelman
Federal Reserve Bank of Atlanta

Yang Yu
Shanghai University of Finance and Economics

Francesco Zanetti*
University of Oxford

February 8, 2021

Abstract

This paper develops a dynamic general equilibrium model with heterogeneous firms that face search complementarities in the formation of vendor contracts. Search complementarities amplify small differences in productivity among firms. Market concentration fosters monopsony power in the labor market, magnifying profits and further enhancing high-productivity firms' output share. Firms want to get bigger and hire more workers, in stark contrast with the classic monopsony model, where a firm aims to reduce the amount of labor it hires. The combination of search complementarities and monopsony power induces a strong "Matthew effect" that endogenously generates superstar firms out of uniform idiosyncratic productivity distributions. Reductions in search costs increase market concentration, lower the labor income share, and increase wage inequality.

Keywords: Market concentration, superstar firms, search complementarities, monopsony power in the labor market.

JEL classification: C63, C68, E32, E37, E44, G12.

*We thank Michael Peters for an outstanding discussion and most helpful suggestions to simplify our analysis, and Luis Garicano and Gustavo Ventura for insightful comments. Ryan Zalla provided outstanding research assistance. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Zanetti gratefully acknowledges financial support from the British Academy (MD20\200025). The usual disclaimer applies.

1 Introduction

Merton (1968) famously identified the "Matthew effect": *For whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them.* Merton's insight was straightforward: small exogenous differences get amplified, often by orders of magnitude, by the endogenous responses of agents to those small differences. In Merton's original example, small differences in scientific productivity are magnified by the extreme inequality in the allocation of limited resources (grant money, graduate students, journal pages). Imagine a national research agency that only has enough money to finance one research lab, but can correctly identify ex-ante differences in scientific productivity among professors. Even if professor A is just 1% more productive than professor B, professor A will get the funds to run a lab, and become famous. In contrast, professor B will linger in obscurity.

This paper argues that a "Matthew effect" drives the high levels of market concentration observed in the data, with a few superstar firms and many small firms, even when the differences in productivity among firms are minor. Our "Matthew effect" operates through strategic complementarities under direct search and monopsony power in the labor market.

Let us unpack these mechanisms. Firms need to sign vendor contracts with their suppliers before producing. This process involves costly search. For example, to operate an ice cream truck company, one needs to find a supplier of milk, a supplier of waffle cones, a supplier of toppings, a supplier of ice cream mixers, a supplier of trucks, etc. This search is costly in terms of time and resources.[1]

Intermediate-goods suppliers search with higher effort when they are more productive because the potential profit from a vendor match is larger. For instance, a high-productivity waffle cone manufacturer will pay the costs in time and resources of attending a trade fair for the restaurant industry, but a low-productivity manufacturer will not. Conversely, final-goods producers send more buying agents when they know that the intermediate-goods suppliers are searching for buyers. This decision is particularly salient with directed search: i.e., the ice cream company receives a directory of the booths at the trade fair and, upon seeing that a high-productivity waffle cone producer is attending the fair, sends an agent to visit that booth right away.

[1] This example is taken from the fascinating tale of how Mister Softee tried and failed to establish an ice cream business in Suzhou in the 2000s. See https://supchina.com/podcast/the-rise-and-fall-of-a-suzhou-soft-serve-baron/.

Hence, high-productivity intermediate-goods vendors will form more matches with final-goods firms. In the terminology of Bulow et al. (1985), a strategic complementarity appears because the stronger the search of the intermediate-goods producers, the stronger the directed search of the final-goods firms and vice versa.

The search complementarity mechanism will induce a highly concentrated distribution of firms' size, vacancies, and output. Small differences in productivities among intermediate-goods firms will result in large differences in firms' size and output. Interestingly, this "Matthew effect" transforms a uniform distribution of idiosyncratic productivity into a highly skewed firm distribution characterized by the presence of superstar firms. In contrast to much of the existing literature, we do not need fat tails in the distribution of firms' idiosyncratic properties (e.g., productivity, demand shifters, etc.) to generate this result.

The process, however, does not end here. In our model, there is a second mechanism reinforcing the "Matthew effect": the labor market power of large firms. If firm A is the only waffle cone producer in a region, it has market power when hiring bakers. This market power creates two effects. First, more productive firms will pay higher wages (as we observe in the data): the surplus of a labor match is larger and the worker will receive part of it. However, conditional on productivity, the wages will be a lower share of the surplus. That is, large firms will have a lower labor income share (again, as we see in the data). The higher the market power, the stronger these two effects will be.

This labor market power will also have a consequence for search under strategic complementarities. Since larger firms will keep a higher share of the surplus of a labor match, larger firms will also have a stronger incentive to search with more intensity (beyond the direct effect of higher productivity). That higher search intensity will be reinforced by the response of final-good producers. That is, search complementarities transform labor market power into significant differences in the market structure and firms' sizes and output.

The mechanisms outlined above also have significant business cycle implications. After a negative aggregate productivity shock, firms will decrease their search effort. This fall in search effort will amplify the original shock and make it more persistent over time. Furthermore, the reduction in aggregate productivity will affect low-productivity firms disproportionately because their profit margins are smaller. Thus, low-productivity firms will reduce their search effort more than high-productivity firms, leading to more market concentration.

To explore these mechanisms formally, we first develop a simple model that isolates the effect of search complementarities and monopsony power for the distribution of firms, market concentration, and the effect of aggregate shocks. The model will not be designed for quantitative work, but it illustrates all our ideas transparently.

Next, we build a dynamic general equilibrium model with heterogeneous firms and frictional labor markets. At the core of our model, we embed the integration of complex production processes that require long-lasting vendor relationships among different intermediate- and final-goods firms. This assumption is motivated by the strong empirical evidence on the existence of sophisticated multi-firm value-chains. In the model, the intermediate-goods firms manage a continuum of product lines and search for buyers of those goods. Final-goods producers assign buying agents to find those product lines and sign contracts with them. The two-sided search among firms leads to strategic complementarities: intermediate-goods firms' optimal search effort increases with the visits of final-goods producers' buying agents, and vice versa.

Our model is enriched with monopsony power in the labor market by adding search and matching frictions that let firms set wages below the marginal product of labor. To do so, we consider a matching technology in the labor market à la Butters (1977), which allows multiple workers to apply to a single vacancy randomly. This environment provides the firm with the power to select one worker among multiple job applicants. A firm operating in concentrated markets can induce workers to accept a low wage, since it could threaten a worker with forgoing future job offers if the worker declines a wage offer. Intuitively, if a few firms dominate a segment of the labor market, a prospective employee who rejects a job offer may be excluded *de facto* from future consideration, since the firm will prefer other job applicants. Top firms exploit their market power to offer a low wage and gain profits, which in turn increases profits and encourages firms with market power to search more actively and attract more visits from potential partners. Labor market power enhances search complementarities in the goods market and it is a critical force in generating market concentration.

We calibrate the model to match quarterly U.S. data and then use it as a measurement device. Given that we want to be consistent with the differences in measured total factor productivity across manufacturing plants, the rate of factory idleness, and labor market observations, how much monopsony power do we need to account for market concentration by the top 10% of firms? From this exercise, we back up a mild degree of monopsony power: the equivalent of

firms being able to punish workers who reject an offer by turning down their future applications for around six months. This reasonable degree of monopsony power enhances our trust in the model as a quantitative laboratory for further exercises.

How can we use our model to think about the recent experience of advanced economies? Several studies have documented a steady increase in market concentration over the past three decades. For example, Autor et al. (2020) show that, starting in the early 1980s, sales moved toward the most productive firms across U.S. industries. At the same time, labor markets became increasingly dominated by fewer players, lowering the bargaining power of workers and deepening income inequality (Wu, 2019). Monopsony power in labor markets has also boosted firm profits and market concentration (Hershbein et al., 2020). Furthermore, the early 1980s witnessed the onset of the Great Moderation, a sustained period with low volatility.

Our model presents a simple mechanism to jointly account for all these observations: a fall in the costs of signing vendor contracts. While all firms increase their search effort when the costs of signing vendor contracts fall, high-productivity firms gain disproportionately, as their search effort decisions are more non-linear on search costs than those of low-productivity firms. Thus, in our model, lower search costs lead to i) higher market concentration, ii) lower labor income shares, iii) more labor market power, and iv) lower responses of output to aggregate shocks. We observe the same four facts in the U.S. data.

What do we have in mind in terms of lower search costs? Improvements in IT technology. The internet has made it much easier to identify vendors and suppliers, to manage them, to run logistics and inventories, to handle sophisticated value-added chains, etc.

While market structure changes the response of the economy to aggregate shocks, the effect also works in the opposite direction: the market structure is endogenously determined by the realization of aggregate shocks. The persistently low search effort and output by low-productivity firms after a large negative aggregate shock result in an increasing market exit among the production lines owned by these firms. As a result, deep slumps render the market structure increasingly concentrated.

This behavior also matches the empirical evidence. Using U.S. Census firm-level panel data, Salgado et al. (2019) show that business cycles are skewed. That is, during recessions, a subset of firms significantly underperforms, leading to a large fat left tail in the production distribution. The process is reversed in expansions, when the right tail becomes fatter.

The market concentration effect of negative aggregate shocks also appears in the customer base literature. Chevalier and Scharfstein (1996) find that, during recessions, small (and liquidity constrained) firms invest less in expanding their customer base and raise prices to boost their liquidity positions. Bigger firms expand their customer base in recessions, which renders the market more concentrated afterward. Investment in the customer base resembles network formation in the context of our model.

Our paper connects with many other different areas of research. First, and most importantly, there is a tradition of papers exploring firms' size distribution that goes back to the span-of-control model by Lucas (1978). One can think about our theory as an endogenous determinant of the span-of-control: production links must be formed either within firms or between firms. Directed search and strategic complementarities determine how many of these links are created in our model.

Our theory has two advantages with respect to a simple span-of-control model. First, we can generate a larger dispersion in firms' size that is compatible with observed differences in measured total factor productivity across plants. Second, our model allows us to have a simple margin to account for the simultaneous increase in market concentration and fall in the labor income share: the reduction in direct search costs, which we link with observed improvements in IT. In a simple span-of-control model, one would need to resort to either production functions getting closer to linear or a change in the underlying distribution of managerial talent to generate similar outcomes (see, however, for more flexible versions of the span-of-control model, Garicano and Rossi-Hansberg, 2006).

Linked with the Lucas tradition, much recent research focuses on growing market concentration. Aghion et al. (2019) find that IT explains the lower cost of production for bigger firms, with newer firms (or less efficient ones) finding it increasingly difficult to contest them. Similarly, Akerman et al. (2013), Bessen (2017), and Unger (2019) attribute this winner-takes-all mechanism to economies of scale arising from intangible capital and advances in information technology, which greatly improve the product and inventory logistics.

A second strand of the literature has been devoted to understanding the recent decline in the labor share of output. See, among many others, Elsby et al. (2013) and Karabarbounis and Neiman (2014). De Loecker and Eeckhout (2018) attribute this phenomenon to a raise in weighted average firm markups, with Gutiérrez and Philippon (2018) emphasizing the role of

weakening antitrust enforcement. Closer to our work is Autor et al. (2020), who argue that the decline in the labor share should be attributed to the reallocation of market share toward "superstar" firms with higher markups. Consistent with this hypothesis Peters (2020) finds that markups vary systematically across firms, with incumbents investing to increase productivity growth (further raising markups). However, a creative destruction mechanism also exists in this last paper, as new and more efficient firms displace incumbents. Higher entry costs or frictions may thus deter this key pro-competitiveness mechanism.

Third, our paper also contributes to the growing theoretical literature on monopsony in labor markets. Examples include Ashenfelter et al. (2010), Berger et al. (2019), Manning (2011), Card et al. (2018), and Lamadon et al. (2019). In turn, empirical papers finding substantial market power in the labor market include Azar et al. (2019), Falch (2010), Ransom and Sims (2010), and Matsudaira (2014).

Fourth, starting with the seminal contributions of Diamond (1982) and Weitzman (1982), several papers have linked strategic complementarities to aggregate fluctuations. See, without being exhaustive, Diamond and Fudenberg (1989), Huo and Ríos-Rull (2013), and Kaplan and Menzio (2016). We depart, though, from those papers in our focus on how strategic complementarities and monopsony power create a "Matthew effect" on market concentration and analyze how those mechanisms interact with aggregate shocks.

Finally, in Fernández-Villaverde et al. (2019), we explore how fiscal policy and strategic complementarities interplay to explain the persistence of both the business cycle and the unemployment rate. Our previous work abstract from firm heterogeneity, market concentration, and monopsonistic labor markets. It focuses, instead, on the possibility of multiple equilibria, which do not play any role in the current paper.

The remainder of the paper is structured as follows. Section 2 develops a simple model to outline the main ideas in our paper. Section 3 extends the simple model to a more fleshed-out dynamic general equilibrium model. Section 4 calibrates the model to U.S. data and uses it to measure monopsony power in the labor market. Section 5 presents our quantitative findings. Section 6 concludes.

2 A simple model

We start our analysis by presenting a simple model, with a closed-form solution, that embodies the central mechanisms we want to explore. The model begets search complementarities through the interplay of directed search and endogenous search effort. We will extend this simple model along two dimensions. First, we will incorporate endogenous variations in market concentration through the entry and exit of product lines. Second, we will introduce monopsony power in the labor market to evaluate how such a power interrelates with market concentration. While neither the simple model nor its extensions are designed for quantitative work (we will impose restrictive functional forms and parametric choices), the mechanisms that drive the results are transparent. In Section 3, we will present an extended model that gives us quantitative predictions.

2.1 Environment

Time is discrete and infinite. The economy is composed of $J+1$ islands. Each island $j \in \{1, 2, ..., J\}$ hosts an intermediate-goods producer (I), such as General Mills or Kellogg's. Each intermediate-goods producer operates a unitary measure of product lines, such as the many food brands manufactured by General Mills, and has an idiosyncratic productivity shock x_j. The central island $J+1$ hosts a representative household and a final-goods producer (F), such as Walmart, that purchases food items from General Mills.

The intermediate-goods producers and the final-goods producer must form a vendor relationship before starting production, e.g., General Mills will not produce breakfast cereals if it does not have access to a supermarket in which to sell them. The intermediate-goods producer either does not have the technology to reach consumers directly or it is too costly for it to do so. In fact, General Mills and similar firms do not sell directly to final consumers.

The process of search to form vendor relationships is directed. At the start of each period t, the firm F decides how many buying agents to send to each island to maximize its total profits. The firm F can pick any positive real number of buying agents.

Production begins when a buying agent from firm F signs a contract with a single product line in firm I. In our example, Walmart decides how many buying agents to send to General Mills. Each Walmart buying agent will work with a General Mills brand manager to reach a vendor contract for that brand. The more buying agents Walmart sends, the more contracts

can be signed. The buying agent, in real life, is a bundle of different workers (from logistics, legal, marketing). For our purposes, we can ignore that margin, since we focus on Walmart's total buying cost. The number of active product lines is equal to the number of buying agents that sign a contract. The total output for each signed contract is $2z_t x_j$, where z_t is an aggregate productivity shock in period t. Output is equally split between firm F and firm I.

Buying agents who fail to sign a contract with a product line in firm I withdraw from the island, while the unmatched product line of firm I stays idle for the period. A law of large numbers holds in the economy and, thus, probabilities equate to realized shares. That is, if the equilibrium implies a 0.32 probability of meeting on any island j, a match occurs in 32% of product lines on this island.

The representative household owns all the firms in the economy, receives the aggregate net profits from them, and consumes. Since our paper focuses on firm heterogeneity, the representative household assumption simplifies our analysis.

At the end of each period t, all the vendor matches are dissolved, buying agents from firm F return to their headquarters, and the searching process restarts *ex novo* in period $t+1$. This assumption transforms the dynamic programming problem of the firms into a sequence of static optimization problems. Figure 1 summarizes the structure of the economy.

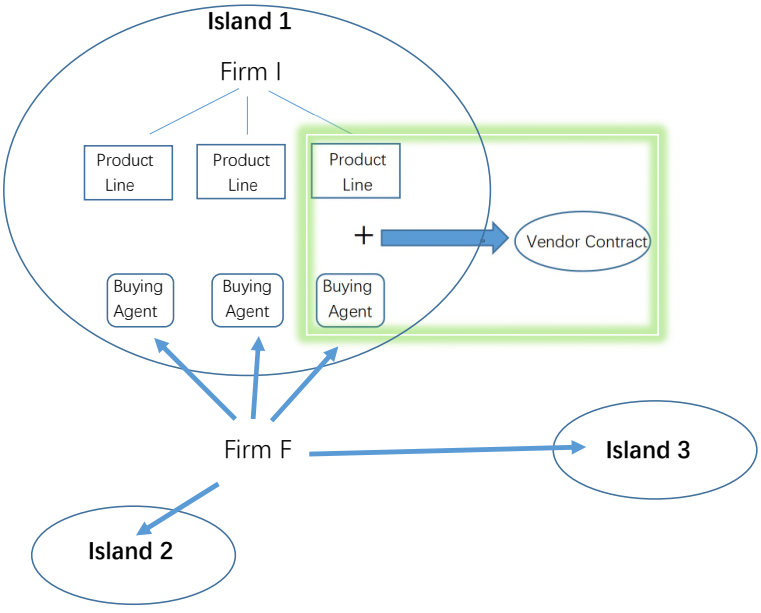

Figure 1: Structure of the economy

A matching function determines the probability of meeting a vendor and signing a contract.

The likelihood of matching on each island j depends on the measure of buying agents from firm F, n_j^F, the measure of product lines owned by firm I, n_j^I, and the effort firm I exerts in finding a buying agent from firm F, $\sigma_j^I \in [0,1]$ (to save on notation, we will only use a subindex t for a variable when needed to avoid confusion). More precisely, the measure of newly formed matches is established by a matching function that is affine on σ_j^I and Cobb-Douglas between n_j^F and n_j^I:

$$M\left(\sigma_j^I, n_j^I, n_j^F\right) = \phi\sigma_j^I \left(n_j^F\right)^{\frac{1}{2}} \left(n_j^I\right)^{\frac{1}{2}}.$$

The matching probability for each product line of firm I is $M\left(\sigma_j^I, n_j^I, n_j^F\right)/n_j^I$ and for each buying agent $M\left(\sigma_j^I, n_j^I, n_j^F\right)/n_j^F$. Since, by assumption, $n_j^I = 1$, the matching probabilities for firms F and I are $\pi^I\left(\sigma_j^I, n_j^F\right) = \phi\sigma_j^I \left(n_j^F\right)^{\frac{1}{2}}$ and $\pi^F\left(\sigma_j^I, n_j^F\right) = \phi\sigma_j^I \left(n_j^F\right)^{-\frac{1}{2}}$, respectively.

Output on each island j is:

$$y_j = 2\phi\sigma_j^I \left(n_j^F\right)^{\frac{1}{2}} z_t x_j. \tag{1}$$

The cost of search effort for firm I on island $j \in \{1, 2, ..., J\}$ is:

$$c\left(\sigma_j^I\right) = \frac{\left(\sigma_j^I\right)^3}{3}. \tag{2}$$

We pick a power of 3 in the function above for algebraic convenience, but all we need is convexity of the search cost.

Firm F pays a unit cost of sending buying agents equal to κ, which we normalize to $\kappa = \phi/2$. Thus, the consumption the representative household gets from island j is:

$$c_j = 2\phi\sigma_j^I \left(n_j^F\right)^{\frac{1}{2}} z_t x_j - \frac{\left(\sigma_j^I\right)^3}{3} - \kappa n_j^F.$$

2.2 Nash equilibria

To find the Nash equilibria, we consider the problem of firm I on island j that takes the measure of buyers from sector F on its island, n_j^F, as given. The profit function for firm I is:

$$J\left(\sigma_j^I, n_j^F \mid x_j, z_t\right) = \phi\sigma_j^I \left(n_j^F\right)^{1/2} z_t x_j - \frac{\left(\sigma_j^I\right)^3}{3}. \tag{3}$$

Maximizing $J\left(\sigma_j^I, n_j^F \mid x_j, z_t\right)$ with respect to σ_j^I, we obtain the best response function for firm I on island j:

$$\sigma_{j,t}^I = \sqrt{\phi \widehat{n}_j^F z_t x_j} \tag{4}$$

where, to simplify notation, we have defined $\widehat{n}_j^F \equiv \left(n_j^F\right)^{1/2}$.

Let us consider now the problem of firm F. Since the search process in the intermediate-goods market is directed, firm F sends enough buyers to visit island j to exploit all profit opportunities. Hence, firm F's income from sending an additional buying agent to an island (the matching probability times the revenue per signed contract) is equal to the unit cost of sending the agent κ, which we normalize to $\kappa = \phi/2$:

$$\widehat{n}_j^F = \sigma_j^I z_t x_j. \tag{5}$$

Equations (4) and (5) show why we have strategic search complementarities in the sense of Bulow et al. (1985): firm I's search effort is (weakly) increasing in firm F's number of buying agents (equation 4) and firm F's number of buying agents is an affine function of firm I's search effort (equation 5). A bigger search effort from firm I on island j increases the profits for firm F and, thus, attracts a larger measure of buying agents to the island, raising the profits for firm I and further stimulating search effort.

Directed search is at the core of this result: firm F's decision depends on firm I on island j's search effort because firm F can direct its buying agents to island j. With random search, an increment in the search effort of firm I on island j would only affect firm F's decision by changing the revenue of an additional contract on island j times the probability that the additional buying agent would arrive at the island. When J is large, the effect would be negligible.

A (within period and island) pure strategy Nash equilibrium is a tuple $\left\{\sigma_j^I, \widehat{n}_j^F\right\}$ that is a fixed point of (4) and (5). The system has two Nash equilibria in pure strategies. One Nash equilibrium, $\left\{\sigma_j^I, \widehat{n}_j^F\right\} = \{0, 0\}$, is not very interesting and we will ignore it. Also, at the cost of some extra notation, we could assume that a minimum number of matches occur even when $\sigma_j^I = 0$ and this equilibrium would disappear.

The other equilibrium is $\left\{\sigma_j^I, \widehat{n}_j^F\right\} = \left\{\phi z_t^2 x_j^2, \phi z_t^3 x_j^3\right\}$. Then, equation (1) implies that the output on island j is $2\phi^3 z_t^6 x_j^6$, with firm I's search cost being $\frac{1}{3}\left(\sigma_j^I\right)^3 = \frac{1}{3}\phi^3 z_t^6 x_j^6$ and firm F's search cost being $n_j^F \kappa = \frac{1}{2}\phi^3 z_t^6 x_j^6$. Thus, consumption, c_j, after the search costs, is $\frac{7}{6}\phi^3 z_t^6 x_j^6$.

By summing over the islands, we get aggregate output y_t:

$$y_t = 2\phi^3 z_t^6 \sum_{j=1}^{J} x_j^6, \tag{6}$$

and aggregate consumption $c_t = \frac{7}{6}\phi^3 z_t^6 \sum_{j=1}^{J} x_j^6$.

Equation (6) reveals how a Δ difference in productivity leads to a Δ^6 difference in output. The degree of amplification, 6, is determined by the curvature of the search cost function (equation 2). We can increase or decrease the amplification effect by adjusting the search cost function.

To illustrate these derivations, we fix the number of islands j to 3 for the rest of this section. We set $\phi = 0.5^{1/3}$, which implies that, when $z_t x_j = 1$, the matching probability for firm I is 0.5. For the moment, $z_t = 1$. With this choice of parameter values, output on island j is x_j^6. Just for simplicity, we assume that productivity across islands is $x_1 = 0.95$, $x_2 = 1$, and $x_3 = 1.05$.

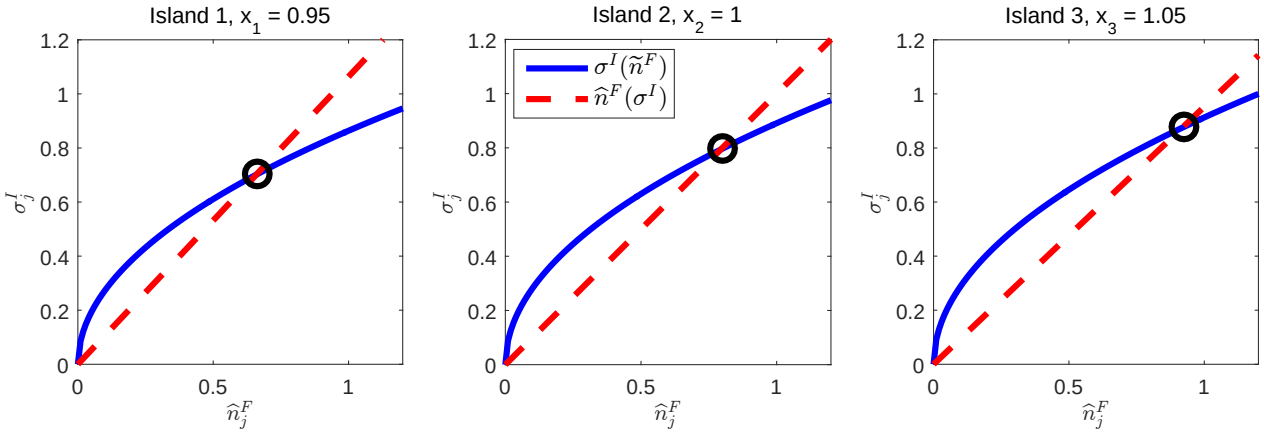

Figure 2: Nash equilibria across islands

Figure 2 plots the best response function of firm I on each island (continuous blue line) and the optimality condition of firm F regarding the number of buying agents sent to the island (discontinuous red line). In the left panel, we plot the functions for island 1; in the center panel, we plot the functions for island 2; and in the right panel, we plot the functions for island 3. The circle markers plot the Nash equilibria, $\left\{\sigma_j^I, \widehat{n}_j^F\right\}$, for each island.

As implied by equations (4) and (5), higher productivity triggers strong strategic complementarities and a "Matthew effect" of degree 6. While island 3 is only 10.5% more productive than island 1, it exerts 22% more search effort and attracts 35% more visits from firm F than

island 1, which generates an output 82% larger. Specifically, $\left(\sigma_1^I, \widehat{n}_1^F, y_1\right) = (0.72, 0.68, 0.74)$, in comparison with $\left(\sigma_3^I, \widehat{n}_3^F, y_3\right) = (0.88, 0.92, 1.34)$.

A similar amplification appears after an aggregate productivity shock. The left panel of Figure 3 plots a one-period aggregate productivity shock that decreases z_t from its original value of 1 to 0.95 in the second period and fully recovers in the third period. The right panel of Figure 3 plots the impulse-response function (IRF) of output to the shock in the left panel in each of our three islands. A reduction of 5% in aggregate productivity results in a 26% fall in output. Given our strong parametric assumptions, the reduction in output is proportional across islands and independent of z_t and κ (the unit cost of sending buying agents). Also, the response of output to aggregate productivity has no persistence. The lack of persistence occurs because, in this version of the model, the distribution of island size is exogenous.

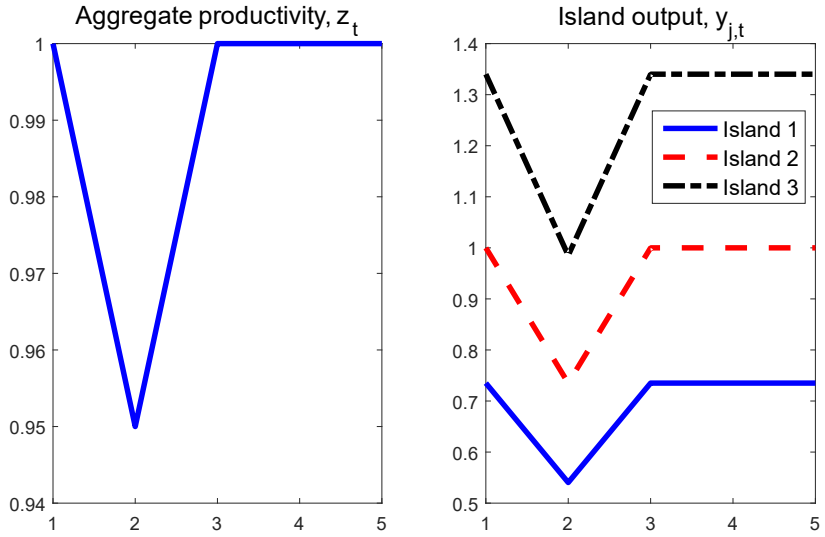

Figure 3: IRFs to negative aggregate productivity shock

In comparison, the extended model with the endogenous entry of product lines we will discuss in the next subsection will generate i) changes in market concentration after a productivity shock or a reduction in κ, and ii) a sluggish adjustment process and a persistent response of output to aggregate productivity shocks. These two phenomena will make the "Matthew effect" even more potent and give us a theory, through the reduction in κ, to account for increased market concentration. Let us, then, enrich our simple model by introducing entry and exit of product lines for intermediate-goods firm I.

2.3 Endogenous market concentration

We show now that the entry and exit margin will deliver three new results: i) the "Matthew effect" becomes even more prominent than before; ii) market concentration will depend on the cost of signing a vendor contract; and iii) aggregate productivity shocks change market concentration and make the effects of short-lived aggregate shocks persistent and asymmetric.

We assume that unmatched product lines of the firm in sector I on each island j become obsolete and exit the economy with probability χ. Conversely, new product lines are created at the constant rate n in each period t. This assumption can be micro-founded with a fixed operation cost with a cash-on-hand constraint: in the absence of a positive cash flow, the product line is forced to close. To simplify, we will assume that firms decide on search effort without accounting for the possibility that forgoing a match may make them obsolete in the next period (we will remove this simplification in Section 3). For simplicity, the entry rate is exogenous. Our results hold, with heavier notation, if entry is endogenous.

The measure of product lines on each island j follows:

$$n^I_{j,t+1} = n^I_{j,t} - \underbrace{\chi \cdot \left[1 - \pi^I\left(\sigma^I_{j,t}, \widehat{n}^F_{j,t}\right)\right] n^I_{j,t}}_{\text{Exit}} + \underbrace{n}_{\text{Entry}}, \tag{7}$$

where $\chi \cdot \left[1 - \pi\left(\sigma^I_{j,t}, \widehat{n}^F_{j,t}\right)\right]$ is the fraction of unmatched product lines that exit island i, and n is the measure of new entrance of product lines. The measure $n^I_{j,t+1}$ increases in the matching probability $\pi^I\left(\sigma^I_{j,t}, \widehat{n}^F_{j,t}\right)$. Thus, the exit rate for product lines is lower on an island with a higher probability of establishing a vendor contract with firm F, leading to a subsequent higher measure of active product lines on the island. Equation (7) implies that the steady-state measure of product lines is:

$$n^I_j = \frac{n}{\chi \cdot \left(1 - \pi^I_j\right)} \tag{8}$$

We set $\chi = 0.282$ to generate a steady-state measure of product lines on island 3 of 1 that is consistent with that in our previous subsection (i.e., a steady-state measure of product lines equals 0.58 and 0.72 on islands 1 and 2, respectively). Figure 4 shows that the steady-state output share on islands 1, 2, and 3 is equal to 0.17, 0.29, and 0.54, respectively. While island 3 is still only 10.5% more productive than island 1 (as in the case without entry-exit), island 3's output is now 209% larger than island 2's output, instead of 82% as without entry-exit. Equation

(8) tells us why. Due to its higher productivity, island 3 searches more actively, attracts more vendors, and accumulates more product lines. As π_j^I gets close to one, this mechanism becomes arbitrarily large. That is, entry-exit generates an even stronger "Matthew effect."

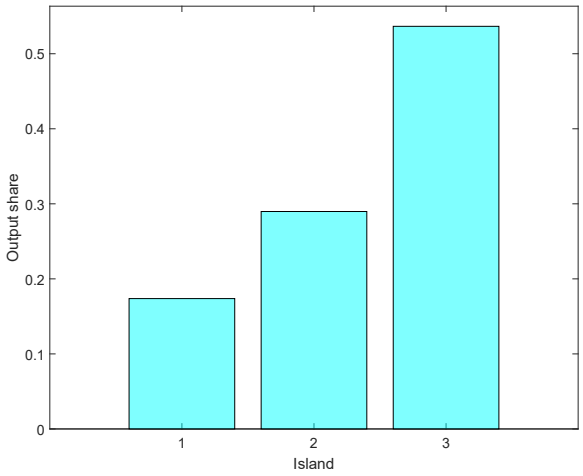

Figure 4: Output share across islands

But market concentration also depends on the cost of signing a vendor contract. For example, imagine that due to the enhancements in search technology (e.g., better logistics software), it becomes cheaper for firm F to send buying agents to each island. Formally, we let the unit cost of visiting each island, κ, decrease at a constant 1% rate per period (i.e., $\kappa_t = 0.99^{t-1} \cdot \phi/2$).

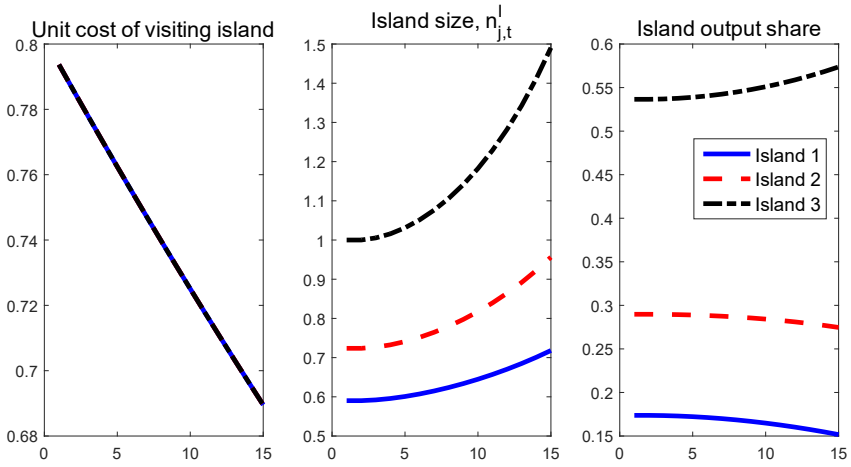

Figure 5: Reduction in search costs

Figure 5 plots the unit cost of visiting each island (left panel), the measure of productive lines for firm I (central panel), and the final output share (right panel) for each island. The decline in unit search cost attracts more buying agents from firm F to all islands and, thus, increases the

probability of forming a vendor relationship and the number of active product lines ($n^I_{j,t}$, middle panel). While all three islands have more active product lines, search complementarities make the increase in $n^I_{j,t}$ proportional to each island's productivity. Therefore, island 3 benefits the most from the decline in κ and the output shares of islands 1 and 2 fall over time. In comparison, in the model without entry and exit, the output on all three islands grows at the same rate, and market concentration remains unchanged. That is, we need both search complementarities *and* entry-exit to transform reductions in search cost into changes in market concentration.

Our result is consistent with the finding in Aghion et al. (2019), who show that the increasing share of output for high-productivity firms is mostly accounted for by a decreasing cost of expanding new businesses. Consider the following example. Historically, each Whole Foods store sourced its products with independent local suppliers (or "local foragers"). Following the Amazon-Whole Foods merger, Amazon took advantage of its leadership in logistics software to revamp the existing Whole Foods vendor contract arrangements and started prioritizing contracts with national, higher-productivity suppliers at the expense of local foragers.

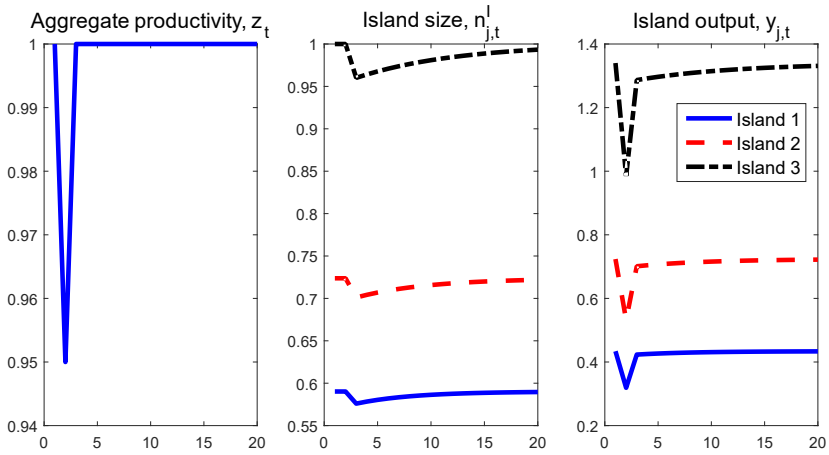

Figure 6: IRFs to negative aggregate productivity shock

Figure 6 shows the IRFs of the measure $n^I_{j,t}$ and output $y_{j,t}$ on the three islands (right panel) to a one-period decrease in aggregate productivity (z_t) from 1 to 0.95 (left panel). Output (aggregate and on each island) falls by 26%, as in the case without entry and exit: both versions of the model behave in the same way at impact. The difference with respect to Figure 3 is that now, through entry and exit, we have i) persistence of the fall in output (even if the productivity shock only lasts for one period) and that ii) such persistence is asymmetric across islands.

2.4 Monopsony power of the labor market

Motivated by the evidence in Berger et al. (2019), Hershbein et al. (2020), and Manning (2020), among others, that links monopsony power in the labor market and market concentration, we investigate how search complementarities and monopsony power interact. We will derive five new results: i) monopsony power lowers wages ceteris paribus; ii) wages grow with the productivity of the firm; iii) monopsony power reduces the marginal effect of the firm's productivity on wages; iv) monopsony power strengthens the "Matthew effect" of productivity differences even further and increases wage inequality; and v) reductions in the cost of signing a vendor contract lower labor income share, but redistributes labor toward higher-productivity jobs.

Our first step before showing these results is to specify labor supply and demand. To keep the model as transparent as possible, we assume that, after the formation of vendor relationships, a measure u_t of workers from the representative household is randomly matched to active product lines ($\sum_j \pi^I_{j,t} n^I_{j,t}$). The labor match lasts for one period and separates at the end of each period. Thus, the meeting probability is equal to one for both sides of the match. A worker's probability of meeting with an active product line on island j is $s_{j,t} = \pi^I_{j,t} n^I_{j,t} / \sum \pi^I_{k,t} n^I_{k,t}$, the share of active product lines on island j.[2]

The wage on island j, $w_{j,t}$, is determined by Nash bargaining between the worker and an active product line. If the worker rejects the wage offer, she becomes unemployed in this period and the active product line receives a zero profit.

To introduce monopsony power on the labor market, we assume that active product lines on the same island negotiate wages in a collective way: if a worker rejects an offer from an active product line on island j, all other active product lines on island j would "punish" the worker by refusing to match with her with probability λ in the next period. For simplicity, we assume that firms have an exogenous commitment to this negotiation rule.

Then, if a worker declines a wage offer from an active product line, she forgoes $w_{j,t} + \lambda s_{j,t+1} \cdot w_{j,t+1}$, the lost wage today plus the probability of losing a wage tomorrow, is proportional to the island's labor market share, $s_{j,t+1}$. Firms will optimally take advantage of this forgone income to increase their profits.

[2]Our modeling choice is equivalent to imposing a Leontief matching function: $M_t = \min(u_t, \sum_j \pi^I_{j,t} n^I_{j,t})$, where we assume $u_t = \sum_j \pi^I_{j,t} n^I_{j,t}$. By doing so, we eliminate the need to keep track of the percentage of unmatched workers or product lines. We can justify the number of workers being a function of the active product lines with the representative household's preferences without wealth effects.

To see this, notice that the total surplus of a labor market match is $LTS_{j,t} = (2z_t x_j - w_{j,t}) + (w_{j,t} + \lambda s_{j,t+1} w_{j,t+1})$, where $(2z_t x_j - w_{j,t})$ and $(w_{j,t} + s_{j,t+1} \cdot w_{j,t+1})$ are the surplus of the active product line and the worker's payoff from the labor market match, respectively (here we implicitly assume linear preferences on income for the worker). Nash bargaining implies that $2z_t x_j - w_{j,t} = \tau \cdot LTS_{j,t}$ and $w_{j,t} + \lambda s_{j,t+1} \cdot w_{j,t+1} = (1-\tau)\, LTS_{j,t}$, where τ and $(1-\tau)$ are the bargaining shares of the active product line and the worker, respectively.

Suppose, first, that labor market punishment is forbidden, i.e., $\lambda = 0$. In this case, the wage in the steady state (with $z_t = z = 1$), $w_j^* = (1-\tau)\, 2x_j$, is a fraction $1-\tau$ of output. The derivative of the wage with respect to the island's productivity x_j is $(1-\tau)\, 2$.

When $\lambda > 0$, the wage in the steady state becomes:

$$w_j = \frac{(1-\tau)2x_j}{1+\tau\lambda s_j} = \frac{1}{1+\tau\lambda s_j} w_j^* < w_j^*. \tag{9}$$

where we can see the monopsony wedge $\frac{1}{1+\tau\lambda s_j} < 1$.[3]

From this expression, we have:

$$\frac{dw_j}{dx_j} = \frac{(1-\tau)\, 2}{1+\tau\lambda s_j} - \frac{\tau\lambda}{(1+\tau\lambda s_j)^2} \frac{\partial s_j}{\partial x_j} < (1-\tau)\, 2. \tag{10}$$

since higher-productivity islands have more active product lines everything else equal ($\frac{\partial s_j}{\partial x_j} > 1$).

Equations (9) and (10) teach us three lessons. First, the monopsony wedge lowers the island's wage i with respect to the case without monopsony power. Second, w_j increases with the island's productivity, but decreases with the island's share of active product lines. The latter change is a general equilibrium effect: the island's share depends on its productivity but also on the productivity of all the other firms in the economy. That is, if firms on other islands are more productive, they will decrease the number of workers on the current island and, therefore, suppress search efforts and wages. Third, wages grow more slowly than productivity in the firms' cross-section.

Figure 7 illustrates these three lessons by plotting the distribution of wages in the steady state of the economy ($z_t = z = 1$) with no monopsony power ($\lambda = 0$) and with monopsony power ($\lambda = 0.1$). Since we calibrate $\tau = 0.5$, we have $(1-\tau)\, 2z_t = 1$. To make our exercise

[3] In particular, $LTS_j = 2x_j + \lambda s_j w_j$ and $2x_j - w_j = \tau \cdot LTS_j$. By combining the two equations, we get: $2x_j - w_j = \tau\,(2x_j + \lambda s_j w_j)$, or $w_j = (1-\tau)\, 2x_j$.

comparable with the previous subsections, we reset $x_1 = 1.9$, $x_2 = 2$, and $x_3 = 2.1$. Then, when labor market punishment is forbidden, firms' profits and the Nash equilibrium are then the same as in subsection 2.3.

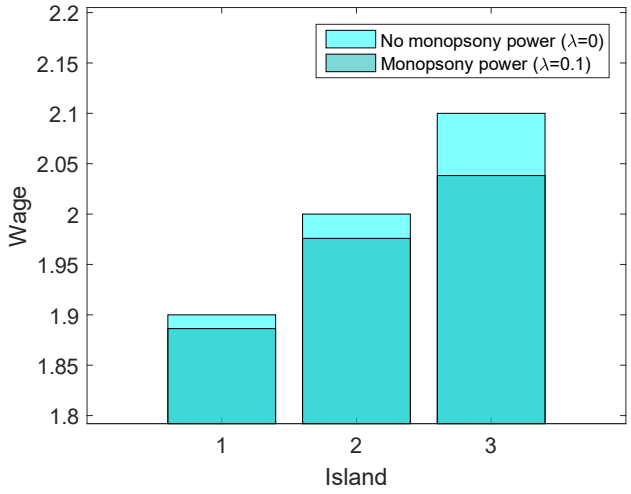

Figure 7: Wage with different λ

Figure 7 shows how, when $\lambda = 0$, wages grow one-to-one with productivity: $w_1 = 1.9$, $w_2 = 2$, and $w_3 = 2.1$. However, under monopsony power, wages i) are lower, ii) and grow more slowly than productivity: $w_1 = 1.89$, $w_2 = 1.98$, and $w_3 = 2.04$. The wedge between wages and productivity is increasing in the island's output share.

We move now to analyze the effects of monopsony power on market concentration. As before, we assume that firm F and firm I evenly split their joint surplus $(2z_t x_j - w_{j,t})$. Equations (4) and (5) become:

$$\sigma^I_{j,t} = \sqrt{\phi \left(z_t x_j - w_{j,t}/2\right) \widehat{n}^F_j} \tag{11}$$

and

$$\widehat{n}^F_{j,t} = \phi \sigma^I_{j,t} \left(z_t x_j - w_{j,t}/2\right). \tag{12}$$

With monopsony power, firms pay a lower wage and achieve a higher profit. This higher profit provides firms with a higher incentive to search. Figure 8 documents this result by plotting the steady-state output share for each island. In the left panel, we plot the distribution of output shares when $\lambda = 0$, which is the same as in Figure 4. In the right panel, we plot the distribution of output shares when $\lambda = 0.1$. The incremental incentive of search is highest for island 3 as it has the greatest effective labor market power due to its size, and is lowest for island 1. As a

result, labor market power intensities market concentration. Island 3's share of output grows from 0.54 to 0.62 and island 2's share falls from 0.17 to 0.14.

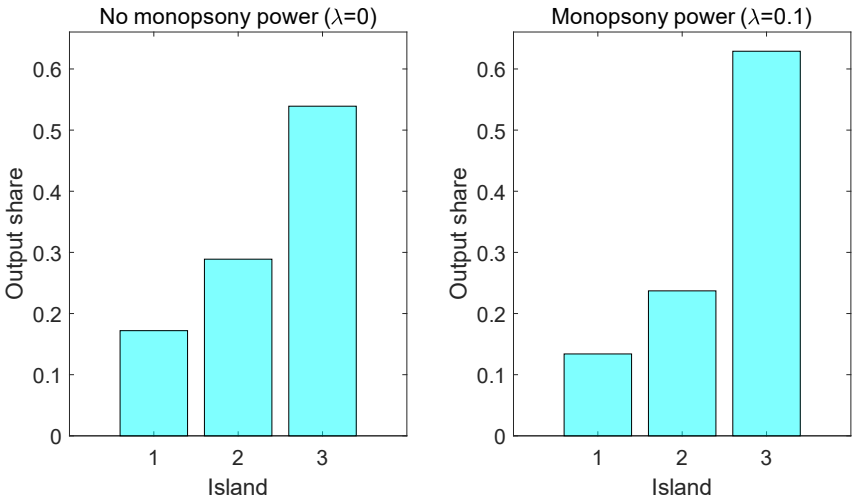

Figure 8: Output share with different λ

This additional strengthening of the "Matthew effect" stands in contrast to the results from a classic model of monopsony in the labor market. In such a classic model, monopsony leads to a smaller firm, since the monopsonist wants to equate the marginal revenue product of labor to the marginal cost of labor by reducing labor hired. In our model, the monopsonist wants to hire more workers, because a larger size allows it to keep more of the total surplus.

Another way to think about this mechanism is that a higher λ leads to a lower labor income share: firms that keep a larger share of the labor surplus grow more in size. When $\lambda = 0$, the labor income share is 0.5 (the Nash bargaining parameter). When $\lambda = 0.1$, the labor income share is 0.49. But, although the share of labor income is lower, the total labor income is 33% higher. Labor income share falls because, when $\lambda = 0.1$, we are providing incentives for higher-productivity firms to scale up and relocate more workers from the low-wage jobs on islands 1 and 2 to the highest-wage jobs on island 3.

We should be careful mapping our results to findings from a cross-sectional regression of wages on labor market power such as those in Marinescu et al. (2020). In our model, all firms have the same monopsony power. Thus, our model's predictions are about two economies with different monopsony power in the labor market (e.g., the U.S. vs. France), not about two firms within the same economy. To think about the latter case, we would need to consider some dimension along which firms diverge, possibly by producing a differentiated good.

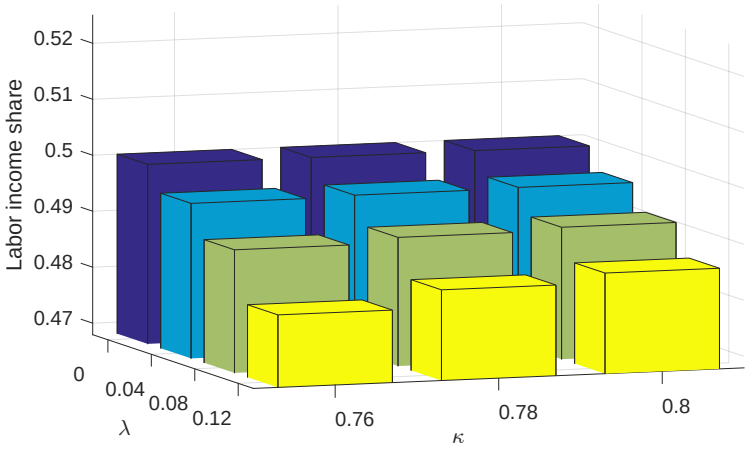

Figure 9: Labor income share with different λ

Figure 9 displays the aggregate labor income share for different values of κ and λ. As we move over the λ-axis, we see the labor share reduction described above. But, interestingly, Figure 9 shows that our model has another mechanism to account for the recent reduction in the labor share in the U.S. economy: a fall in κ. As we move over the κ-axis, the labor income share falls, but output and productivity increase. Since a fall in the cost of signing vendor contracts leads to higher market concentration, it will also lead to firms' higher market power.

Thus, our model predicts that changes such as better software and other technologies to manage vendors and suppliers deliver i) more market concentration and ii) lower labor income share but also iii) higher average wages and iv) higher productivity. More pointedly, our model also suggests that the differences observed between the U.S. and Europe over the last decades in terms of market concentration, labor income shares, and wage and productivity growth may be due to differences in the speed of adopting information technologies that allow for a cheaper scale-up of businesses on each side of the Atlantic.[4] Also, European labor market regulations might limit the extent to which European firms can exert their monopsony power in the labor market, limiting their ability to scale up production.

Figure 10 displays the wage distribution for different values of κ and λ. In each plot, the vertical top-circled line presents workers' density for each wage, and the vertical discontinuous line, the average wage. Either a higher λ or a lower κ makes the market structure more concentrated and, therefore, allocates more workers to more productive firms (i.e., an increase in the height of the vertical line at the right). However, λ and κ have different effects on the level

[4]For some empirical documentation of these differences, see Cette et al. (2019) and Covarrubias et al. (2019).

of wages. A higher λ generally decreases the wage for every worker (i.e., shifts all the vertical lines to the left) and increases wage inequality: more workers move to higher-wage jobs. In contrast, a lower κ reduces the highest wage, but increases the medium and the lowest wages.

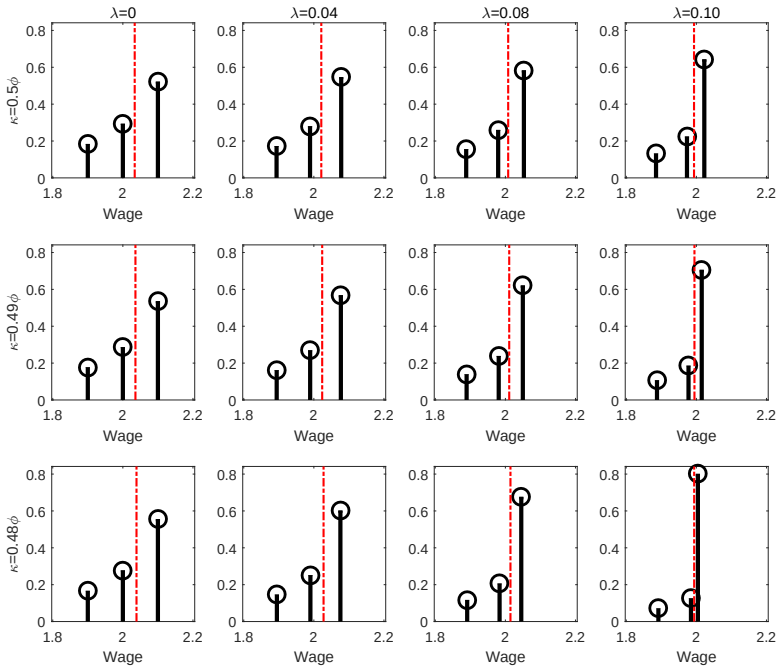

Figure 10: Wage distribution with different λ

We finish this subsection with Figure 11, which is the analogous to Figure 6 but with monopsony power. Aggregate output falls 27% at impact and, as before, the IRFs show persistence. As in Figure 6, island 3 is the one that experiences the largest output over time. However, our simple model ignores an important factor of wage bargaining. The stronger market power of high-productivity firms can increase the outside option value of low-productivity firms' employees by making it easier to find high-paying jobs, which lowers the low-productivity firms' profit margin. In the extended model, this mechanism can make low-productivity firms more responsive to productivity shocks.

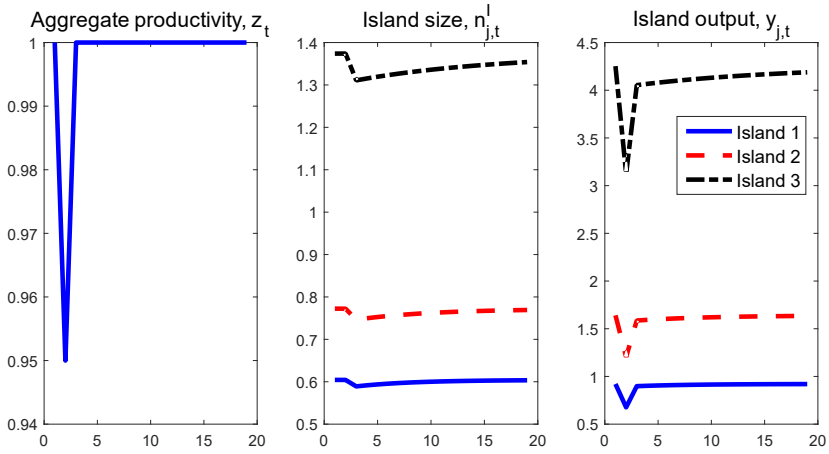

Figure 11: IRFs to a negative aggregate productivity shock

2.5 Taking stock

We can now summarize the eight main takeaways from our simple model:

1. Search complementarities, under directed search, result in a "Matthew effect" that transforms small differences in productivity into large output differences.

2. Entry and exit make the "Matthew effect" even more prominent.

3. Entry and exit make the degree of market concentration depend on the cost of signing a vendor contract. This observation gives us a theory of why market concentration has been growing in the U.S. economy: the fall in search costs related to business relationships (better logistics software, improved inventory control, easier database management).

4. Monopsony power in the labor market strengthens the "Matthew effect" across firms. In our model, search complementarities imply that firms want to get bigger and hire more workers to keep more of the surplus. This result stands in stark contrast with the classic monopsony model, in which firms want to reduce the amount of labor they hire, which leads to relatively smaller firms.

5. Monopsony power lowers wages for a given level of productivity, and the percentage of reduction grows with the firm size.

6. Higher monopsony power increases wage inequality by redistributing workers to larger, more productive firms.

7. Reductions in the cost of signing a vendor contract lower the labor income share, but shift workers' distribution toward high-wage jobs.

8. With entry and exit and monopsony power, aggregate productivity shocks change market concentration, generating a long persistence of the effects of short-lived aggregate shocks.

Let us analyze how these takeaways appear in our extended, quantitative model.

3 Extended model

In this section, we enrich our simple model along three important dimensions. First, we broaden the analysis to general equilibrium by including utility-maximizing households that choose consumption and labor supply and allowing for a richer heterogeneity in intermediate-goods producers. Second, we introduce persistence in vendor relationships. Third, we flesh out the monopsony power in the labor market to be consistent with the granular search theory in Jarosch et al. (2019).

3.1 The representative household

The economy is populated by a continuum of households of size one. Each household has preferences represented by:

$$\sum_{t=0}^{\infty} \beta^t \left[\log\left(C_t\right) + \xi\left(1 - h_t\right)\right], \tag{13}$$

where $\beta \in (0,1)$ is the discount factor, $\xi \geq 0$ is the marginal disutility of labor, C_t is consumption of final goods, and h_t is total hours worked in the household (defined below). The time constraint is normalized to one. Total hours worked is equal to $h_t = \sum_j \widehat{n}_{j,t} h_{j,t}$, where $\widehat{n}_{j,t}$ represents the fraction of households working in a j-type product line. The household's budget constraint is $C_t = \sum_j \widehat{n}_{j,t} w_{j,t} h_{j,t} + \Pi_t$, where $w_{j,t}$ and $h_{j,t}$ are the wage rate and the labor supply in a j-type product line, respectively. Π_t is the per-capita profit from ownership of firms. The wage is different across product lines because of the search and matching frictions.

3.2 The labor market and the goods market: An overview

There are $j = 1, 2, ..., J$ types of firms in the intermediate-goods sector I, and each j-type firm manufactures identical intermediate goods using a technology with different productivity. We denote the idiosyncratic productivity for firm I of type j as x_j. Without loss of generality, we assume strictly increasing idiosyncratic productivity in the index of firm type (i.e., $x_1 < x_2 < ... < x_J$). Each firm I manages a positive measure of product lines, which we interpret as firm size. The distribution of firm size is endogenously determined by search-matching and entry-exit processes, as we describe below. A law of large numbers holds in this economy, equating individual probabilities with realized shares.

To manufacture goods, a product line must first form a vendor relationship with a final-goods producer (firm F) and match with a worker. Firms in the final-goods sector F have the same productivity. Each firm sends buying agents to form vendor relationships with product lines that supply intermediate goods to them. Search is directed, and each firm in sector F optimally chooses the j-type firm in sector I to visit. Since J types of firm I exist, there are J segmented inter-firm submarkets, indexed by j. Sending a buying agent to submarket j incurs the unit cost κ. Each firm I in submarket j chooses the costly search effort, denoted by $\sigma^I_{j,t}$, to maximize profits. Variable search effort and directed search generate strategic complementarities since the optimal search effort exerted by firm I is increasing in the measure of buying agents sent by firm F. Similarly, the optimal measure of buying agents sent by firm F will also be increasing in the search effort exerted by firm I.

After a vendor relationship is formed, each vendor relationship without a worker posts one vacancy (without any costs) in the labor market and stays idle. At the end of each period, vendor relationships and labor market matches separate exogenously with probability $\widehat{\delta}$ and δ, respectively, and in either case, workers become unemployed.

Figure 12 summarizes the timeline for firm I. At the beginning of each period, product lines search for buying agents to establish a vendor relationship. Next, vendor relationships search for workers. If successfully matched with a worker, the vendor relationship enters the production stage; otherwise it stays idle. Vendor relationships and labor market matches separate randomly at the end of each period.

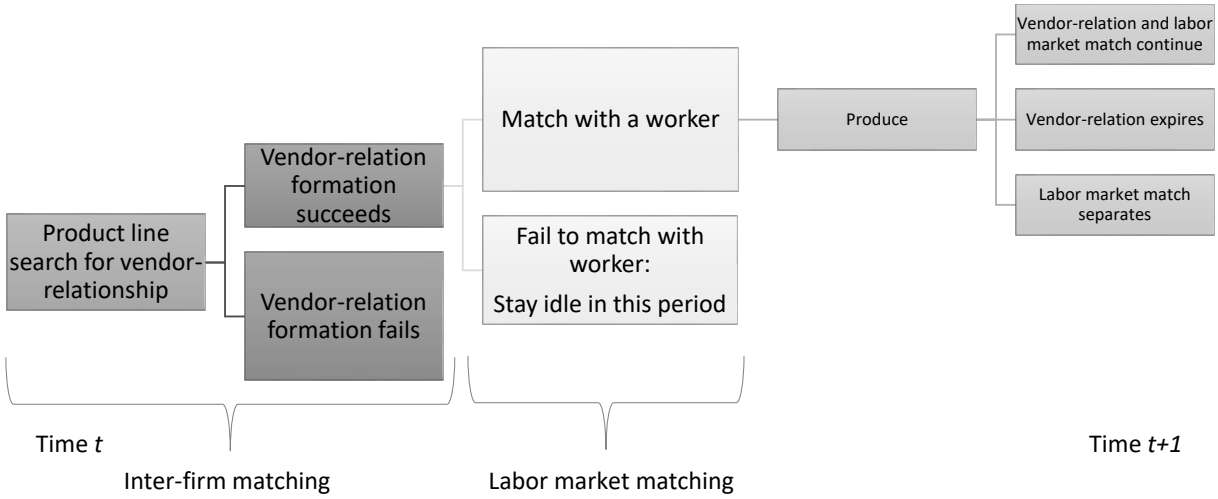

Figure 12: Timeline for firm I

3.3 The labor market: Search frictions and monopsony power

3.3.1 Matching function

We assume a frictional labor market. We depart from the DMP framework by allowing multiple workers to apply for a single vacancy randomly. With this simple variation, we give firms monopsony power in the labor market because they can threaten to preclude workers who decline a wage offer from future job offers. Therefore, the bargained wage may be below the marginal product of labor.

The matching technology is formulated by the process of randomly placing balls in urns as in Butters (1977). Product lines play the role of urns and workers the role of balls. An urn becomes "productive" when it has a ball in it. Even with the same number of urns and balls, a random placing of the balls in the urns will not match all the pairs exactly because of a coordination failure by those placing the balls in the urns. Some urns will end up with more than one ball and some with none. In the context of the labor market, if only one worker could occupy each job, an uncoordinated application process by workers will lead to overcrowding in some jobs and to no applications in others. As illustrated by Petrongolo and Pissarides (2001), the imperfection that leads to unemployment in this environment is the lack of information about other workers' action.

In the simplest version of this process, we assume that workers and vendor relationships are discrete. There are u_t number of unemployed workers who know the location of v_t number

of unmatched product lines. If a product line receives one or more job applications, it selects one applicant and forms a match (the selection criterion is specified below), while the other applicants become unemployed in the current period t.

Given that each product line receives a worker's application with probability $1/\widetilde{n}_t$, and there are u_t applicants, with probability $(1-1/v_t)^{u_t}$ a given product line will receive no applications. Thus, the number of labor market matches formed in each period is $E_t = v_t\left[1-(1-1/v_t)^{u_t}\right]$.

We let the measure of each product line and worker be infinitely small, such that $\widetilde{n}_t$ and u_t tends to infinity, in which case we have that $\lim_{v_t,u_t\to\infty} E_t = v_t\left(1-e^{v_t/u_t}\right)$. Then, the vacancy filling rate is:

$$p_t^n = \frac{E_t}{v_t} = 1 - e^{-u_t/v_t},$$

and the job finding rate is:

$$p_t^u = \frac{E_t}{u_t} = v_t/u_t \cdot \left(1-e^{-u_t/v_t}\right).$$

To introduce labor market power, we adopt a "granular search" approach proposed by Jarosch et al. (2019), who show that large firms hold a strong bargaining power by threatening workers with future job refusals, since workers can hardly avoid large employers and they are likely to re-apply for job openings from the same firms in the future. We denote the measure of vendor relationships managed by the type-j firms that are not matched with a worker as $v_{j,t}$, which we interpret as a proxy for the size of the labor market. Thus, the number of unmatched product lines is $v_t = \sum_{j=1}^{J} v_{j,t}$. We define the relative labor market size, $s_{j,t}$, as the fraction of unfilled product lines owned by type-j firms with $s_{j,t} = v_{j,t}/v_t$ and $\sum_{j=1}^{J} s_{j,t} = 1$, which is endogenously determined. In general, a more productive firm and a firm that searches more actively gain a higher $s_{j,t}$ that generates stronger labor market power.

The dynamics in the model depend on three probabilities. First, the conditional probability p^n that a worker meeting a product line is not the only applicant for the job opening. Second, the probability $s \cdot p^u$ that a worker meets with a product line owned by a large firm (that possesses the fraction s of matched product lines of the economy). Third, the probability that a worker contacts a product line owned by the same firm and that the product line has more than one job applicant: $s \cdot p^u \cdot p^n$.

Labor market matches separate exogenously with probability δ. In addition, if a product line becomes obsolete (with probability $\widehat{\delta}$), the labor market match terminates.

3.3.2 Monopsony power and value functions

The wage is determined by Nash bargaining. The bargaining set is within the product line's output $y_{j,t}$ and the disutility of working $\xi h_{j,t}$. When multiple homogeneous workers apply for a single vacancy, the product line offers a wage contract to one candidate. The product line exerts its monopsony power by threatening the worker with forgoing future hiring if the current offer is rejected. This threat is potent when the product line belongs to a firm of large size, since job applicants are likely to re-encounter the same firm in the future with a probability proportional to relative labor market size (s_j). Thus, more productive (and therefore larger) firms retain a stronger threatening power.

The firm that operates the product line precludes workers who reject a current job offer from future hiring with probability $\widetilde{\delta}$, such that the mean duration of the punishment is $1/\widetilde{\delta}$ periods. To rule out the complicated case in which a worker is punished by multiple firms, we assume that firms withdraw punishment to workers once a worker is hired by other product lines.

We now define the Bellman equations that determine the value of an unemployed worker without punishment U_t, of an unemployed worker punished by a type-j firm $\widetilde{U}_{j,t}$, of an employed worker in a type-j firm $W_{j,t}$, of a product line owned by a type-j firm that is matched with a worker $J_{j,t}$, and of a product line that is not matched with a worker $X_{j,t}$.

The value of an unemployed worker without punishment is:

$$U_t = \xi + \beta \left(\frac{C_t}{C_{t+1}} \right) \left[p_t^u \sum_k s_{k,t} W_{k,t+1} + (1 - p_t^u) U_{t+1} \right], \tag{14}$$

where ξ is the flow of utility from being unemployed in period t. In $t+1$, with probability $p_t^u \cdot s_{k,t}$, the worker finds a job in a k-type product line or, with probability $1 - p_t^u$, remains unemployed. The continuation value is discounted by the stochastic discount factor, $\beta (C_t/C_{t+1})$.

The value of an unemployed worker under punishment by a type-j firm is:

$$\begin{aligned} \widetilde{U}_{j,t} = {} & \widetilde{\delta} U_{j,t} \\ & + \left(1 - \widetilde{\delta}\right) \left\{ \xi + \beta \left(\frac{C_t}{C_{t+1}} \right) \left\{ \begin{array}{c} p_t^u \left[\sum_{k \neq j} s_{k,t} W_{k,t+1} + s_{j,t} (1 - p_t^n) W_{j,t+1} \right] \\ + (1 - p_t^u + p_t^u s_{j,t} p_t^n) \widetilde{U}_{j,t+1} \end{array} \right\} \right\}, \end{aligned} \tag{15}$$

where, with probability $\widetilde{\delta}$, the punishment of the worker is forgiven and the value of unemployment

becomes $U_{j,t}$. Otherwise, the worker continues under punishment. In this case, with probability $p_t^u \cdot s_{k,t}$, the worker finds a job in a type-k product line ($k \neq j$) and, with probability $p_t^u \cdot s_{j,t}$, the worker is hired by a type-j product line. This last hiring occurs because either the firm is not the one enforcing the punishment, or if it is, the firm has no other applications for the job. Finally, with probability $(1 - p_t^u + p_t^u s_{j,t} p_t^n)$, the worker remains unemployed. This occurs because either the worker fails to meet any vacancy or the worker meets a type-j product line, but the product line has alternative applicants and, thus, rejects the worker.

By multiplying equation (14) by $(1 - \widetilde{\delta})$ and subtracting equation (15) from it, we obtain the loss of value associated with labor market punishment:

$$U_t - \widetilde{U}_{j,t} = \left(1 - \widetilde{\delta}\right) \beta \left(\frac{C_t}{C_{t+1}}\right) \left[\begin{array}{c} s_{j,t} p_t^u p_t^n \left(W_{j,t+1} - U_{t+1}\right) \\ + \left(1 - p_t^u + s_{j,t} p_t^u p_t^n\right) \left(U_{t+1} - \widetilde{U}_{j,t+1}\right) \end{array}\right]. \tag{16}$$

In the deterministic steady state, equation (16) reduces to:

$$U - \widetilde{U}_j = \frac{\left(1 - \widetilde{\delta}\right) \beta s_j p^u p^n \left(W_j - U\right)}{1 - \beta \left(1 - p^u + s_j p^u p^n\right) \left(1 - \widetilde{\delta}\right)}. \tag{17}$$

Equation (17) shows that if $\widetilde{\delta} = 1$, there is no labor market punishment and $U = \widetilde{U}_j$. If $\widetilde{\delta} < 1$, however, equation (15) implies that $U > \widetilde{U}_j$, that is, labor market punishment generates a loss to the worker, since she prefers working to being unemployed (i.e., $W_j > U$). Moreover, equation (17) shows that the loss of value due to labor market punishment strictly increases with the firm's relative labor market size (s_j), and strictly decreases with the probability of forgiving ($\widetilde{\delta}$). When the firm's labor market size is zero, $U = \widetilde{U}_j$.

The value of an employed worker in a vendor relationship is:

$$W_{j,t} = w_{j,t} + \beta \left(\frac{C_t}{C_{t+1}}\right) \left[\left(1 - \delta - \widehat{\delta}\right) W_{j,t+1} + \left(\delta + \widehat{\delta}\right) U_{t+1}\right], \tag{18}$$

where the first term on the right-hand side (RHS) of equation (18) is the current period wage $w_{j,t}$ to be determined by Nash bargaining. The job relationship terminates randomly because either the job separates with probability δ or the vendor relationship dissolves with probability $\widehat{\delta}$. In both instances, the worker becomes unemployed and gains value U_{t+1}. Otherwise, the

worker continues the job relationship and earns value $W_{j,t+1}$.

Similarly, the value of firms in a vendor relationship with a worker is:

$$J_{j,t}^{k} = \Pi_{j,t}^{k} + \beta \left(\frac{C_t}{C_{t+1}} \right) \left[\left(1 - \delta - \widehat{\delta} \right) J_{j,t+1}^{k} + \delta X_{j,t+1}^{k} + \widehat{\delta} \widetilde{J}_{j,t+1}^{k} \right], \; k \in \{I, F\}, \tag{19}$$

where the first term on the RHS of equation (19) is current profit, and the second term is the continuation value in the next period $t+1$, in which the job separates with probability δ and the idle vendor relationship gets $X_{j,t+1}^{k}$ (defined below), or the vendor relationship dissolves with probability $\widetilde{\delta}$ and each firm gets $\widetilde{J}_{t+1}^{k}$.

The value of an idle product line without a worker is:

$$X_{j,t}^{k} = \beta \left(\frac{C_t}{C_{t+1}} \right) \left[p_t^n J_{j,t+1}^{k} + \left(1 - p_t^n \right) X_{j,t+1}^{k} \right], \; k \in \{I, F\}. \tag{20}$$

Equation (20) shows that an idle product line produces zero profits in period t, but by hiring a worker, with probability p_t^n, it receives the value $J_{j,t+1}^{k}$. Otherwise, with probability $(1 - p_t^n)$, the product line remains unmatched and earning $X_{j,t+1}^{k}$ in the next period $t+1$.

The value of a product line without a vendor relationship is:

$$\widetilde{J}_{j,t}^{I} = -c\left(\sigma_{j,t} \right) + \beta \left(\frac{C_t}{C_{t+1}} \right) \left[\pi_{j,t}^{I} X_{j,t+1}^{I} + \left(1 - \pi_{j,t}^{I} \right) \left(1 - \chi \right) \widetilde{J}_{j,t+1}^{I} \right]. \tag{21}$$

Equation (21) shows that a product line without a vendor relationship exerts search effort in period t, and in the next period $t+1$, finds a firm in sector I with probability $\pi_{j,t}^{I}$ that yields a value $X_{j,t+1}^{I}$. Otherwise, and if it survives obsolescence with probability $(1 - \chi)$, it remains without a vendor relationship and yields a value of $\widetilde{J}_{j,t+1}^{I}$.

Lastly, when a vendor relationship terminates, the buying agent of firm F returns to the central island and receives zero value:

$$\widetilde{J}_{j,t}^{F} = 0 \tag{22}$$

Firms split the joint profit from the match by Nash bargaining, which yields:

$$\frac{X_{j,t}^{I} - \widetilde{J}_{j,t}^{I}}{\tau} = \frac{X_{j,t}^{F} - \widetilde{J}_{j,t}^{F}}{1 - \tau} \tag{23}$$

where $X^k_{j,t} - \widetilde{J}^k_{j,t}$ is the capital gain by signing a vendor contract. The parameter τ is the bargaining share of firm I.

3.3.3 Wage determination

The wage is negotiated between the worker and the vendor relationship by Nash bargaining. The total surplus from forming a match in the labor market ($LTS_{j,t}$) is equal to:

$$LTS_{j,t} = (J_{j,t} - X_{j,t}) + \left(W_{j,t} - \widetilde{U}_{j,t}\right), \tag{24}$$

where $J_{j,t}$ and $X_{j,t}$ are joint values of a vendor relationship with $J_{j,t} = J^I_{j,t} + J^F_{j,t}$, and $X_{j,t} = X^I_{j,t} + X^F_{j,t}$. Equation (24) departs from the standard bargaining protocols because the worker surplus depends on $\widetilde{U}_{j,t}$ rather than U_t, and the additional surplus $(U_t - \widetilde{U}_{j,t})$ arises from the firm's credible threat of future rejection.

Thus, given a vendor relationship's bargaining share $\widetilde{\tau}$, the bargained wage ($w_{j,t}$) satisfies:

$$W_{j,t} - \widetilde{U}_{j,t} = (1 - \widetilde{\tau})\, LTS_{j,t}, \tag{25}$$

and

$$J_{j,t} - X_{j,t} = \widetilde{\tau} LTS_{j,t}. \tag{26}$$

In the online appendix, we prove the following proposition.

Proposition 1. *In the steady state, ceteris paribus, the wage decreases with the firm's vacancy share (s_j) and increases with the probability of forgiveness ($\widetilde{\delta}$).*

Proposition 1 shows that, conditional on a level of productivity, greater market power –either because a firm represents a larger share in the labor market or because a firm has a lower probability of forgiveness– implies a lower wage.

3.4 The goods market: Vendor contract formation

As in the simple model, the matching process in each submarket is governed by a technology with variable search intensity. Following Burdett and Mortensen (1980), the number of newly formed vendor relationships in market j is $M\left(\widetilde{n}^F_{j,t}, \widetilde{n}^I_{j,t}, \sigma^I_{j,t}\right) = \psi \sigma^I_{j,t} H\left(\widetilde{n}^F_{j,t}, \widetilde{n}^I_{j,t}\right)$, where $\sigma^I_{j,t}$ is

firm I's variable search effort, $\widetilde{n}_{j,t}^{F}$ is the measure of firm F's buying agents, and $\widetilde{n}_{j,t}^{I}$ is the measure of product lines owned by type-j firm I. The parameter ψ controls the efficiency in matching. The function $H(\cdot)$ has constant returns to scale and it is strictly increasing in both arguments.

Each submarket j has a tightness ratio $\theta_{j,t}$, defined as $\theta_{j,t} = n_{j,t}^{F}/n_{j,t}^{I}$. The probability that a product line forms a joint venture with a firm in sector F is:

$$\pi_{j,t}^{I} = \frac{M\left(\widetilde{n}_{j,t}^{F}, \widetilde{n}_{j,t}^{I}, \sigma_{j,t}^{I}\right)}{\widetilde{n}_{j,t}^{I}} = \psi \sigma_{j,t}^{I} \mu\left(\theta_{j,t}\right),$$

and the probability that a firm in sector F forms a vendor relationship with a type-j firm in sector I is:

$$\pi_{j,t}^{F} = \frac{M\left(\widetilde{n}_{j,t}^{F}, \widetilde{n}_{j,t}^{I}, \sigma_{j,t}^{I}\right)}{\widetilde{n}_{j,t}^{F}} = \psi \sigma_{j,t}^{I} q\left(\theta_{j,t}\right),$$

where $\mu\left(\theta_{j,t}\right) = H\left(\theta_{j,t}, 1\right)$ and $q\left(\theta_{j,t}\right) = H\left(1, 1/\theta_{j,t}\right)$. Then, $\mu'\left(\theta_{j,t}\right) > 0$ and $q'\left(\theta_{j,t}\right) < 0$.

Each firm in sector I faces the cost of searching with intensity $\sigma_{j,t}^{I}$ equal to:

$$c\left(\sigma_{j,t}^{I}\right) = \frac{\left(\sigma_{j,t}^{I}\right)^{1+\nu}}{1+\nu}, \quad j \in \{1, 2, ..., J\}.$$

3.4.1 Production technology

A product line manufactures intermediate goods according to the production technology:

$$\widetilde{y}_{j,t} = x_j h_{j,t}, \tag{27}$$

where $\widetilde{y}_{j,t}$ is the output for firms in the intermediate-goods sector (a tilde indicates intermediate-goods sector variables), and x_j is the idiosyncratic productivity for type-j intermediate-goods producer. Each product line matches with one worker and hours are fixed to one (i.e., $h_{j,t} = 1$).

Final-goods producers transform the intermediate goods into the final goods $y_{j,t}$ with the linear production technology:

$$y_{j,t} = z\widetilde{y}_{j,t} = z x_j, \tag{28}$$

where z is the level of aggregate productivity.

Total output is split $y_{j,t} = w_{j,t} + \Pi_{j,t}^{I} + \Pi_{j,t}^{F}$, where $W_{j,t}$, $\Pi_{j,t}^{I}$, and $\Pi_{j,t}^{F}$ are the wage of the

worker, the profits of the product line, and the profits of the final-goods producer (conditional on vendor relationship formation and labor market matching), respectively.

3.4.2 Optimal search effort for intermediate goods producers

The product line chooses the optimal search effort by maximizing the value $\widetilde{J}_{j,t}$:

$$\max_{\sigma^I_{j,t}\geq 0} -c\left(\sigma_{j,t}\right)+\beta\left(\frac{C_t}{C_{t+1}}\right)\left[\pi^I_{j,t}X^I_{j,t+1}+\left(1-\pi^I_{j,t}\right)\left(1-\chi\right)\widetilde{J}^I_{j,t+1}\right], \tag{29}$$

where $\pi^I_{j,t}$ is the probability of forming a vendor relationship. $J_{j,t}\left(0\right)$ and $J_{j,t}\left(1\right)$ are the ex-post value of a product line defined in equation (19), conditional on the success and failure of a vendor contract, respectively. The interior solution to the problem in equation (29) is:

$$\left(\sigma^I_{j,t}\right)^{\nu}=\beta\left(\frac{C_t}{C_{t+1}}\right)\psi\mu\left(\theta_{j,t}\right)\Delta J^I_{j,t+1}, \tag{30}$$

where $\Delta J_{j,t}$ is the capital gain due to the establishment of a vendor contract:

$$\Delta J_{j,t+1}=X^I_{j,t+1}-\left(1-\chi\right)\widetilde{J}^I_{j,t+1}, \tag{31}$$

which includes the capital gain $X^I_{j,t+1}-\widetilde{J}^I_{j,t+1}$, and the gain $\chi\widetilde{J}^I_{j,t+1}$ from a product line with a vendor contract avoiding obsolescence.

The left-hand side (LHS) of equation (30) is the marginal cost of exerting search effort to form a vendor relationship for a j-type firm in sector I, and the RHS of the equation is the benefit of signing a vendor contract, which increases in tightness $\theta_{j,t}$ (since $\mu'\left(\theta_{j,t}\right)>0$) and in the capital gain from forming a vendor relationship.

The solution to the optimization problem is:

$$\sigma^I_{j,t}=\left[\beta\left(\frac{C_t}{C_{t+1}}\right)\psi\mu\left(\theta_{j,t}\right)\Delta J^I_{j,t+1}\right]^{\frac{1}{\nu}}. \tag{32}$$

Since $\nu>1$ and $\mu\left(\cdot\right)$ is an increasing function, equation (32) shows that the optimal search intensity $\sigma^I_{j,t}$ increases with the tightness ratio $\theta_{j,t}$, implying that $\sigma^I_{j,t}>0$.

In the online appendix, we show that strong market power –either because a firm owns a larger share in the labor market, or because it exercises a lower probability of forgiveness–

implies a greater search effort (conditional on a level of productivity and the number of agents visiting from sector F).

Proposition 2. *In the steady state, ceteris paribus, firm I's search effort increases with the firm's vacancy share s_j, and it decreases with the probability of forgiveness $\widetilde{\delta}$.*

Intuitively, Proposition 2 establishes that strong labor market power enables firms to offer a lower wage to the worker, which expands the firm's profit for every signed vendor contract, which stimulates an active search. As we will see later, a critical implication of Proposition 2 is that labor market power entails a more concentrated market structure.

3.4.3 Buying agents and search complementarity

The value of sending a buying agent for a firm in sector F is:

$$V_t^F = \max_j \left\{ -\kappa + \beta \left(\frac{C_t}{C_{t+1}} \right) \pi_{j,t}^F \left(X_{j,t+1}^F - \widetilde{J}_{j,t+1}^F \right) \right\}. \tag{33}$$

Equation (33) shows that each firm in sector F pays a unit cost κ for each agent who visits submarket j that may establish a vendor relationship with probability $\pi_{j,t}^F = \psi \sigma_{j,t}^I q\left(\theta_{j,t}\right)$, and brings a capital gain $X_{j,t+1}^F - \widetilde{J}_{j,t+1}^F$.

Firms in sector F send buying agents to visit prospective intermediate-goods suppliers at the optimal submarkets until the value of forming a vendor relationship collapses to zero (recall that a law of large numbers holds and, thus, conditional on the aggregate states, expected and realized profits are equated): $\mathbb{E}_t\left(V_t^F\right) = 0$.

Substituting this last condition into equation (33), we get:

$$\max_j \left\{ -\kappa + \beta \left(\frac{C_t}{C_{t+1}} \right) \psi \sigma_{j,t}^I q\left(\theta_{j,t}\right) \left(X_{j,t+1}^F - \widetilde{J}_{j,t+1}^F \right) \right\} = 0,$$

such that the capital gain in each submarket j is equal to the cost κ:

$$q\left(\theta_{j,t}\right) \sigma_{j,t}^I \beta \left(\frac{C_t}{C_{t+1}} \right) \psi \cdot \left(X_{j,t+1}^F - \widetilde{J}_{j,t+1}^F \right) = \kappa, \tag{34}$$

and consequently the submarkets with a higher capital gain, $X_{j,t+1}^F - \widetilde{J}_{j,t+1}^F$, attract more buying agents to visit. The inflow of buying agents increases the tightness ratio in each submarket,

which decreases the matching probability for those buying agents. In equilibrium, the tightness ratio adjusts to make the gain from entering into all submarkets equal to the cost κ.

Equation (34) implies that, because $q(\cdot)$ is a decreasing function, the tightness ratio $\theta_{j,t}$ increases with intermediate-goods producers' search effort $\sigma^I_{j,t}$:

$$\theta_{j,t} = q^{-1}\left[\frac{\kappa}{\sigma^I_{j,t}\beta\left(\frac{C_t}{C_{t+1}}\right)\psi\cdot\left(X^F_{j,t+1}-\widetilde{J}^F_{j,t+1}\right)}\right]. \tag{35}$$

As in our simple model, directed search is key to generating search complementarities.

3.5 Period equilibrium

The period equilibrium of submarket j is a tuple of $\left\{\sigma^I_{j,t},\theta_{j,t}\right\}$ that is a fixed point of the product of the best response function (32) and the optimality condition (35). As before, we ignore the trivial equilibrium with zero output. The whole dynamic equilibrium of the economy is a repetition of these period equilibria as linked by the value functions outlined above.

To determine the measure of firms and aggregate output, we assume that new product lines are created at the constant rate n in each period t. The measure of product lines that remain unmatched with a final-goods producer in the next period $t+1$ ($\widetilde{n}^I_{j,t+1}$) is equal to those lines that fail to sign a vendor contract and do not become obsolete ($\left(1-\pi^I_{j,t}\right)(1-\chi)\widetilde{n}^I_{j,t}$), plus those that recently separated from a vendor relationship ($\widehat{\delta}n^I_{j,t}$), and the new product line (n), such that:

$$\widetilde{n}^I_{j,t+1} = \left(1-\pi^I_{j,t}\right)(1-\chi)\widetilde{n}^I_{j,t} + \widehat{\delta}n_{j,t} + n. \tag{36}$$

Using the definition of the tightness ratio $\theta_{j,t}$, the measure of buying agents sent to submarket j is $\widetilde{n}^F_{j,t} = \widetilde{n}^I_{j,t}\theta_{j,t}$, and the measure of vendor relationship ($n_{j,t+1}$) comprises those that survive separation ($(1-\widehat{\delta})n_{j,t}$) plus new vendor relationship formation ($\pi^I_{j,t}\widetilde{n}^I_{j,t}$), such that:

$$n_{j,t+1} = \left(1-\widehat{\delta}\right)n_{j,t} + \pi^I_{j,t}\widetilde{n}^I_{j,t}. \tag{37}$$

The measure of vendor relationships matched with a worker ($\widehat{n}_{j,t+1}$) comprises those that do not separate with a worker and do not dissolve ($(1-\delta-\widehat{\delta})\widehat{n}_{j,t}$) plus the new labor market

matches ($p_t^n v_{j,t}$):

$$\widehat{n}_{j,t+1} = \left(1 - \delta - \widehat{\delta}\right) \widehat{n}_{j,t} + p_t^n v_{j,t}. \tag{38}$$

The measure of vendor relationships that are unmatched with workers is $v_{j,t} = n_{j,t} - \widehat{n}_{j,t}$, and vacancies are equal to the measure of vendor relationships unmatched with workers $v_t = \sum v_{j,t}$.

Unemployment is equal to $u_{t+1} = (1 - p_t^u)\, u_t + \left(\delta + \widehat{\delta}\right) \sum \widehat{n}_{j,t}$, where the first term on the RHS shows the unemployment outflow induced by job creation ($p_t^u u_t$), and the second term shows the unemployment inflow from random job and vendor-relationship separation.

Aggregate output is a weighted sum of final goods produced across submarkets $Y_t = \sum_{j=1}^{J} \widehat{n}_{j,t} y_{j,t}$, where $\widehat{n}_{j,t}$ is the measure of vendor relationships matched with a worker, determined by equation (38), and $y_{j,t}$ is the final output of vendor relationships, determined by equations (27) and (28), respectively. Aggregate output is used for aggregate consumption, C_t, search costs, and entry costs:

$$Y_t = C_t + \sum_{j=1}^{J} \widetilde{n}_{j,t}^I \frac{\left(\sigma_{j,t}^I\right)^{1+\nu}}{1+\nu} + \kappa \sum_{j=1}^{J} \widetilde{n}_{j,t}^F.$$

4 Calibration and measurement

We calibrate our model by matching its steady state to post-WWII U.S. data at a quarterly frequency. A discount factor β of 0.987 (equivalent to 0.95 at a yearly frequency) replicates an average annual interest rate of 5% over the sample period.

We pick 20 productivity types, J, such that each type of firm I corresponds to a vigintile of the productivity distribution. Hence, type-1 firms are the bottom 5% of the productivity distribution and type-20 firms the top 5%. In our model, the measured total factor productivity (mTFP) of firms I results from the combination of the exogenous productivity, x_j, and the endogenous product line utilization rate, $n_j / \left(n_j + \widetilde{n}_j^I\right)$. Thus, we calibrate the dispersion of x_j to match the observations by Syverson (2011) that the average ratio of mTFP between industry plants at the 90th and 10th percentiles of the productivity distribution using four-digit SIC industries in the U.S. manufacturing sector is 1.92. We match this ratio by assuming that $log\,(x_j)$ is uniformly distributed between -0.12 and 0.12. We normalize the level of aggregate productivity z equal to 1.

With respect to the search cost function, we set $\nu = 3$, implying that the marginal search cost

is a quadratic function of the search effort. We normalize the cost of signing a vendor contract to be equal to the average productivity of vendor relationships, i.e., $\kappa = 1$ (the parameter ψ, to be calibrated below, varies to compensate for this normalization).

We calibrate $\widehat{\delta} = 1/16$ to replicate the average duration of 4 years in vendor relationships in the Compustat Customer Segment data (which report the major customers for a subset of U.S. listed companies on a yearly basis). For the $H(\cdot)$ function, we assume a Cobb-Douglas form, $\psi \left(\widetilde{n}_j^F\right)^{\alpha} \left(\widetilde{n}_j^I\right)^{1-\alpha}$, where $\alpha = 0.5$ imposes symmetry. By setting $\psi = 0.54$, we get that 88% of product lines for the medium firms are active in the steady state, matching the observed 12% average rate of idleness in the U.S. non-manufacturing and manufacturing sectors before the Great Recession (Michaillat and Saez, 2015, and Ghassibe and Zanetti, 2020).

Following Shimer (2005) and Thomas and Zanetti (2009), the flow value of unemployment ξ (the marginal value of leisure in our model) is set to 40% of the mean labor productivity. The worker's bargaining share $\widetilde{\tau}$ is set to 0.65, such that the labor income share of output is equal to 0.66, consistent with the long-run average of labor share in the U.S. economy. With $\tau = 0.5$, the remaining 34% of total income is evenly distributed between firms I and F.

We normalize population to one. Following Shimer (2005), we target the quarterly job finding rate $p^u = 0.7$, an unemployment rate, $u = 0.055$, and labor market tightness $v/u = 1.3$. These targets imply that the probability of filling a vacancy, equal for all firms, is $p^n = \left(1 - e^{v/u}\right) = 0.54$, the employment-to-unemployment (EU) transition rate is 0.041 ($0.041/\left(0.041 + 0.7\right) = 0.055$), and the EU transition probability from vendor-contract dissolutions is $\left(1 - p^u\right)\widehat{\delta} = 0.019$. Thus, δ, the exogenous job separation rate, is $0.041 - 0.019 = 0.022$. We set the creation rate of new product lines, $\widehat{n}$, equal to 0.0017 to be consistent with this calibration.

In our model, the rate of obsolescence of a product line can be interpreted as the rate of plant exit. Lee and Mukoyama (2015) estimate the average exit rate of manufacturing plants equal to 5.5% on a yearly basis (1.4% on a quarterly basis) using the Longitudinal Research Database (LRD) from the U.S. Census Bureau (see Hamano and Zanetti, 2017, for a discussion on the empirical estimates of plant entry and exit rates). Hence, we set the rate of product line obsolescence $\chi = 0.13$ and get an average obsolescence rate equal to 1.4%.

Finally, we use the model to measure the probability of labor market forgiveness, $\widetilde{\delta}$, equal to 0.51, which matches the output share of the top 10% of firms of 0.64 reported by Autor et al. (2020). A value $\widetilde{\delta} = 0.51$ means that firms forgive workers on average after 2 periods (i.e.,

after six months). In our numerical analysis below, we will vary $\widetilde{\delta}$ to assess monopsony power's non-linear effect on market concentration. Table 1 summarizes our model's calibration.

Description	Parameter	Value
Discount factor	β	0.987
Number of firm type	J	20
Productivity	$\log(x_j)$	$\mathcal{U}$[-0.12,0.12]
Search cost function, curvature	ν	3
Cost of sending a buying agent	κ	1
Vendor contract expiration rate	$\widehat{\delta}$	1/16
Matching elasticity	α	0.5
Matching efficiency	ψ	0.54
Flow value of unemployment	ξ	0.4
Worker's bargaining share	$\widetilde{\tau}$	0.65
Final goods firm's bargaining share	τ	0.5
Exogenous job separation rate	δ	0.022
Inflow of product line	$\widehat{n}$	0.0017
Rate of product line obsolescence	χ	0.14
Probability of labor market forgiveness	$\widetilde{\delta}$	0.51

Table 1: Calibration

5 Quantitative results

In this section, we report nine quantitative findings from our extended model. First, search effort increases with productivity. Second, search complementarities induce market concentration. Third, monopsony power in the labor market reinforces market concentration. Four, monopsony power lowers wages and the labor income share, but it also moves workers toward high-wage jobs and increases wage inequality. Five, in the absence of strategic complementarities, monopsony power in the labor market has a limited effect on market concentration. Six, lower search costs increase market concentration. Seven, search complementarities amplify the effect of negative aggregate productivity shocks and make them more persistent. Eight, negative aggregate productivity shocks increase market concentration because they disproportionally affect the output of low-productivity firms. Nine, lower search costs reduce the volatility of the economy. These findings mirror our simple model's main takeaways in Section 2. Let us review each of them in more detail.

5.1 Search effort increases with productivity

Figure 13 plots, for each productivity level (j), the search efforts ($\{\sigma_j^I\}_{j=1}^J$, top panel) of firms I, the measure of buying agents sent by firm F to each island ($\{\widetilde{n}_j^F\}_{j=1}^J$, middle panel), and the probability that product lines form vendor relationships ($\{\pi_j^I\}_{j=1}^J$, bottom panel). Higher-productivity intermediate-goods producers search more intensively, attract more buying agents, and enjoy a higher matching probability.

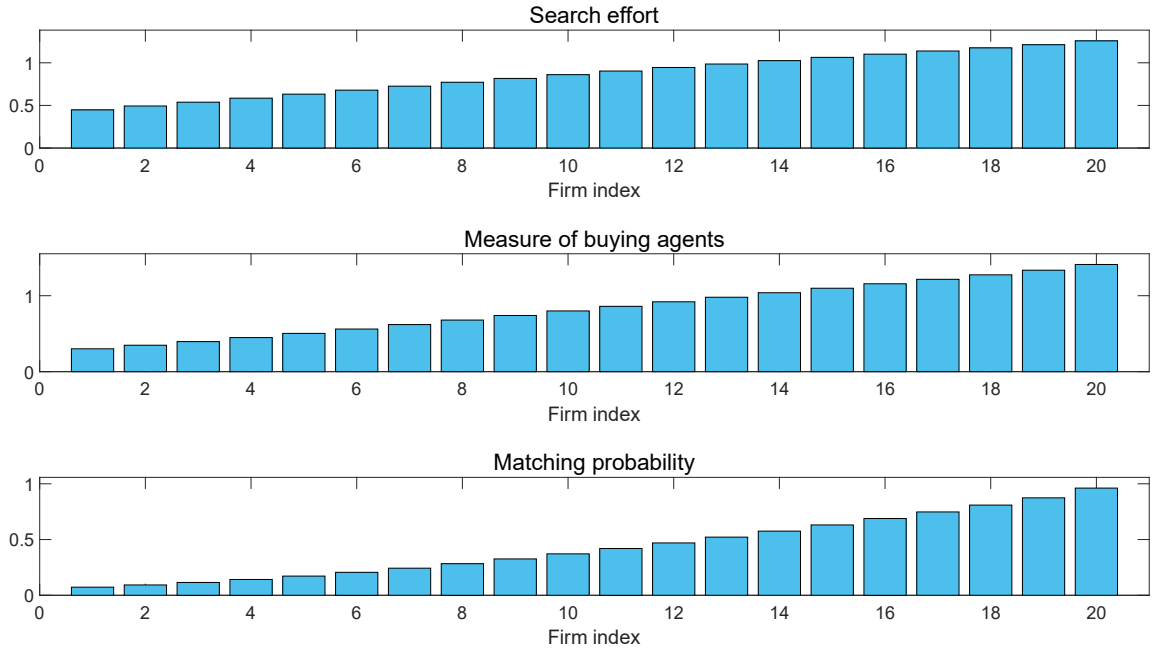

Figure 13: Search effort

5.2 Search complementarities induce market concentration

We turn now to market concentration. Equations (36) and (37) give us the measures of production lines that are unmatched:

$$\widetilde{n}_j^I = \frac{n}{\left(1-\pi_{j,t}^I\right)\chi} \tag{39}$$

and matched:

$$n_j = \frac{n\pi_{j,t}^I}{\widehat{\delta}\left(1-\pi_{j,t}^I\right)\chi}. \tag{40}$$

Therefore, firm size is:

$$n_j + \widetilde{n}_j^I = \frac{n\left(1+\pi_j^I/\widehat{\delta}\right)}{\left(1-\pi_j^I\right)\chi},$$

which is strictly increasing in the probability of forming vendor relationships π_j^I, and strictly decreasing in the rate of product line obsolescence χ.

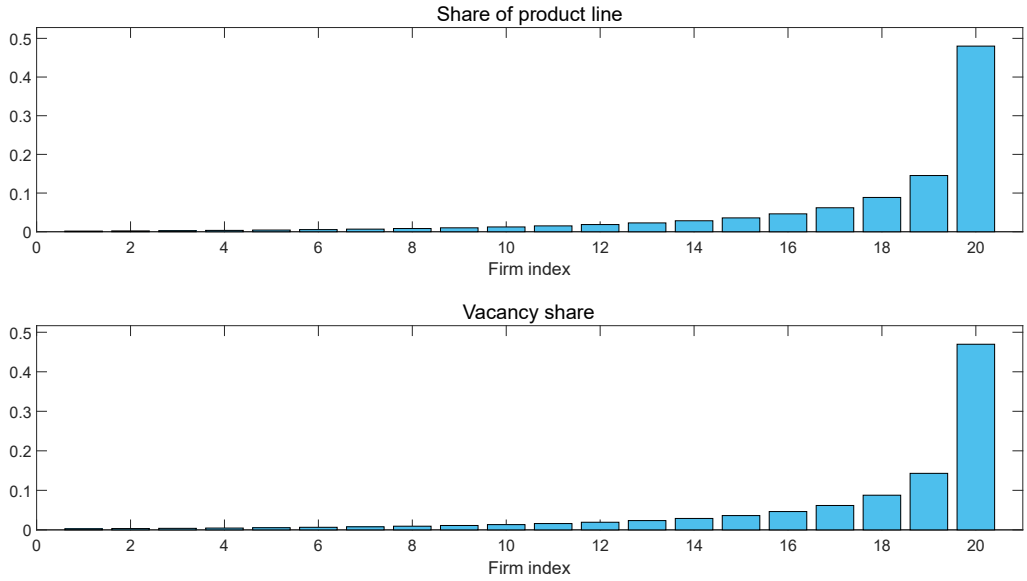

Figure 14: Firm size and vacancies

The upper panel of Figure 14 plots the distribution of the measure of product lines matched with a worker. This measure is increasing in the firm's productivity and highly concentrated among the "superstar firms": the top 5% own 48% of product lines and the next 5% own an additional 14.5%. The high concentration of firm size appears despite a moderate productivity dispersion in our calibration. The top 5% of firms own around three times more product lines as the next 5% of firms, although the former ones are only 1.3% more productive than the latter. The reason is that equations (39)-(40) imply that the measure of product lines is non-linear in π_j^I. The non-linearity becomes stronger when the measure gets close to one, which is the case for the most productive firms.

A popular measure of firms' size is labor market concentration, as measured by the Herfindahl-Hirschman Index (HHI) for employment, $HHI^{emp} = \sum_{j=1}^{20} \widehat{n}_j$. Berger et al. (2019) document that the unweighted average HHI for employment across labor markets in the U.S. is about 0.42, while the average HHI becomes 0.12 once it is weighted by the labor market's share of total employment. The HHI implied by our model is 0.27, exactly at the mean between the unweighted and weighted HHI measured by Berger et al. (2019). This result is a strong validation of our theory, since we did not target this observation.

With respect to the distribution of vacancies (bottom panel of Figure 14), since all firms I

have the same vacancy-filling rates, vacancies are a measure of labor demand. The number of vacancies posted by the firm of j-type productivity:

$$v_j = \frac{n_j}{1 + p^n / \left(\delta + \widehat{\delta}\right)} = \frac{\left(\delta + \widehat{\delta}\right) n \pi_j^I}{\widehat{\delta}\left(1 - \pi_j^I\right) \chi \left(p^n + \delta + \widehat{\delta}\right)}.$$

is strictly increasing in π_j^I, but strictly decreasing in χ, for a given probability of matching with a worker, p^n. Intuitively, a higher π_j^I or a lower χ decreases the rate of product line obsolescence and, hence, raises the measure of product lines $(n_j + \widetilde{n}_j^I)$. As a firm gains more product lines, it also has a greater need to expand hiring (v_j). For instance, the top 5% of firms post 47% of all vacancies and the next 5%, 14.3%.

With respect to the distribution of output, a type-j firm produces:

$$y_j = \underbrace{zx_j}_{\text{Output per active product line}} \cdot \underbrace{\frac{n_j}{n_j + \widetilde{n}_j^I}}_{\text{prod. line util. rate}} \cdot \underbrace{n_j + \widetilde{n}_j^I}_{\text{Measure of prod. line}} \cdot \underbrace{\frac{\widehat{n}_j}{n_j}}_{\text{labor market matching}} \tag{41}$$

Equation (41) embodies the four channels that generate market concentration in our model. First, high-productivity firms produce more per active product line (zx_j). Given the small calibrated productivity dispersion in our model, this channel explains a minor fraction of industry concentration. The difference in exogenous idiosyncratic productivity between the top 5% of "superstar firms" and the bottom 5% of "lightweight firms" is 24%. Yet, our model generates a ratio between the outputs of the top and bottom firms of 372. Second, high-productivity firms have a higher product line utilization rate $(n_j/(n_j + \widetilde{n}_j^I))$ because they search more actively for partners, and due to search complementarities, potential partners send more buying agents to them. While the bottom 5% of firms have 54% of their product lines active, the top 5% operates 95% of their product lines. Third, since the product lines $(n_j + \widetilde{n}_j^I)$ of high-productivity firms are active more frequently, fewer of them become obsolete. Fourth, firms have monopsony power, which increases their profit share of vendor matches and provides incentives to expand their output. The next subsection will elaborate on this channel, which affects $\widetilde{n}_j^I$, and n_j^I non-linearly in the firm's size.

The quantitative effect of equation (41) is that the most productive firms account for a disproportionate share of output: the top 5% produce 49% of output and the next 5% produce

an additional 15%, while the bottom 5% only generate 0.13% of output. These numbers are in line with the observations documented by Autor et al. (2020).

We can compare this result with a simple span-of-control model à la Lucas. With a production function xl^{γ} where x is managerial talent and l is hours worked, the output ratio in such a model between two firms with $x = e^{-0.12}$ and $x = e^{0.12}$ (the same dispersion in managerial abilities as the dispersion in productivities in our model) is $e^{0.24\frac{1}{1-\gamma}}$. To replicate an output ratio of 372 between the top and bottom firms, we would need $\gamma = 0.96$, which is much higher than other estimates of returns to scale. Atkeson and Kehoe (2005) argue that, in a span-of-control model, we should calibrate $\gamma = 0.85$, while Guner et al. (2018) estimate $\gamma = 0.77$. An alternative way to think about this is that a span-of-control model with $\gamma = 0.96$ would generate differences in mTFP much larger than the ratio of 1.92 documented by Syverson (2011).

Our results also challenge the classic prediction that firms with market power operate under excess capacity in equilibrium (an idea that goes back to Wicksell, 1934). Our model delivers the opposite result: top firms operate at a higher utilization rate, eliminating a key source of inefficiency in the economy—the more concentrated the market, the greater the rate of utilization of product lines.

5.3 Monopsony power: Market structure and wages

We mentioned before that monopsony power affects $\widetilde{n}_j^I$ and n_j^I non-linearly. To see this, Figure 15 shows the market structure for three alternative degrees of threatening power $\widetilde{\delta}$: 0.51 (our benchmark calibration), 0.75, and 1.

In the top panel of Figure 15, the top 10% of firms produce 64.1% of output, our calibration target. The middle panel of Figure 15 documents that, as we increase $\widetilde{\delta}$ to 0.75 (equivalent to an average rate of forgiveness of 1.3 periods), the share of output of the top 10% of firms falls to 56.9%. When we completely eliminate monopsony power (i.e., $\widetilde{\delta} = 1$), the share of output of the top 10% of firms becomes 53.1%. Figure 15 also illustrates Proposition 2: firms with a larger market share search more intensely, but this intensity decreases with $\widetilde{\delta}$.

Figure 15 justifies why we can think about our model as a measurement device: the model tells how much monopsony power we need to account for market concentration that is consistent with mTFP, rate of idleness, and labor market observations. Our model measures a moderate

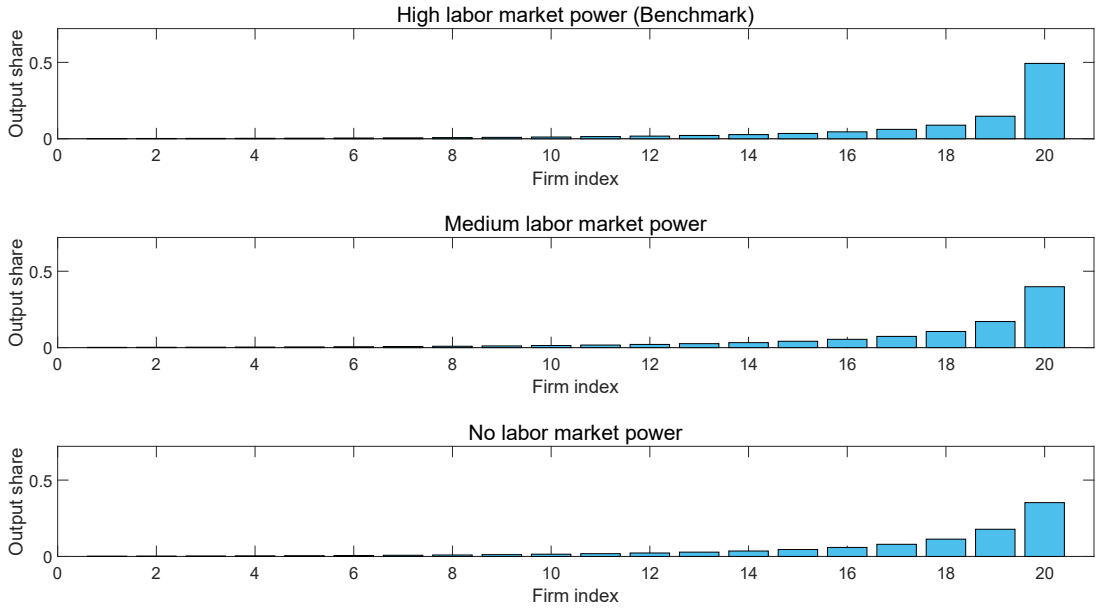

Figure 15: Market structure

amount of monopsony power (a punishment that lasts only six months on average), but such a market power can increase concentration at the top 10% of firms from 53.1% to 64.1% of output.

We move now to wages. In our model, monopsony power affects the wage distribution via two channels. In the first channel, the equilibrium wage decreases with the threatening power. We can see this effect in Figure 16, which plots the distribution of wages for j-type firms with high labor market power (dark-green histograms), medium labor market power (green histograms), and no labor market power cases (light-green histograms). As stated in Proposition 1, wages are increasing in the productivity of the firm, but decreasing with monopsony power. Moreover, the wage differential within the j-type firm is non-linear: the more productive the firm, the stronger the threatening power and, therefore, the larger the share of the surplus kept by the firm.

Our result agrees with a large empirical literature on the negative effect of market concentration on wage compensation. For example, see Dube et al. (2016), Benmelech et al. (2018), Qiu and Sojourner (2019), and Naidu et al. (2018). Also, Berger et al. (2019) find that markdowns, the ratio between the marginal revenue product of labor and its wage, are increasing in firm size. Jarosch et al. (2019) show that employer market power is also boosted by search and matching frictions. Finally, in Peters (2020), firms' market power is endogenous and the distribution of markups emerges as an equilibrium outcome.

The second channel through which monopsony power affects the wage distribution is that it

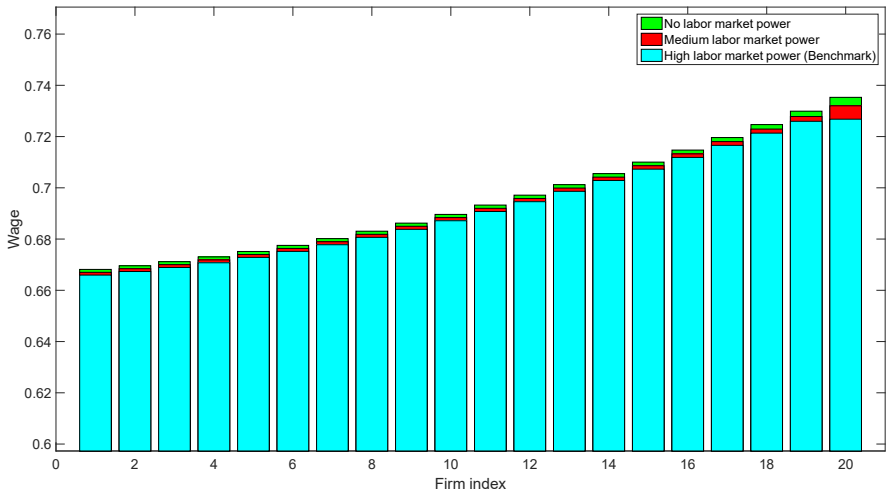

Figure 16: Wage with different labor market power

reallocates workers toward high-productivity firms that have lower firm labor income shares. This last point is a standard feature of search and matching models. Intuitively, when a firm's productivity is sufficiently low, the firm still needs to compensate the workers that have the outside value of finding another job and receives zero profit. In this case, the labor income share is close to one. As we increase the firm's productivity, the outside value becomes less binding and the labor income share decreases. In addition, by reallocating workers toward high-productivity firms, monopsony power increases wage inequality.

These results agree with the evidence. De Loecker et al. (2020) find that the decline in the economy-wide labor share is mainly driven by large, high-markup firms that have individually low labor shares. Similar findings appear in Autor et al. (2020) and Kehrig and Vincent (2017).

But labor income share also falls when $\widetilde{\delta}$ decreases. In our benchmark calibration ($\widetilde{\delta} = 0.51$), the labor income share is 0.66 and it increases to 0.663 when $\widetilde{\delta} = 0.75$ and 0.667 when $\widetilde{\delta} = 1$. While this effect is modest, we could substantially increase it if we were to assume (as is likely to be the case in the real world) that high-productivity firms also have a higher $\widetilde{\delta}$ (for example, through better HR processes to "punish" workers that do not accept low wage offers).

Finally, we can compare the wage markdown implied by our model with the empirical evidence. Given the linear production function, the wage markdown is equal to x_j/w_j. The average wage markdown weighted by the employment share in our calibration is 1.52. The difference in the wage markdowns across firms is sizeable (Figure 19 in Appendix A.3 plots the distribution of wage markdown across firms). For example, the type-20 firm generates a wage

markdown of 1.55, 16.5% larger than the wage markdown charged by the type-1 firm, whose wage markdown is 1.33. Hershbein et al. (2019) document that the average U.S. establishment sets a wage markdown of 1.53, and the establishments in the top decile of the distribution charge roughly a 20% higher wage markdown than the establishments in the bottom decile. Hence, our model roughly agrees with the data regarding the average level and the dispersion of the wage markdown, even though we did not use any information regarding these observations in our calibration.

5.4 Monopsony power without search complementarities

We just saw how monopsony power in the labor market amplifies the effect of search complementarities on market concentration. Does monopsony power generate market concentration in the absence of search complementarities? Yes, but the effect is mild.

To see this, we compute the market structure without search complementarities. To make our analysis comparable to our benchmark results, we fix all tightness ratios $\theta_{j,t}$ at their values in the benchmark model, but let the search effort (σ_j^I) vary. Thus, when $\widetilde{\delta} = 0.51$, the distribution of output shares across would be the same with or without search complementarities.

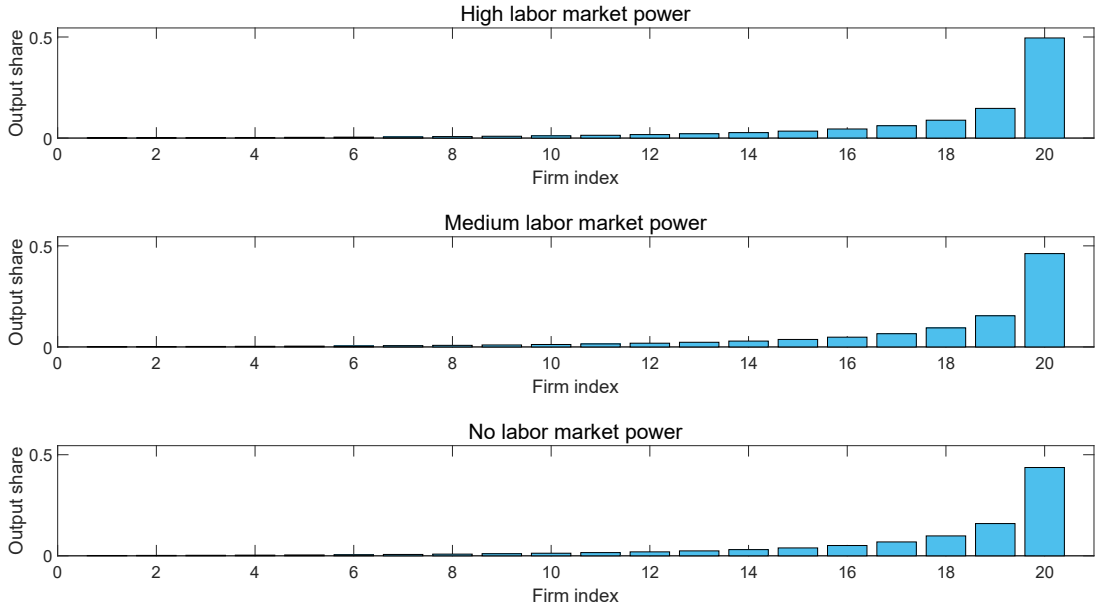

Figure 17: Market structure with no search complementarity

Figure 17 shows the output shares for the same three levels of monopsony power as in Figure 15: 0.51 (our benchmark calibration), 0.75, and 1. By construction, the top panel of Figure 17

is identical to the top panel of Figure 15. The middle and lower panels of Figure 17 show that the role of monopsony power in market concentration becomes milder as the incentives to scale up production become smaller. For example, the top 5% of firms decrease their output share from 49.5% to 46.2% when $\widetilde{\delta}$ increases from 0.51 to 0.75.

5.5 Lower search costs increase market concentration

We study the dependence of market structure on the cost of signing a vendor contract by considering a permanent decline in the unit cost of visiting each submarket, κ, from 1 to 0.98. A lower unit search cost induces all firms to search more actively, attracting a larger number of buying agents from sector F to visit them, increasing the probability of forming a vendor relationship and raising the number of product lines at the new steady state. However, the top 5% of "superstar firms" benefit the most from the reduction in κ, growing from producing 49.3% of output to producing 65.6% (see Figure 20 in Appendix A.3).

As in the case of the basic model, we interpret these results as suggesting that improvements in IT over the last few decades (or, more generally, in the ability to scale up production) have been a critical factor behind the recent increase in market concentration documented by Autor et al. (2020) and others.

5.6 Response of output to aggregate shocks

Our last exercise explores how the model responds to an aggregate shock by assuming that the level of aggregate productivity is stochastic. The production technology in equation (28) becomes $y_{j,t} = z_t x_j$, where z_t follows $log\,(z_t) = \rho_z log\,(z_{t-1}) + \sigma_z \epsilon_t$, and $\epsilon_t \sim \mathcal{N}(0,1)$. We calibrate ρ_z and σ_z to 0.95 and 0.011, respectively, which are standard values in the literature. We also modify the conditional expectations in all the relevant value functions of the agents of the model.

We implement a negative aggregate productivity shock, which reduces all firms' log-productivity. To ease the computational burden of keeping track of 20 different types of firms, we simplify our problem by assuming, for this subsection only, that the utility function of the household is linear in consumption.

Figure 18 shows the IRFs of the output of type-1 firms (the bottom 5%; continuous blue line), the type-10 firms (the median firms; discontinuous black line), and the type-20 firms

(the top 5%; firms; discontinuous red line) to an aggregate shock that reduces log aggregate productivity by 10%. Aggregate productivity reverts back to the steady state with a persistence of 0.95. We express the IRFs in percentage deviations with respect to the deterministic steady state. At impact, all firms' output drops by 9.52%. The recovery after this drop is slow because the productivity shock reduces firms' incentive to search. Thus, more product lines remain idle and they become obsolete at a higher rate. The process of reducing product lines is protracted and induces a lot of endogenous persistence in output.

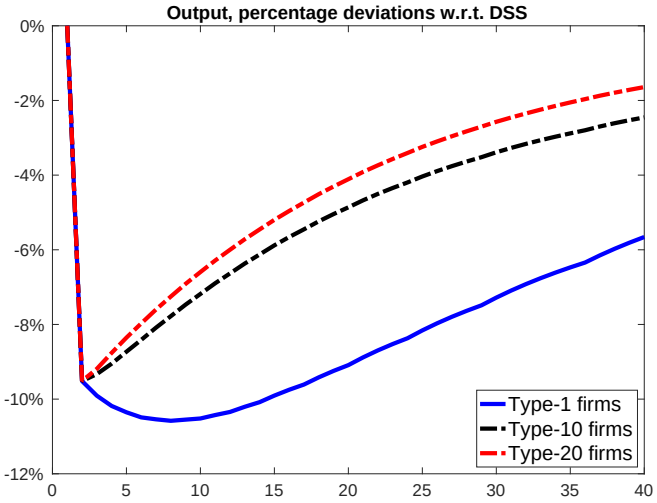

Figure 18: Response of output to 10% TFP shock

Interestingly, the recovery process is uneven across firms and increases market concentration. Specifically, it takes longer for low-productivity firms to recover. Intuitively, low-productivity firms' profit share is lower due to the worker's outside option of finding higher-paying jobs. Consequently, the low-productivity firms' search effort is more sensitive to productivity shocks, and these low-productivity firms lose more product lines in relative terms. In fact, for a few periods after the shock, the higher obsolescence rate of the non-active product lines of these low-productivity firms is such a powerful mechanism that their output continues dropping even if aggregate productivity is reverting to its mean. This mechanism accounts for the U-shaped IRF of low-productivity firms. In comparison, high-productivity firms recover faster, and market concentration increases.

Our findings agree, in sign and size, with Şahin et al. (2011), who documented that between December 2007 and December 2009, jobs declined 10.4% in small firms (those with fewer than fifty employees), compared with 7.5% in large ones.

Figure 18 also links our model with the Great Moderation of the U.S. economy after 1984. Since larger firms react less to negative aggregate shocks, the fall in search costs that we have argued above has occurred during the last decades helps us understand the growth in the size of superstar firms and the lower aggregate volatility of the economy.

6 Conclusion

Search complementarities have enormous consequences for market structure and the firm's size distribution. Through a "Matthew effect," small differences in productivity are transformed into large differences in firms' size, vacancies, and output. The key to this "Matthew effect" lies with the endogenous search decisions of intermediate- and final-goods producers under directed search: higher productivity leads to higher search effort by the intermediate-goods producers and more buying agents by the final-goods producers. The presence of monopsony power in the labor market reinforces the process even more. The forces combine to generate superstar firms with output shares that match empirical observations.

Our model also suggests that a reduction in search costs (which can be more generally understood as a fall in the cost of scaling a business up) leads to i) higher market concentration; ii) lower labor income shares; and iii) more monopsony power by firms. We interpret the IT revolution since the 1980s in the U.S. and other advanced economies as a reduction in search costs (better logistics software, improved inventory control, easier database management, etc.). Thus, our model offers a simple and parsimonious explanation of several important aspects of the data.

There is much scope for further investigation. We want to look at microdata to cross-validate the forces we highlight in our theoretical and quantitative analysis. We want to incorporate a firm's life-cycle. We want to think about innovation and technological adoption within the context of strategic complementarities. We want to think more about heterogeneity among different industry sectors. Are search costs as relevant in heavy manufacturing as in consumer services? Do the differences among industries in terms of market structure and the firm's size distribution align with our model? Finally, we also want to think about the policy implications of our model. We hope to explore some of these avenues of research shortly.

References

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2019): "A Theory of Falling Growth and Rising Rents," Working Paper 26448, National Bureau of Economic Research.

Akerman, A., E. Helpman, O. Itskhoki, M.-A. Muendler, and S. Redding (2013): "Sources of wage inequality," *American Economic Review*, 103, 214–219.

Ashenfelter, O., H. Farber, and M. Ransom (2010): "Labor market monopsony," *Journal of Labor Economics*, 28, 203–210.

Atkeson, A. and P. Kehoe (2005): "Modeling and Measuring Organization Capital," *Journal of Political Economy*, 113, 1026–1053.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): "The fall of the labor share and the rise of superstar firms," *Quarterly Journal of Economics*, 135, 645–709.

Azar, J., E. Huet-Vaughn, I. E. Marinescu, B. Taska, and T. Von Wachter (2019): "Minimum wage employment effects and labor market concentration," SSRN Scholarly Paper, Social Science Research Network, Rochester, NY.

Benmelech, E., N. Bergman, and H. Kim (2018): "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?" Working Paper 24307, National Bureau of Economic Research.

Berger, D. W., K. F. Herkenhoff, and S. Mongey (2019): "Labor Market Power," Working Paper 25719, National Bureau of Economic Research.

Bessen, J. E. (2017): "Industry concentration and information technology," SSRN Scholarly Paper ID 3044730, Social Science Research Network, Rochester, NY.

Bulow, J. I., J. D. Geanakoplos, and P. D. Klemperer (1985): "Multimarket oligopoly: Strategic substitutes and complements," *Journal of Political Economy*, 93, 488–511.

Burdett, K. and D. T. Mortensen (1980): "Search, layoffs, and labor market equilibrium," *Journal of Political Economy*, 88, 652–672.

Butters, G. R. (1977): "Equilibrium distributions of sales and advertising prices," *Review of Economic Studies*, 44, 465–491.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36, S13 – S70.

Cette, G., L. Koehl, and T. Philippon (2019): "Labor shares in some advanced economies," Working Paper 26136, National Bureau of Economic Research.

Chevalier, J. A. and D. S. Scharfstein (1996): "Capital-Market Imperfections and Countercyclical Markups: Theory and Evidence," *American Economic Review*, 86, 703–725.

Covarrubias, M., G. Gutiérrez, and T. Philippon (2019): "From good to bad concentration? U.S. industries over the past 30 years," Working Paper 25983, National Bureau of Economic Research.

De Loecker, J. and J. Eeckhout (2018): "Global market power," Working Paper 24768, National Bureau of Economic Research.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The rise of market power and the macroeconomic implications," *Quarterly Journal of Economics*, 135, 561–644.

Diamond, P. (1982): "Aggregate demand management in search equilibrium," *Journal of Political Economy*, 90, 881–894.

Diamond, P. and D. Fudenberg (1989): "Rational expectations business cycles in search equilibrium," *Journal of Political Economy*, 97, 606–619.

Dube, A., T. W. Lester, and M. Reich (2016): "Minimum wage shocks, employment flows, and labor market frictions," *Journal of Labor Economics*, 34, 663–704.

Elsby, M., B. Hobijn, and A. Sahin (2013): "The decline of the U.S. labor share," *Brookings Papers on Economic Activity*, 44, 1–63.

Falch, T. (2010): "The elasticity of labor supply at the establishment level," *Journal of Labor Economics*, 28, 237–266.

Fernández-Villaverde, J., F. Mandelman, Y. Yu, and F. Zanetti (2019): "Search complementarities, aggregate fluctuations, and fiscal policy," Working Paper 26210, National Bureau of Economic Research.

Garicano, L. and E. Rossi-Hansberg (2006): "Organization and inequality in a knowledge economy," *Quarterly Journal of Economics*, 121, 1383–1435.

Ghassibe, M. and F. Zanetti (2020): "State Dependence of Fiscal Multipliers: the Source of Fluctuations Matters," Economics Series Working Papers 930, University of Oxford.

Guner, N., A. Parkhomenko, and G. Ventura (2018): "Managers and Productivity Differences," *Review of Economic Dynamics*, 29, 256–282.

Gutiérrez, G. and T. Philippon (2018): "Ownership, concentration, and investment," *AEA Papers and Proceedings*, 108, 432–437.

Hamano, M. and F. Zanetti (2017): "Endogenous turnover and macroeconomic dynamics," *Review of Economic Dynamics*, 26, 263–279.

Hershbein, B., C. Macaluso, and C. Yeh (2019): "Concentration in U.S. local labor markets: Evidence from vacancy and employment data," 2019 Meeting Papers 1336, Society for Economic Dynamics.

——— (2020): "Monopsony in the U.S. Labor Market," Tech. rep., Working Paper.

Huo, Z. and J.-V. Ríos-Rull (2013): "Paradox of thrift recessions," Working Paper 19443, National Bureau of Economic Research.

Jarosch, G., J. S. Nimczik, and I. Sorkin (2019): "Granular search, market structure, and wages," Working Paper 26239, National Bureau of Economic Research.

Kaplan, G. and G. Menzio (2016): "Shopping Externalities and Self-Fulfilling Unemployment Fluctuations," *Journal of Political Economy*, 124, 771 – 825.

Karabarbounis, L. and B. Neiman (2014): "The global decline of the labor share," *Quarterly Journal of Economics*, 129, 61–103.

Kehrig, M. and N. Vincent (2017): "Growing productivity without growing wages: The micro-level anatomy of the aggregate labor share decline," CESifo Working Paper Series 6454, CESifo.

Lamadon, T., M. Mogstad, and B. Setzler (2019): "Imperfect competition, compensating differentials and rent sharing in the U.S. labor market," Working Paper 25954, National Bureau of Economic Research.

Lee, Y. and T. Mukoyama (2015): "Entry and exit of manufacturing plants over the business cycle," *European Economic Review*, 77, 20–27.

Lucas, R. (1978): "On the size distribution of business firms," *Bell Journal of Economics*, 9, 508–523.

Manning, A. (2011): "Imperfect Competition in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4B, chap. 11, 973–1041, 1 ed.

——— (2020): "Monopsony in labor markets: A review," *ILR Review*.

Marinescu, I., I. Ouss, and L.-D. Pape (2020): "Wages, Hires, and Labor Market Concentration," Working Paper 28084, National Bureau of Economic Research.

Matsudaira, J. D. (2014): "Monopsony in the low-wage labor market? Evidence from minimum nurse staffing regulations," *Review of Economics and Statistics*, 96, 92–102.

Merton, R. K. (1968): "The Matthew Effect in Science," *Science*, 159, 56–63.

Michaillat, P. and E. Saez (2015): "Aggregate demand, idle time, and unemployment," *Quarterly Journal of Economics*, 130, 507–569.

Naidu, S., E. A. Posner, and G. Weyl (2018): "Antitrust remedies for labor market power," *Harvard Law Review*, 132, 536.

Peters, M. (2020): "Heterogeneous Markups, Growth, and Endogenous Misallocation," *Econometrica*, 88, 2037–2073.

Petrongolo, B. and C. A. Pissarides (2001): "Looking into the black box: A survey of the matching function," *Journal of Economic Literature*, 39, 390–431.

Qiu, Y. and A. Sojourner (2019): "Labor-market concentration and labor compensation," *Available at SSRN 3312197*.

Ransom, M. and D. Sims (2010): "Estimating the firm's labor supply curve in a "new monopsony" framework: School teachers in Missouri," *Journal of Labor Economics*, 28, 331–355.

Şahin, A., S. Kitao, A. Cororaton, and S. Laiu (2011): "Why small businesses were hit harder by the recent recession," *Current Issues in Economics and Finance*, 17, 1–7.

Salgado, S., F. Guvenen, and N. Bloom (2019): "Skewed business cycles," Working Paper 26565, National Bureau of Economic Research.

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

Syverson, C. (2011): "What determines productivity?" *Journal of Economic Literature*, 49, 326–65.

Thomas, C. and F. Zanetti (2009): "Labor market reform and price stability: An application to the Euro Area," *Journal of Monetary Economics*, 56, 885–899.

Unger, R. M. (2019): *The Knowledge Economy*, Verso Books.

Weitzman, M. (1982): "Increasing returns and the foundations of unemployment theory," *Economic Journal*, 92, 787–804.

Wicksell, K. (1934): *Lectures on Political Economy*, Macmillan Company.

Wu, L. (2019): "Partially Directed Search in the Labor Market," *University of Chicago, mimeo.*

A Appendix

A.1 Proof of Proposition 1

In the steady state, *ceteris paribus*, the wage decreases with the firm's vacancy share (s_j) and increases with the probability of forgiveness ($\widetilde{\delta}$)

Proof. We begin our proof by showing that the *ex-ante* value of employment W_j decreases with the firm's vacancy share, s_j, and it increases with the probability of forgiveness, $\widetilde{\delta}$. We denote the total surplus in a labor market without labor market power as:

$$LTS_j^* = W_j - U + J_j - X_j,$$

so that the following equality holds:

$$LTS_j = LTS_j^* + U - \widetilde{U}_j.$$

Equation (25) implies that:

$$W_j = \widetilde{U}_j + (1 - \widetilde{\tau})\, LTS_j,$$

or, equivalently,

$$W_j - U = (1 - \widetilde{\tau})\, LTS_j^* - \widetilde{\tau}\left(U - \widetilde{U}_j\right). \tag{42}$$

Equation (17) entails that:

$$U - \widetilde{U}_j = \Gamma\left(s_j, \widetilde{\delta}\right)(W_j - U) \tag{43}$$

with

$$\Gamma\left(s_j, \widetilde{\delta}\right) = \frac{\left(1 - \widetilde{\delta}\right)\beta s_j p^u p^n (W_j - U)}{1 - \beta\left(1 - p^u + s_j p^u p^n\right)\left(1 - \widetilde{\delta}\right)}. \tag{44}$$

Notice that $\partial\Gamma/\partial s_j > 0$ and $\partial\Gamma/\partial\widetilde{\delta} < 0$.

Substituting equation (43) into equation (42), it yields the following value for employment:

$$W_j = U + \frac{1 - \widetilde{\tau}}{1 + \widetilde{\tau}\Gamma\left(s_j, \widetilde{\delta}\right)} LTS_j^*, \tag{45}$$

which implies that W_j decreases with s_j, and it increases with $\widetilde{\delta}$. Since changes in s_j or $\widetilde{\delta}$ determine the split of the total surplus between firms and workers, they involve a variation in $\Gamma(s_j, \widetilde{\delta})$, and do not have a first-order effect on the value of U and LTS_j^*.

Next, we show that the current period wage, w_j, decreases with s_j and increases with $\widetilde{\delta}$. Equation (18) implies that:

$$W_j = w_j + \beta\left[\left(1 - \delta - (1 - \pi_j)\chi\right) W_j + \left(\delta + (1 - \pi_j)\chi\right) U\right],$$

or:

$$w_j = (1-\beta)\, W_j + \beta\left(\delta + (1-\pi_j)\,\chi\right)(W_j - U)\,, \tag{46}$$

which shows that w_j strictly increases with W_j. Therefore, we have that w_j decreases with s_j and increases with $\widetilde{\delta}$.

□

A.2 Proof of Proposition 2

In the steady state, *ceteris paribus*, firm I's search effort increases with the firm's vacancy share s_j, and it decreases with the probability of forgiveness $\widetilde{\delta}$.

Proof. We begin our proof by showing that the value of a firm matched to a worker (J_j) increases with s_j and decreases with $\widetilde{\delta}$.

Equation (20) implies:

$$X_j = \alpha_{XJ} J_j, \tag{47}$$

where $\alpha_{XJ} = \frac{\beta p^n}{1-\beta(1-p^n-\chi)} < 1$. We rewrite equation (26) as:

$$(1-\alpha_{XJ})\, J_j = \widetilde{\tau} LTS_j,$$

or, equivalently:

$$J_j = \frac{\widetilde{\tau}}{1-\alpha_{XJ}}\left(LTS_j^* + U - \widetilde{U}_j\right). \tag{48}$$

Substituting equations (43) and (45) into equation (48), we find:

$$J_j = \frac{\widetilde{\tau}}{1-\alpha_{XJ}} \cdot \frac{(1-\widetilde{\tau})\,\Gamma\left(s_j, \widetilde{\delta}\right)}{1+\widetilde{\tau}\Gamma\left(s_j, \widetilde{\delta}\right)} \cdot LTS_j^*, \tag{49}$$

where $\Gamma(s_j, \widetilde{\delta})$ is defined by equation (44). Equation (49) implies that J_j increases with $\Gamma(\cdot)$. Since $\partial\Gamma/\partial s_j > 0$ and $\partial\Gamma/\partial\widetilde{\delta} < 0$, J_j increases with s_j, and decreases with $\widetilde{\delta}$. Consequently, equation (47) implies that X_j increases with s_j and decreases with $\widetilde{\delta}$. From equations (22) and (23), it is straightforward to show that $X_j^I = X_j/2 + \widetilde{J}_j^I/2$, which implies that X_j^I increases with X_j, and it thus increases with s_j and decreases with $\widetilde{\delta}$.

Next, we show that $\Delta J_j^I = X_j^I - (1-\chi)\,\widetilde{J}_j^I$ increases with X_j^I, and, thus, it increases with s_j and decreases with $\widetilde{\delta}$. We prove $d\Delta J_j^I/dX_j^I > 0$ in two steps.

In the first step, we show that $\widetilde{J}_j^I$ increases with X_j^I. Specifically, by denoting the optimal search effort with σ^*, and expressing $\widetilde{J}_j^I$ and σ^* as functions of X_j, we re-write equation (29) in the steady state as:

$$\widetilde{J}_j^I\left(X_j^I\right) = -c\left(\sigma^*\left(X_j^I\right)\right) + \beta\left[\pi_j^I\left(\sigma^*\left(X_j^I\right)\right)\cdot X_j^I + \left(1-\pi_j^I\left(\sigma^*\left(X_j\right)\right)\right)\cdot \widetilde{J}_j^I\left(X_j^I\right)\right], \tag{50}$$

which we solve explicitly for $\widetilde{J}_j^I\left(X_j^I\right)$:

$$\widetilde{J}_j^I\left(X_j^I\right)=\frac{\beta\pi_j^I\left(\sigma^*\left(X_j^I\right)\right)\cdot X_j^I-c\left(\sigma^*\left(X_j^I\right)\right)}{1-\beta\left(1-\pi_j^I\left(\sigma^*\left(X_j\right)\right)\right)}. \tag{51}$$

An increase of X_j^I by Δ is equal to:

$$\begin{aligned}\widetilde{J}_j^I\left(X_j^I+\Delta\right)&=-c\left(\sigma^*\left(X_j^I+\Delta\right)\right)+\\&\beta\left[\pi_j^I\left(\sigma^*\left(X_j^I+\Delta\right)\right)\cdot\left(X_j^I+\Delta\right)+\left(1-\pi_j^I\left(\sigma^*\left(X_j+\Delta\right)\right)\right)\cdot\widetilde{J}_j^I\left(X_j^I+\Delta\right)\right] \qquad (52)\\&>-c\left(\sigma^*\left(X_j^I\right)\right)+\beta\left[\pi_j^I\left(\sigma^*\left(X_j^I\right)\right)\cdot\left(X_j^I+\Delta\right)+\left(1-\pi_j^I\left(\sigma^*\left(X_j\right)\right)\right)\widetilde{J}_j^I\left(X_j^I+\Delta\right)\right], \qquad (53)\end{aligned}$$

which implies:

$$\widetilde{J}_j^I\left(X_j^I+\Delta\right)>\frac{\beta\pi_j^I\left(\sigma^*\left(X_j^I\right)\right)\cdot\left(X_j^I+\Delta\right)-c\left(\sigma^*\left(X_j^I\right)\right)}{1-\beta\left(1-\pi_j^I\left(\sigma^*\left(X_j\right)\right)\right)}. \tag{54}$$

Comparing equation (54) to equation (51) yields:

$$\widetilde{J}_j^I\left(X_j^I+\Delta\right)>\widetilde{J}_j^I\left(X_j^I\right),$$

clearly implying that $\widetilde{J}_j^I$ increases with X_j^I.

In the second step, we show that ΔJ_j^I increases with X_j^I. From equation (29), we have that:

$$\widetilde{J}_j^I=\frac{\beta\pi\left(\sigma_j\right)\Delta J_j^I-c\left(\sigma_j\right)}{1-\beta\left(1-\chi\right)}. \tag{55}$$

We denote $G\left(X_j^I\right)=\beta\pi\left(\sigma_j\left(X_j^I\right)\right)\Delta J_j^I\left(X_j^I\right)-c\left(\sigma_j\left(X_j^I\right)\right)$, and we treat σ_j and ΔJ_j^I as functions of X_j^I. Since $\partial\widetilde{J}_j^I/\partial X_j^I>0$, the following holds:

$$G'\left(X_j^I\right)=\beta\pi'\frac{d\sigma_j}{dX_j}\Delta J_j^I+\beta\left(\sigma_j\left(X_j^I\right)\right)\frac{d\Delta J_j^I}{dX_j^I}-c'\frac{d\sigma_j}{dX_j}>0. \tag{56}$$

The optimality condition for firm I's problem (equation (29)) implies that:

$$\beta\pi'\Delta J_j^I-c'\left(\sigma_j\right)=0, \tag{57}$$

and by substituting equation (57) into equation (56), we get:

$$\frac{d\Delta J_j^I}{dX_j^I}>0, \tag{58}$$

which shows that ΔJ_j^I increases with X_j^I, and consequently it increases with s_j and decreases with $\widetilde{\delta}$. By using these findings in equation (32), we have that firm I's search effort increases with the firm's labor market share s_j and decreases with the probability of forgiveness $\widetilde{\delta}$. □

A.3 Additional Figures

Figure 19 plots the distribution of the wage markdown in the deterministic steady state. We can see how the markdown is increasing with the firm's productivity.

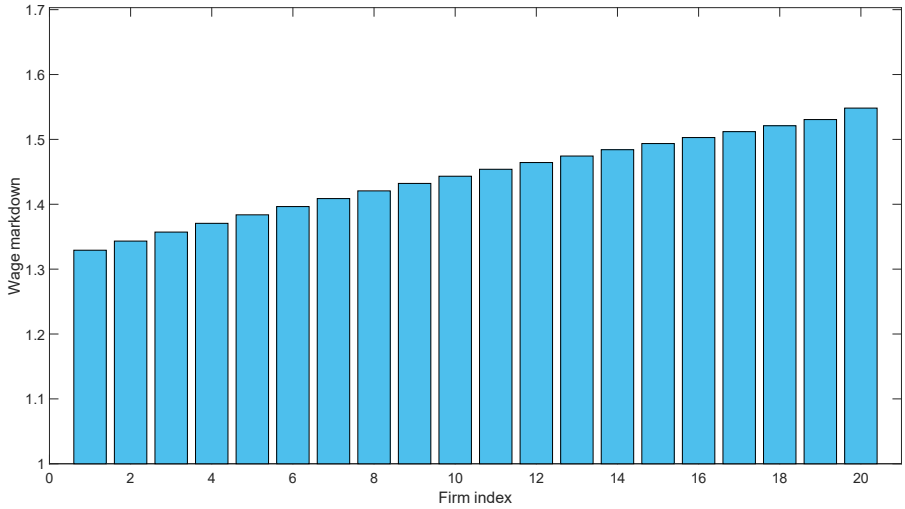

Figure 19: Distribution of the wage markdown

Figure 20 plots the distribution of output share for two different values of κ. The bottom panel shows our benchmark case of $\kappa = 1$, while the top panel shows the firms' output when $\kappa = 0.98$.

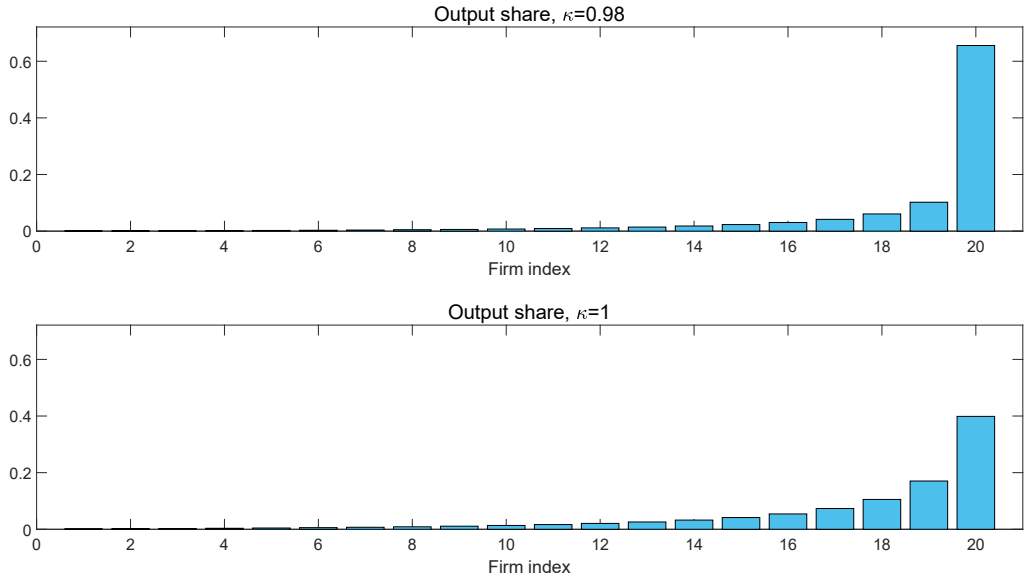

Figure 20: Output shares with different κ